Simulator training for endobronchial ultrasound: a randomised controlled trial

Lars Konge[1], Paul Frost Clementsen[2], Charlotte Ringsted[3], Valentina Minddal[2], Klaus Richter Larsen[4] and Jouke T. Annema[5,6]

Affiliations: [1]Centre for Clinical Education, University of Copenhagen and The Capital Region of Denmark, Copenhagen, Denmark. [2]Dept of Pulmonology, Gentofte Hospital, University of Copenhagen, Hellerup, Denmark. [3]The Wilson Centre and Dept of Anesthesiology, University of Toronto and University Health Network, Toronto, ON, Canada. [4]Dept of Pulmonology, Bispebjerg Hospital, University of Copenhagen, Copenhagen, Denmark. [5]Dept of Pulmonology, Leiden University Medical Center, Leiden, The Netherlands. [6]Dept of Pulmonology, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands.

Correspondence: Lars Konge, Centre for Clinical Education, Department 5404, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen, Denmark. E-mail: lkonge@yahoo.dk

ABSTRACT Endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) is very operator dependent and has a long learning curve. Simulation-based training might shorten the learning curve, and an assessment tool with solid validity evidence could ensure basic competency before unsupervised performance.

A total of 16 respiratory physicians, without EBUS experience, were randomised to either virtual-reality simulator training or traditional apprenticeship training on patients, and then each physician performed EBUS-TBNA procedures on three patients. Three blinded, independent assessor assessed the video recordings of the procedures using a newly developed EBUS assessment tool (EBUSAT).

The internal consistency was high (Cronbach's α=0.95); the generalisability coefficient was good (0.86), and the tool had discriminatory ability ($p<0.001$). Procedures performed by simulator-trained novices were rated higher than procedures performed by apprenticeship-trained novices: mean±SD are 24.2±7.9 points and 20.2±9.4 points, respectively; $p=0.006$. A pass/fail standard of 28.9 points was established using the contrasting groups method, resulting in 16 (67%) and 20 (83%) procedures performed by simulator-trained novices and apprenticeship-trained novices failing the test, respectively; $p<0.001$.

The endobronchial ultrasound assessment tool could be used to provide reliable and valid assessment of competence in EBUS-TBNA, and act as an aid in certification. Virtual-reality simulator training was shown to be more effective than traditional apprenticeship training.

@ERSpublications
Virtual-reality simulation-based training shortens the learning curve for endobronchial ultrasound (EBUS) http://ow.ly/OazC9

Received: Dec 23 2014 | Accepted after revision: May 14 2015 | First published online: July 09 2015

This article has supplementary material available from erj.ersjournals.com

Conflict of Interest: None declared.

Support statement: Simbionix (3D Systems, Cleveland, OH, USA) donated €8000 to cover expenses related to data collection (transportation between Denmark and the Netherlands). The company had no influence on the study design or presentation of data.

Introduction

Accurate staging of mediastinal lymph nodes is essential to ensure the correct treatment of patients with potentially resectable nonsmall cell lung cancer (NSCLC). Endobronchial ultrasonography (EBUS) and transoesophageal ultrasonography (EUS) have replaced surgical mediastinoscopy as the first choice to obtain tissue confirmation [1, 2]. Consequently, the availability of EBUS equipment has increased exponentially [3]. However, this rapid dissemination has occurred without a consensus on how operators should be trained and how procedural competency should be assessed. The diagnostic yield is highly operator-dependent, and the learning curve shows substantial variation between individual operators [4, 5]. The traditional apprenticeship model of EBUS training is not optimal, as trainee participation increases procedure time, amount of sedation used, and shows a trend towards increased complication rates [6].

The use of virtual reality simulators in training is gaining ground in educational environments for health professions. In comparison to no intervention, simulation is consistently associated with large effects for outcomes of knowledge, skills, and behaviours [7]. Two studies on virtual reality EBUS simulators have also shown promising results, but no randomised controlled trials comparing apprenticeship training to virtual-reality simulator training have been performed [8, 9]. Performing experimental studies in medical education is challenging [10]. Multicentre studies are often necessary in order to reach a sufficient sample size, and the validity of the outcome measure requires great attention; what defines competent performance of an EBUS procedure? A dichotomous outcome measure, such as diagnostic yield is not adequate to assess performance of individual procedures and is not a viable option in a supervised training environment, as yield is influenced by supervisors' assistance. Guidelines from the British Thoracic Society recommend focussing on monitoring an individual's performance, and state that standards for assessment of competency should be determined [11]. Ideally, evidence of the validity of these assessments should be gathered from all five sources in Messick's unitary framework of validity, which are: content, response process, internal structure, relationship to other variable, and consequences [12].

The aims of this study were to develop an assessment tool for measuring competency in EBUS-guided transbronchial needle aspiration (TBNA) and to establish evidence of validity for the tool, along with comparing the competency of trainees after traditional apprenticeship training on patients and virtual-reality simulator training, respectively.

Material and methods

Development of the assessment tool

The assessment tool was developed by a group consisting of two respiratory physicians, a thoracic surgeon, and a professor of medical education, all with considerable experience in performance, teaching, and validation of endoscopic ultrasound and other procedures [13–15]. The tool was designed according to the original format for "objective structured assessment of technical skills", in which each item is rated on a scale from 1 to 5, with descriptive anchors in the middle and at the ends, and re-coded into a score from 0 to 4 points [16]. Six items were designed to assess knowledge of the mediastinal anatomy, by requesting the operators to identify six anatomic landmarks: lymph node stations 4L, 7, 10L or 11L, 10R or 11R; the azygos vein; and lymph node station 4R. Four items related to the technical skills necessary to handle the scope and perform TBNA were defined: insertion of the endoscope, positioning of the transducer, use of sheath, and use of needle. Finally, two items were added to allow assessors to give their overall opinion on anatomic orientation and biopsy sampling, respectively. After pilot testing in Denmark and the Netherlands using both direct observation and video-based assessment, the 12-item endobronchial ultrasound assessment tool (EBUSAT) was finalised and a copy can be found in the online supplementary material.

Participants

Two experts and 16 trainees in endosonography were enrolled in the study. Figure 1 shows a flowchart of the study. The trainees were respiratory physicians in Denmark (n=8) and the Netherlands (n=8). Inclusion criteria were knowledge of mediastinal staging and experience in flexible bronchoscopy; exclusion criterion was former EBUS training. Both experts had been actively engaged in EBUS-TBNA for >10 years, thereby fulfilling the criteria for international expertise [17]. All participants were volunteers and signed informed consent at the time of inclusion. All procedures were performed in a supervised fashion similar to daily practice. All data was kept confidential, and according to the national legislation of both countries, the study was exempt from full ethical approval.

Training programme

All trainees attended a full-day theoretical EBUS course in either Denmark or the Netherlands with identical lectures and equipment demonstration given by the same faculty, and with no hands-on training. Special attention was given to making all participants familiar with the standardised approach to EBUS, as described earlier, and with the assessment tool that would be used to assess their competence. Thereafter,

DOI: 10.1183/13993003.02352-2015

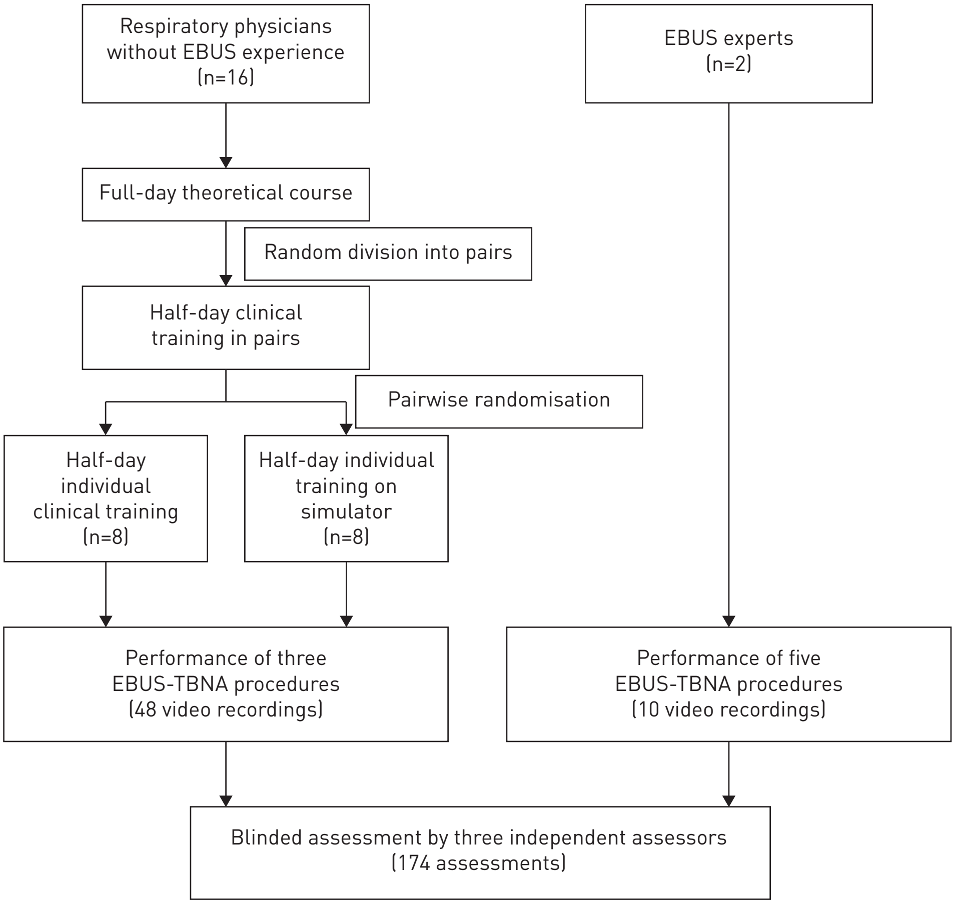

Respiratory physicians without EBUS experience (n=16)
EBUS experts (n=2)
Full-day theoretical course
Random division into pairs
Half-day clinical training in pairs
Pairwise randomisation
Half-day individual clinical training (n=8)
Half-day individual training on simulator (n=8)
Performance of three EBUS-TBNA procedures (48 video recordings)
Performance of five EBUS-TBNA procedures (10 video recordings)
Blinded assessment by three independent assessors (174 assessments)

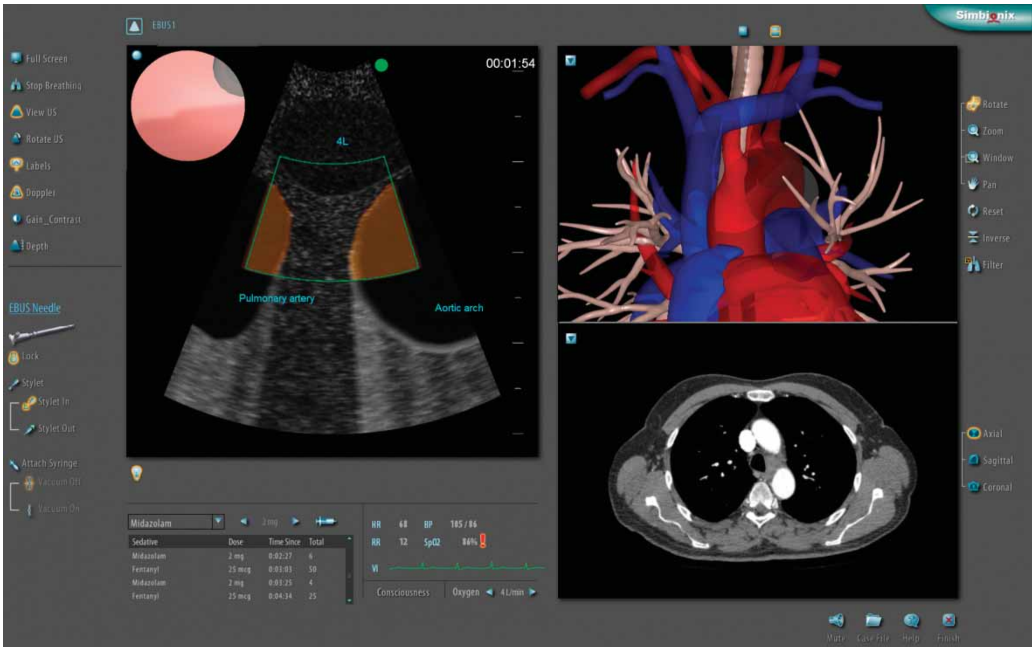
FIGURE 1 A flowchart demonstrating the study design showing the randomised controlled trial and the procedures by experts in endosonography and the additional data collection with regard to the validation study. EBUS-TBNA: endobronchial ultrasound-guided transbronchial needle aspiration.

the participants were randomly divided into eight pairs, and each pair had a half-day of clinical hands-on training, supervised by one of the two experts. Through a drawing of sealed envelopes by an independent nurse, the participants were then randomised into individual training on either patients (apprenticeship training) or a virtual reality simulator.

Apprenticeship training consisted of half a day of focussed supervised performance of EBUS-TBNA procedures. Each trainee performed two to four procedures. Whereas the virtual-reality simulator training consisted of half a day of hands-on training on the GI Bronch Mentor EBUS Simulator (Simbionix, Cleveland, OH, USA) (figure 2). The simulator consists of a proxy EBUS scope and TBNA needle, an interface to track the motions of the equipment, and a computer that generates endoscopic and ultrasound images. Each participant completed each of the six different training cases at least once. The same thoracic surgeon supervised all training sessions, in order to standardise the simulator training (that is the intervention). The total training time equalled the time for clinical training in the control group.

Testing of competence

Testing was performed as retention tests between 1 and 8 weeks after training completion. All test sessions were scheduled before the participants were randomised. The participants were not allowed to practice or perform EBUS procedures in the interval between training and retention testing. The test consisted of the performance of three EBUS-TBNA procedures in three consecutive patients. After introduction of the scope, the trainee had to identify the six anatomical landmarks (described previously) in the predefined order, followed by two transbronchial fine-needle aspirations of one lymph node station. All procedures were supervised by one of the two EBUS experts. The supervisor told the trainee which lymph node station to puncture, but otherwise did not interfere during the procedure, unless interference was essential for the patient or the equipment. Any verbal or manual intervention was noted. After testing all 16 trainees, the two experts performed five consecutive procedures each, in the exact same way as described above.



Midazolam 2 mg

Sedative	Dose	Time Since	Total
Midazolam	2 mg	0:02:27	6
Fentanyl	25 mcg	0:03:03	50
Midazolam	2 mg	0:03:25	4
Fentanyl	25 mcg	0:04:34	25

HR 68 BP 105 / 86
RR 12 SpO2 86%
VI
Consciousness
Oxygen 4 L/min

FIGURE 2 A screenshot from the virtual reality simulator used for training the simulator group.

The ultrasound images and the endoscopic images of all procedures were video recorded using the picture-in-picture function.

Scoring

The assessment process started 3 months after the last test was completed. We used three assessors: one EBUS expert involved in the development of the EBUSAT and two independent, external EBUS experts who only received written instructions on the use of the assessment tool. Each assessor received a portable hard drive with anonymised video recordings of the procedures, and independently assessed the procedures using the corresponding EBUSAT forms. If the trainee had required verbal assistance, the appropriate item was given a score of 1 point. If the supervisor had manually assisted the trainee, the score was reduced to 0 points for that item.

Statistical analysis

Internal consistency of the EBUSAT form was explored using Cronbach's α. Generalisability theory was used to give a combined estimate of the reliability of the assessment tool and to explore the different sources of variance [18]. A "decision study" was performed to explore the effect of changing the number of assessors and procedures assessed. We followed the recommendations by Downing [19] for all reliability indices: coefficients >0.7 were considered sufficient for formative assessment, coefficients >0.8 were considered good (suitable for summative assessment), and coefficients >0.9 were considered excellent. EBUSAT scores of procedures performed by different groups were compared using the Mann–Whitney test. Item scores were compared using independent samples t-tests. All p-values <0.05 were considered statistically significant. A pass/fail score was established using the contrasting-groups method [20]. The consequences of the standard regarding pass/fail within the three groups were reported using frequencies and explored using Fisher's exact test.

The G-string IV statistical software package (Papaworx, Hamilton, ON, Canada) was used for the generalisability analyses; all other analyses were performed using PASW, version 20.0 (SPSS Inc, Chicago, IL, USA).

Results

Evidence for validity

A summary of all gathered evidence of validity for the EBUSAT form is shown in table 1. The internal consistency of the EBUSAT was high, Cronbach's α=0.95. The correlations between the two overall items ("orientation overall" and "biopsy sampling overall") and the underlying specific items were high: Pearson's r=0.88 and 0.86, respectively (p<0.001 for both correlations). The generalisability coefficient was good for our setup, 0.86. Table 2 shows the different sources of variance. More than half of the variance

TABLE 1 Overview of the validity evidence for the endobronchial ultrasound (EBUS) assessment tool (EBUSAT) using Messick's framework of validity

Source of evidence for validity	Examples of questions related to each source of evidence	Validity evidence for the EBUSAT
Content	Is the construct of interest covered by the assessment tool?	Uniform agreement between experts in the field on design, items and anchors
Response process	Does the assessment process eliminate sources of error to the maximum extent possible?	Allows blinded (unbiased) assessment Uses global (not checklist) rating scores
Internal structure	Does the assessment tool have good psychometric characteristics?	Excellent internal consistency Cronbach's α=0.95 Good reliability: generalisability coefficient=0.86
Relations to other variable	Does the assessment score correlate with known measures of competence?	EBUS experts score significantly better than EBUS novices: p<0.001
	Can the validity evidence be generalised?	The tool was used with success in two different countries and with three different assessors
Evidence based on consequences of testing	What are the consequences of the assessment in terms of pass/fail decisions?	17% of procedures by apprenticeship-trained novices passed, 33% of procedures by simulator-trained novices passed, and 90% of procedures by EBUS experts passed

originated from differences among participants (the facet of interest), and the second largest source of variance was the difference in patient cases' difficulty with disagreement between assessors only accounting for a small part of the variance. Figure 3 shows the results of the D-study, demonstrating the reliability of the EBUSAT when one, two, or three assessors assess one to eight procedures.

Procedures performed by EBUS experts received a significantly higher score than procedures performed by novices, mean±SD scores were 35.2±9.4 points *versus* 22.3±9.0 points, respectively, p<0.001. Performance of TBNA was more difficult than the identification of anatomical landmarks; the items "orientation overall" and "biopsy sampling overall" scored 2.0 points and 1.6 points on average, respectively, p<0.001. Stations 4R, 10R, and 10L were relatively difficult to identify, resulting in mean item scores of 1.9, 1.7, and 1.9 points, respectively. Stations 4L, 7, and the azygos vein were easier, receiving item scores of 2.5, 2.3, and 2.3 points, respectively. There was also a considerable difference in the item scores concerning biopsy sampling: "use of needle" received the lowest score of all items, with 1.4 points; "positioning of transducer" scored 1.7 points; and "use of sheath" scored 2.6 points.

A pass/fail standard of 28.9 points was established using the contrasting group method (figure 4). Only one of the procedures performed by experts (10%) scored below this standard, whereas 20 (83%) and 16 (67%) procedures performed by apprenticeship-trained novices and simulator-trained novices failed the test, respectively, p<0.001.

TABLE 2 Results from the G-Study indicating the contribution of each source of variance

Source of variance	Description	Estimated VC	Relative contribution %	Interpretation of results
Physicians	Systematic variation among physicians	57.8	52.1	Most of the variance derives from different competence levels between the physicians
Assessors	Systematic variability among assessors	4.77	4.3	The assessors had a high degree of agreement
Interaction between physician and assessor	Consistent trend for a assessor to access a particular physician differently	2.08	1.9	There was no bias between assessor and physician due to successful blinding
Interaction between cases and physicians	Systematic variability among cases	12.30	11.1	Some difference in the difficulty of the EBUS-TBNA cases
Interaction between case, assessor, and physician	All remaining variability	34.11	30.7	Expected unexplained error

VC: variance component; EBUS-TBNA: endobronchial ultrasound-guided transbronchial needle aspiration.

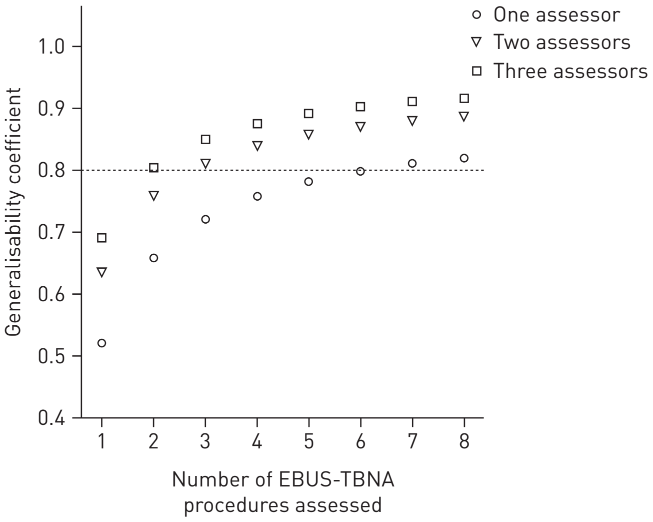


FIGURE 3 Resulting generalisability coefficient when one, two, or three assessors assess from one to eight endobronchial ultrasound (EBUS)-guided transbronchial needle aspiration procedures using the EBUS assessment tool. The dotted line indicates the reliability needed for summative assessment (0.8).

Virtual-reality simulator training versus *traditional apprenticeship training*

Procedures performed by simulator-trained novices were rated higher than procedures performed by apprenticeship-trained novices: mean±SD 24.2±7.9 points and 20.2±9.4 points, respectively, p=0.006 (figure 5). Simulator training resulted in a higher score for anatomical orientation (14.8±6.0 points *versus* 12.0±6.5 points, p=0.007), as well as for technical skills 9.6±3.8 points *versus* 8.2±4.1 points, p=0.023.

Discussion

We developed a tool for assessing competence in EBUS-TBNA (EBUSAT) and gathered evidence of validity from the five sources in the unitary framework (table 1 and online supplementary material). Physicians randomised to virtual-reality simulator training received higher EBUSAT scores on blinded assessments of EBUS-TBNA procedures than traditional apprenticeship-trained physicians (figure 5).

Validity evidence for the assessment tool

Development by a faculty with expertise in both endosonography and assessment tools lends credibility to the content of the EBUSAT; the items are representative of the important issues defining an EBUS procedure. The content validity is further supported by the good correlation between the overall items (“orientation overall” and “biopsy sampling overall”) and the underlying specific items.

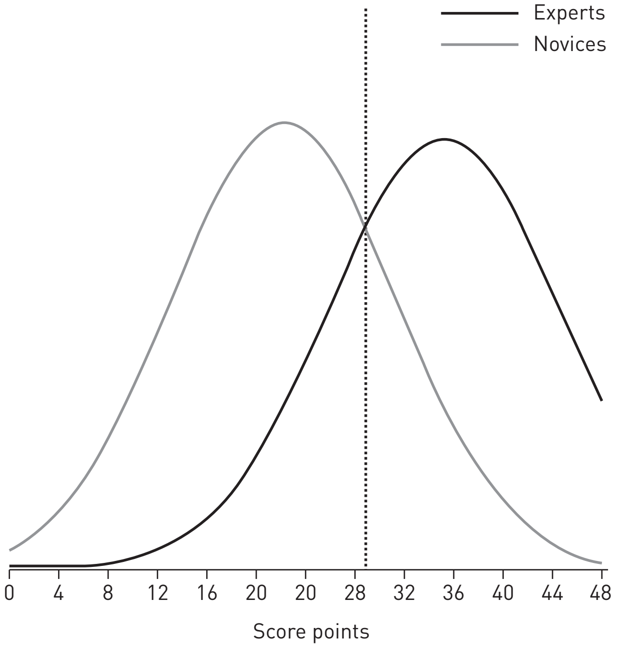


FIGURE 4 Establishing a pass/fail standard using the contrasting groups method. The dotted line represents the pass/fail standard at 28.9 points.

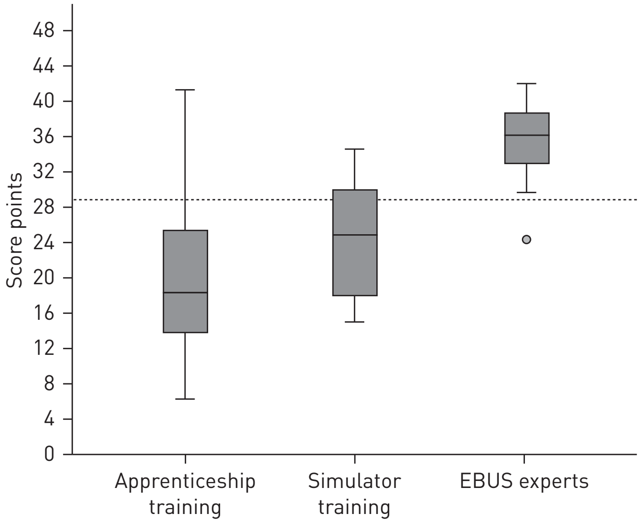


FIGURE 5 Endobronchial ultrasound (EBUS) assessment tool scores of procedures performed by the apprenticeship-trained novices, simulator-trained novices, and EBUS experts, respectively. Boxplot showing outliers, minimum, first quartile, median, third quartile and maximum. Dotted line represents pass/fail standard at 28.9 points.

We took great care to eliminate sources of error in the "response process". The thoroughly tested objective structured assessment of technical skill format allows for graduated judgments of competence, as well as registration of an overall impression of incompetence caused by one or more errors that only result in a minimal reduction in the checklist score. Furthermore, the dichotomous nature of checklists introduces a significant ceiling effect that is unsuitable for measuring nuances of proficiency [21, 22]. Wahidi *et al.* [23] used a checklist to assess when 13 trainees could independently perform a successful EBUS-TBNA procedure, and found that 25%, 50%, and 75% of the trainees did so after an average of five, nine, and 13 procedures, respectively. A study on central venous catheterisation skills found that a number of incompetent trainees committing serious procedural errors still managed to achieve a high checklist score (⩾80%) [24]. EBUSAT was developed to allow blinded assessments based on anonymised recordings of EBUS-TBNA procedures, and the generalisability analysis (table 2) showed that this blinding was successful; interaction between assessor and physician accounted for <2% of the variance. Blinding eliminated a major threat to validity, that of bias in the assessor–trainee relationship during direct observation. A study on assessment of performance in EUS showed that consultants achieved significantly higher scores when the assessors knew their identity, whereas the opposite was true for trainees [15].

The psychometric characteristics ("internal structure") of the EBUSAT tool proved satisfactory. Cronbach's α showed good internal consistency, and generalisability analysis showed that disagreement between assessors only accounted for 4.3% of the variance (showing a good inter-assessor reliability). The variance due to performance of different cases (test–retest reliability) was considerably larger; this was expected, as we used consecutive patients of differing difficulty. This finding underlines the importance of assessing multiple procedures to reach a reliable judgment regarding the competence of a trainee [13]. The decision study showed that several feasible combinations of the number of assessors and the number of procedures resulted in acceptable generalisability coefficients (figure 3). Assessments of three procedures by a single assessor, *e.g.* the supervisor, results in a coefficient >0.7 and is sufficient for a formative assessment (feedback). Two assessors assessing three procedures (or three assessors assessing two procedures) are necessary to achieve a coefficient >0.8 for high-stakes summative assessment (certification). This corresponds to similar studies on performance assessment of oesophageal and abdominal ultrasound skills supporting the generalisability of our results [15, 25].

Procedures that were performed by experienced operators scored significantly higher than procedures performed by novices (p<0.001), providing important validity evidence regarding discriminatory ability. Our findings are based on data from two countries and three different assessors who only received written instructions. Thus, the EBUSAT instrument is probably feasible for use in other institutions.

The shift towards competency-based medical education, with the introduction of assessment tools, makes it important to explore the "consequences of testing" with regard to pass/failure [26]. We used a credible standard-setting method to establish a pass/fail score; only one of the procedures performed by the experts received a lower score, whereas most procedures performed by trainees failed to meet the criterion.

DOI: 10.1183/13993003.02352-2015

Virtual-reality simulator training versus *traditional apprenticeship training*

An important part of this study was to explore if virtual-reality simulator training could replace apprenticeship training in the initial part of the learning curve. We found that the simulator-trained novices scored significantly higher than novices who had trained on real patients and were supervised by EBUS experts (p=0.006). The current study is the largest EBUS-training study, and the first randomised study to use performance on real patients as an outcome parameter. While we did not investigate why simulation-based training was more efficient than apprenticeship training, we suspect that it is due to the different nature of the two training modalities: virtual reality simulators allow trainees maximum hands-on time in a standardised and relatively stress-free environment, whereas clinical training is naturally dependent on the available patients. Moreover, some cases can be too challenging for novices, and there will often exist some waiting time between procedures. Also, especially in the early part of the learning curve, the supervisor will often take control over the procedure due to concerns regarding the patient, the equipment, or time constraints.

A systematic EBUS-TBNA training programme should not be based on virtual-reality simulator training alone; simulator training can only replace the initial part of the learning curve (figure 6). Our results confirm that trainees should not be considered fully competent after training on a virtual reality simulator (figure 5). We propose a three-step approach consisting of learning the necessary anatomy and theory (step one), simulation-based training (step two), and supervised practice on patients (step three), before performing independent procedures. Testing can ensure basic competency and has been shown to accelerate learning and improve retention [27]. Thus, we propose that all three steps should end with a test of competence before proceeding to the next step. Tests with validity evidence have been published regarding theoretical knowledge [28], performance on EBUS-TBNA simulators [29, 30] and performance on patients EBUS-STAT [21] and EBUSAT (current study).

Our study has several limitations. Even though it is the largest EBUS training study to date, we acknowledge that 16 respiratory physicians is still a relatively small number. Unfortunately, this is often the case in medical education research due to feasibility issues and scarcity of participants suitable for inclusion, *e.g.* the two randomised studies performed on virtual reality bronchoscopy simulators included six and 10 participants, respectively [31, 32]. Our study had sufficient power to detect the differences in performance between the two groups (which was ∼20%). Another limitation relates to the outcome measure (EBUSAT score). Ideally, training studies should show better patient outcomes in terms of morbidity and mortality or use clinical outcome measures such as diagnostic yield. However, this would require a large number of trainees performing unsupervised procedures, which does not seem feasible or ethically acceptable. For this limitation, we believe that the best possible solution is the assessment of multiple procedures by multiple blinded assessors using an assessment tool with solid evidence for validity from multiple sources. The validity of an assessment tool is dependent on the context in which it is being used, and the issue of generalisability should always be contemplated. A recent review of assessment tools found that a vast majority of studies used “an outdated framework on the basis of types of validity”; the systematic exploration of the EBUSAT, using an accepted framework, is a major strength of our study [33]. However, it is important to acknowledge that the EBUSAT was only developed to test anatomical orientation and technical skills, and other important competences such as theoretical knowledge, communication with the patient, and the ability to work in a team, should also be assessed.

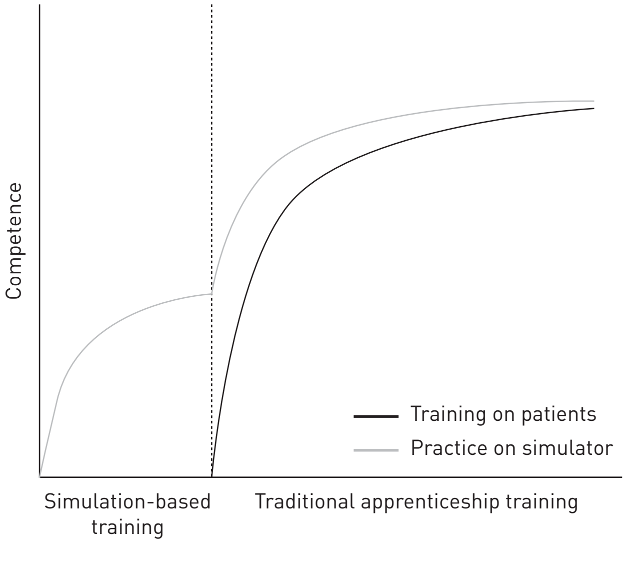


FIGURE 6 Graph illustrating two approaches to procedural training: practicing on simulators before performing procedures on patients (dotted line) and initial training on patients (solid line). The area between the curves represents the potential benefit of simulation-based training.

Conclusion

EBUSAT is the first assessment tool that allows for a blinded assessment of clinical EBUS-TBNA performance. This study gathered evidence from all five sources of evidence for validity in two different countries and using three independent assessors, making it highly probable that our findings can be generalised to other settings. A credible pass/fail standard was established, making it possible to use the EBUSAT as an aid in certification. Virtual-reality simulator training was more effective than traditional apprenticeship training in the initial part of the learning curve.

Acknowledgements

The authors would like to thank all the respiratory physicians who participated in the study

References

1 Vilmann P, Clemensten PF, Colella S, *et al.* Combined endobronchial and oesophageal endosonography for the diagnosis and staging of lung cancer. *Eur Respir J* 2015; 46: 40–60.

2 Silvestri GA, Gonzalez AV, Jantz MA, *et al.* Methods for staging non-small cell lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013; 143: Suppl. 5, e211S–e250S.

3 Tanner NT, Pastis NJ, Silvestri GA. Training for linear endobronchial ultrasound among US pulmonary/critical care fellowships: a survey of fellowship directors. *Chest* 2013; 143: 423–428.

4 Kemp SV, El Batrawy SH, Harrison RN, *et al.* Learning curves for endobronchial ultrasound using cusum analysis. *Thorax* 2010; 65: 534–538.

5 Steinfort DP, Hew MJ, Irving LB. Bronchoscopic evaluation of the mediastinum using endobronchial ultrasound: a description of the first 216 cases carried out at an Australian tertiary hospital. *Intern Med J* 2009; 41: 815–824.

6 Stather DR, Maceachern P, Chee A, *et al.* Trainee impact on advanced diagnostic bronchoscopy: an analysis of 607 consecutive procedures in an interventional pulmonary practice. *Respirology* 2013; 18: 179–184.

7 Cook DA, Hatala R, Brydges R, *et al.* Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA* 2011; 306: 978–988.

8 Stather DR, Maceachern P, Rimmer K, *et al.* Assessment and learning curve evaluation of endobronchial ultrasound skills following simulation and clinical training. *Respirology* 2011; 16: 698–704.

9 Stather DR, Maceachern P, Chee A, *et al.* Evaluation of clinical endobronchial ultrasound skills following clinical *versus* simulation training. *Respirology* 2012; 17: 291–299.

10 Ringsted C, Hodges B, Scherpbier A. 'The research compass': an introduction to research in medical education: AMEE Guide No. 56. *Med Teach* 2011; 33: 695–709.

11 Du Rand IA, Barber PV, Goldring J, *et al.* British Thoracic Society guideline for advanced diagnostic and therapeutic flexible bronchoscopy in adults. *Thorax* 2011; 66: Suppl. 3, iii1–iii21.

12 Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ* 2003; 37: 830–837.

13 Konge L, Larsen KR, Clementsen P, *et al.* Reliable and valid assessment of clinical bronchoscopy performance. *Respiration* 2012; 83: 53–60.

14 Konge L, Lehnert P, Hansen HJ, *et al.* Reliable and valid assessment of performance in thoracoscopy. *Surg Endosc* 2012; 26: 1624–1628.

15 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: 928–933.

16 Martin JA, Regehr G, Reznick R, *et al.* Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997; 84: 273–278.

17 Ericsson KA. Enhancing the development of professional performance: implications from the study of deliberate practice. *In*: Ericsson KA, ed. Development of Professional Expertise. 1st Edn. New York, Cambridge University Press, 2009; pp. 405–431.

18 Brennan RL. Generalizability Theory. 1st Edn. New York, Springer-Verlag, 2001.

19 Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ* 2004; 38: 1006–1012.

20 Yudkowsky R, Downing SM, Tekian A. Standard setting. *In*: Downing SM, Yudkowsky R, eds. Assessment in Health Professions Education. 1st Edn. New York, Routledge, 2009; pp. 119–148.

21 Davoudi M, Colt HG, Osann KE, *et al.* Endobronchial ultrasound skills and tasks assessment tool: assessing the validity evidence for a test of endobronchial ultrasound-guided transbronchial needle aspiration operator skill. *Am J Respir Crit Care Med* 2012; 186: 773–779.

22 Ilgen JS, Ma IW, Hatala R, *et al.* A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ* 2015; 49: 161–173.

23 Wahidi MM, Hulett C, Pastis N, *et al.* Learning experience of linear endobronchial ultrasound among pulmonary trainees. *Chest* 2014; 145: 574–578.

24 Ma IW, Zalunardo N, Pachev G, *et al.* Comparing the use of global rating scale with checklists for the assessment of central venous catheterization skills using simulation. *Adv Health Sci Educ Theory Pract* 2012; 17: 457–470.

25 Todsen T, Tolsgaard MG, Olsen BH, *et al.* Reliable and valid assessment of point-of-care ultrasonography. *Ann Surg* 2015; 261: 309–315.

26 Reznick RK, MacRae H. Teaching surgical skills--changes in the wind. *N Engl J Med* 2006; 355: 2664–2669.

27 Kromann CB, Jensen ML, Ringsted C. The effect of testing on skills learning. *Med Educ* 2009; 43: 21–27.

28 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.

29 Konge L, Annema J, Clementsen P, *et al.* Using virtual-reality simulation to assess performance in endobronchial ultrasound. *Respiration* 2013; 86: 59–65.

30 Stather DR, Maceachern P, Rimmer K, *et al.* Validation of an endobronchial ultrasound simulator: differentiating operator skill level. *Respiration* 2011; 81: 325–332.

DOI: 10.1183/13993003.02352-2015

31 Ost D, DeRosiers A, Britt EJ, *et al.* Assessment of a bronchoscopy simulator. *Am J Respir Crit Care Med* 2001; 164: 2248–2255.

32 Blum MG, Powers TW, Sundaresan S. Bronchoscopy simulator effectively prepares junior residents to competently perform basic clinical bronchoscopy. *Ann Thorac Surg* 2004; 78: 287–291.

33 Ghaderi I, Manji F, Park YS, *et al.* Technical skills assessment toolbox: a review using the unitary framework of validity. *Ann Surg* 2015; 261: 251–256.