



# Assessment of competence in local anaesthetic thoracoscopy: development and validity investigation of a new assessment tool

Leizl Joy Nayahangan<sup>1^</sup>, Morten Bo Søndergaard Svendsen<sup>1^</sup>, Uffe Bodtger<sup>2,3</sup>, Najib Rahman<sup>4^</sup>, Nick Maskell<sup>5^</sup>, Jatinder Singh Sidhu<sup>3</sup>, Jonathan Lawaetz<sup>1,6^</sup>, Paul Frost Clementsen<sup>1,2,7^</sup>, Lars Konge<sup>1,7^</sup>

<sup>1</sup>Copenhagen Academy for Medical Education and Simulation, Centre for HR and Education, The Capital Region of Denmark, Copenhagen, Denmark; <sup>2</sup>Department of Internal and Respiratory Medicine, Zealand University Hospital, Roskilde, Denmark; <sup>3</sup>Department of Respiratory Medicine, Næstved Hospital, Næstved, Denmark; <sup>4</sup>Nuffield Department of Medicine, Oxford Respiratory Trials Unit, University of Oxford, Oxford, UK; <sup>5</sup>Academic Respiratory Unit, School of Clinical Sciences, University of Bristol, Bristol, UK; <sup>6</sup>Department of Vascular Surgery, Rigshospitalet, Copenhagen, Denmark; <sup>7</sup>Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

**Contributions:** (I) Conception and design: LJ Nayahangan, MBS Svendsen, U Bodtger, JS Sidhu, J Lawaetz, N Rahman, N Maskell, PF Clementsen, L Konge; (II) Administrative support: LJ Nayahangan, U Bodtger, JS Sidhu, J Lawaetz, PF Clementsen; (III) Provision of study materials: LJ Nayahangan, MBS Svendsen, U Bodtger, JS Sidhu, J Lawaetz, N Rahman, N Maskell, PF Clementsen, L Konge; (IV) Collection and assembly of data: LJ Nayahangan, MBS Svendsen, U Bodtger, JS Sidhu, J Lawaetz, PF Clementsen, L Konge; (V) Data analysis and interpretation: LJ Nayahangan, PF Clementsen, L Konge; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

**Correspondence to:** Leizl Joy Nayahangan. Copenhagen Academy for Medical Education and Simulation, Copenhagen, Denmark.

Email: Leizl.joy.nayahangan@regionh.dk.

**Background:** The aims of the study were to develop an assessment tool in local anaesthetic thoracoscopy (LAT), investigate validity evidence, and establish a pass/fail standard.

**Methods:** Validity evidence for the assessment tool was gathered using the unified Messick framework. The tool was developed by five experts in respiratory medicine and medical education. Doctors with varying experience performed two consecutive procedures in a standardized, simulation-based setting using a newly developed thorax/lung silicone model. Performances were video-recorded and assessed by four expert raters using the new tool. Contrasting groups' method was used to set a pass/fail standard.

**Results:** Nine novices and 8 experienced participants were included, generating 34 recorded performances and 136 expert assessments. The tool had a high internal consistency (Cronbach's alpha =0.94) and high inter-rater reliability (Cronbach's alpha =0.91). The total item score significantly correlated with the global score ( $r_s=0.86$ ,  $P<0.001$ ). Participants' first performance correlated to second performance (test-retest reliability) with a Pearson's  $r$  of 0.93,  $P<0.001$ . Generalisability (G) study showed a G-coefficient of 0.92 and decision (D) study estimated that one performance assessed by two raters or four performances assessed by one rater are needed to reach an acceptable reliability, i.e., G-coefficient  $>0.80$ . The tool was able to discriminate between the two groups in both performances: experienced mean score =30.8±4.2; novice mean score =15.8±2.3,  $P<0.001$ . Pass/fail standard was set at 22 points.

**Conclusions:** The newly developed assessment tool showed solid evidence of validity and can be used to ensure competence in LAT.

**Keywords:** Assessment; training; simulation training; thoracoscopy; pleuroscopy; validation study

Submitted Dec 22, 2020. Accepted for publication Feb 24, 2021.

doi: 10.21037/jtd-20-3560

View this article at: <https://dx.doi.org/10.21037/jtd-20-3560>

<sup>^</sup> ORCID: Leizl Joy Nayahangan, 0000-0002-6179-1622; Morten Bo Søndergaard Svendsen, 0000-0002-4492-3750; Najib Rahman, 0000-0002-1231-9209; Nick Maskell, 0000-0002-1276-6500; Jonathan Lawaetz, 0000-0001-7661-4058; Paul Frost Clementsen, 0000-0001-6624-4620; Lars Konge, 0000-0002-1258-5822.

## Introduction

Diagnostic thoracoscopy in the management of exudative pleural effusion of unknown origin can be performed either as video-assisted thoracoscopy surgery in general anaesthesia or as local anaesthetic thoracoscopy (LAT) under conscious sedation. For LAT (formerly called medical thoracoscopy), the semi-rigid thoracoscope with increased maneuverability has replaced the rigid thoracoscope in many institutions and was found to be a safe and simple procedure with acceptable sensitivity for malignancy (1,2). LAT allows both to take biopsies and perform pleurodesis and is one of the techniques with the highest diagnostic yield in cytology negative, exudative pleural effusion. It is increasingly being used by respiratory physicians (3,4) but certain recommendations must be followed to maintain and enhance safety of the procedure (5) and there is an increasing need to train respiratory physicians in LAT (6). However, in contrast to other technical procedures for lung cancer diagnostics such as endoscopic ultrasound (7), the dissemination of LAT has occurred without consensus on how operators should be trained and how their competences should be assessed. The current training approach to achieve the competences is through the traditional apprenticeship method. The American College of Chest Physicians stated that aside from extensive knowledge of pleural and thoracic anatomy, ample experience is also a requirement. To attain basic competency, physicians are recommended to perform at least 20 procedures in a supervised setting and 10 procedures to maintain competency (8). However, doctors learn at different paces and clinical training opportunities might be scarce in some departments (9). In place of training on patients and sheer (arbitrary) volume requirements to reach competency, front-line medical educators recommend to shift to competency-based assessment following mastery learning in a simulated environment (10). Defensible and sound assessment is the cornerstone for mastery learning and is increasingly being applied in medical education (11,12).

There is a paucity in the literature on the availability of evidence-based assessment tools for LAT. On this background we decided to develop an assessment tool for competences in LAT and to investigate its validity.

The aims of this study were:

- (I) To develop an assessment tool for LAT;
- (II) To explore validity evidence of the assessment tool in a simulated-based setting;
- (III) To establish a pass/fail standard that can be used to

ensure competence in a mastery learning training program.

We present the following article in accordance with the MDAR checklist (available at <https://dx.doi.org/10.21037/jtd-20-3560>).

## Methods

The study was conducted prospectively, and performed at the Copenhagen Academy for Medical Education and Simulation from September 2018 to June 2019. Validity investigation of the assessment tool was based on the unified framework proposed by Messick, drawing evidence of validity from five sources: content, response process, internal structure, relations to other variables, and consequences (13). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). It is an educational study that did not include patient participation and therefore did not require ethical approval in accordance with Danish law. Informed consent to participate in the study was obtained.

### *Test development*

The content of the assessment tool was developed by a collaboration between four respiratory physicians with expertise in LAT (PFC, UB, NM, and NR) and a professor in medical education with expertise regarding assessment tools (LK). This was based on the important competences that should be considered when assessing a trainee performing LAT: sterile approach, instillation of local anesthesia, skin incision, insertion of trocar, insertion of videoscope, systematic exploration of the thoracic cavity, location of possible tumors, biopsy of these, avoiding lung lesions, and insertion of chest tube (*Figure 1*). Once the test content and competences were agreed upon, a pilot rating was initiated to ensure its usability.

### *The simulated set-up*

A thorax/lung model was developed and built in silicone for the study providing the possibility to perform a completely standardized LAT procedure from administration of local anaesthesia to inspection of the thoracic cavity and biopsy (*Figure 2*). Five predetermined tumors were distributed in the pleural cavity. All necessary equipment was available including the thoracoscope (LTF-160 and EVIS Exera II, Olympus, Tokyo, Japan).

Local anaesthesia	1 Administers inappropriate amount of lidocaine AND does not aspirate air	2	3 Administers enough lidocaine BUT does not aspirate air	4	5 Aspirates air AND administers appropriate amount of lidocaine in all layers
Skin incision	1 Incision too wide AND too superficial/deep	2	3 Incision too wide OR superficial	4	5 Stab incision with appropriate depth and size
Insertion of Trocar	1 Inserts without rotation through chest wall and stops before or advances further after loss of resistance	2	3 Uses inappropriate force without rotating through chest wall	4	5 Uses rotation through chest wall with appropriate force until loss of resistance
Handling of endoscope	1 Superfluous movements of endoscope handle	2	3 Intermittent movements of endoscope handle		5 Deliberate movements of endoscope handle
360° inspection and search for tumors	1 Fails to visualize several areas of available cavity AND found only one tumor	2	3 Uses random movement with frequent backtracking but manages to visualize entire available cavity AND found three tumors.	4	5 Visualizes entire available area in a systematic manner AND found all five tumors
Biopsies from five predetermined tumors	1 Fails to obtain biopsies from visible tumor OR takes biopsies from visceral pleura.	2	3 Takes at least one biopsy from three out of five tumors	4	5 Takes at least one biopsy from five tumors using correct technique
Insertion of chest tube	1 Fails to insert chest tube OR side holes visible	2	3 Removes trocar prior to insertion of tube	4	5 Inserts tube aiming at the apical part of the thorax and removes trocar in a fluent manner
Use of sterile technique throughout the procedure	1 Fails to use sterile technique during the procedure	2	3 Inconsistently uses sterile technique during the procedure	4	5 Consistently uses and maintains sterile technique during the entire procedure
( ) <b>FAIL</b> ( ) <b>BORDERLINE</b> ( ) <b>PASS</b>					
<b>PASS:</b> Correct and safe examination with sufficient biopsies from predefined tumors					
<b>BORDERLINE:</b> Patient is not at risk, BUT more training needed					
<b>FAIL:</b> Performed risky procedure, e.g., biopsy of visceral pleura, insufficient material					

**Figure 1** The Local Anaesthetic Thoracoscopy Assessment Tool.

### *Participants*

Two groups of participants were enrolled. The inclusion criteria for the novice group were junior doctors who are in their clinical clerkship, with no experience in LAT, flexible bronchoscopy, or laparoscopy. For the experienced group, the inclusion criteria include respiratory consultants having performed more than 20 LAT procedures.

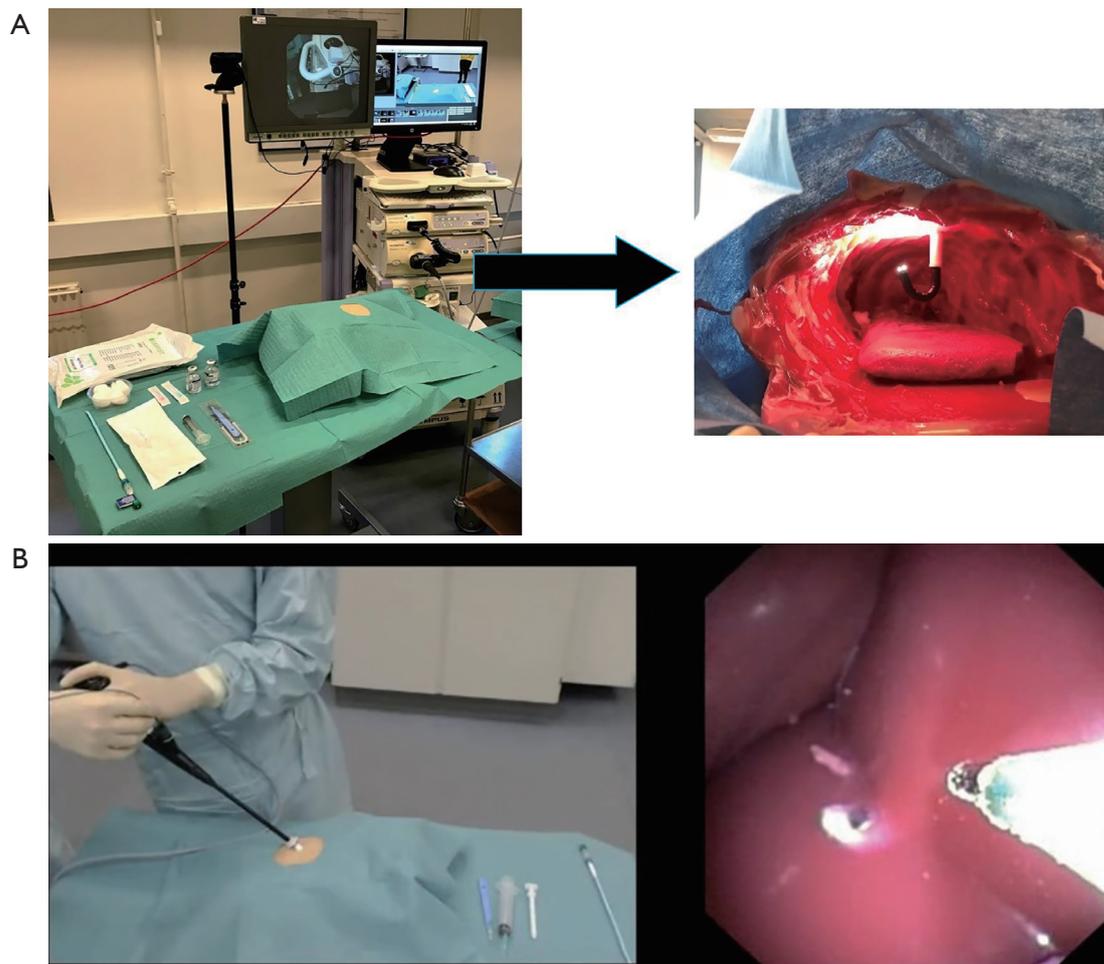
### *Preprocedural preparation*

The participants were provided with written material

outlining the stepwise approach to the procedure (14) as well as oral information on the set-up given by the primary investigator (LJN).

### *Video recording*

The procedures were video recorded with two cameras—one focused on the hands of the operator, while the other (Mediacapture USB200 device connected to the Olympus endoscopic tower), captured images from the screen. The videos were processed and merged using the VSDC Free



**Figure 2** Set-up of the simulation setting including video image presented to raters. (A) Set-up of the simulation setting; (B) video image presented to the raters.

*Video Editor 3.3.0.394* (Flash-Integro LLC). These were saved in an online hosting service (Vimeo, Inc., New York, USA) and were embedded in an online rating platform (Research Electronic Data Capture software [REDCap (15)], hosted at the Capital Region of Denmark). All videos were anonymized to allow for mutual blinded rating.

### *The simulated procedure*

Prior to the test, the participants performed a warm-up exercise including the insertion of the thoracoscope and exploration of the “cavity” using a plastic box specially designed by an engineer (MBSS) for endoscopy practice. The participants were instructed to find four letters (T, E, S, T) that were strategically placed inside the box. Thereafter, the participants performed LAT, starting from

instillation of local anesthesia, skin incision and insertion of trocar, insertion of scope and systematic exploration of the thoracic cavity, location and biopsy of tumors, and ending with insertion of the chest tube. They had to perform two procedures consecutively, with 5 minutes rest in between. LJN assisted as a nurse during all procedures, but did not offer any kind of advice to the doctors.

### *Pilot rating and rater calibration*

Four consultants with extensive experience in LAT were invited to assess all video performances (NR, NM, PFC, UB). An initial pilot rating of three movies was carried out by two of the raters (PFC, UB) to assess whether they have understood how the assessment tool is being used. After assessing each video individually, the raters compared their

**Table 1** Baseline characteristics of the participants enrolled in the study

Group	Age in years, median [range]	Number of bronchoscopies performed	Number of medical thoracoscopies performed unassisted
Novice (n=9)	29 [27–46]	0	0
Experienced (n=8)	48 [37–70]	50+	20–50+

scoring and discussed any discrepancies. The test items and anchor descriptors for 1, 3 and 5 of the rating scale were reworded as a result of the discussions. A final rater training followed where all four raters assessed the same three videos using the revised tool. A rater instruction was also sent out to all raters to guide them through the process.

### Ratings

Individual links were sent to each of the four raters through REDCap. Access to each other's rating profile was not possible, and the order of the 34 videos was individually randomised. All 136 assessment were automatically stored in REDCap and later transferred to SPSS Statistics version 25.0 (IBM, NY, USA) for statistical analysis.

### Statistical analysis

Cronbach's alpha was calculated to explore internal consistency of the test items and interrater reliability. The correlation between the total item score and the global rating was calculated using Spearman's rank correlation coefficient (16). Pearson's correlation coefficient was used to measure the correlation between the first and second performances, i.e., test-retest reliability. A generalisability study (G) and decision study (D) were performed to explore generalisability of the LAT assessment tool and the most optimal combinations of assessed procedures and raters (17). Differences in performance scores between the two groups were tested with independent samples *t*-test. We used the contrasting groups' method to determine a pass/fail standard, which is defined as the intercept between the normalised distribution of performance scores of the novices and experienced participants (18). All statistical analysis was performed using SPSS.

## Results

Ten novice and eight experienced participants were enrolled in the study. However, the videos of one novice were

corrupted and therefore excluded. In consequence, 9 novice participants were included. *Table 1* presents characteristics of the participants. All participants performed two consecutive procedures, generating a total of 34 video recordings that were rated by four raters resulting in 136 assessments for analysis. Evidence of validity from the five sources are summarised in *Table 2*.

### Content

The Local Anaesthetic Thoracoscopy Assessment Tool consists of eight procedural items and a global rating scale (*Figure 2*). The item regarding sterile technique was added after rater training. The global rating scale was included to ensure that the included items covered all important aspects of the procedure.

### Response process

Standardized data collection was ensured using the same researcher (LJN) to conduct the data collection throughout the study. Furthermore, the simulation-based setting provided a controlled environment for testing without the difference in procedural difficulty that is unavoidable in a clinical setting. Finally, the response process was also optimized by providing rater training and calibration before the actual rating of performances started.

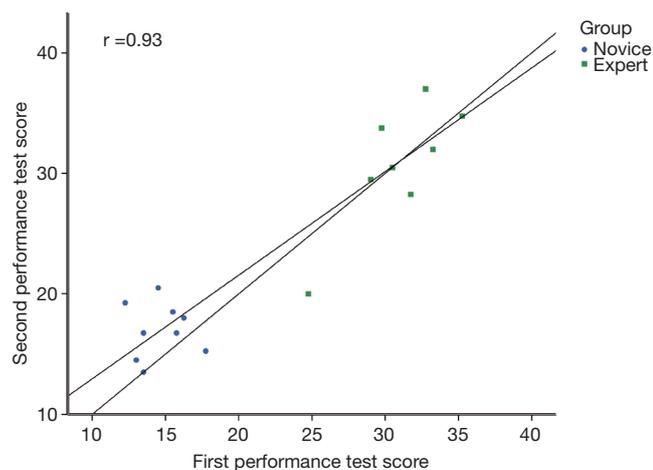
### Internal structure

The internal consistency reliability of the test items was high with a Cronbach's alpha =0.94, while the inter-rater reliability was 0.91. The total item score to global score analysis showed strong, significant correlation with a Spearman's rho of  $r_s=0.86$ ,  $P<0.001$ . The participants' first performance correlated to their second performance with a Pearson *r* of 0.93,  $P<0.001$  (*Figure 3*).

A generalisability study determined a generalisability G-coefficient of 0.92. The D-study estimated that four performances assessed by one rater is needed to be

**Table 2** Different sources of validity for the Local Anaesthetic Thoracoscopy (LAT) Assessment Tool

Sources of evidence for validity	Description of source	Validity evidence for the LAT Assessment tool
Content	The test content should measure the intended goals and objectives	The content of the test was developed by five experts in thoracoscopy and assessment in medical education. The test developers agreed on the design, items and descriptors for the anchors
Response process	Integrity of data should be maintained at all times. Test administration should be controlled or standardized at a maximum level possible	The same investigator facilitated the test to all participants, from providing instructions prior to the test to acting as a nurse during the procedure. All raters participated in an extensive rater training. All video recordings of the procedure were blinded and randomized
Internal structure	This refers to the reliability of the test results. The internal consistency of the test items should result to similar scores when a measuring the same construct. For high stakes assessment, a Cronbach's alpha of >0.90 is sound	The internal consistency and inter-rater reliability of the LAT assessment tool is high with Cronbach's alpha values of 0.94 and 0.91, respectively. The total item score is highly correlated with the global score, a Spearman's rho of $r_s=0.86$ , $P<0.0001$ . The generalisability G coefficient is 0.92, with the D-study estimating four performances rated by one rater to reach a G-coefficient of >0.80
Relations to other variables	Assessment scores should correlate with known measures of competence	The LAT assessment tool was able to significantly discriminate the performances between the different levels of experience with mean scores for novices and experienced of $30.80\pm 4.20$ and $15.83\pm 2.32$ respectively, with $P<0.0001$
Consequences	Consequences of testing is supported by the pass/fail standard that is set	The contrasting groups' method was used to set the pass/fail score at 22 points. None of the novices passed the test in the first and second performances. All experienced participants passed the first performance, while one did not pass in the second performance

**Figure 3** Correlation of first performance to the second performance.

sufficiently reliable (G-coefficient of >0.80). If there is only one video-recorded performance of the procedure, two raters are needed to reach an acceptable G-coefficient (Figure 4).

### Relations to other variables

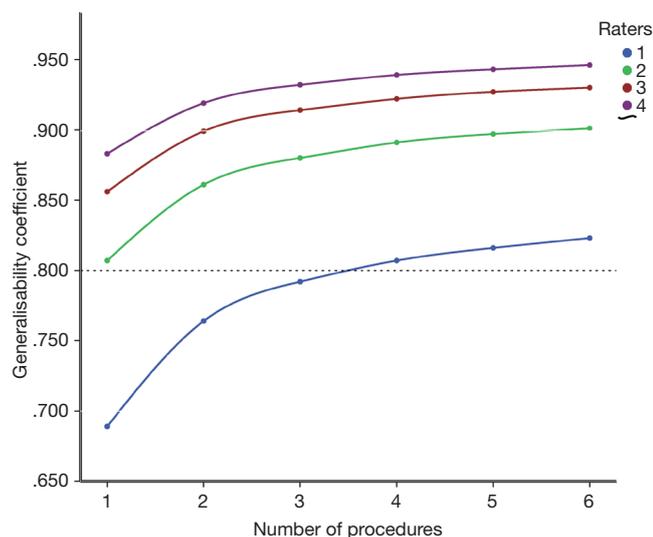
The LAT assessment tool was able to discriminate between the two groups in both performances. The experienced group performed significantly better with a mean score of  $30.8\pm 4.2$ , while the mean score for the novice group was  $15.8\pm 2.3$ ,  $P<0.001$ . Table 3 present a summary of the performances of both groups.

### Consequences

The pass/fail standard as determined by the contrasting groups' method was 22 points (Figure 5). None of the novices passed the test in either of the two performances. All experienced passed the test in the first performance, while one failed in the second performance (Figure 6).

### Discussion

We designed a procedure-specific assessment tool to assess competences in LAT and established validity evidence based



**Figure 4** Effect on G coefficient when an independent rater is added to assess LAT performance. LAT, local anaesthetic thoracoscopy.

on five sources to support its intended use, that is to assess competence in performing LAT. A pass/fail standard was also established to ensure that trainees have acquired the necessary skills to competently perform the procedure. This is the first study to describe an assessment tool for LAT, with extensive validity and a pass/fail standard to ensure proficiency.

Focused training and assessment is needed in LAT. Simulation is a safe and effective training ground for the acquisition of procedural skills and competences, as well as an ideal environment for structured assessment. In this study, we report a high validity in the interpretation of performance scores based on the newly developed LAT assessment tool.

### Validity of the assessment tool

Validity evidence for *content* was gathered by involving a group of experts within pulmonary medicine and medical education. One of the ways to devise test items is to involve experts with relevant knowledge and experience, to ensure that the content is representative of the construct that is intended to be investigated (16). This could be in a group discussion or using a more systematic approach such as the Delphi method as performed in other studies (19,20). The expert group included important steps when performing LAT. Being set in a simulation-based environment, sterile

technique was not initially considered, however after pilot rating, the raters highlighted the importance of maintaining sterility during the procedure, and added this as an additional item. Additionally, a global rating scale was added to allow us to test the completeness of the devised items.

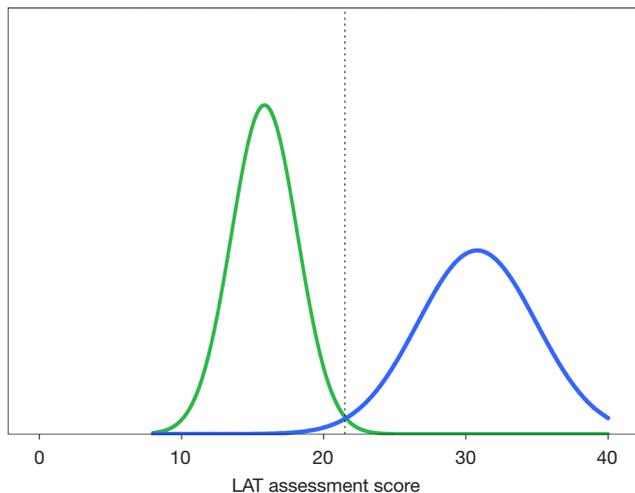
We optimized the *response process* to minimize any biases during the testing and assessment processes. Data collection was performed in a controlled and standardised environment using the same investigator who also acted as the nurse in all procedures. Furthermore, meticulous rater training was performed to improve reliability (21,22).

The *internal structure* of the LAT assessment tool was reflected by a high internal consistency reliability, indicating that the items correlate to the purpose of the assessment, which is competency in LAT. This corresponds with high stakes summative assessments such as certification or licensure examinations, where a reliability >0.90 is preferred (23). Additionally, the total item score correlated significantly with the global rating score, indicating that all items represent the same construct, that the items cover all important parts of the procedure, and that none should be discarded. Performing the procedure twice and being assessed by four raters increases the reliability, as reflected by a high G-coefficient. However, the D-study showed that to reach an acceptable reliability of >0.80, four video-recorded performances are needed if there is only one rater. Alternatively, one performance is needed when assessed by two raters to reach sufficient reliability. These approaches are more feasible, less expensive, and easier to administer than our experimental use of four raters.

The LAT assessment tool was able to significantly discriminate between the performances of the novices and experienced participants, indicating its capability to stratify different levels of experience. This strongly supports evidence for *relations to other variables*. The experienced group scored superiorly as compared to the novice group. None of the novice participants passed the test, however their performance scores improved in the second procedure as compared to the first attempt, signifying a learning effect. The novices were all doctors in their first year of clerkship who volunteered to participate despite having no experience in the procedure. The interest to learn was seen in their performances, from struggling to manipulate the thoracoscope in the first attempt to an evident improvement in the second performance. All the experienced participants passed the test in their first performance, however as shown in the test scores, some of them performed inferiorly with one out of nine failing in the second performance. One of

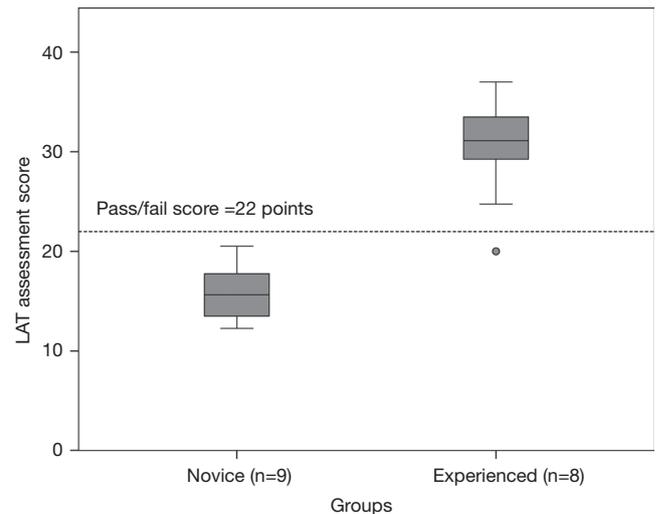
**Table 3** Summary of total mean performance scores of each group

Group	No. of video-recorded performances	Mean	Std. deviation	Std. error mean
Novice	18	15.83	2.32	0.55
Expert	16	30.80	4.17	1.04

**Figure 5** Establishing a pass/fail score using the contrasting groups' method to between the novice and experienced participants performance scores. A pass/fail score was set at 22 points.

the reasons could be that the experienced participants were highly competent and had already reached the plateau of the learning curve. Furthermore, the experienced participants were not accustomed to the simulation-based setting and consequently, were unable to apply the habits and tricks that they would do during real LAT procedures. Another reason could also be the differences in practices regarding LAT, including the continued use of rigid thoroscopes in some institutions. These findings are consistent with other simulation-based studies, where novices perform towards achieving a higher level of competency with training, while experienced or experts tend to underperform (21,24).

Establishing a pass/fail standard to define competence is a strength of the study. Standard setting serves as a benchmark or cutpoint between those who perform competently and those who do not (25). The changing education landscape towards competency-based education is accompanied by a demand for assessment methods to include standards for competence that will help with consequence decisions. We used the contrasting groups' method to define the pass/fail standard for LAT in a simulation-based setting, a method

**Figure 6** Box and whiskers plot showing the scores of the two groups, including the median, the minimum and maximum scores, and the pass/fail score.

that is well-accepted and is commonly used to establish standards for procedural skills (26).

The small sample size in this study is a limitation. The enrolment of more participants in each group might have given the study more power, increasing the reliability and generalisability of the results (27). Another limitation is the inclusion of novice participants who do not have any experience in LAT. Comparing novices to experts is one of the most common methods in validation studies but can inflate the reliability of the assessment (10,28). Nevertheless, the participants were not complete novices (ex. students) but were doctors during their first year of clerkship, and were assumed to have some skills and experiences in performing technical procedures. One way to minimise this spectrum bias is to include participants who are representative of the intended target population, i.e., respiratory physicians in different stages of the thoracoscopy education. The model used in the study did not allow ultrasonography to detect pleural effusion, which is a limitation since the use of pleural ultrasonography facilitates the insertion of trocar. However, this simulation set-up is directed towards novice trainees to

train the critical procedural steps before progressing on to the next stages that could include management and evaluation of complications such as bleeding, unexpected septation and etc.

### Future implications

The newly developed assessment tool in LAT is targeted towards a standardized performance of thoracoscopy and this procedural specificity makes it easier, more reliable, and more informative to use for feedback and certification. However, variations in how the procedure is performed (e.g., using ultrasound to ensure pneumothorax or using the camera during trocar insertion) can make it necessary to make smaller revisions of the tool to fit the local context. It can be used in a simulation-based environment and possibly also for assessment of supervised procedures on patients. Simulation with mastery learning provides an excellent opportunity to learn and train until a certain level of competency before supervised training with patients. Several studies have reported the implementation of simulation-based training programmes in respiratory medicine, including summative assessment as end-of-course assessment or certification (29-31). Although more research is needed such as its use in the clinical environment, the study is an important first step. We hope to implement a LAT simulation-based training programme for novices to train the different steps and skills until they pass the pass/fail standard towards competency. The overall goal is to increase the quality of LAT to improve the diagnosis and staging of patients with malignant pleural disease for better prognosis.

### Conclusions

The newly developed local anaesthetic assessment tool is an objective, procedure-specific assessment tool with solid evidence of validity. It can be used to provide structured, formative feedback for trainees (during simulation-based training), as a summative assessment at the end of the course, and possibly for certification before supervised practice in the clinical environment.

### Acknowledgments

*Funding:* None.

### Footnote

*Reporting Checklist:* The authors have completed the MDAR

checklist. Available at <https://dx.doi.org/10.21037/jtd-20-3560>

*Data Sharing Statement:* Available at <https://dx.doi.org/10.21037/jtd-20-3560>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://dx.doi.org/10.21037/jtd-20-3560>). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). It is an educational study that did not include patient participation and therefore did not require ethical approval in accordance with Danish law. Informed consent to participate in the study was obtained.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

### References

1. Rahman NM, Ali NJ, Brown G, et al. Local anaesthetic thoracoscopy: British Thoracic Society Pleural Disease Guideline 2010. *Thorax*. 2010;65 Suppl 2:ii54-60.
2. Willendrup F, Bodtger U, Colella S, et al. Diagnostic accuracy and safety of semirigid thoracoscopy in exudative pleural effusions in Denmark. *J Bronchology Interv Pulmonol* 2014;21:215-9.
3. Halifax RJ, Corcoran JP, Psallidas I, et al. Medical thoracoscopy: Survey of current practice-How successful are medical thoracoscopists at predicting malignancy? *Respirology* 2016;21:958-60.
4. Roberts ME, Neville E, Berrisford RG, et al. Management of a malignant pleural effusion: British Thoracic Society pleural disease guideline 2010. *Thorax* 2010;65 Suppl 2:ii32-ii40.

5. Rodríguez-Panadero F. Medical thoracoscopy. *Respiration* 2008;76:363-72.
6. Nayahangan LJ, Clementsen PF, Paltved C, et al. Identifying Technical Procedures in Pulmonary Medicine That Should Be Integrated in a Simulation-Based Curriculum: A National General Needs Assessment. *Respiration* 2016;91:517-22.
7. Konge L, Colella S, Vilmann P, et al. How to learn and to perform endoscopic ultrasound and endobronchial ultrasound for lung cancer staging: A structured guide and review. *Endosc Ultrasound* 2015;4:4-9.
8. Ernst A, Silvestri GA, Johnstone D. Interventional pulmonary procedures: Guidelines from the American College of Chest Physicians. *Chest* 2003;123:1693-717.
9. Scally G, Donaldson LJ. Clinical governance and the drive for quality improvement in the new NHS in England. *BMJ* 1998;317:61-5.
10. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)* 2016;1:31.
11. Lineberry M, Park YS, Cook DA, et al. Making the case for mastery learning assessments: Key issues in validation and justification. *Acad Med* 2015;90:1445-50.
12. Yudkowsky R, Park YS, Downing SM, editors. *Assessment in health professions education*. Routledge. 2019.
13. Messick S. 1989: Validity. In Linn RL, editor. *Educational measurement*. New York: Macmillan, 1989:13-103.
14. Loddenkemper R, Lee P, Noppen M, et al. Medical thoracoscopy/pleuroscopy: step by step. *Breathe* 2011;8:156-67.
15. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377-81.
16. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use. Devising Items*. 5th ed. Oxford: Oxford University Press, 2015:19-37.
17. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Med Teach* 2012;34:960-92.
18. Yudkowsky R, Park YS, Lineberry M, et al. Setting mastery learning standards. *Acad Med* 2015;90:1495-500.
19. Strøm M, Lönn L, Bech B, et al. Assessment of competence in EVAR procedures: a novel rating scale developed by the Delphi technique. *Eur J Vasc Endovasc Surg* 2017;54:34-41.
20. Savran MM, Clementsen PF, Annema JT, et al. Development and validation of a theoretical test in endosonography for pulmonary diseases. *Respiration* 2014;88:67-73.
21. Rasmussen KMB, Hertz P, Laursen CB, et al. Ensuring Basic Competence in Thoracentesis. *Respiration* 2019;97:463-71.
22. Robertson RL, Vergis A, Gillman LM, et al. Effect of rater training on the reliability of technical skill assessments: a randomized controlled trial. *Can J Surg* 2018;61:15917.
23. Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ* 2004;38:1006-12.
24. Dyre L, Norgaard LN, Tabor A, et al. Collecting Validity Evidence for the Assessment of Mastery Learning in Simulation-Based Ultrasound Training. *Ultraschall Med* 2016;37:386-92.
25. Norcini JJ. Setting standards on educational tests. *Med Educ* 2003;37:464-69.
26. Goldenberg MG, Garbens A, Szasz P, et al. Systematic review to establish absolute standards for technical performance in surgery. *Br J Surg* 2017;104:13-21.
27. Cook DA, Hatala R. Got power? A systematic review of sample size adequacy in health professions education research. *Adv Health Sci Educ Theory Pract* 2015;20:73-83.
28. Norman G. Data dredging, salami-slicing, and other successful strategies to ensure rejection: twelve tips on how to not get your paper published. *Adv Health Sci Educ Theory Pract* 2014;19:1-5.
29. Konge L, Vilmann P, Clementsen P, et al. Reliable and valid assessment of competence in endoscopic ultrasonography and fine-needle aspiration for mediastinal staging of non-small cell lung cancer. *Endoscopy* 2012;44:928-33.
30. Konge L, Arendrup H, Von Buchwald C, et al. Using performance in multiple simulated scenarios to assess bronchoscopy skills. *Respiration* 2011;81:483-90.
31. Konge L, Clementsen P, Larsen KR, et al. Establishing pass/fail criteria for bronchoscopy performance. *Respiration* 2012;83:140-46.

**Cite this article as:** Nayahangan LJ, Svendsen MBS, Bodtger U, Rahman N, Maskell N, Sidhu JS, Lawaetz J, Clementsen PF, Konge L. Assessment of competence in local anaesthetic thoracoscopy: development and validity investigation of a new assessment tool. *J Thorac Dis* 2021;13(7):3998-4007. doi: 10.21037/jtd-20-3560