Validity evidence for procedural competency in virtual reality robotic simulation, establishing a credible pass/fail standard for the vaginal cuff closure procedure

Lisette Hvid Hovgaard, BSc[1,2]; Steven Arild Wuyts Andersen, MD, PhD[3,4]; Lars Konge, MD, PhD[2,3]; Torur Dalsgaard, MD, PhD[5]; Christian Rifbjerg Larsen, MD, PhD[1,2]

1. Centre for Robotic Surgery, Department of Gynaecology, Copenhagen University Hospital; Herlev-Gentofte, Denmark
2. Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
3. The Simulation Centre at Rigshospitalet, Copenhagen Academy for Medical Education and Simulation (CAMES), Centre for HR, the Capital Region of Denmark
4. Department of Otorhinolaryngology-Head & Neck Surgery, Copenhagen University Hospital; Rigshospitalet-Glostrup, Denmark
5. Endometriosis Team and Robotic Surgery Section, Department of Gynaecology, Copenhagen University Hospital; Rigshospitalet-Glostrup, Denmark

Correspondence: Lisette Hvid Hovgaard, Centre for Robotic Surgery, Department of Gynaecology G-115, Copenhagen University Hospital; Herlev-Gentofte, Herlev Ringvej 75, DK-2730 Herlev. Phone: +45 38 68 98 80. E-mail: lisette@hovgard.dk

Funding: None

Disclosures: Lisette H. Hovgaard and Drs. Steven A. W. Andersen, Lars Konge, Torur Dalsgaard, and Christian R. Larsen have no conflicts of interest or financial ties to disclose.

Full citation: Hovgaard LH, Andersen SA, Dalsgaard T, Konge L, Rifbjerg C. Validity evidence for procedural competency in virtual reality robotic simulation, establishing a credible pass/fail standard for the vaginal cuff closure procedure. Surg Endosc. 2018 Oct;32(10):4200–4208.

DOI: 10.1007/s00464-018-6165-5

ABSTRACT

Background: The use of robotic surgery for minimally invasive procedures has increased considerably over the last decade. Robotic surgery has potential advantages compared to laparoscopic surgery but also requires new skills. Using virtual reality (VR) simulation to facilitate the acquisition of these new skills could potentially benefit training of robotic surgical skills and also be a crucial step in developing a robotic surgical training curriculum. The study objective was to establish validity evidence for a simulation-based test for procedural competency for the vaginal cuff closure procedure, that can be used in a future simulation-based, mastery-learning training curriculum.

Methods: Eleven novice gynaecological surgeons without prior robotic experience and 11 experienced gynaecological robotic surgeons (>30 robotic procedures) were recruited. After familiarization with the VR simulator participants completed the module 'Guided Vaginal Cuff Closure' six times. Validity evidence was investigated for 18 preselected simulator metrics. The internal consistency was assessed using Cronbach's alpha and a composite score was calculated based on metrics with significant discriminative ability between the two groups. Finally, a pass/fail standard was established using the contrasting groups' method.

Results: The experienced surgeons significantly outperformed the novice surgeons on 6 of the 18 metrics. The internal consistency was 0.58 (Cronbach's alpha). The experienced surgeons' mean composite score for all six repetitions were significantly better than the novice surgeons' (76.1 vs. 63.0, respectively, $p<0.001$). A pass/fail standard of 75/100 was established. Four novice surgeons passed this standard (false positives) and three experienced surgeons failed (false negatives).

Conclusion: Our study has gathered validity evidence for a simulation-based test for procedural robotic surgical competency in the vaginal cuff closure procedure and established a credible pass/fail standard for future proficiency-based training.

Keywords: Robotic surgery · virtual reality simulation · gynaecology · assessment · proficiency-based training

INTRODUCTION

Robotic surgery with the da Vinci® Surgical System (Intuitive Surgical, California, USA) was approved by US Food and Drug Administration (FDA) in 2000[1]. The use of robotic surgery has increased considerably during the last decade[1] and is used for minimally invasive procedures in most specialities including gynaecology[2], urology[3], otorhinolaryngology[4], cardiothoracic surgery[5], and general surgery[6]. Although comparisons of robotic and laparoscopic surgery are still inconclusive concerning perioperative outcomes[7,8] the robotic platform provides some advantages for the operating surgeon such as improved visualization, wider range of instrument movement, elimination of hand tremor and better ergonomics[9].

The skills necessary for robotic surgery differ from the skills required for open and laparoscopic surgery. The surgeon works at the console and has no physical contact with the patient or the surgical instruments, which results in a lack of tactile feedback - an essential component of open and to some extend laparoscopic surgery. Also, robotic surgery has unique features such as the clutch for repositioning of the master controls, instrument arms with 7 degrees of freedom (EndoWrist®) and camera navigation different from laparoscopic surgery[10].

The introduction of robotic surgery is associated with initial challenges due to the new skills required for this type of minimally invasive surgery[11,12]. Virtual reality (VR) simulation of robotic surgery can be used to train these new skills and can potentially optimize the early learning curve in robotic surgery[13,14]. VR simulation enables novice robotic surgeons to practice in a patient-free and safe learning environment allowing also for repeated and distributed practice, which is optimal for learning[15]. Finally, for the experienced surgeon, re-familiarization with the robotic console prior to a procedure as a 'warm-up' with the VR simulator improves subsequent performance[16].

In the field of laparoscopic surgery, the value of VR simulators and box trainers as valuable educational tools to acquire laparoscopic skills has been established[17]. VR simulation training of complete operative procedures seems to be more efficient than basic skills training[18]. Skill transfer studies have to some extent demonstrated the same efficacy of using VR simulation to learn robotic skills[19,20].

Until now most studies on VR simulation of robotic surgery have focused on basic skills training[21-24]. However, advances in simulation and the introduction of complete procedural training for specialities such as gynaecology, urology and general surgery[25] may be the next step for inclusion of VR simulators into training curricula[21].

A standardized and evidence-based training curriculum in robotic surgery is necessary to ensure adequate training and competencies of prospective robotic surgeons. However, no standardized training curriculum for robotic surgery has been widely accepted[26]. Proficiency-based training seems to be the way forward[27] and such mastery learning enables the individual trainee to practice to the predefined proficiency level for each specified task. The trainees' learning time might vary but all trainees will reach the same objectives contrary to fixed training based on time or number of repetitions where the learning outcomes vary[28]. Mastery learning in simulation-based surgical skills training is dependent on a test with solid validity evidence using a contemporary framework of validity[29,30] as well as establishing a credible pass/fail standard for the test.

In this study, we therefore aimed to develop a simulation-based test for procedural competency in robotic surgery for the vaginal cuff closure procedure, gather validity evidence for this test, and establish a credible pass/fail standard for proficiency-based training.

MATERIALS AND METHODS

Participants and setting

Eleven robotic surgical novices and 11 experienced surgeons subspecialized in robotic surgery enrolled for this prospective, cohort study. Participants were recruited from two different institutions: the Copenhagen University Hospitals of Herlev-Gentofte and Rigshospitalet-Glostrup. The included robotic surgical novices were gynaecological residents, fellows and consultants with no prior experience in real-life or simulated robotic surgery. The included experienced surgeons were gynaecologists subspecialized in robotic surgery with > 30 performed robotic procedures.

The study was conducted at the Simulation Centre at Copenhagen Academy for Medical Education and Simulation (CAMES) from March to June 2017.

The simulator and metrics

The Robotix Mentor (3D Systems, Colorado, USA) VR simulator was used for this study. The simulator closely mimics the da Vinci® Surgical System and has an adjustable 3D stereoscopic display, non-fixed hand-controls, and adjustable foot pedals. An additional monitor mirrors the surgical view and allows the instructor to observe the procedure (Figure 1).

A number of gynaecologic procedures are available in the simulator and for this study we chose the 'Guided Vaginal Cuff Closure with a Barbed Suture' module (Figure 2) because it represents important gynaecologic robotic skills.

For this module, 30 different simulator metrics were recorded and these were divided into five domains: 1) general, 2) time and economy, 3) safety and tissue handling, 4) needle handling and 5) suture handling. Á priori, 18 of these metrics (Appendix 1) were selected by two experienced robotic surgeons who judged them to be clinically relevant to the procedure (content validity). Also, the metrics were chosen based on experience from simulation-based training in laparoscopic surgery[31].

Study design and data collection

First, all participants completed a demographic questionnaire for baseline characteristics including surgical experience. Next, participants received 15 minutes of general introduction to the simulator. This included a short demonstration on how to operate the simulator, followed by completion of two introductory tasks (Robotic Basic Skills: Manipulation Level 1 and Camera 30) for familiarization. Finally, they watched a short demonstration video of the procedure. After this introduction, participants performed three repetitions of the test followed by a 20-minute break before performing another three repetitions. The test entailed closing the vaginal cuff with a barbed suture using EndoWrist® needle driver. Using a barbed suture is considered gold standard for both laparoscopic and robotic assisted laparoscopic vaginal cuff closure in Denmark[32].

Participants received technical assistance with the simulator but did not receive help or feedback on the test. LH technically assisted all participants and ensured that they received the same information (providing validity evidence for response process, see below).

Validity evidence

Validity evidence for procedural competency in robotic surgery for the vaginal cuff closure procedure was assessed in accordance with Messick's framework of validity as it is considered best standard in medical educational research[33]. Messick's framework considers five sources of evidence and our study was designed accordingly (Table 1).

Data analysis

Sample size was in advance determined to be more than 10 participants in each group to assume normal distribution of test scores[34].

Internal structure: The internal consistency of the simulator metrics for each repetition was assessed using Cronbach's alpha. In general, a coefficient > 0.7 is considered acceptable for lower stakes tests, a coefficient > 0.8 is considered acceptable for moderate stakes tests, and a coefficient > 0.9 is considered necessary for high stakes tests[35].

Relationships with other variables: Independent t-tests were used for comparison of test scores between the two groups for each of the 18 predetermined simulator metrics. Metrics without a significant discriminative ability between the two groups were excluded together with metrics where the novice surgeons scored significantly better than the experienced surgeons. Next, a composite score from 0 to 100 for the remaining metrics was calculated using linear normalization and mean of included scores. Finally, independent t-tests were used to compare the composite scores of the two groups for each repetition and all six repetitions combined. P-values < 0.05 were considered statistically significant.

Consequences: A pass/fail standard was established based on the composite score using the contrasting groups' method[35]. This method considers the intersection between the distribution of composite scores of the novice and experienced surgeons, respectively to have as few false positives (passed novice surgeons) and as few false negatives (failed experienced surgeons) as possible.

The data analysis was performed using IBM SPSS Statistics for Windows version 24 (IBM, New York, USA).

Ethics

All participants provided written consent prior to participation and were assigned a unique identification number. Data were kept according to local guidelines. The study was deemed exempt by the regional ethics committee (protocol no. H-17008815).

RESULTS

All 22 participants recruited for the study completed the simulation test the required number of times. An overview of participant demographics is provided in Table 2.

Internal structure: The internal consistency for the simulator metrics with discriminative ability was 0.58($p<0.001$).

Relationships with other variables: Scores on the 18 predetermined metrics for the robotic surgical novices and experienced robotic surgeons were compared (Table 3). The experienced surgeons performed significantly better on six of the 18 metrics: Total time, path length (right instrument), path length (left instrument), distance by camera, instrument collisions, and number of unnecessary needle piercing points. The novice surgeons improved significantly in their score between first and sixth repetition on three of the six metrics: total time, path length (left instrument) and path length (right instrument) (appendix 2). However, the experienced surgeons still outperformed the novices. The novices performed significantly better on four of the 18 metrics: Total path of instruments travelled out of view, number of times instruments are out of view, total time instruments are out of view, and number of times the needle was held outside the visible field.

The experienced surgeons did not improve significantly on these metrics between the first and sixth repetition (appendix 2). The remaining eight metrics were without significant discriminative ability.

For the composite score, we included metrics where the experienced surgeons performed better than the novices. The experienced surgeons performed significantly better on the metric 'distance by camera'. However, this metric was excluded since the novices did not use the camera and therefore scored better in metrics concerning instruments out of view.

The experienced surgeons' mean composite score for all six repetitions was 76.1 (SD 17.0) and significantly higher than the novice surgeons' mean composite score of 63.0 (SD 19.9) ($p<0.001$) (Table 4 and Figure 3). For the first four repetitions, there were no significant differences between the two groups' composite scores, but the scores for the experienced surgeons were consistently higher than for the novices. For the final fifth and sixth repetition, the experienced surgeons' composite scores were 84.1 (SD 8.1) and 83.1 (SD 6.3), respectively, and significantly better than the novice surgeons' composite scores of 65.1 (SD 15.6) and 70.6 (SD 9.7), respectively ($p=0.002$ for both comparisons).

Consequences: A pass/fail standard for the composite score of 75 was determined using data from the fifth and sixth repetition and the contrasting groups' method (Figure 4). This pass/fail standard resulted in four out of 11 novices passed the test (false positives): three passed in both their fifth and sixth repetition whereas the last novice only passed the sixth repetition. For the experienced surgeons, three out of 11 failed the test (false negatives): two experienced surgeons failed the fifth repetition but passed their sixth whereas one failed to achieve the cut-off score in all six repetitions.

DISCUSSION

In this prospective study, we investigated the performance of novice and experienced robotic surgeons to gather validity evidence for a simulation-based test for procedural robotic competency in the vaginal cuff closure procedure. A credible pass/fail standard was defined by a composite score of 75/100, which allows for future proficiency-based training to this standard.

The composite score was based on five of 18 predetermined simulator metrics of relevance to the procedure and based on the metrics where the experienced surgeons statistically significantly outperformed the novices: total time, path length (right instrument), path length (left instrument), instrument collisions and number of unnecessary needle piercing points. The novices performed better than experienced surgeons on metrics concerning having the needle or instruments out of sight. However, data confirmed that the novices rarely even used the camera and performed the entire procedure without zooming in on the surgical field and as a result never lost view of the instruments (the ‘distance by camera’ metric). In contrast, the experienced surgeons worked close to the tissue and therefore frequently lost sight of the instruments and the needle especially when tightening the suture. Similarly, another study found that experienced surgeons accumulated errors (instruments out of view) due to their camera’s close proximity to the tissue[24]. Consequently, we excluded the ‘distance by camera’ metric from the composite score.

One previous study on basic robotic skills training using the same VR simulator similarly reported discriminative ability for the metrics: time to complete task, path length and instrument collisions[24]. In addition, studies from other fields also reported validity evidence for only a minority of built-in metrics[36,37] and further supported that validity evidence is needed for robust standard setting in all simulation-based technical skills training. Non-discriminative metrics could, however, have a place in formative feedback in the context of a specific procedural test, but should be avoided in the test score.

We found a relatively low internal consistency of the simulation-based test (Cronbach’s alpha = 0.58) with a plausible explanation being that some of the included metrics concerned time whereas the others concerned precision. In other words, participants that were either fast but had low precision or slow but with good precision lowered the internal consistency. In addition, participants’ performances had a high variability as indicated in figure 3, which could also reduce the Cronbach’s alpha.

A pass/fail standard of 75 as established in our study failed only seven out of 11 robotic surgical novices but at the same time failed three out of 11 experienced surgeons. At first glance this could seem problematic. However, a representative sample of proficient robotic surgeons should be used for standard setting because otherwise the established standard could be unachievable for novice surgeons or too low, allowing inferior skills[38]. Furthermore, competency in one procedural test is just a single component in a larger training curriculum and other procedures and training modalities should complement training.

Both groups demonstrated a learning curve as seen in Figure 3 with statistically similar scores in the first four repetitions for both the novices and experienced robotic surgeons. This phenomenon is well known: the novices needed to learn both the procedure and the simulation equipment and therefore demonstrated a prolonged learning curve whereas the learning curve for the experienced surgeons was caused by their need to learn how to use the simulator and adjust to any differences between the simulated environment and real-life. Several experienced surgeons explicitly mentioned how both the suture and tissue differed from reality, causing frustration and underperformance whereas novices simply accepted the premises of the simulation. It is therefore important to allow for enough repetitions to achieve familiarization and a stable performance in the simulation setting. Further studies are needed to investigate the learning curves of novice surgeons as well as the time needed to achieve proficiency.

Similar to previous studies on robotic simulation[39,22,40,41], our study had a small number of participants mainly because of the limited number of experienced robotic surgeons and even within the experienced group, the number of robotic procedures performed by each surgeon varied. Regardless, our sample-size was sufficient to detect statistically significant differences between the two groups. Other limitations to our study related to the simulation-based test not encompassing all potentially relevant robotic surgical skills such as the use of the fourth arm, the scissors and cautery instruments. As previously discussed, one test is just part of the battery needed for robotic surgical skills training before supervised surgery. Other skills besides technical competency should also be considered including non-technical skills such as communication and teamwork[42].

A strength of our study was that the robotic surgical novice group consisted of gynaecological residents, fellows, and consultants. Often, medical students are recruited for the novice group, which artificially improves reliability and discriminatory ability of the test but reduces the generalizability of the study[43]. Furthermore, the use of simulator-generated metrics for a procedural test eliminates rating bias in performance assessment and provides immediate feedback.

Several other gynaecologic procedures can be simulated in the VR environment and these needs to be further explored and pass/fail standards set. Further, the integration of several procedures into a future robotic gynaecologic training curriculum as well as transfer of skills to the operating room performance need future research.

CONCLUSION

We have established validity evidence for a simulation-based test for procedural robotic surgical competency in the vaginal cuff closure procedure. A credible pass/fail standard with reasonable consequences has been established and can be used for proficiency-based training. This should be embedded into an evidence-based robotic surgical training curriculum so that novice robotic surgeons have achieved optimal training before commencing supervised surgery.

ACKNOWLEDGEMENTS

The research group sincerely thank all doctors from the Centre for Robotic Surgery at Copenhagen University Hospital, Herlev-Gentofte and from Robotic Surgery Section at Department of Gynaecology, Copenhagen University Hospital, Rigshospitalet-Glostrup for participation.

REFERENCE LIST

1. Schreuder HW, Verheijen RH (2009) Robotic surgery. BJOG 116 (2):198-213. doi:10.1111/j.1471-0528.2008.02038.x
2. Weinberg L, Rao S, Escobar PF (2011) Robotic surgery in gynecology: an updated systematic review. Obstet Gynecol Int 2011:852061. doi:10.1155/2011/852061
3. Yates DR, Vaessen C, Roupret M (2011) From Leonardo to da Vinci: the history of robot-assisted surgery in urology. BJU Int 108 (11):1708-1713; discussion 1714. doi:10.1111/j.1464-410X.2011.10576.x
4. Maan ZN, Gibbins N, Al-Jabri T, D'Souza AR (2012) The use of robotics in otolaryngology-head and neck surgery: a systematic review. Am J Otolaryngol 33 (1):137-146. doi:10.1016/j.amjoto.2011.04.003
5. Veronesi G (2013) Robotic surgery for the treatment of early-stage lung cancer. Curr Opin Oncol 25 (2):107-114. doi:10.1097/CCO.0b013e32835daf4f
6. Baek SK, Carmichael JC, Pigazzi A (2013) Robotic surgery: colon and rectum. Cancer J 19 (2):140-146. doi:10.1097/PPO.0b013e31828ba0fd
7. Liu Z, Li X, Tian S, Zhu T, Yao Y, Tao Y (2017) Superiority of robotic surgery for cervical cancer in comparison with traditional approaches: A systematic review and meta-analysis. Int J Surg 40:145-154. doi:10.1016/j.ijsu.2017.02.062
8. Xie W, Cao D, Yang J, Shen K, Zhao L (2016) Robot-assisted surgery versus conventional laparoscopic surgery for endometrial cancer: a systematic review and meta-analysis. J Cancer Res Clin Oncol 142 (10):2173-2183. doi:10.1007/s00432-016-2180-x
9. Foote JR, Valea FA (2016) Robotic surgical training: Where are we? Gynecol Oncol 143 (1):179-183. doi:10.1016/j.ygyno.2016.05.031
10. Bric J, Connolly M, Kastenmeier A, Goldblatt M, Gould JC (2014) Proficiency training on a virtual reality robotic surgical skills curriculum. Surg Endosc 28 (12):3343-3348. doi:10.1007/s00464-014-3624-5
11. Sandadi S, Gadzinski JA, Lee S, Chi DS, Sonoda Y, Jewell EL, Brown CL, Gardner GJ, Barakat RR, Leitao MM, Jr. (2014) Fellowship learning curve associated with completing a robotic assisted total laparoscopic hysterectomy. Gynecol Oncol 132 (1):102-106. doi:10.1016/j.ygyno.2013.11.017
12. Seamon LG, Fowler JM, Richardson DL, Carlson MJ, Valmadre S, Phillips GS, Cohn DE (2009) A detailed analysis of the learning curve: robotic hysterectomy and pelvic-aortic lymphadenectomy for endometrial cancer. Gynecol Oncol 114 (2):162-167. doi:10.1016/j.ygyno.2009.04.017

13. Balasundaram I, Aggarwal R, Darzi A (2008) Short-phase training on a virtual reality simulator improves technical performance in tele-robotic surgery. Int J Med Robot 4 (2):139-145. doi:10.1002/rcs.181
14. Schreuder HW, Wolswijk R, Zweemer RP, Schijven MP, Verheijen RH (2012) Training and learning robotic surgery, time for a more structured approach: a systematic review. BJOG 119 (2):137-149. doi:10.1111/j.1471-0528.2011.03139.x
15. Andersen SA, Konge L, Caye-Thomasen P, Sorensen MS (2015) Learning Curves of Virtual Mastoidectomy in Distributed and Massed Practice. JAMA Otolaryngol Head Neck Surg 141 (10):913-918. doi:10.1001/jamaoto.2015.1563
16. Lendvay TS, Brand TC, White L, Kowalewski T, Jonnadula S, Mercer LD, Khorsand D, Andros J, Hannaford B, Satava RM (2013) Virtual reality robotic surgery warm-up improves task performance in a dry laboratory environment: a prospective randomized controlled study. J Am Coll Surg 216 (6):1181-1192. doi:10.1016/j.jamcollsurg.2013.02.012
17. Larsen CR, Soerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C, Schroeder TV, Ottesen BS (2009) Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. BMJ 338:b1802. doi:10.1136/bmj.b1802
18. Larsen CR, Oestergaard J, Ottesen BS, Soerensen JL (2012) The efficacy of virtual reality simulation training in laparoscopy: a systematic review of randomized trials. Acta Obstet Gynecol Scand 91 (9):1015-1028. doi:10.1111/j.1600-0412.2012.01482.x
19. Culligan P, Gurshumov E, Lewis C, Priestley J, Komar J, Salamon C (2014) Predictive validity of a training protocol using a robotic surgery simulator. Female Pelvic Med Reconstr Surg 20 (1):48-51. doi:10.1097/spv.0000000000000045
20. Kiely DJ, Gotlieb WH, Lau S, Zeng X, Samouelian V, Ramanakumar AV, Zakrzewski H, Brin S, Fraser SA, Korsieporn P, Drudi L, Press JZ (2015) Virtual reality robotic surgery simulation curriculum to teach robotic suturing: a randomized controlled trial. J Robot Surg 9 (3):179-186. doi:10.1007/s11701-015-0513-4
21. Hung AJ, Jayaratna IS, Teruya K, Desai MM, Gill IS, Goh AC (2013) Comparative assessment of three standardized robotic surgery training methods. BJU Int 112 (6):864-871. doi:10.1111/bju.12045
22. Kenney PA, Wszolek MF, Gould JJ, Libertino JA, Moinzadeh A (2009) Face, content, and construct validity of dV-trainer, a novel virtual reality simulator for robotic surgery. Urology 73 (6):1288-1292. doi:10.1016/j.urology.2008.12.044
23. Raza SJ, Froghi S, Chowriappa A, Ahmed K, Field E, Stegemann AP, Rehman S, Sharif M, Shi Y, Wilding GE, Kesavadas T, Kaouk J, Guru KA (2014) Construct validation of the key

components of Fundamental Skills of Robotic Surgery (FSRS) curriculum--a multi-institution prospective study. J Surg Educ 71 (3):316-324. doi:10.1016/j.jsurg.2013.10.006

24. Whittaker G, Aydin A, Raison N, Kum F, Challacombe B, Khan MS, Dasgupta P, Ahmed K (2016) Validation of the RobotiX Mentor Robotic Surgery Simulator. J Endourol 30 (3):338-346. doi:10.1089/end.2015.0620
25. Simbionix. http://simbionix.com/simulators/robotix-mentor/robotix-library-of-modules/. Accessed 02.21. 2017
26. Fisher RA, Dasgupta P, Mottrie A, Volpe A, Khan MS, Challacombe B, Ahmed K (2015) An over-view of robot assisted surgery curricula and the status of their validation. Int J Surg 13:115-123. doi:10.1016/j.ijsu.2014.11.033
27. Bric JD, Lumbard DC, Frelich MJ, Gould JC (2016) Current state of virtual reality simulation in robotic surgery training: a review. Surg Endosc 30 (6):2169-2178. doi:10.1007/s00464-015-4517-y
28. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R (2013) Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. Acad Med 88 (8):1178-1186. doi:10.1097/ACM.0b013e31829a365d
29. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R (2014) What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. Adv Health Sci Educ Theory Pract 19 (2):233-250. doi:10.1007/s10459-013-9458-4
30. Ghaderi I, Manji F, Park YS, Juul D, Ott M, Harris I, Farrell TM (2015) Technical skills assessment toolbox: a review using the unitary framework of validity. Ann Surg 261 (2):251-262. doi:10.1097/sla.0000000000000520
31. Larsen CR, Grantcharov T, Aggarwal R, Tully A, Sorensen JL, Dalsgaard T, Ottesen B (2006) Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. Surg Endosc 20 (9):1460-1466. doi:10.1007/s00464-005-0745-x
32. Tulandi T, Einarsson JI (2014) The use of barbed suture for laparoscopic hysterectomy and myomectomy: a systematic review and meta-analysis. J Minim Invasive Gynecol 21 (2):210-216. doi:10.1016/j.jmig.2013.09.014
33. American Educational Research Association APA, & National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (2014) Standards for educational and psychological testing. American Educational Research Association, Washington, DC

34. Bloch R, Norman G (2012) Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. Med Teach 34 (11):960-992. doi:10.3109/0142159x.2012.703791
35. Downing SM, Yudkowsky R (2009) Assessment in Health Professions Education. Routledge, New York
36. Dyre L, Norgaard LN, Tabor A, Madsen ME, Sorensen JL, Ringsted C, Tolsgaard M (2016) Collecting Validity Evidence for the Assessment of Mastery Learning in Simulation-Based Ultrasound Training. Ultraschall Med 37 (4):386-392. doi:10.1055/s-0041-107976
37. Madsen ME, Konge L, Norgaard LN, Tabor A, Ringsted C, Klemmensen AK, Ottesen B, Tolsgaard MG (2014) Assessment of performance measures and learning curves for use of a virtual-reality ultrasound simulator in transvaginal ultrasound examination. Ultrasound Obstet Gynecol 44 (6):693-699. doi:10.1002/uog.13400
38. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, Smith CD, Satava RM (2005) Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. Ann Surg 241 (2):364-372
39. Connolly M, Seligman J, Kastenmeier A, Goldblatt M, Gould JC (2014) Validation of a virtual reality-based robotic surgical skills curriculum. Surg Endosc 28 (5):1691-1694. doi:10.1007/s00464-013-3373-x
40. Lee JY, Mucksavage P, Kerbl DC, Huynh VB, Etafy M, McDougall EM (2012) Validation study of a virtual reality robotic simulator--role as an assessment tool? J Urol 187 (3):998-1002. doi:10.1016/j.juro.2011.10.160
41. Sethi AS, Peine WJ, Mohammadi Y, Sundaram CP (2009) Validation of a novel virtual reality robotic simulator. J Endourol 23 (3):503-508. doi:10.1089/end.2008.0250
42. Yule S, Flin R, Paterson-Brown S, Maran N (2006) Non-technical skills for surgeons in the operating room: a review of the literature. Surgery 139 (2):140-149. doi:10.1016/j.surg.2005.06.017
43. Lyons C, Goldfarb D, Jones SL, Badhiwala N, Miles B, Link R, Dunkin BJ (2013) Which skills really matter? proving face, content, and construct validity for a commercial robotic simulator. Surg Endosc 27 (6):2020-2030. doi:10.1007/s00464-012-2704-7

Table 1. *Messick's framework of validity in relation to the simulation-based test.*

Source of evidence	Definition	Method
Content	The relationship between the content of the simulation-based test and the construct of interest	Two experienced robotic surgeons assessed if the content of the simulation-based test related to the construct of interest hence if the test could be used the measure procedural competency in robotic surgery for the vaginal cuff closure procedure.
Response process	Elimination or control of potential sources of bias	The main investigator (LH) carried out all the data collection ensuring identical instruction of each participant. The simulator automatically generated all simulator metrics.
Internal structure	Assessment of the reliability of the simulation test	The internal consistency of the simulator metrics for each attempt completed by the participants was assessed.
Relationships with other variables	Assessment of test scores of participants with different levels of experience	Comparison of test scores for the 18 predetermined metrics between the novice surgeons and the experienced surgeons.
Consequences	The intended and unintended effects of the test	Establishment of a pass/fail standard. The objective was to identify a score allowing as few false positives (passed novice surgeons) and as few false negatives (failed experienced surgeons) as possible.

Table 2. *The demographics of the novice and experienced surgeons.*

	Novice surgeons N = 11	**Experienced surgeons N = 11**
Sex, male:female	3:8	5:6
Age, mean (range)	36 (28-48)	49 (42-59)
Dominant hand, n:		
Right	11	9
Left	-	1
Ambidexterity	-	1
Profession, n:		
Resident	4	-
Fellow	5	-
Consultant	2	11
Years of experience in gynaecology, mean (range)	4 (0-16)	16 (7-25)
Experience with robotic simulation, yes:no	0:11	11:0
Experience from robotic surgery, mean (range):		
Years of experience	-	4 (1-8)
Number of vaginal cuff closure procedures	-	102 (40-250)
Total number of procedures	-	143 (40-350)
Experience with laparoscopic simulation, yes:no	9:2	9:2
Experience from laparoscopic surgery, mean (range):		
Years of experience	3 (0-9)	12 (6-20)
Number of vaginal cuff closure procedures	2 (0-18)	56 (2-160)
Total number of procedures	64 (0-330)	323 (50-700)

Table 3. *Overview of test scores in the 18 simulator metrics for the novice and experienced surgeons. Metrics with significant discriminative ability are marked *.*

Objective metrics	**Novice group** Mean score (SD)	**Experienced group** Mean score (SD)	**p-value**
Total time (sec) *	428 (209)	290 (160)	<0.001
Path length, left instrument (mm) *	5739 (2816)	4604 (2124)	0.010
Path length, right instrument (mm) *	4343 (2432)	3311 (2001)	0.009
Distance by camera (mm) *	25 (46)	110 (107)	<0.001
Instrument collisions (n) *	19 (14)	12 (14)	0.004
Total path of instruments travelled out of view (mm) *	408 (423)	802 (695)	<0.001
Number of times instruments are out of view (n) *	14 (15)	35 (27)	<0.001
Total time instruments are out of view (sec) *	16 (22)	36 (37)	<0.001
Clutch usage (n)	2 (4)	2 (2)	0.883
Percentage of accurate needle passages (%)	86 (14)	85 (19)	0.753
Number of unnecessary needle piercing points (n) *	16 (16)	8 (9)	0.002
Number of precise needle passages - entrance points (n)	12 (4)	12 (3)	0.830
Percentage of needle passages at an approach angle of 45 degrees to 90 degrees relative to the tissue surface (%)	75 (20)	74 (20)	0.829
Number of precise needle passages - exit dots (n)	14 (4)	13 (3)	0.360
The total number of entrance and exit points through which the needle has passed (n)	31 (8)	31 (9)	0.951
Time the needle was held outside the visible field (sec)	1 (2)	1(2)	0.942
Number of passages the needle was inserted into the tissue at an approach angle of 45 degrees to 90 degrees relative to the tissue surface (n)	12 (5)	11 (3)	0.613
Number of times the needle was held outside the visible field (n) *	1 (2)	2 (3)	0.007

Table 4. *The composite scores of the novice and experienced surgeons for each test attempt and all six test attempts.*

Test attempt	Novice surgeons' composite score Mean (SD)	Experienced surgeons' composite score Mean (SD)	p-value
Attempt 1	52.8 (21.0)	62.9 (18.4)	0.244
Attempt 2	54.4 (24.9)	74.8 (21.8)	0.055
Attempt 3	62.9 (24.3)	71.6 (23.7)	0.406
Attempt 4	72.3 (14.9)	80.1 (6.0)	0.122
Attempt 5	65.1 (15.6)	84.1 (8.1)	0.002
Attempt 6	70.6 (9.7)	83.1 (6.3)	0.002
All six attempts	63.0 (19.9)	76.1 (17.0)	<0.001

Figure 1. The Robotix Mentor Simulator.

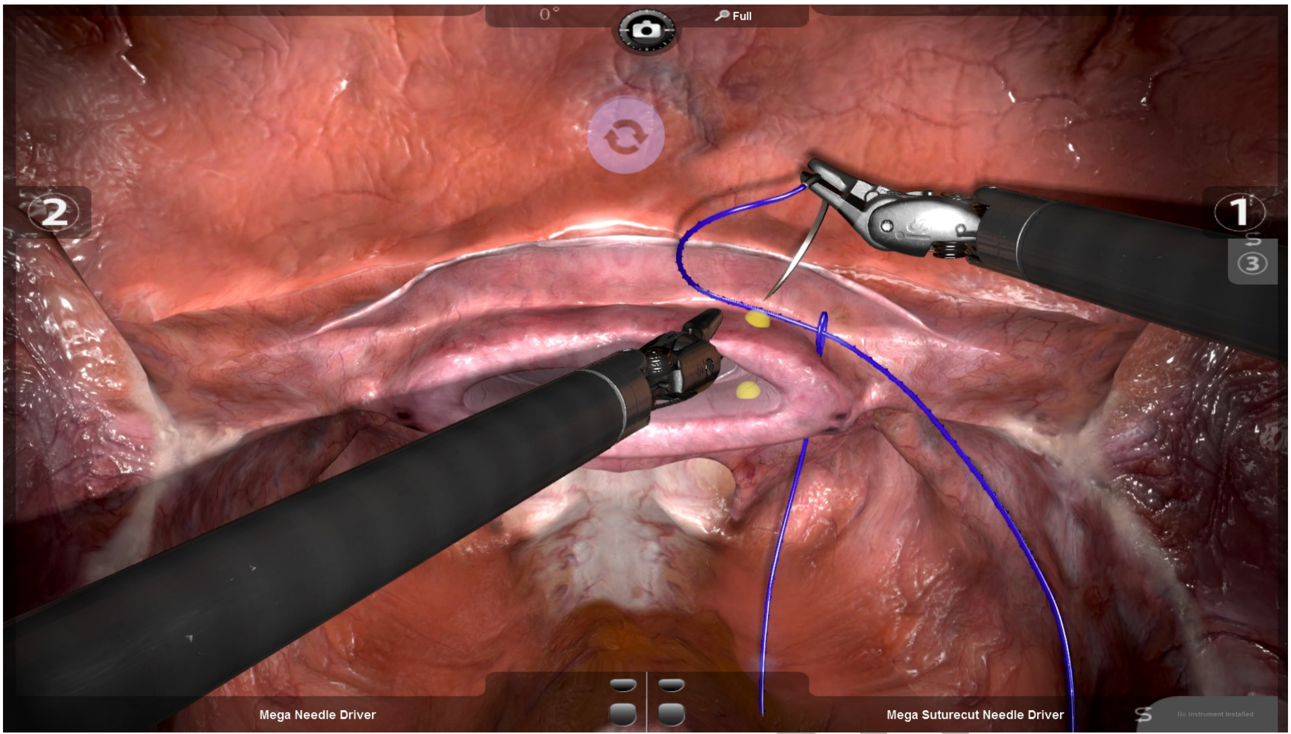

Figure 2. Screenshot from the simulator of the 'Guided Vaginal Cuff Closure with a Barbed Suture' procedure.

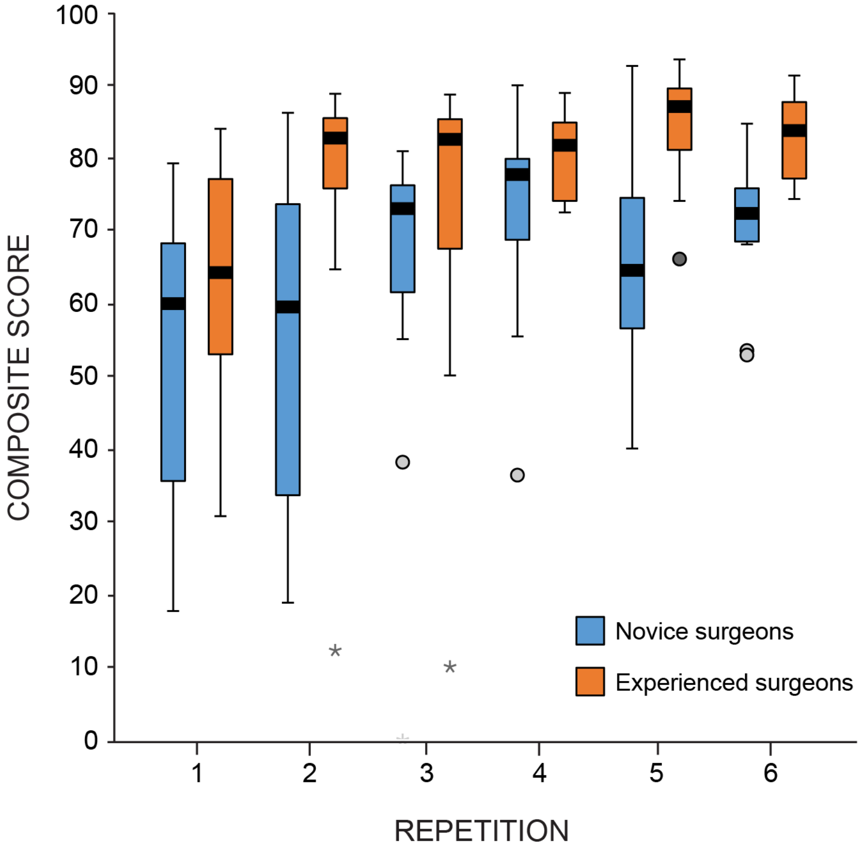

Figure 3. Distribution of the novice and experienced surgeons' composite scores for each repetition. Box-plot showing outliers, minimum, first quartile, median, third quartile and maximum.

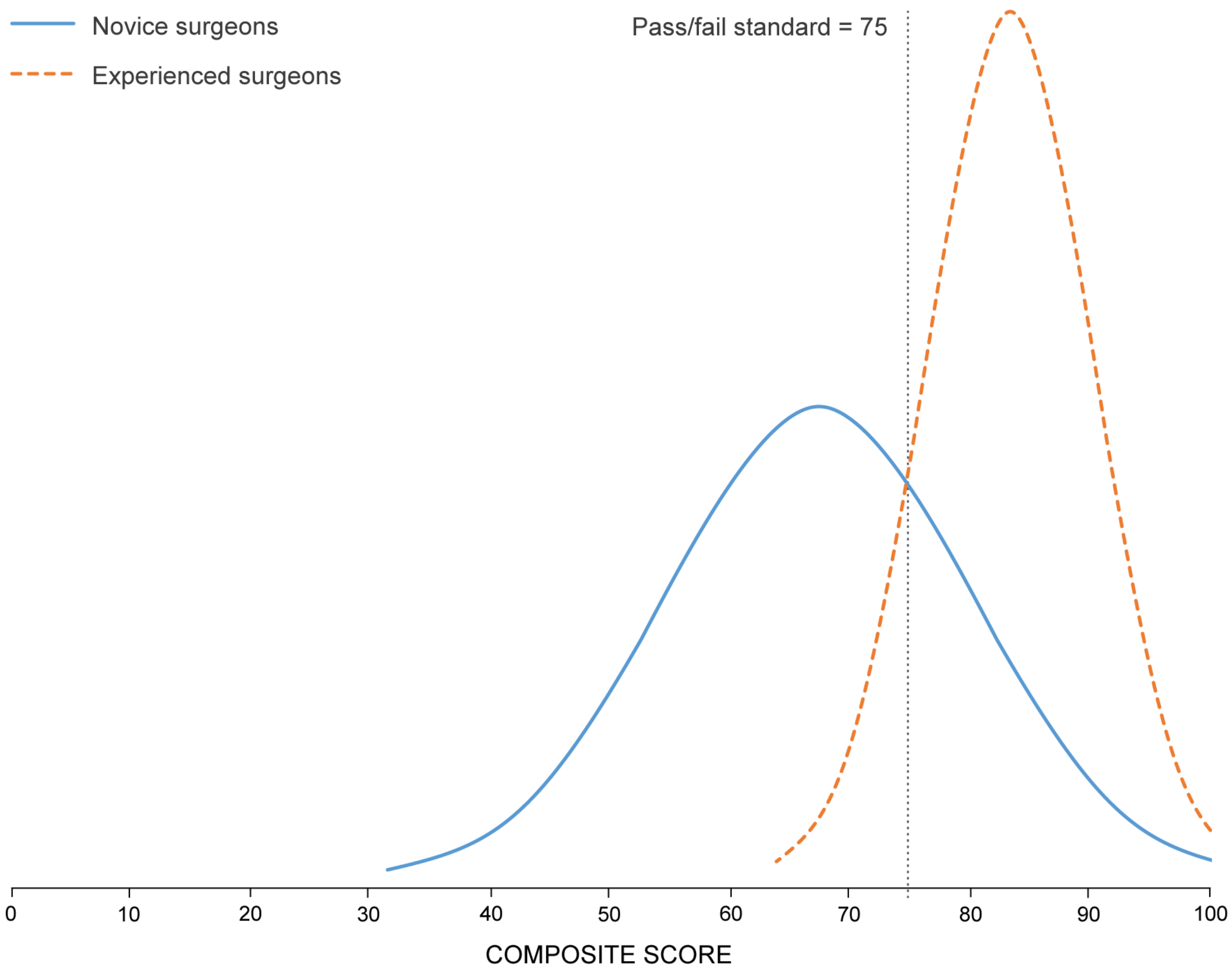

Figure 4. Establishing a pass/fail standard using the contrasting groups' method. The intersection between the distribution of the novice and experienced surgeons' composite scores were used to determine a proficiency level allowing as few passing non-competent novice surgeons and failing as few competent surgeons as possible.