Surg Endosc
DOI 10.1007/s00464-016-5254-6

Using virtual reality simulation to assess competence in video-assisted thoracoscopic surgery (VATS) lobectomy

Katrine Jensen[1,2] · Flemming Bjerrum[3] · Henrik Jessen Hansen[1] · René Horsleben Petersen[1] · Jesper Holst Pedersen[1] · Lars Konge[2]

Received: 30 May 2016 / Accepted: 14 September 2016

Abstract

Background The societies of thoracic surgery are working to incorporate simulation and competency-based assessment into specialty training. One challenge is the development of a simulation-based test, which can be used as an assessment tool. The study objective was to establish validity evidence for a virtual reality simulator test of a video-assisted thoracoscopic surgery (VATS) lobectomy of a right upper lobe.

Methods Participants with varying experience in VATS lobectomy were included. They were familiarized with a virtual reality simulator (LapSim®) and introduced to the steps of the procedure for a VATS right upper lobe lobectomy. The participants performed two VATS lobectomies on the simulator with a 5-min break between attempts. Nineteen pre-defined simulator metrics were recorded.

Results Fifty-three participants from nine different countries were included. High internal consistency was found for the metrics with Cronbach's alpha coefficient for standardized items of 0.91. Significant test–retest reliability was found for 15 of the metrics (p-values <0.05). Significant correlations between the metrics and the participants VATS lobectomy experience were identified for seven metrics (p-values <0.001), and 10 metrics showed significant differences between novices (0 VATS lobectomies performed) and experienced surgeons (>50 VATS lobectomies performed). A pass/fail level defined as approximately one standard deviation from the mean metric scores for experienced surgeons passed none of the novices (0 % false positives) and failed four of the experienced surgeons (29 % false negatives).

Conclusion This study is the first to establish validity evidence for a VATS right upper lobe lobectomy virtual reality simulator test. Several simulator metrics demonstrated significant differences between novices and experienced surgeons and pass/fail criteria for the test were set with acceptable consequences. This test can be used as a first step in assessing thoracic surgery trainees' VATS lobectomy competency.

Keywords Virtual reality simulation · Video-assisted thoracoscopic surgery · VATS · Lobectomy · Validity

Simulation-based training can teach the technical aspects of a procedure and accelerate the learning curve, since the surgical trainee can practice technical skills in a highly deliberate and time-efficient manner to achieve surgical competence [1–3]. Skills training in a simulated, risk-free environment allows for the exploration of possible outcomes and reduces the time required to train surgeons in complex procedures [4]. With simulation-based training, the trainee can adapt to the continually increasing complexity and variety of procedures being performed.

With the implementation of simulation models for each procedure, thoracic surgeons will have the opportunity to practice a procedure to proficiency before performing it on

✉ Katrine Jensen
katrine.jensen@rh.regionh.dk

1 Department of Cardiothoracic Surgery, Sect. 2152, University Hospital of Copenhagen, Rigshospitalet, Copenhagen, Denmark

2 Copenhagen Academy for Medical Education and Simulation (CAMES), Section 5404, University of Copenhagen and Capital Region, Rigshospitalet, Blegdamsvej 9, 2100 København Ø Copenhagen, Denmark

3 JMC Simulation Unit, The Juliane Marie Centre, Section 4704, University Hospital of Copenhagen, Rigshospitalet, Copenhagen, Denmark

a patient. Training using virtual reality simulators has been shown to improve performance in the operating room [5, 6]; however, there has been limited development of models and virtual reality programs for thoracoscopic procedures [3, 7]. Although laparoscopic surgery has experienced the rapid growth of simulation models and virtual reality systems for the training and assessment of technical skills [8], these models have shown not to be transferrable to thoracoscopic procedures, and thus, there is need for further procedure-specific development [9, 10].

The societies of thoracic surgery have recognized the need for change in the methods used to educate future thoracic surgeons and are working towards using simulation-based training and assessment [11–19]. The implementation of competency-based curricula in thoracic surgery requires the teaching of technical skills, with subsequent assessment to confirm proficiency, prior to the transition from the simulated environment to the operating room [14, 15]. Defining competencies in VATS lobectomy and developing tests which accurately assess surgical competence is, therefore, of great importance for surgical training, maintenance of surgical skills, evaluation and certification of surgical competence, and the development of standardized surgical curricula [20–22].

The increased desire to use simulators as an assessment tool for the evaluation of competence comes with a need for solid evidence of validity for the test [8, 23, 24], and thoracic surgery faces challenges in the development and incorporation of simulation into resident training [11–13]. The use of simulators as assessment tools has proven to be challenging [8, 25, 26]: In order for a simulator to be used to assess surgical competence, it must be evaluated vigorously to determine its sources of evidence for validity [27].

The objective of the study was to establish validity evidence for a simulated virtual reality test of a VATS lobectomy of a right upper lobe.

Materials and methods

Sample size calculation

A pilot test with four novices and four experienced surgeons was conducted to estimate the sample size, based on their mean procedure time for two performed tests. Approximate procedure time for novices (1650 s) and experienced surgeons (1250 s) and a standard deviation of 375 s with alpha set at 95 %, and a power of 0.8, 14 novices and 14 experienced thoracic surgeons were required to show a significant difference between the groups.

Inclusion

A convenience sample of medical students with no surgical experience, trainees in thoracic surgery with varying experience (1–50 VATS lobectomy procedures), and experienced thoracic surgeons in VATS lobectomy with experience from 50 procedures or more were included between September 2015 and January 2016. The novices were recruited from Danish medical students, and the trainees and experienced surgeons were recruited from various thoracic centres from all over the world as part of a training program at the University Hospital of Copenhagen, Rigshospitalet, Denmark. Participants were excluded if they had experience in using virtual reality simulation. All participants were assigned a unique trial identification number, and data were kept anonymous for all members of the study team except for the main investigator (KJ).

Testing

A simulated virtual reality module was developed for a video-assisted thoracoscopic surgery lobectomy of a right upper lobe based on the standardized anterior approach. The first version was presented in June 2014 at the European Society for thoracic surgery conference [7, 28]. The test was subsequently revised from feedback obtained at the conference, with corrections made to the anatomy and further development of the built-in instruments and metrics. New handles for performing VATS, which are opened and closed with the lower branch instead of the upper as the normal laparoscopic handles, were manufactured especially for this scenario. Four identical virtual reality simulators (LapSim®, Surgical Science, Gothenburg, Sweden) at the Copenhagen Centre for Medical Education and Simulation (CAMES), Copenhagen, Denmark [29] were used for testing (Fig. 1). Each participant started by reviewing the built-in instructions for the procedural steps for the simulated right upper lobe VATS lobectomy. The participants were then instructed on how to use the simulator by KJ, while having the opportunity to familiarize themselves with the simulator and the VATS scenario for exactly 15 min. Participants were allowed to ask the instructor questions about how to operate the simulator, such as how to change instruments, but not about how to complete the test. They then performed a VATS lobectomy of the right upper lobe on the simulator two times in a row with a 5-min break in between attempts (Fig. 2). The instructor was present throughout the procedure. The simulator automatically recorded all metrics.

Simulator metrics

The simulator had different metrics built-in for the test to measure the surgeon's performance (Table 1). The metrics

Fig. 1 LapSim® system (www.surgical-science.com)

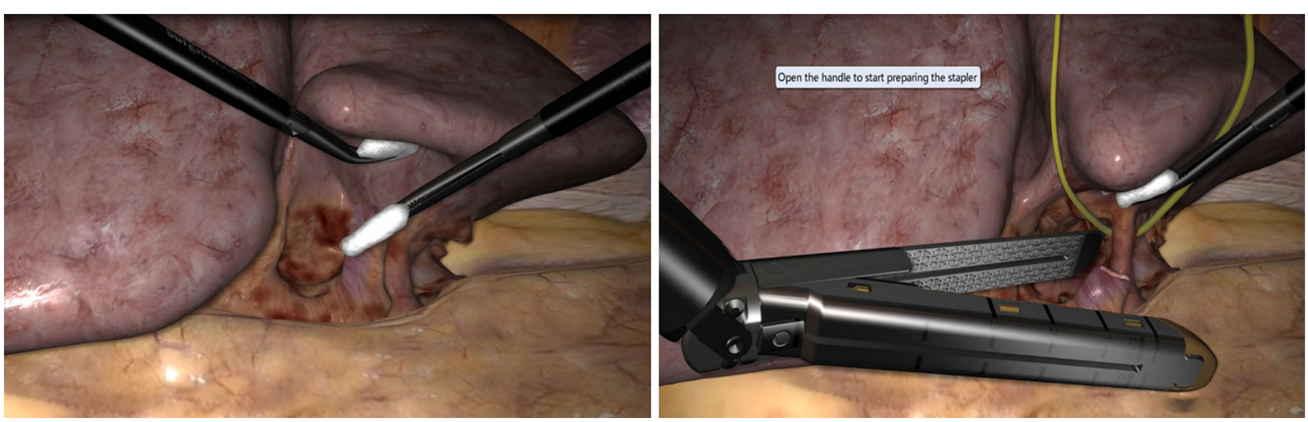

Fig. 2 Screenshots from the virtual reality VATS lobectomy test: dissection of the hilum and stapling of the main artery to the right upper lobe

include standard metrics such as total procedure time, bleeding, and hand movements, as well as procedure-specific metrics [7]. The basic metrics were chosen based on metrics with evidence of validity from research in virtual reality simulation in laparoscopic surgery [30–32], and the new metrics were pre-defined by the authors in an attempt to best evaluate skills necessary for performing a VATS lobectomy.

Table 1 The 19 built-in simulator metrics and used abbreviations for either first, second, or mean metric values

Total procedure time (s)	TIMEtot (1/2/mean)
Right instrument pathway length (m)	RIGHTlength (1/2/mean)
Left instrument pathway length (m)	LEFTlength (1/2/mean)
Right instrument angle (degrees cumulated)	RIGTHang (1/2/mean)
Left instrument angle (degrees cumulated)	LEFTang (1/2/mean)
Number of tool switches	TOOLSwitches (1/2/mean)
Blood loss (mL)	BLOODloss (1/2/mean)
Stretch damage on vessels to middle lobe	STRETCHvesmid (1/2/mean)
Stretch damage on vessels to superior lobe	STRETCHvessup (1/2/mean)
Stretch damage on bronchus to middle lobe	STRETCHbronchmid (1/2/mean)
Stretch damage on bronchus to superior lobe	STRETCHbronchsup (1/2/mean)
Stretch damage on middle lobe	STRETCHlobmid (1/2/mean)
Stretch damage on superior lobe	STRETCHlobsup (1/2/mean)
Number of vessels severed	VESSELSsevered (1/2/mean)
Number of times stapled on lobes	STAPLEDonlobes (1/2/mean)
Number of structures stapled without removing yellow band	STAPLEDwyellow (1/2/mean)
Number of times bronchus is crushed before stapling	BRONCHUSCrushed (1/2/mean)
Number of incorrect staple reload used for bronchus	STAPLERincbronch (1/2/mean)
Number of incorrect staple reload used for vessels	STAPLERincves (1/2/mean)

Outcome measures and data analysis

Validity evidence was described using Messick's unitary framework [33]:

Content: Ensured in a previously published study [7].

Response process: Standardized data collection was ensured with the use of an instructional element on the simulator prior to the attempt and identical data collection procedures in each case.

Internal structure: Internal consistency of the metrics was assessed for each of the attempts undertaken by the participants by calculating Cronbach's alpha. For test/retest reliability Spearman's rho was calculated for all simulator scores between the first and second try.

Relation to other variables: Correlations between simulator metrics and the surgeons' VATS lobectomy experience were analysed using Pearson's r. Metrics with significant differences between novices, intermediates, and experienced surgeons were identified by testing them one by one between the groups using independent t test and ANOVA test with post hoc tests (Bonferroni).

Consequences of testing: The pass/fail level on the simulator was chosen for each metric as the experienced surgeons' mean score plus one standard deviation. The objective was to identify a score that ensures as few false positives (passing novices) and false negatives (failing experienced surgeons) as possible. The intermediate group's performance was evaluated based on this pass/fail threshold.

IBM SPSS statistics version 23 (IBM, NY, USA) was used for the data analysis.

Results

A total of 53 participants from nine different countries with varying experience with VATS were included. Mean group characteristics are shown in Table 2.

Response process: One person (KJ) carried out all data collection in a standardized way by using the instructional element on the simulator before the participants' testing. The simulator automatically recorded all metrics.

Internal structure: High internal consistency was found for the 19 metrics for the 53 participants' first and second attempts and for the calculated mean score for the first and second attempt with Cronbach's alpha coefficients for standardized items of 0.91, 0.90, and 0.91, respectively. A linear relationship between metrics for the first and second attempt was found, and significant test/retest reliability was found for 15 of the metrics as shown in Table 3, where values for Spearman's rho (ranked data) for each metric are shown and significant results are highlighted.

Relation to other variables: Seven metrics (the procedure time, the moved length and the total angulation of the left instrument, number of severed vessels, stretching of the bronchus to the middle lobe, blood loss, and number of times a vessel or bronchus was stapled without removing the rubber band) correlated with the participant's experience with $p < 0.001$. Ten metrics demonstrated significant differences between the groups (Table 4). When performing a one-way ANOVA with Bonferroni's post hoc analysis for the groups, the intermediate group performed significantly ($p < 0.05$) better than the novices with respect to stretch damage to the bronchus for the upper and middle

Table 2 Group descriptives

Group		Novices	Intermediates	Experienced
Numbers of participants	#	17	22	14
Age	Years	24	39	43
Approximate numbers of performed VATS procedures (VATStotal)	#	0	163	616
Approximate numbers of performed VATS lobectomies (VATSlob)	#	0	10	126
Approximate numbers of performed lobectomies by thoracothomy (THORlob)	#	0	144	348
Self-assessed knowledge of VATS lobectomies (theoretical)	1 = Very limited knowledge - 5 = Expert knowledge	1	3	4
Self-assessed performance in VATS lobectomies (practical)	1 = Novice - 5 = Experienced	1	2	4

Numbers are given as means for the age and numbers of procedures, and as medians for self-assessed scores

Table 3 Test–retest for all 19 metrics

	Spearman's Rho	*p*-value
TIMEtot1 TIMEtot2	0.328	0.016
RIGHTlength1 RIGHTlength2	0.486	<0.001
LEFTlength1 LEFTlength2	0.252	0.069
RIGHTang1 RIGHTang2	0.491	<0.001
LEFTang1 LEFTang2	0.275	0.049
ToolSwitches1 ToolSwitches2	0.346	0.011
VesselsSevered1 VesselsSevered2	0.540	<0.001
STRETCHvessup1 STRETCHvessup2	0.227	0.102
STRETCHvesmid1 STRETCHvesmid2	0.445	<0.001
STRETCHlobsup1 STRETCHlobsup2	0.373	<0.001
STRETCHlobmid1 STRETCHlobmid2	0.365	<0.001
STRETCHbronchsup1 STRETCHbronchsup2	0.402	<0.001
STRETCHbronchmid1 STRETCHbronchmid2	0.288	0.037
BLOODloss1 BLOODloss2	0.580	<0.001
STAPLEDwyellow1 STAPLEDwyellow2	0.295	0.032
BronchusCrushed1 BronchusCrushed2	0.293	0.033
STAPLEDonlobes1 STAPLEDonlobes2	0.173	0.214
STAPLERincves1 STAPLERincves2	−0.083	0.552
STAPLERincbronch1 STAPLERincbronch2	0.332	0.015

lobe, numbers of vessels severed, and blood loss. When intermediates and experienced surgeons were compared, only severed vessels and blood loss were significantly different.

Consequences of testing: In order to establish a pass/-fail level, the ten metrics with significant differences between novices and experienced surgeons were chosen and metrics for the angulation of the right instrument were also included. The pass/fail level for each metric was determined as the experienced surgeons' mean metric score plus approximately one standard deviation and is shown for the selected metrics in Table 4. If all of these metrics are equally weighted and have to be passed simultaneously, then none of the novices in our sample would pass (0 % false positives) and four of the experienced surgeons would fail (29 % false negatives).Two (9 %) from the intermediate group would pass all 11 metrics simultaneously, and the rest (91 %) would fail for one or more metrics.

Discussion

This study has established validity evidence for a virtual reality test for a right upper lobe VATS lobectomy. Our results revealed a Cronbach's alpha of 0.91 and statistically

Table 4 Metrics with significant differences between novices and experienced participants and the pass/fail level for significant metrics

Novices versus experienced	Numbers	Mean	Standard deviation	*p*-value	Pass/fail level
Meantime	17	1708.0	458.8		
	14	1223.2	301.3	0.001	1500
Meanrightlength	17	30.9	14.8		
	14	21.9	6.2	0.034	28
Meanleftlength	17	32.1	12.2		
	14	20.5	4.5	0.002	25
Meanrightang	17	7334.1	3710.8		
	14	6117.7	1648.3	0.237	7700
Meanleftang	17	6948.4	2729.6		
	14	4710.9	1071.1	0.005	5800
Meantoolswitch	17	39.9	11.0		
	14	41.5	9.6	0.657	
Meanvessev	17	3.5	1.0		
	14	1.5	1.2	<0.001	2.5
Meanstretchvessup	17	29.7	8.4		
	14	29.6	8.8	0.961	
Meanstretchvesmid	17	31.1	15.3		
	14	21.9	10.0	0.061	
Meanstretchlobsup	17	57.3	17.3		
	14	49.3	14.5	0.180	
Meanstretchlobmid	17	46.7	17.0		
	14	34.9	12.5	0.038	45
Meanstretchbronchsup	17	50.0	22.3		
	14	33.7	12.2	0.016	45
Meanstretchbronchmid	17	41.6	21.3		
	14	21.9	5.9	0.002	28
Meanbloodloss	17	365.7	104.6		
	14	158.1	118.5	<0.001	250
Meanstapledwyellow	17	0.6	0.4		
	14	0.1	0.3	0.003	0
Meanbronchcrush	17	4.3	1.6		
	14	3.8	1.8	0.356	
Meanstapledonlob	17	2.5	1.8		
	14	1.8	1.6	0.251	
Meanincves	17	0.1	0.2		
	14	0.1	0.2	0.483	
Meanincbronch	17	0.4	0.4		
	14	0.2	0.3	0.052	

significant test–retest reliability. For simulators assessing technical or skill competency, the cost or possible harmful effect of misclassification is high; therefore, an internal consistency of $\geq$0.8 is considered reasonable [34, 35]. Ten of the 19 metrics showed a significant difference between novices and experienced surgeons. Metrics are often found to be more applicable to simple/basic tasks, rather than to an entire complex procedure, such as VATS lobectomy which is associated with a long procedure time [36]. As this study also discovered, Woodrum et al. [37] found that only some metrics on the LapSim could discriminate between novices and experienced, and Rivard et al. [36] found that more metrics from simple tasks than complex tasks on the LapVR showed a difference between groups.

The performance of experienced surgeons is often used to develop performance metrics and determine a standard or pass/fail level [38, 39]. A pass/fail level for 11 of the metrics with discriminatory abilities was set using the experienced surgeons' mean score plus approximately one standard deviation, yielding appropriate consequences. The participants in this study performed two procedures/tests, and it is known from other studies that a learning curve exists on the simulator [40, 41]. This factor may explain why four experienced surgeons and nearly all intermediate participants failed, and it is likely that more of them would pass once they were more familiar with how to operate the simulator.

Virtual reality simulators generate metrics that are available for immediate feedback without the presence of a supervisor, and they are independent of an assessor's opinion [39]. When using proficiency-based training with pre-defined relevant goals, trainees are able to monitor their progress towards these goals. They can easily see which of their metric scores lie below the desired score and, thus, are given direct feedback on where they need to improve. The advantage of the simulator used for this study is that performance curves for all users are automatically stored on the device and individual progress can be easily reviewed. In addition, weights can be given to each metric and a total score calculated. This score provides the trainees with a clear parameter for how close they are to simultaneously passing the chosen metrics.

Metrics may not all be equally relevant for trainees at different stages of their education/training. In the initial phase of learning VATS lobectomy, procedure time and how much one manipulates the instruments can in the initial phase of learning VATS lobectomy considered less important than severed vessels, blood loss, stretch damage to structures, and choosing the correct stapler. However, these metrics are very important as the trainee advances his/her skills [37]. When one looks at the metrics with validity evidence from this study, they make good overall clinical sense: In order to pass this test you have to finish the procedure in reasonable time, so one must be economic in their movements, one must not stretch the adjacent middle lobe and its structures since it is to remain undamaged in the chest cavity, one must not sever vessels and cause major bleeding, and one must not staple on the yellow band used to sling the structures, as this would be difficult to repair if stapled to the structures.

Several studies have shown transfer of skills from a virtual reality simulator to real surgical procedure [5], and simulators have the potential to standardize the assessment of technical skills. They could, therefore, be used for resident selection and advancement, as well as for certification and credentialing. Implementation of national and international simulator-based training standards will allow for standardized objective assessment of technical skill for all trainees. The McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) is a system that relies on a specific set of metrics to score performance and was the basis for the technical skill component of the Fundamentals of Laparoscopic Surgery program (FLS). Vassiliou et al. [34] found a very high internal consistency and test–retest reliability for these metrics, while Zendejas et al. [42] from their review only found that validity evidence for relations with other variables and acceptable inter-rater reliability for the FLS as an assessment tool. These studies emphasize the importance of rigorously evaluating an assessment tool, before it is used for high-stakes evaluation. The disadvantage of using simulation-based procedural tests is that this approach is time-consuming. Significant effort and time investment are necessary to develop a useful training model and gather sufficient validity evidence, since experienced surgeons must be identified, and validity evidence established for metrics, assessment tools and tests.

The strength of this study lies in the objectivity of the design with the use of metrics that are automatically stored and, therefore, facilitate replication and retesting of the research. The sample size is acceptable and the study participants are from different countries, thereby, increasing the generalization of the results. A limitation is that the participants only performed two procedures each, and their performance curve was not examined. The complexity of the VATS lobectomy scenario resulted in different operation techniques by surgeons, making it harder to compare metrics between surgeons, as seen in other studies [25, 43, 44]. In addition, the right instrument angulation did not demonstrate a statistically significant difference between novices and experienced surgeons, but we included it in the pass/fail criteria since it is "paired" with the right path length and the left instrument movement metrics. The way the groups are divided poses another limitation. Since having done one or 49 VATS lobectomies both qualify as an intermediate experienced surgeon, one could argue that the groups need to be further stratified to explore a statistically significant differences. The group divisions were based on learning curves from the published literature [45–48], and the groups were defined per protocol. The sample size was not considered large enough to divide the intermediate group into smaller groups i.e. 1–25 VATS lobectomies performed and 26–49 VATS lobectomies performed. Only severed vessels and blood loss were significantly different when comparing the intermediate and experienced surgeons and this could result from the variable amount of experience in the intermediate group. A fourth limitation is that the test cannot simulate lymph node dissection and primarily focuses on technical skills. Non-technical skills are equally important, but must be assessed using other instruments [49]. Only the right upper lobe can be simulated at present, but with further development and research, there is the potential to simulate removal of all five lung lobes along with lymph nodes. This development will allow tests that are needed for the staging of lung cancer [50] to be incorporated into the simulator.

Traditional medical education with learning from clinical experience and apprenticeship is not an effective approach to achieve key competencies. The newest guidelines for other technical procedures now recommend that trainees should follow a structured training curriculum, including simulation-based training, and that they must demonstrate basic competency using available validated assessment tools prior to performing procedures independently [51]. Simulation-based mastery learning together with new technology, validated assessment tools, standardized tests, and standard settings will help achieve better and faster competency acquisition and, therefore, improved quality of care and patient safety [7, 14, 50–57].

Conclusion

This study succeeded in providing validity evidence for a VATS right upper lobe lobectomy virtual reality test. Metrics showed significant differences between novices and experienced surgeons and pass/fail criteria for the test were set with acceptable consequences. This test can be used as a first step in assessing thoracic surgery trainees' VATS lobectomy competency.

Acknowledgments We would like to thank medical student Julia Dagnæs for her invaluable help with managing the simulators during the period of data collection.

Compliance with ethical standards

Disclosures Katrine Jensen's Ph.D.project is partly funded by a grant from The Danish Cancer Society (Kræftens Bekæmpelse, "Knæk Cancer"). Henrik Jessen Hansen and René Horsleben Petersen are at the Speakers Bureau of Covidien. Flemming Bjerrum, Jesper Holst Pedersen and Lars Konge have no conflicts of interest or financial ties to disclose.

References

1. McGaghie WC (2008) Research opportunities in simulation-based medical education using deliberate practice. Acad Emerg Med 15:995–1001. doi:10.1111/j.1553-2712.2008.00246.x
2. Price J, Naik V, Boodhwani M, Brandys T, Hendry P, Lam BBK (2011) A randomized evaluation of simulation training on performance of vascular anastomosis on a high-fidelity in vivo model: the role of deliberate practice. J Thorac Cardiovasc Surg 142:496–503. doi:10.1016/j.jtcvs.2011.05.015
3. Marshall MB (2012) Simulation for technical skills. J Thorac Cardiovasc Surg 144:S43–S47. doi:10.1016/j.jtcvs.2012.06.004
4. Crochet P, Aggarwal R, Dubb SS, Ziprin P, Rajaretnam N, Grantcharov T, Ericsson KA, Darzi A (2011) Deliberate practice on a virtual reality laparoscopic simulator enhances the quality of surgical technical skills. Ann Surg 253:1216–1222. doi:10.1097/SLA.0b013e3182197016
5. Seymour NE (2008) VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. World J Surg 32:182–188. doi:10.1007/s00268-007-9307-9
6. Schijven MP, Jakimowicz JJ, Broeders IAMJ, Tseng LNL (2005) The Eindhoven laparoscopic cholecystectomy training course—improving operating room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum. Surg Endosc Other Interv Tech 19:1220–1226. doi:10.1007/s00464-004-2240-1
7. Jensen K, Bjerrum F, Hansen HJ, Petersen RH, Pedersen JH, Konge L (2015) A new possibility in thoracoscopic virtual reality simulation training: development and testing of a novel virtual reality simulator for video-assisted thoracoscopic surgery lobectomy. Interact Cardiovasc Thorac Surg 21:420–426. doi:10.1093/icvts/ivv183
8. Seymour NE, Røtnes JS (2006) Challenges to the development of complex virtual reality surgical simulations. Surg Endosc Other Interv Tech 20:1774–1777. doi:10.1007/s00464-006-0107-3
9. Jensen K, Ringsted C, Hansen HJ, Petersen RH, Konge L (2014) Simulation-based training for thoracoscopic lobectomy: a randomized controlled trial: virtual-reality versus black-box simulation. Surg Endosc Other Interv Tech 28:1821–1829. doi:10.1007/s00464-013-3392-7
10. Bjerrum F, Sorensen JL, Konge L, Rosthøj S, Lindschou J, Ottesen B, Strandbygaard J (2016) Randomized clinical trial to examine procedure-to-procedure transfer in laparoscopic simulator training. Br J Surg 103:44–50. doi:10.1002/bjs.9966
11. Carpenter AJ, Yang SC, Uhlig PN, Colson YL (2008) Envisioning simulation in the future of thoracic surgical education. J Thorac Cardiovasc Surg 135:477–484. doi:10.1016/j.jtcvs.2007.12.005
12. Verrier ED (2011) Joint Council on Thoracic Surgical Education: an investment in our future. J Thorac Cardiovasc Surg 141:318–321. doi:10.1016/j.jtcvs.2010.10.013
13. Moon MR (2014) Technical skills assessment in thoracic surgery education: we won't get fooled again. J Thorac Cardiovasc Surg 148:2497–2498. doi:10.1016/j.jtcvs.2014.09.057
14. Lodge D, Grantcharov T (2011) Training and assessment of technical skills and competency in cardiac surgery. Eur J Cardiothorac Surg 39:287–293. doi:10.1016/j.ejcts.2010.06.035
15. Aggarwal R, Darzi A (2011) Simulation to enhance patient safety: Why aren't we there yet? Chest 140:854–858. doi:10.1378/chest.11-0728
16. Stefanidis D, Sevdalis N, Paige J, Zevin B, Aggarwal R, Grantcharov T, Jones DB (2015) Simulation in surgery: What's needed next? Ann Surg 261:846–853. doi:10.1097/SLA.0000000000000826
17. Baumgartner WA (2003) Cardiothoracic surgery: a specialty in transition—Good to great? Ann Thorac Surg 75:1685–1692. doi:10.1016/S0003-4975(03)00538-1
18. Chitwood WR, Spray TL, Feins RH, Mack MJ (2008) Mission critical: thoracic surgery education reform. Ann Thorac Surg 86:1061–1062. doi:10.1016/j.athoracsur.2008.08.047
19. Crawford FA (2003) Presidential address: thoracic surgery education—responding to a changing environment. J Thorac Cardiovasc Surg 126:1235–1242. doi:10.1016/S0022-5223(03)00814-6
20. Scott DJ, Bergen PC, Rege RV, Laycock R, Tesfay ST, Valentine RJ, Euhus DM, Jeyarajah DR, Thompson WM, Jones DB (2000) Laparoscopic training on bench models: better and more cost effective than operating room experience? J Am Coll Surg 191:272–283. doi:10.1016/S1072-7515(00)00339-2
21. Torkington J, Smith SGT, Rees BI, Darzi A (2001) Skill transfer from virtual reality to a real laparoscopic task. Surg Endosc 15:1076–1079. doi:10.1007/s004640000233
22. Paisley AM, Baldwin PJ, Paterson-Brown S (2001) Validity of surgical simulation for the assessment of operative skill. Br J Surg 88:1525–1532. doi:10.1046/j.0007-1323.2001.01880.x
23. Reznick RK, MacRae H (2006) Teaching surgical skills—changes in the wind. N Engl J Med 355:2664–2669. doi:10.1016/S0084-392X(08)70199-0
24. Lee R, Enter D, Lou X, Feins RH, Hicks GL, Gasparri M, Takayama H, Young JN, Calhoon JH, Crawford FA, Mokadam

NA, Fann JI (2013) The joint council on thoracic surgery education coronary artery assessment tool has high interrater reliability. Ann Thorac Surg 95:2064–2070. doi:10.1016/j.athoracsur.2012.10.090

25. Van Bruwaene S, Schijven MP, Miserez M (2014) Assessment of procedural skills using virtual simulation remains a challenge. J Surg Educ 71:654–661. doi:10.1016/j.jsurg.2014.01.005
26. Seagull FJ, Rooney DM (2014) Filling a void: developing a standard subjective assessment tool for surgical simulation through focused review of current practices. Surgery 156:718–722. doi:10.1016/j.surg.2014.04.048
27. Van Hove PD, Tuijthof GJM, Verdaasdonk EGG, Stassen LPS, Dankelman J (2010) Objective assessment of technical surgical skills. Br J Surg 97:972–987. doi:10.1002/bjs.7115
28. Hansen HJ, Petersen RH, Christensen M (2011) Video-assisted thoracoscopic surgery (VATS) lobectomy using a standardized anterior approach. Surg Endosc Other Interv Tech 25:1263–1269. doi:10.1007/s00464-010-1355-9
29. Konge L, Ringsted C, Bjerrum F, Tolsgaard MG, Bitsch M, Sørensen JL, Schroeder TV (2015) The simulation centre at Rigshospitalet, Copenhagen, Denmark. J Surg Educ 72:362–365. doi:10.1016/j.jsurg.2014.11.012
30. Larsen CR, Grantcharov T, Aggarwal R, Tully A, Sørensen JL, Dalsgaard T, Ottesen B (2006) Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. Surg Endosc 20:1460–1466. doi:10.1007/s00464-005-0745-x
31. Van Dongen KW, Tournoij E, Van Der Zee DC, Schijven MP, Broeders IAMJ (2007) Construct validity of the LapSim: Can the LapSim virtual reality simulator distinguish between novices and experts? Surg Endosc Other Interv Tech 21:1413–1417. doi:10.1007/s00464-006-9188-2
32. Eriksen JR, Grantcharov T (2005) Objective assessment of laparoscopic skills using a virtual reality stimulator. Surg Endosc 19:1216–1219. doi:10.1007/s00464-004-2154-y
33. 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 261:251–262. doi:10.1097/SLA.0000000000000520
34. Vassiliou MC, Ghitulescu GA, Feldman LS, Stanbridge D, Leffondré K, Sigman HH, Fried GM (2006) The MISTELS program to measure technical skill in laparoscopic surgery: evidence for reliability. Surg Endosc Other Interv Tech 20:744–747. doi:10.1007/s00464-005-3008-y
35. Downing SM (2003) Validity: on the meaningful interpretation of assessment data. Med Educ 37:830–837. doi:10.1046/j.1365-2923.2003.01594.x
36. Rivard JD, Vergis AS, Unger BJ, Hardy KM, Andrew CG, Gillman LM, Park J (2014) Construct validity of individual and summary performance metrics associated with a computer-based laparoscopic simulator. Surg Endosc Other Interv Tech 28:1921–1928. doi:10.1007/s00464-013-3414-5
37. Woodrum DT, Andreatta PB, Yellamanchilli RK, Feryus L, Gauger PG, Minter RM (2006) Construct validity of the LapSim laparoscopic surgical simulator. Am J Surg 191:28–32. doi:10.1016/j.amjsurg.2005.10.018
38. Fried GM, Feldman LS (2008) Objective assessment of technical performance. World J Surg 32:156–160. doi:10.1007/s00268-007-9143-y
39. MacRae HM (2014) Technical skills assessment: time to take it seriously. Dis Colon Rectum 57:141–142. doi:10.1097/DCR.0000000000000032
40. Hogle NJ, Briggs WM, Fowler DL (2007) Documenting a learning curve and test–retest reliability of two tasks on a virtual reality training simulator in laparoscopic surgery. J Surg Educ 64:424–430. doi:10.1016/j.jsurg.2007.08.007
41. Loukas C, Nikiteas N, Schizas D, Lahanas V, Georgiou E (2012) A head-to-head comparison between virtual reality and physical reality simulation training for basic skills acquisition. Surg Endosc 26:2550–2558. doi:10.1007/s00464-012-2230-7
42. Zendejas B, Ruparel RK, Cook DA (2015) Validity evidence for the fundamentals of laparoscopic surgery (FLS) program as an assessment tool: a systematic review. Surg Endosc Other Interv Tech 30:512–520. doi:10.1007/s00464-015-4233-7
43. Våpenstad C, Buzink SN (2013) Procedural virtual reality simulation in minimally invasive surgery. Surg Endosc 27:364–377. doi:10.1007/s00464-012-2503-1
44. Rocco G, Internullo E, Cassivi SD, Van Raemdonck D, Ferguson MK (2008) The variability of practice in minimally invasive thoracic surgery for pulmonary resections. Thorac Surg Clin 18:235–247. doi:10.1016/j.thorsurg.2008.06.002
45. Petersen RH, Hansen HJ (2012) Learning curve associated with VATS lobectomy. Ann Cardiothorac Surg 1:47–50. doi:10.3978/478
46. Li X, Wang J, Ferguson MK (2014) Competence versus mastery: the time course for developing proficiency in video-assisted thoracoscopic lobectomy. J Thorac Cardiovasc Surg 147:1150–1154. doi:10.1016/j.jtcvs.2013.11.036
47. Ferguson J, Walker W (2006) Developing a VATS lobectomy programme—Can VATS lobectomy be taught? Eur J Cardiothorac Surg 29:806–809. doi:10.1016/j.ejcts.2006.02.012
48. McKenna RJ (2008) Complications and learning curves for video-assisted thoracic surgery lobectomy. Thorac Surg Clin 18:275–280. doi:10.1016/j.thorsurg.2008.04.004
49. Gjeraa K, Spanager L, Konge L, Petersen RH, Ostergaard D (2016) Non-technical skills in minimally invasive surgery teams: a systematic review. Surg Endosc. doi:10.1007/s00464-016-4890-1
50. Savran MM, Hansen HJ, Petersen RH, Walker W, Schmid T, Bojsen SR, Konge L (2015) Development and validation of a theoretical test of proficiency for video-assisted thoracoscopic surgery (VATS) lobectomy. Surg Endosc 29:2598–2604. doi:10.1007/s00464-014-3975-y
51. Vilmann P, Frost Clementsen P, Colella S, Siemsen M, De Leyn P, Dumonceau JM, Herth FJ, Larghi A, Vazquez-Sequeiros E, Hassan C, Crombag L, Korevaar DA, Konge L, Annema JT (2015) Combined endobronchial and esophageal endosonography for the diagnosis and staging of lung cancer: European Society of gastrointestinal endoscopy (ESGE) guideline, in cooperation with the European Respiratory Society (ERS) and the European Society of thor. Eur J Cardiothorac Surg 48:1–15. doi:10.1093/ejcts/ezv194
52. Aggarwal R, Mytton OT, Derbrew M, Hananel D, Heydenburg M, Issenberg B, MacAulay C, Mancini ME, Morimoto T, Soper N, Ziv A, Reznick R (2010) Training and simulation for patient safety. Qual Saf Health Care 19(Suppl 2):i34–i43. doi:10.1136/qshc.2009.038562
53. Barsuk JH, Cohen ER, Wayne DB, Siddall VJ, McGaghie WC (2016) Developing a simulation-based mastery learning curriculum. Simul Healthc J Soc Simul Healthc 11:52–59. doi:10.1097/SIH.0000000000000120
54. Mcgaghie WC (2015) Mastery learning : it is time for medical education to join the 21st century. Acad Med 90:1–4. doi:10.1097/ACM.0000000000000911
55. Konge L, Vilmann P, Clementsen P, Annema JT, Ringsted C (2012) Reliable and valid assessment of competence in endoscopic ultrasonography and fine-needle aspiration for mediastinal staging of non-small cell lung cancer. Endoscopy 44:928–933. doi:10.1055/s-0032-1309892
56. Tong BC, Gustafson MR, Balderson SS, D'Amico TA, Meyerson SL (2012) Validation of a thoracoscopic lobectomy simulator. Eur J Cardiothorac Surg 42:364–369. doi:10.1093/ejcts/ezs012
57. Meyerson SL, Tong BC, Balderson SS, D'Amico TA, Phillips JD, Decamp MM, Darosa DA (2012) Needs assessment for an errors-based curriculum on thoracoscopic lobectomy. Ann Thorac Surg 94:368–373. doi:10.1016/j.athoracsur.2012.04.023