Assessment of competence in video-assisted thoracoscopic surgery lobectomy

A Danish nationwide study

Petersen, René Horsleben; Gjeraa, Kirsten; Jensen, Katrine; Møller, Lars Borgbjerg; Hansen, Henrik Jessen; Konge, Lars

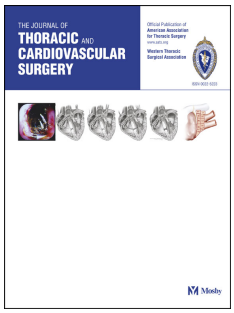
Published in:
The Journal of Thoracic and Cardiovascular Surgery

DOI (link to publication from Publisher):
10.1016/j.jtcvs.2018.04.046

Publication date:
2018

Document Version
Accepted author manuscript, peer reviewed version

Accepted Manuscript



Assessment of competence in Video Assisted Thoracoscopic Surgery (VATS) Lobectomy: A Danish nationwide study

René Horsleben Petersen, MD, PhD, Kirsten Gjeraa, MD, PhD, Katrine Jensen, MD, PhD, Lars Borgbjerg Møller, MD, Henrik Jessen Hansen, MD, Lars Konge, MD, PhD

PII: S0022-5223(18)31030-4

DOI: 10.1016/j.jtcvs.2018.04.046

Reference: YMTC 12917

To appear in: *The Journal of Thoracic and Cardiovascular Surgery*

Received Date: 4 January 2018

Revised Date: 5 April 2018

Accepted Date: 7 April 2018

Please cite this article as: Petersen RH, Gjeraa K, Jensen K, Møller LB, Hansen HJ, Konge L, Assessment of competence in Video Assisted Thoracoscopic Surgery (VATS) Lobectomy: A Danish nationwide study, *The Journal of Thoracic and Cardiovascular Surgery* (2018), doi: 10.1016/j.jtcvs.2018.04.046.

Assessment of competence in Video Assisted Thoracoscopic Surgery (VATS) Lobectomy: A Danish nationwide study

René Horsleben Petersen, MD, PhD [1], Kirsten Gjeraa, MD, PhD[2], Katrine Jensen, MD, PhD[1], Lars Borgbjerg Møller, MD[3], Henrik Jessen Hansen, MD[1], Lars Konge, MD, PhD[2]

[1]Department of Cardiothoracic Surgery, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

[2]Copenhagen Academy for Medical Education and Simulation (CAMES), University of Copenhagen and the Capital Region of Denmark, Copenhagen, Denmark

[3]Department of Cardiothoracic Surgery, Aalborg University Hospital, Aalborg, Denmark

Corresponding Author:

René Horsleben Petersen, Department of Cardiothoracic Surgery, Copenhagen University Hospital, Rigshospitalet, Inge Lehmannsvej 7, 2100 Copenhagen, Denmark.

Email: rene.horsleben.petersen@regionh.dk

Disclosures: René Horsleben Petersen and Henrik Jessen Hansen have received speaker honoraria from Medtronic; the remaining authors have no disclosures.

Funding: No funding was received for this study.

Word count: 2422

20

Central message

Validity evidence using Messick´s framework was provided for a newly developed specific assessment tool (VATSAT) allowing for structured and objective assessment of VATS lobectomy competence.

Perspective

In the surgical societies around the world there is an increasing focus on ensuring continuous education and credentialing of surgical skills according to stringent quality criteria.

This study provides validity evidence for a newly developed specific assessment tool for VATS lobectomy (VATSAT), which may be an important aid the future training and certification of thoracic surgeons.

Central picture

Revised figure 1

Central picture legend

Box-and-whiskers plot showing relation between the experience level of the thoracic surgeons and the VATSAT score. Beginners n=10 procedures, n=6 surgeons (red dots), intermediates n=28 procedures, n= 9 surgeons (green dots), experts n=20 procedures, n=3 surgeons (blue dots). Colored bar: median VATSAT score.

Glossary of abbreviations:

ANOVA	Analysis of Variance
GOALS	Global Operative Assessment of Laparoscopic Skills
SD	Standard deviation
VATS	Video-Assisted Thoracoscopic Surgery
VATSAT	Video-Assisted Thoracoscopic Surgery Lobectomy Assessment Tool

Abstract:

Background: Competence in VATS lobectomy has previously been established based on numbers of procedures performed but this approach does not ensure competence. Specific assessment tools like the newly developed VATSAT allow for structured and objective assessment of competence. Our aim was to provide validity evidence for VATSAT.

Methods: Video recordings of 60 VATS lobectomies performed by 18 thoracic surgeons were rated using the VATSAT. All four centers of thoracic surgery in Denmark participated in the study. Two VATS experts rated the videos. They were blinded to surgeon and center.

Results: The total internal consistency reliability, Cronbach's Alpha was 0.93. Inter-rater reliability between the two raters was Pearson's r=0.71 ($p< 0.001$). The mean VATSAT score for the 10 procedures performed by beginners were 22.1 (SD 8.6), for the 28 procedures performed by the intermediate surgeons 31.2 (SD 4.4) and for the 20 procedures performed by experts 35.9 (SD 2.9); $p<0.001$. Bonferroni post-hoc tests showed that experts were significantly better than intermediates ($p < 0.008$) and beginners ($p< 0.001$). Intermediates' mean scores were significantly better than beginners ($p< 0.001$). The pass/fail standard calculated using the contrasting group's method was 31 points. One of the beginners passed and two procedures performed by experts failed the test.

Conclusion: Validity evidence was provided for a newly developed assessment tool for VATS lobectomy (VATSAT) in a clinical setting. The discriminatory ability between expert surgeons, intermediate surgeons, and beginners proved highly significant. VATSAT could be an important aid in the future training and certification of thoracic surgeons.

Background

Lung cancer is the most deadly cancer worldwide and it is estimated that 1.7 million people died from lung cancer in 2015 (1). Surgical resection remains the mainstay in curing localized lung cancer (2). Traditionally, the approach for surgical resection has been a thoracotomy. Video Assisted Thoracoscopic Lobectomy was introduced 25 years ago and is now the recommended approach for early stage lung cancer (3). The potential benefits include less postoperative pain, shorter length of stay, better quality of life (4), better shoulder function, fewer complications (5), better tolerance of adjuvant chemotherapy (6), and maybe even improved survival (7). Despite the obvious advantages of this approach, the adoption of the procedure has been slow. Performing a VATS lobectomy requires a different set of skills compared to thoracotomy, such as overcoming the fulcrum effect when operating through ports and transforming the 2-dimensional images on the monitor into a 3-dimensional understanding. The potential risk of hemorrhage due to injury of the pulmonary artery requires experience and skills to handle in a VATS scenario without causing a catastrophic intraoperative complication (8). Several papers have addressed the issue of learning how to perform a VATS lobectomy (9). Recommendations so far have been to attend courses in VATS lobectomy, visit centers with a substantial experience in VATS lobectomy and then begin in a step wise manner preferably supervised by an experienced VATS surgeon (mentor) until competency was achieved (10). Traditionally, competency has been established based on numbers of procedures performed and experts in VATS surgery have proposed 50 VATS lobectomies as a threshold for competency (11, 12). However, procedural experience does not ensure competence (13). Specific assessment tools have been developed to allow for structured and objective assessment of competence, but it is essential that these provide valid measures (14). The aim of this study was to provide validity evidence for a newly developed VATS lobectomy assessment tool (VATSAT) (15).

Methods

Data collection

An independent investigator (KG) recorded the videos from VATS lobectomies performed at all four thoracic centers in Denmark. Only unedited videos were used for assessment and the surgeons did not have access to their videos. The investigator was present in the operating theatre throughout the operations to make detailed notes of who performed the single parts of the procedure. The surgeons were divided into three groups according to their previous experience in VATS lobectomy at the beginning of the study. Surgeons having performed between one and 49 VATS lobectomies were grouped as beginners. Surgeons having performed between 50 and 499 VATS lobectomies were labeled intermediates, and finally experts were surgeons having performed 500 VATS lobectomies or more. Two independent thoracic surgeons with a solid experience in VATS lobectomy rated the videos using a newly developed VATS Lobectomy Assessment Tool (VATSAT) for technical scoring of VATS lobectomies (15). VATSAT score was developed using the Delphi method as a structured process for collecting and distilling knowledge from a group of international experts in VATS lobectomy (15, 16). The eight items in the VATSAT are: 1.Localization of tumor and other pathological tissue, 2. Dissection of the hilum and veins, 3. Dissection of the arteries, 4. Dissection of the bronchus, 5. Dissection of lymph nodes, 6. Retrieval of lobe in bag, 7. Respect for tissue and structures, 8. Technical skills in general. Each item was rated one to five, where five were the best score, giving a minimum score of 8 and a maximum score of 40. The two raters were blinded to the surgeon and the center where the procedure was performed.

Validity evidence

Validity evidence was established based on Messick's framework (17) as recommended by the Standards of Educational and Psychological Testing (18) with the following five major sources of evidence:

Content: Content validity for the VATSAT tool was established in a previous study from our research group (15). The content was thoroughly evaluated in three rounds by a large group of internationally recognized VATS lobectomy experts using the Delphi method (16).

Response process: The two raters were carefully instructed on how to rate the videos using the VATSAT tool. Both raters rated all videos independently according to their instructions.

Internal structure: The degree to which the items in the VATSAT fit the underlying construct was reported by internal consistency reliability and inter-rater reliability.

Relations to other variables: VATSAT´s discriminatory ability between beginners, intermediates, and experts was calculated using mean scores and ANOVA (analysis of variance) with Bonferroni post hoc tests. The correlation coefficient between the number of VATS lobectomies performed (expressed in the logarithmic scale) and the VATSAT score was calculated.

Consequences: Impact of the VATSAT scores was assessed using the contrasting group's method (a method to identify a cut score based on overlapping frequency distributions of two groups) to calculate the mean pass/fail VATSAT score and explore false positives and false negatives.

Ethics

An application was send to the local ethics committee (journal no H-16041772), but was waived. According to Danish law, educational studies do not need approval. Written and oral informed consent was obtained from all participating surgeons.

Statistics

Cronbach's alpha, Pearson's r, and ANOVA with post hoc analysis were calculated using IBM SPSS statistics version 23 (IBM, New York, USA). A value of $p<0.05$ was considered statistically significant.

Results

From December 19th 2016 until July 5th 2017, 60 VATS lobectomies performed at the four thoracic centers in Denmark were video recorded and enrolled into the study. Eighteen thoracic surgeons performed the 60 procedures. Their personal experience in VATS lobectomy ranged from 9 to 1200 procedures completed at the beginning of data collection. Fifteen of the 18 surgeons were specialists in Cardio-thoracic Surgery and the remaining three surgeons were senior residents in Cardio-thoracic Surgery. A specialist supervised all procedures (n=8) performed by residents. If the supervisor had to interfere in the procedure and perform part of it, the investigator noted this and the corresponding item received the minimum score of one point. Two VATS lobectomies were converted to open surgery during the procedure. They were excluded from the study, since the raters were unable to use the assessment tool (VATSAT), which is constructed for VATS specific issues only. The remaining 58 VATS lobectomies were included in the final data analysis. Patient characteristics and surgical outcome are listed in table 1.

157 *Internal structure:* The total internal consistency reliability, Cronbach's Alpha was 0.93 with a
158 value of 0.89 for rater 1 and 0.91 for rater 2. Inter-rater reliability between the two raters was
159 Pearson's r=0.71 ($p< 0.001$).

160 *Relation to other variables:* The mean VATSAT score for the 10 procedures performed by
161 beginners were 22.1 (SD 8.6; range 8.0-34.0), for the 28 procedures performed by the intermediate
162 surgeons 31.2 (SD 4.4; range 24.0-38.0), and for the 20 procedures performed by experts 35.9 (SD
163 2.9; range 29.0-39.5); $p<0.001$, presented as a Box plot in figure 1.

164 ANOVA with Bonferroni post-hoc tests revealed that mean scores for experts were significantly
165 better than for intermediates and beginners, $p < 0.008$ and $p< 0.001$, respectively. Intermediates'
166 mean scores were significantly better than beginners ($p< 0.001$). The logarithmic relation between
167 number of VATS lobectomies performed and the mean VATSAT score is shown in figure 2. The
168 Pearson's Correlation is r=0.68 ($p< 0.001$).

169 *Evidence based on consequences of testing:* The pass/fail mean standard calculated using the
170 contrasting group's method was 31 points. One procedure performed by a beginner passed the test
171 with a mean score of 34 (false positive) and two procedures performed by experts failed the test
172 with mean scores of 29 and 30.5 points (false negatives). See figure 3.

173

174 **Discussion**

175 Validity evidence has previously been demonstrated for the VATSAT used in a simulated
176 environment (15). In this study, validity evidence for the VATSAT used in a clinical situation with
177 live surgical cases from four different centers being video recorded and the raters blinded for the
178 institution and the surgeon is demonstrated.

Our group has previously published an assessment tool targeted towards VATS wedge resections, but VATSAT is the first assessment tool developed specifically to assess VATS lobectomy (19). A systematic review published in 2015 identified 29 articles focused on procedural tasks. The majority of studies addressed tasks related to general surgery and the remaining to obstetrics/gynecology, vascular surgery, orthopedics, cardiac surgery, plastic surgery, and minor surgical procedures by family physicians (20). Minimally invasive thoracic surgery and perhaps especially VATS lobectomies are highly specialized procedures and there is a need for dedicated assessment tools (21).

The total internal consistency reliability, Cronbach's Alpha of 0.93 shows that the eight items in the VATSAT measure the same trait and thereby provides evidence for the well-aligned content of the tool. A high Cronbach's Alfa indicates a very strong correlation between the eight individual items in the VATSAT. Surgeons who have a high score high in one item also have a high score in the other items (22). This internal consistency reliability is similar to what have been demonstrated for the Global Operative Assessment of Laparoscopic skills (GOALS) that were developed by *Vassilou et al* in 2005. They found an internal consistency reliability of 0.91-0.93 by assessing 21 participants performing laparoscopic cholecystectomies by two trained observers present in the operating theatre and by the attending surgeon assisting the procedures (23).

In our study the inter-rater reliability between the two blinded raters at a Pearson's r 0.71 was highly significant ($p< 0.001$), meaning that there was a significant agreement in the total score between the two raters. An inter-rater reliability in the range of 0.70 to .079 may be applied for formative assessments such as feedback after a completed training course. For moderate stake summative assessments as end of year examination in medical school an inter-rater reliability between 0.80 and 0.89 is expected. High stake tests as a board certification and licensure require

inter-rater reliability above 0.90 (22). The inter-rater reliability can be improved by increasing the number of rated procedures per surgeon or by increasing the number of raters (24). Our results clearly show, that certification aided by VATSAT scores should be based on assessment of more than one procedure per trainee.

Rating of VATS lobectomy is a time consuming task and the use of a VATS specialist is costly. Therefore it is important that a potential test do not need too many raters. The use of video recordings has several advantages compared to direct observation in the operating theatre that will always be prone to bias. A previous study showed that direct observation favored operators well known by the rater or considered competent due to their position (25). Another advantage of video recordings is that the VATS expert rater can schedule the rating to an appropriate time and place (26). Using non-experts or novice raters may be considered, since the availability is easier and the costs are less. This approach should be used with some caution but recent work has shown good inter-rater reliability between expert and non-expert raters (27, 28).

The logarithmic relation between the experience level of the thoracic surgeons and the mean VATS score shows good consistency. However, a Pearson's Correlation of $r=0.68$ ($p< 0.001$) is not a perfect correlation. In figure 1 it can be seen that it is not possible to precisely predict competence based on the VATSAT score from a certain experience level – a threshold of e.g. 50 procedures will not ensure that all surgeons are competent. The VATSAT score is increasing with increasing experience level and at the same time the variance in performance is decreasing (figure 3). This is in accordance with the model for skills acquisition by Fitts and Posner: Performance is variable in the beginning of the learning process but as the performance improves the *variability* also decreases and the performance characteristics become more similar (29). The use of volume cut off to determine the beginner, intermediate and expert surgeons are not ideal, but a necessary step at this point.

The VATSAT test was able to discriminate between expert surgeons and surgeons with an intermediate experience and between intermediates and beginners using the ANOVA with Bonferroni post-hoc test, and this were highly significant ($p < 0.008$ and $p< 0.001$, respectively). In the simulation study we were not able to discriminate between intermediate surgeons and expert surgeon. This may be due to the challenging and maybe impossible task to make simulators reflect every aspect of real live surgery (30). The pass/fail standard of 31 points, calculated using the contrasting groups' method established good validity evidence for consequences (31). One of the beginners passed the test and two of the experts failed the test. In a simulation study case difficulty can be standardized (30). Other studies with live surgical procedures have tried to reduce the effect of disease and patient variability (32). This was not possible due to the nationwide design in this study. VATS lobectomies were heterogeneous in terms of difficulty level and a considerable bias may reflect the variation observed. This underlines the point of basing important decisions on more than one procedure – even a beginner can do okay on a very easy patient and competency cannot be determined based on a single case.

The strengths of this study are the nationwide participation of thoracic surgeons from all departments of thoracic surgery in Denmark. This adds to the generalizability of our findings. Having 18 thoracic surgeons perform 60 VATS lobectomies reduced construct underrepresentation. The fact that we were able to show a highly significant discrimination of experts, intermediate surgeons, and beginners with only two raters is an important finding indicating that the VATSAT may aid in training and credentialing of VATS surgeons.

Limitations of the study and a threat to validity evidence of the assessments may be construct-irrelevant variance in the form of rater errors. Raters may express restriction of range in their rating (33). Interestingly none of the participating expert surgeons scored the mean maximum of 40 and only one of the beginners scored the mean minimum score of 8. Despite the raters were blinded, we

cannot rule out the possibility that special instruments or certain movements were recognizable. Therefore possible identification of a center or individual surgeons may have biased the raters. It is important to emphasize that the focus was technical skills. Non-technical skills are import and may interfere with the overall performance (34, 35), but this was not captured in the current study.

Conclusion

Validity evidence was provided for a newly developed assessment tool for VATS lobectomy (VATSAT) in a clinical setting with 18 surgeons representing all thoracic units in Denmark based on video recording of 60 VATS lobectomies and two blinded raters. Internal consistency reliability was high and inter-rater reliability acceptable. The discriminatory ability between expert surgeons, intermediate surgeons, and beginners was highly significant with a pass/fail standard of 31 points. One of the beginners passed the test (false positive) and two experts failed the test (false negatives). We believe that the VATSAT can be a valid and important tool to aid in deciding when thoracic surgeons are competent to perform VATS lobectomies.

Acknowledgements

The authors wishes to express their gratitude to the participating surgeons and theatre staff at the departments of Cardiothoracic Surgery at Odense University Hospital, Aarhus University Hospital (Skejby), Aalborg University Hospital, and Copenhagen University Hospital (Rigshospitalet) for their thoughtful input and support to the study.

References

1. WHO Media centre: http://www.who.int/mediacentre/factsheets/fs297/en/
2. Rosen JE, Keshava HB, YaoXet al. The natural history of operable non-small cell lung cancer in the National Cancer Database. *Ann Thorac Surg* 2016; 101: 1850–1855.
3. Detterbeck FC, Lewis SZ, Diekemper R, Addrizzo-Harris D, Alberts WM Executive Summary: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013 May; 143(5 Suppl): 7S-37S.
4. Bendixen M, Jørgensen OD, Kronborg C, Andersen C, Licht PB Postoperative pain and quality of life after lobectomy via video-assisted thoracoscopic surgery or anterolateral thoracotomy for early stage lung cancer: a randomised controlled trial. Lancet Oncol. 2016 Jun; 17(6): 836-844.
5. Falcoz PE, Puyraveau M, Thomas PA, Decaluwe H, Hürtgen M, Petersen RH, Hansen H, Brunelli A; ESTS Database Committee and ESTS Minimally Invasive Interest Group Video-assisted thoracoscopic surgery versus open lobectomy for primary non-small-cell lung cancer: a propensity-matched analysis of outcome from the European Society of Thoracic Surgeon database. Eur J Cardiothorac Surg. 2016 Feb; 49(2): 602-9.
6. Petersen RP, Pham D, Burfeind WR, Hanish SI, Toloza EM, Harpole Jr DH, D'Amico TA. Thoracoscopic lobectomy facilitates the delivery of chemotherapy after resection for lung cancer. Ann Thorac Surg 2007; 83(4): 1245-9.
7. Whitson BA, Groth SS, Duval SJ, Swanson SJ, Maddaus MA. Surgery for early-stage non-small cell lung cancer: a systematic review of the video assisted Thoracoscopic surgery versus thoracotomy approaches to lobectomy. Ann Thorac Surg 2008; 86: 2008-18.
8. Decaluwe H, Petersen RH, Hansen H, Piwkowski C, Augustin F, Brunelli A, Schmid T, Papagiannopoulos K, Moons J, Gossot D; ESTS Minimally Invasive Thoracic Surgery Interest Group (MITIG). Major intraoperative complications during video-assisted thoracoscopic anatomical lung resections: an intention-to-treat analysis. Eur J Cardiothorac Surg. 2015 Oct; 48(4): 588-98; discussion 599.

9. Petersen RH, Hansen HJ. Learning thoracoscopic lobectomy. Eur J Cardiothorac Surg. 2010 Mar; 37(3): 516-20.
10. Petersen RH, Hansen HJ. Learning curve associated with VATS lobectomy. Ann Cardiothorac Surg. 2012 May; 1(1): 47-50.
11. McKenna Jr RJ. Complications and learning curves for video-assisted thoracic surgery lobectomy. Thorac Surg Clin 2008; 18(3): 275-80.
12. Yan TD, Cao C, D'Amico TA, Demmy TL, He J, Hansen H, Swanson SJ, Walker WS; International VATS Lobectomy Consensus Group. Video-assisted thoracoscopic surgery lobectomy at 20 years: a consensus statement. Eur J Cardiothorac Surg. 2014 Apr; 45(4): 633-9.
13. Barsuk JH, Cohen ER, Feinglass J, McGaghie WC, Wayne DB. Residents` procedural experience does not ensure competence: A research synthesis. J Grad Med Educ. 2017 Apr; 9(2): 201-208.
14. Dearani JA, Gold M, Leibovich BC, Ericsson KA, Khabbaz KR, Foley TA, Julsrud PR, Matsumoto JM, DalyRC. The role of imaging, deliberate practice, structure, and improvisation in approaching surgical perfection. J Thorac Cardiovasc Surg 2017 Oct;154(4):1329-1336.
15. Jensen K, Petersen RH, Hansen HJ, Walker W, Pedersen JH, Konge L. A novel assessment tool for evaluating competence in video-assisted thoracoscopic surgery lobectomy. Surg Endosc. 2018 Mar 30. doi: 10.1007/s00464-018-6162-8. [Epub ahead of print]
16. Day J, Bobeva M. A generic toolkit for the successful management of delphi studies. Electron J Bus Res Methods. 2005;3(2):103–16.
17. Downing SM, Yudkowsky R. Assessment in Health Professions Education. New York, NY: Routledge; 2009.
18. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association 1999.
19. Konge L, Lehnert P, Hansen HJ, Petersen RH, Ringsted C. Reliable and valid assessment of performance in thoracoscopy. Surg Endosc. 2012 Jun; 26(6): 1624-8.

20. Hatala R, Cook DA, Brydges R, Hawkins R. Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): a systematic review of validity evidence. Adv in Health Sci Educ (2015) 20:1149–1175
21. Ferguson MK, Umanskiy K, Warnes C, Celauro AD, Vigneswaran WT, Prachand VN. Robot Training in minimally invasive lobectomy: thoracoscopic versus robotic approach. Ann Thorac Surg. 2014 Jun; 97(6): 1885-92.
22. Downing SM. Reliability: on the reproducibility of assessment data. Med Educ. 2004 Sep; 38(9): 1006-12.
23. Vassiliou MC et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg. 2015, 190:107–113.
24. Konge L, Annema J, Clementsen P, Minddal V, Vilmann P, Ringsted C. Using virtual-reality simulation to assess performance in endobronchial ultrasound. Respiration. 2013,86(1), 59–65.
25. Vogt VY, Givens VM, Keathley CA, Lipscomb GH, Summitt RL. Is a resident's score on a videotaped objective structured assessment of technical skills affected by revealing the resident's identity? In American Journal of Obstetrics and Gynecology, 2003, Vol. 189, pp. 688–691.
26. Konge L, Vilmann P, Clementsen P, Annema JT, Ringsted C. Reliable and valid assessment of competence in endoscopic ultrasonography and fine-needle aspiration for mediastinal staging of non-small cell lung cancer. Endoscopy. 2012 Oct; 44(10): 928-33
27. Mahmood O, Dagnæs J, Bube S, Rohrsted M, Konge L. Nonspecialist Raters Can Provide reliable assessments of procedural skills. Journal of Surgical Education 2017 in press: http://dx.doi.org/10.1016/j.jsurg.2017.07.003
28. Lendvay TS, White L, Kowalewski T. Crowdsourcing to assess surgical skill. JAMA surg 2015; 150(11):1086-87.
29. Magill R, Anderson D "Motor Learning and Control Motor Learning and Control: Concepts and Applications 10th Edition 2015.
30. Jensen K, Bjerrum F, Hansen HJ, Petersen RH, Pedersen JH, Konge L. 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. 2015 Oct;21(4):420-6.

31. Norcini JJ (2003). Setting standards on educational tests. Medical Education, 37(5), 464–469.
32. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Darzi A. Toward Feasible, Valid, and Reliable Video-Based Assessments of Technical Surgical Skills in the Operating Room. Ann Surg 2008;247: 372–379.
33. Iramaneerat C, Yudkowsky R. Rater errors in a clinical skills assessment of medical students. Eval Health Prof. 2007; 30(3):266-283.
34. Gjeraa K, Mundt AS, Spanager L, Hansen HJ, Konge L, Petersen RH, Østergaard D. Important Non-Technical Skills in Video-Assisted Thoracoscopic Surgery Lobectomy: Team Perspectives. Ann Thorac Surg. 2017 Jul; 104(1):329-335.
35. Gjeraa K, Spanager L, Konge L, Petersen RH, Østergaard D. Non-technical skills in minimally invasive surgery teams: a systematic review. Surg Endosc. 2016 Dec; 30(12):5185-5199.

365 **Figure Legends**

366 **Figure 1:** Box-and-whiskers plot showing relation between the experience level of the thoracic
367 surgeons and the VATSAT score. Beginners n=10 procedures, n=6 surgeons (red dots),
368 intermediates n=28 procedures, n= 9 surgeons (green dots), experts n=20 procedures, n=3 surgeons
369 (blue dots). Colored bar: median VATSAT score.

370 **Figure 2:** Logarithmic relation between number of VATS lobectomies performed (n=58) and the
371 VATSAT score. Red dots represents VATSAT scores for beginners, green dots represents
372 VATSAT scores for intermediates and blue dots represents VATSAT scores for experts. Black
373 dotted line is the pass/fail ratio of 31.

374 **Figure 3:** Pass/fail VATSAT score assessed using the contrasting group's method for beginners
375 (red, n=10 procedures) and expert thoracic surgeons (blue, n=20 procedures).

376

377 **Video legend**

378 Video: Right lower VATS lobectomy rated using the VATSAT tool.

Table 1. Patient characteristics and surgical outcome.

Patient characteristics, n=58		
		Mean (SD)
Age		70 (8.2)
Gender	Male/female	29/29
FEV_1*		89 (22.0)
Tumor size		26 (11.9)
Type of lobectomy	Lower lobes	21 (36%)
	Middle lobe	5 (9%)
	Upper lobes	30 (52%)
	Bi-lobectomy	2 (3%)
Surgical outcome, n=58		
		Median (interquartile range)
Procedural time (minutes)		101 (88; 123)
Procedural bleeding (ml)		100 (20; 150)

* FEV_1 (forced expiratory volume in one second)

Table 1. Patient characteristics and surgical outcome

Patient characteristics, n=58		
		Mean (SD)
Age		70 (8.2)
Gender	Male/female	29/29
FEV_1*		89 (22.0)
Tumor size		26 (11.9)
Type of lobectomy	Lower lobes	21 (36%)
	Middle lobe	5 (9%)
	Upper lobes	30 (52%)
	Bi-lobectomy	2 (3%)
Surgical outcome, n=58		
		Median (interquartile range)
Procedural time (minutes)		101 (88; 123)
Procedural bleeding (ml)		100 (20; 150)

* FEV_1 (forced expiratory volume in one second)