Development of a Bronchoscopy-Radiologic Skills and Task Assessment Tool (BRadSTAT): A Tool for Evaluating the Radiological Skills of Bronchoscopists with Different Experience

Elaine L.C. Yap[a]   Alain C. Vandal[b, c]   Jonathan P. Williamson[d, e]   Phan Nguyen[f]
Henri Colt[g, h]

[a]Department of Respiratory Medicine, Middlemore Hospital, Counties Manukau District Health Board, Auckland, New Zealand; [b]Department of Statistics, The University of Auckland, Auckland, New Zealand; [c]Ko Awatea Research and Evaluation Office, Counties Manukau Health, Auckland, New Zealand; [d]South Western Sydney Clinical School, Liverpool Hospital, University of NSW, Kensington, NSW, Australia; [e]MQ Health Respiratory and Sleep, Macquarie University Hospital, Macquarie Park, NSW, Australia; [f]The Royal Adelaide Hospital, Adelaide, SA, Australia; [g]Emeritus Professor, University of California Irvine, Irvine, CA, USA; [h]Founder, Bronchoscopy International, Laguna Beach, CA, USA

Abstract

Background: Competency using radiologic images for bronchoscopic navigation is presumed during subspecialty training, but no assessments objectively measure combined knowledge of radiologic interpretation and ability to maneuver a bronchoscope into peripheral airways. ***Objectives:*** The objectives of this study were (i) to determine whether the Bronchoscopy-Radiology Skills and Tasks Assessment Tool (BRadSTAT) discriminates between bronchoscopists of various levels of experience and (ii) to improve construct validity using study findings. ***Methods:*** BRadSTAT contains 10 questions that assess chest X-ray and CT scan interpretation using multiple images per question and 2 technical skill assessments. After administration to 33 bronchoscopists (5 Beginners, 9 Intermediates, 10 Experienced, and 9 Experts), discriminative power was strengthened using differential weighting on CT-related questions, producing the BRadSTAT-CT score. Cut points for both scores were determined via cross-validation. ***Results:*** Mean BRadSTAT scores for Beginner, Intermediate, Experienced, and Expert were 74 (±13 SD), 78 (±14), 86 (±9), and 88 (±8), respectively. Statistically significant differences were noted between Expert and Beginner, Expert and Intermediate, and Experienced and Beginner (all $p \leq 0.05$). Mean BRadSTAT-CT scores for Beginner, Intermediate, Experienced, and Expert were 63 (±14), 74 (±15), 82 (±13), and 90 (±9), respectively, all statistically significant ($p \leq 0.03$). Cut points for BRadSTAT-CT had lower sensitivity but greater specificity and accuracy than for BRadSTAT. ***Conclusion:*** BRadSTAT represents the first validated assessment tool measuring knowledge and skills for bronchoscopic access to peripheral airways, which discriminates between bronchoscopists of various experience levels. Refining BRadSTAT produced the BRadSTAT-CT, which had higher discriminative power. Future studies should focus on their usefulness in competency-based bronchoscopy programs.

© 2022 The Author(s).
Published by S. Karger AG, Basel

Correspondence to:
Elaine L.C. Yap, eyap@middlemore.co.nz

Introduction

Technological advances, associated with modern imaging techniques such as radial probe endobronchial ultrasound (EBUS) and multi-row detector computed tomography (CT), render the ability to accurately select endobronchial paths to peripheral pulmonary lesions and skillfully navigate peripheral airways more important than ever before. Competency using radiologic images to navigate the airways is presumed during hands-on subspecialty education, but no existing validated assessment tools measure knowledge and technical skill in this area. Current assessment tools focus on dexterity and handling of the bronchoscope and knowledge of the bronchoscopic airway anatomy [1–3] but do not address the radiological component. The purpose of this study was to evaluate a novel instrument, the Bronchoscopy-Radiology Skills and Tasks Assessment Tool (BRadSTAT), that objectively tests a user's combined knowledge of radiologic interpretation and ability to maneuver a flexible bronchoscope into peripheral airway regions.

Methods

BRadSTAT Design and Scoring

BRadSTAT is designed for bronchoscopists with passing scores on the Bronchoscopy Skills and Task Assessment (BSTAT) [1] to ensure documented knowledge of airway anatomy and precise bronchoscopic handling. The assessment tool (Fig. 1) contains 10 questions that assess chest X-ray (CXR) (questions 1, 2, and 7) and CT scan (questions 3–6, 8–10) interpretation, using multiple images per question. Each image is scored separately and equally for a total possible score of 10 for each question. For example, if a question contains five images, each has a score of 2 if answered correctly and 0 if answered incorrectly. The maximum possible BRadSTAT score therefore is 100. Airway and bronchopulmonary segmental anatomy of the right and left lungs are addressed separately. All questions require users to mentally convert 2-dimensional images to a 3-dimensional perspective. Users must (1) match images of normal lobar and segmental anatomy with their corresponding anatomic descriptions (questions 1–6); (2) match abnormalities on CXR and CT with their corresponding descriptions of location (questions 7 and 8); and (3) identify the location of an abnormality on CT before navigating the bronchoscope to the appropriate pulmonary segment (questions 9 and 10).

BRadSTAT Administration

Radiologic images for questions 5 and 6 are viewed as still photos on the assessment tool or in the slideshow mode (online suppl. Questions 5, 6; see www.karger.com/doi/10.1159/000526011 for all online suppl. material) to mimic how users scroll through images in real life. Technical skill (questions 9 and 10) is measured using an airway model or simulator. We used the commercially available, previously validated ORSIM high-fidelity simulator (Auckland, New Zealand) [4, 5] made of a proxy flexible broncho-scope, an interface device into which the bronchoscope is inserted, and a laptop with simulation software. Sensors detect lever movements and bronchoscope rotation. Maneuvering feels realistic and virtual airway images appear true to life.

Study Participants

Participants were trainees in medical thoracic training programs and pulmonary consultants. All were asked to complete a questionnaire regarding bronchoscopy experience (number of procedures and whether they performed transbronchial lung biopsy, radial EBUS, or navigational bronchoscopy independently) and year of training. Participants were categorized into four groups: *Beginners* were trainees in the first half of their 3-year program who had completed <100 bronchoscopies; *Intermediates* were in the second half of their training who had completed between 100 and 200 bronchoscopies; *Experienced* were consultants who had performed >200 bronchoscopies; and *Experts* were interventional consultants who performed navigational bronchoscopy and radial EBUS.

BRadSTAT was administered over a 6-month period by examiners (E.Y., J.W., P.N.) at three different hospitals (Middlemore Hospital, Auckland, Liverpool and Macquarie Hospitals, Sydney, and Royal Adelaide Hospital, Adelaide) with interventional pulmonology units accredited for advanced training in thoracic medicine with the Royal Australasian College of Physician (RACP). Some Experts were recruited in Auckland during the June 2019 Australia-New Zealand Interventional Pulmonology (ANZIP) meeting.

Testing Protocol

BRadSTAT was administered to participants during a single session. The time for each participant to complete the assessment was noted. Questions 1–8 were addressed using a combination of paper-based and laptop computer images without examiner assistance. For questions 9 and 10 (technical skill), examiners asked participants to verbally identify the location of an abnormality before maneuvering the bronchoscope to that target location. Questions were scored as incorrect if either verbal or maneuvering portions were erroneous.

Study Aims

Our primary aim was to determine whether BRadSTAT scores discriminate between bronchoscopists of varying levels of experience, thus establishing construct validity. The secondary aim was to analyze data to further refine or improve BRadSTAT as an assessment tool.

Statistical Methods

Statistical analysis was performed in three phases by an investigator (A.V.) who neither knew nor observed study participants. In the first phase, BRadSTAT scores were assessed visually, overall, and question by question. Descriptive models were produced with total and question scores linearly regressed on the group. The BRadSTAT score's association with the groups was tested using the Jonckheere-Terpstra procedure [6]. Questions were assessed for internal consistency using Cronbach's α.

A second phase analysis was performed to determine if BRadSTAT could be refined. Different weightings to each question were assessed. Thirteen different methods were assessed, including the original BRadSTAT score weighting which allotted equal weight to

Bronchoscopy education

Bronchoscopy/Radiology Skills and Tasks Assessment Tool
(BRadSTAT)

Name: _____ Training Year _____ Date_____

Educational item Tick ☑ each correct answer	Score
1. Chest radiograph – identify lobe or segment: right lung (match images) ☐ Image A ☐ Image B ☐ Image C ☐ Image D ☐ Image E	Each answer 2 points Score____/10
2. Chest radiograph – identify lobe or segment: left lung (match images) ☐ Image A ☐ Image B ☐ Image C ☐ Image D ☐ Image E	Each answer 2 points Score____/10
3. CT – identify airways: right lung (match images) ☐ Image A ☐ Image B ☐ Image C ☐ Image D ☐ Image E	Each answer 2 points Score____/10
4. CT – identify airways: left lung (match images) ☐ Image F ☐ Image G ☐ Image H ☐ Image I ☐ Image J	Each answer 2 points Score____/10
5. CT – identify bronchopulmonary segments: right lung (match images) ☐ RB 1 apical RUL ☐ RB2 posterior RUL ☐ RB3 anterior RUL ☐ RB4 lateral RML ☐ RB5 medial RML ☐ RB6 superior RLL ☐ RB7 mediobasal RLL ☐ RB8 anterobasal RLL ☐ RB9 laterobasal RLL ☐ RB10 posterobasal RLL	Each answer 1 point Score____/10
6. CT – identify bronchopulmonary segments: left lung (match images) ☐ LB1+2 apical/posterior LUL ☐ LB3 anterior LUL ☐ LB1+2 apical/posterior LUL ☐ LB3 anterior LUL ☐ LB4 superior lingula ☐ LB5 inferior lingula ☐ LB6 superior ☐ LB7/8 anteromedial ☐ LB9 laterobasal ☐ LB10 posterobasal	Each answer 1 point Score____/10
7. Circle the response that correctly identifies these five CXR lesions ☐ Image A ☐ Image B ☐ Image C ☐ Image D ☐ Image E	Each answer 2 points Score____/10
8. Circle the response that correctly identifies these five CT lesions ☐ Image A ☐ Image B ☐ Image C ☐ Image D ☐ Image E	Each answer 2 points Score____/10
9. CT: correctly identify the location of these lesions and maneuver the bronchoscope to the correct segment: right lung ☐ Image A ☐ Image B ☐ Image C ☐ Image D ☐ Image E	Each answer 2 points Score____/10
10. CT: correctly identify the location of these lesions and maneuver the bronchoscope to the correct segment: left lung ☐ Image A ☐ Image B ☐ Image C ☐ Image D ☐ Image E	Each answer 2 points Score____/10

Score _____/100

Fig. 1. BRadSTAT. Questions 1–10.   *(Figure continued on next pages.)*

each question. The assessment criterion was misclassification error, i.e., the proportion of misclassified participants per total number of participants. Weights were applied as regression coefficients in an adjacent-category logit proportional odds model. This model called for the estimation of thresholds on the log-odds scale between every pair of adjacent categories. Thresholds could be translated as maximum likelihood cut points for the corresponding scores, conditionally on the weighting scheme. Misclassification was minimized using leave-one-out cross-validation in all cases (online supplementary).

Question 1

Instructions
Match each highlighted chest radiograph to it's corresponding right lung lobe.
There is one correct response to each image (i.e., there will be one incorrect response remaining)
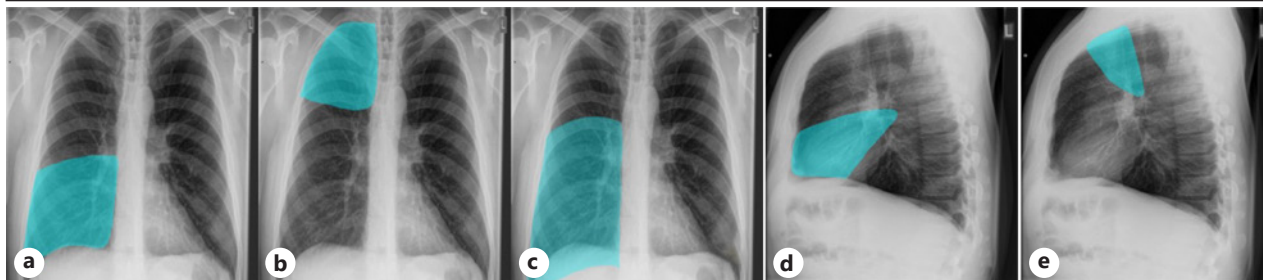
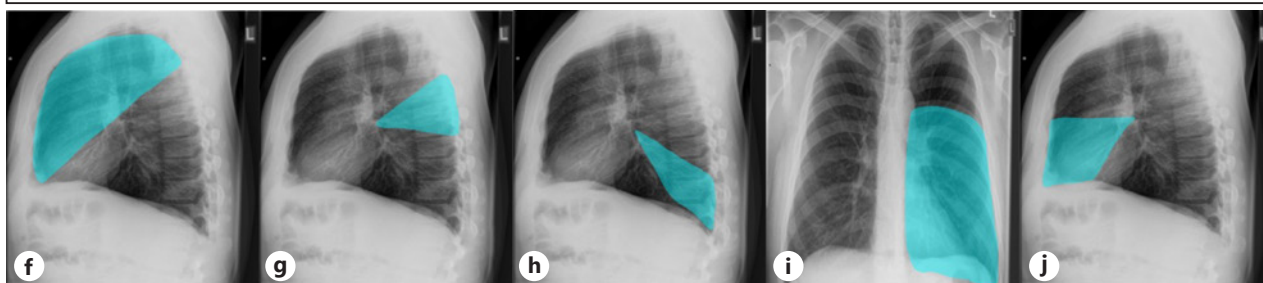| a	b	c	d	e	
Right middle lobe	Right lower lobe	Right upper lobe anterior	Right middle lobe	Right upper lobe apical	Right upper lobe

Question 2

Instructions
Match each highlighted chest radiograph to it's corresponding left lung lobe.
There is one correct response to each image (i.e., there will be one incorrect response remaining)

f	g	h	i	j	
Left lower lobe superior	Left upper lobe	Left lower lobe posterior	Left lingula	Left lower lobe lateral	Left lower lobe

1

In the third phase, we selected the weighting method yielding the smallest misclassification error as well as being most clinically relevant. The first phase was then repeated with the corresponding score. The conditional maximum likelihood cut points of this score and the BRadSTAT score were assessed by producing cross-validated sensitivity, specificity, and accuracy estimates by comparing Beginner to non-Beginner, Beginner or Intermediate to Experienced or Expert, and non-Expert to Expert, respectively. Sensitivity and specificity here correspond to correct classification in the lower ability and the higher ability groups, respectively.

Ethics
No ethics approval was required according to the Health and Disability Ethics Committee in New Zealand as no patient data were collected for this study. Participants provided written informed consent prior to participation in the study.

Question 3

Instructions
Match each computed tomography image to it's corresponding RIGHT lung airway*
There is one correct response for each image (i.e., there will be one incorrect response remaining)

| Right main bronchus | Right lower lobe superior (RB6) | Right upper lobe posterior (RB2) | Right middle lobe lateral (RB4) | Right lower lobe lateral (RB9) | Right lower lobe posterior (RB10) |

Question 4

Instructions
Match each computed tomography image to it's corresponding LEFT lung airway*
There is one correct response for each image (ie., there will be one incorrect response remaining)
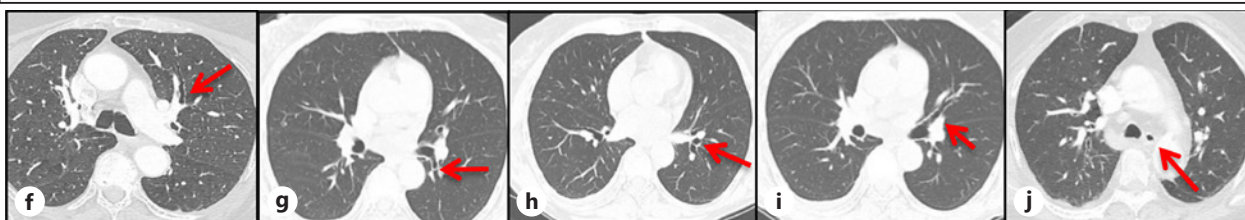
| Left lower lobe lateral (LB9) | Left upper lobe anterior (LB3) | Left lower lobe posterior (LB10) | Left lingula | Esophagus | Left lower lobe superior (LB6) |

*Jackson-Huber nomenclature and Japanese system numbering are used throughout BRadSTAT

1

Results

Phase 1
BRadSTAT was administered to 33 participants: Beginners (5), Intermediates (9), Experienced (10), and Experts (9). Eight were from Sydney, four from Adelaide, and thirteen from Auckland. Eight Experts were recruited during the ANZIP conference.

Mean scores for each of the four groups (Beginner, Intermediate, Experienced, and Expert) were 74 (±13 SD), 78 (±14), 86 (±9), and 88 (±8), respectively (Fig. 2). Statistically significant differences were noted between Expert and Beginner ($p = 0.02$), Expert and Intermediate ($p = 0.04$), and Experienced and Beginner groups ($p = 0.05$) (Table 1). Experts' scores were less variable than those of Beginners (SD 8 vs. 13,

Question 5.

Right lung bronchopulmonary segments

Instructions
Match the images (A–J) to the corresponding 10 CT segments*
There is one correct response for each image

Refer to Appendix S2 for supporting CT slides
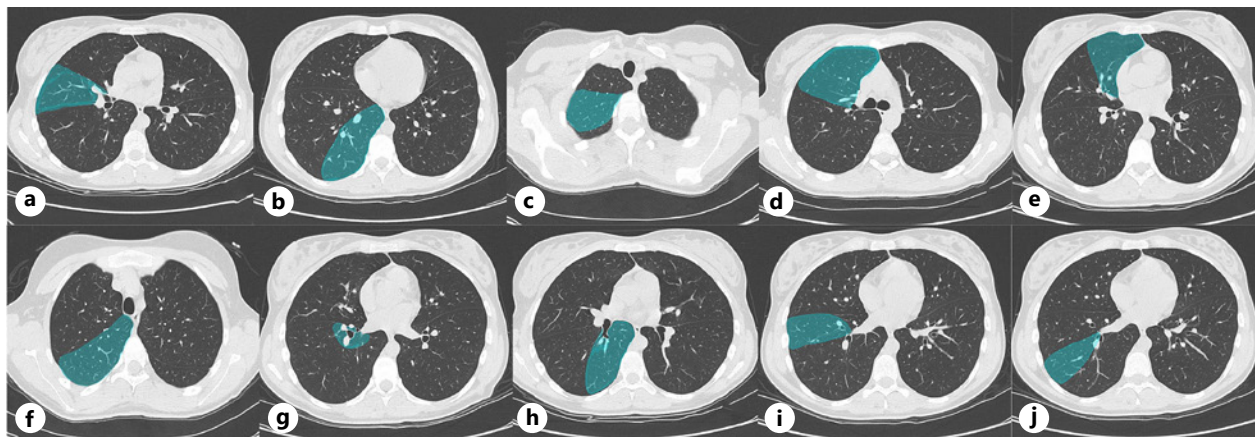
| Right upper lobe apical (RB1) | Right upper lobe posterior (RB2) | Right upper lobe anterior (RB3) | Right middle lobe lateral (RB4) | Right middle lobe medial (RB5) |
| Right lower lobe superior (RB6) | Right lower lobe medial (RB7) | Right lower lobe anterior (RB8) | Right lower lobe lateral (RB9) | Right lower lobe posterior (RB10) |

1

$p = 0.02$). When participants were grouped into trainees ($n = 14$) or consultants ($n = 19$), differences between mean scores remained statistically significant (76 vs. 87, respectively, $p = 0.014$), The median time to complete the BRadSTAT differed significantly between the Expert group (median 29 min, range [21, 35]) and others ($p = 0.0028$): Beginner (median 50 min, range [30, 53]), Intermediate (median 47 min, range [35, 55]), and Experienced (median 40 min, range [30, 64]) (Fig. 3). The median time to complete the assessment was also statistically significant between trainees and consultants (45.2 vs. 36.6 min, respectively, $p = 0.014$).

Table 1. Between group differences for BRadSTAT and BRadSTAT-CT

Contrast	Δ	*p* value	Δ	*p* value
Intermediate-Beginner	3.8	0.54	10.8	0.20
Experienced-Beginner	12.1	0.048*	19.0	0.03*
Expert-Beginner	14.4	0.02*	27.0	0.007*
Experienced-Intermediate	8.3	0.10	8.2	0.21
Expert-Intermediate	10.7	0.043*	16.2	0.015*
Expert-Experienced	2.3	0.65	8.0	0.13

Δ: estimated difference between groups; the standard deviation is 25% larger for BRadSTAT-CT compared to BRadSTAT. *p* value: for the stated contrast in a generalized least-squares heteroscedastic model. *Significant at the 0.05 level.
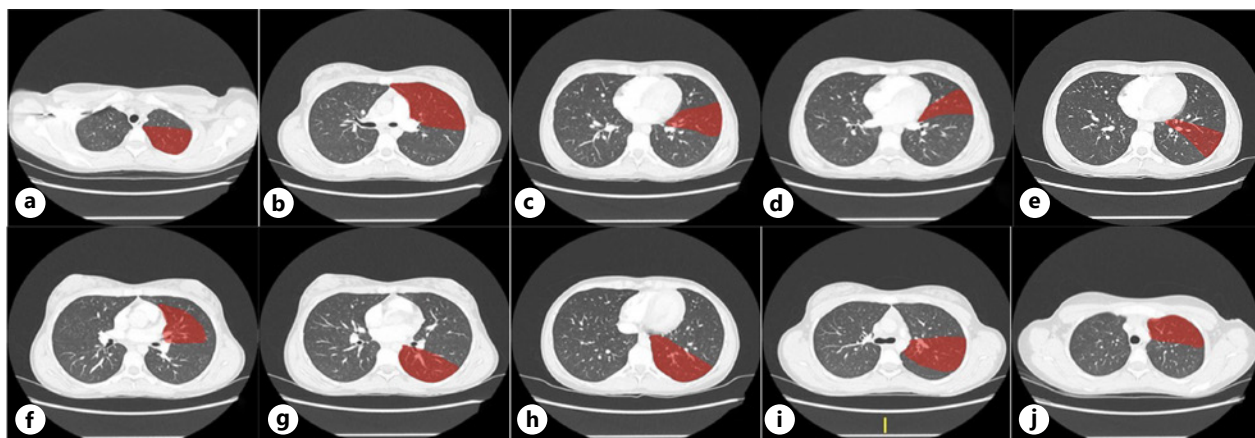
Question 6.

Left lung bronchopulmonary segments

Instructions
Match the images (A–J) to the corresponding 10 CT segments*
There is one correct response for each image
Note: There are 4 images corresponding to left upper lobe
Refer to Appendix S3 for supporting CT slides

_____	_____	_____	_____	_____
Left upper lobe apicoposterior (LB1+2)	Left upper lobe anterior (LB3)	Left upper lobe apicoposterior (LB1+2)	Left upper lobe anterior (LB3)	Superior lingula (LB4)
_____	_____	_____	_____	_____
Inferior lingula (LB5)	Left lower lobe superior (LB6)	Left lower lobe anteromedial (LB7/8)	Left lower lobe lateral (LB9)	Left lower lobe posterior (LB10)

1

Table 2. Observed significance level of scores as predictors of group, level, and total test time

Score	Group*	Level*	Total test time
BRadSTAT	0.004	0.005	0.056
BRadSTAT-CT	0.0001	0.001	

* Group represents Beginner, Intermediate, Experienced, Expert.
* Level represents trainee and consultant.

Phase 2 and 3

Analysis of the ten individual questions showed good discriminative power of questions 3, 4, 6, 8, 9, and 10, which were CT-related questions (Fig. 4), whereas CXR-related questions 1, 2, and 7 were less discriminative. Question 5 had lower discriminative power compared to other CT-related questions.

Based on these findings and our examination of weighting methods, the weighting scheme yielding the smallest misclassification error as well as being most clinically relevant was selected. We named this weighting scheme the BRadSTAT-CT because it included only CT-related questions. A score of 10 was assigned for questions 3, 4, 5, 6, and 8 each if answered correctly, and a

Yap/Vandal/Williamson/Nguyen/Colt

Question 7

Each CXR has a lobar or segmental abnormality.
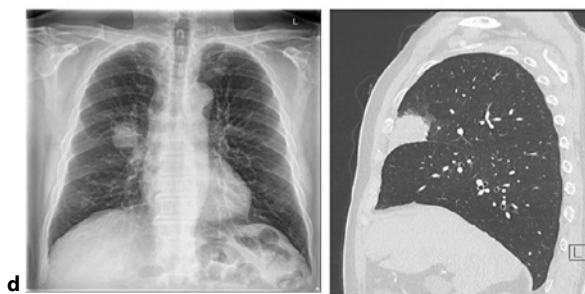Circle the response that corresponds to each CXR abnormality
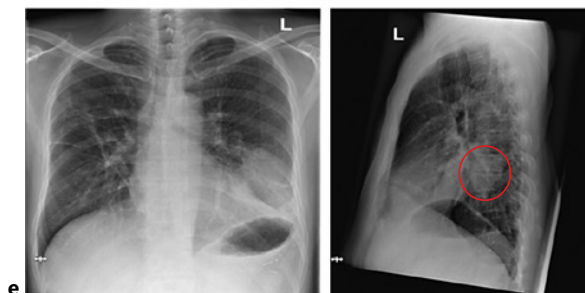
Left lower lobe
Or
Lingula
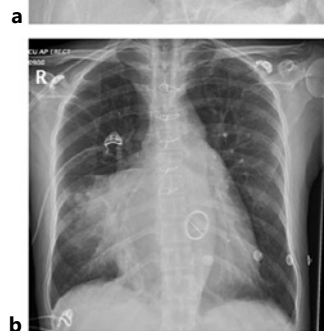
Right lower lobe
Or
Right middle lobe

Left upper lobe
Or
Superior lingula
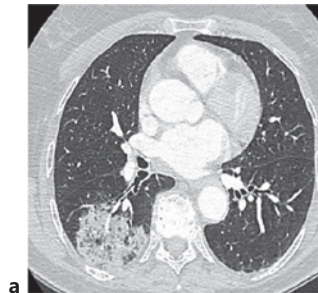
Right upper lobe anterior
Or
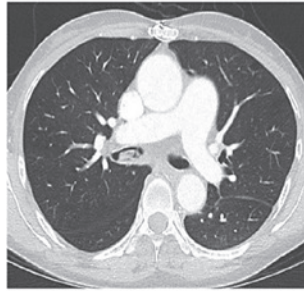Right middle lobe

Left lower lobe
Or
Lingula

Question 8

Each CT has a lobar or segmental abnormality.
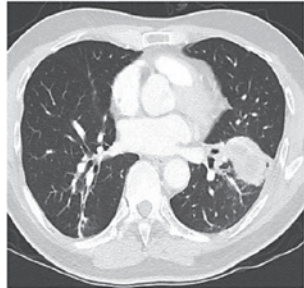Circle the correct answer corresponding to each CT abnormality
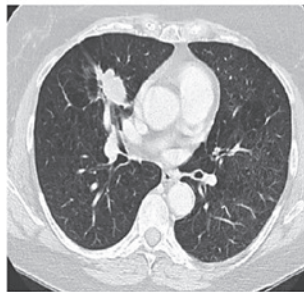

Right lower lobe lateral
Or
Right lower lobe superior

Right lower lobe common basal
Or
Bronchus intermedius
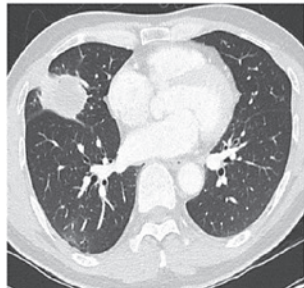
Inferior lingula
Or
Left lower lobe

Right middle lobe medial
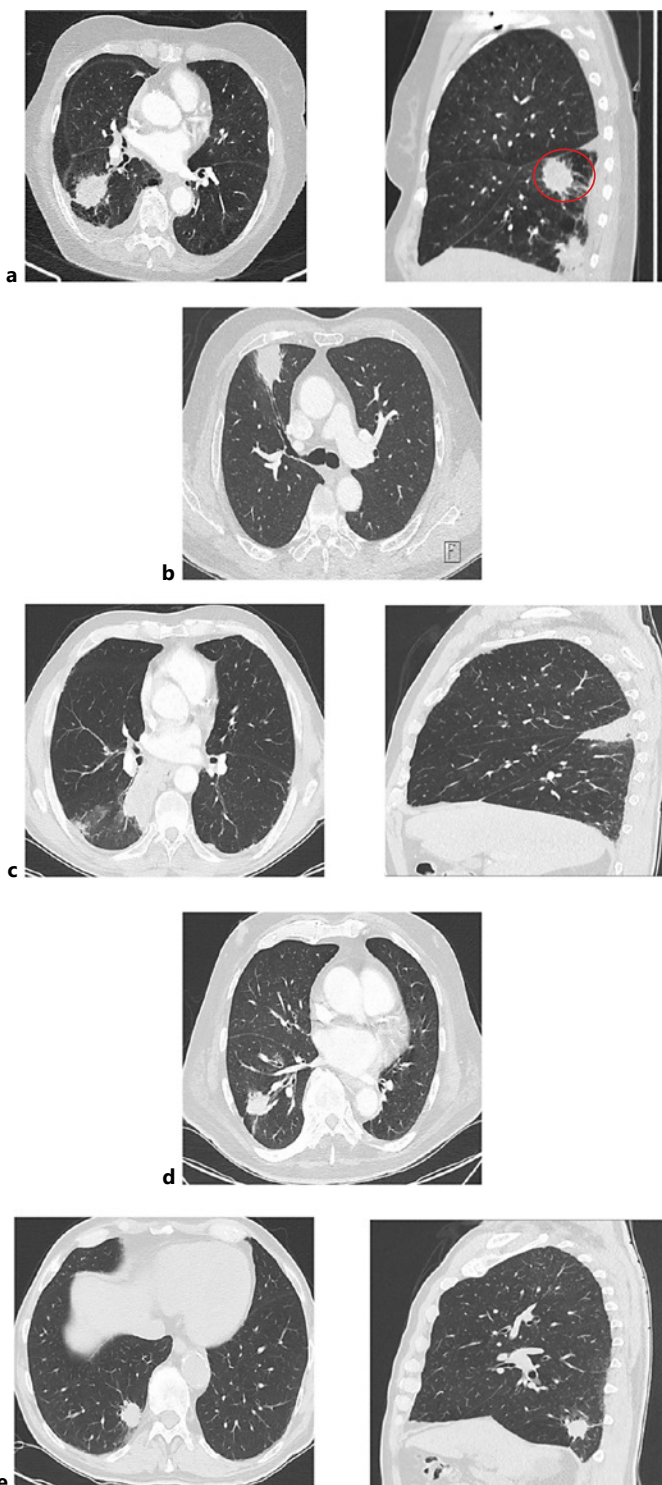Or
Right upper lobe anterior

Right middle lobe medial
Or
Right middle lobe lateral

1

Question 9 — Identify the abnormal segment on CT, then navigate the bronchoscope to the correct segment in the simulator/model

Question 10

Identify the abnormal segment on CT, then navigate the bronchoscope to the correct segment in the simulator/model
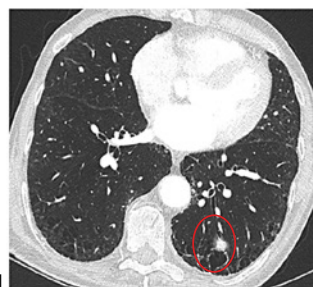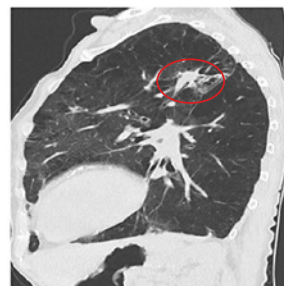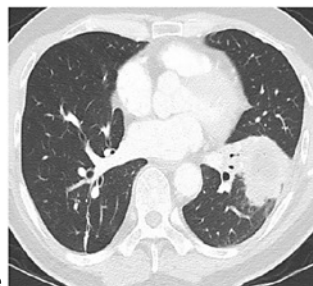
Yap/Vandal/Williamson/Nguyen/Colt

Fig. 2. Comparing BRadSTAT and BRad-STAT-CT scores across the four groups.
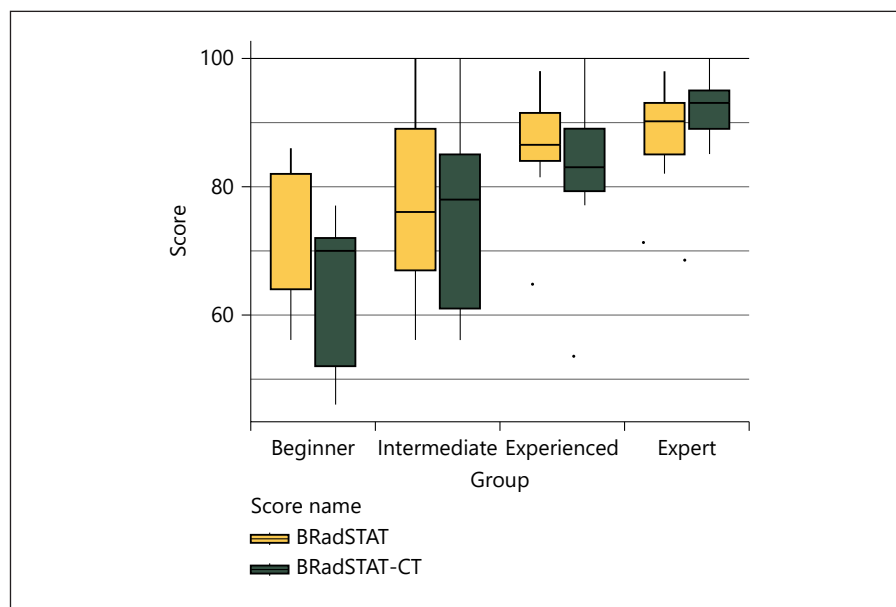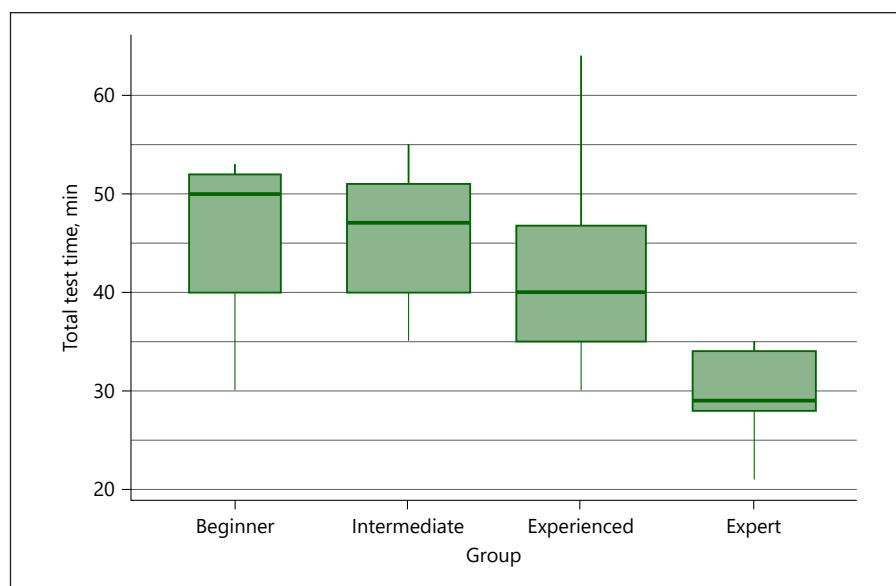

Fig. 3. Test time required by groups to complete BRadSTAT.

score of 25 was assigned for questions 9 and 10 each. The rationale for the higher weighted questions 9 and 10 was that they required a combination of cognition (localizing the abnormality on CT) and technical skills (navigation to the correct bronchopulmonary segment) as compared to CT interpretation alone.

The BRadSTAT-CT score was determined for all 33 participants. Mean BRadSTAT-CT scores for the four groups (Beginner, Intermediate, Experienced, and Expert) were 63 (±14), 74 (±15), 82 (±13), and 90 (±9), respectively (Fig. 2). Statistically significant differences were noted between Expert and Beginner ($p = 0.007$), Expert and Intermediate ($p = 0.01$), and Experienced and Beginner scores ($p = 0.03$) (Table 1). A statistically significant difference between the mean BRadSTAT-CT scores for trainees and consultants (70.4 vs. 86.2, $p = 0.002$) was noted. The Jonckheere-Terpstra test relating the group order to the BRadSTAT and BRadSTAT-CT scores as a predictor yielded p values of $p = 0.004$ and $p = 0.0001$, respectively (Table 2). Median time to complete BRadSTAT-CT was not measured because BRadSTAT-CT was determined a posteriori.

BRadSTAT: A Bronchoscopy-Radiologic Assessment Tool
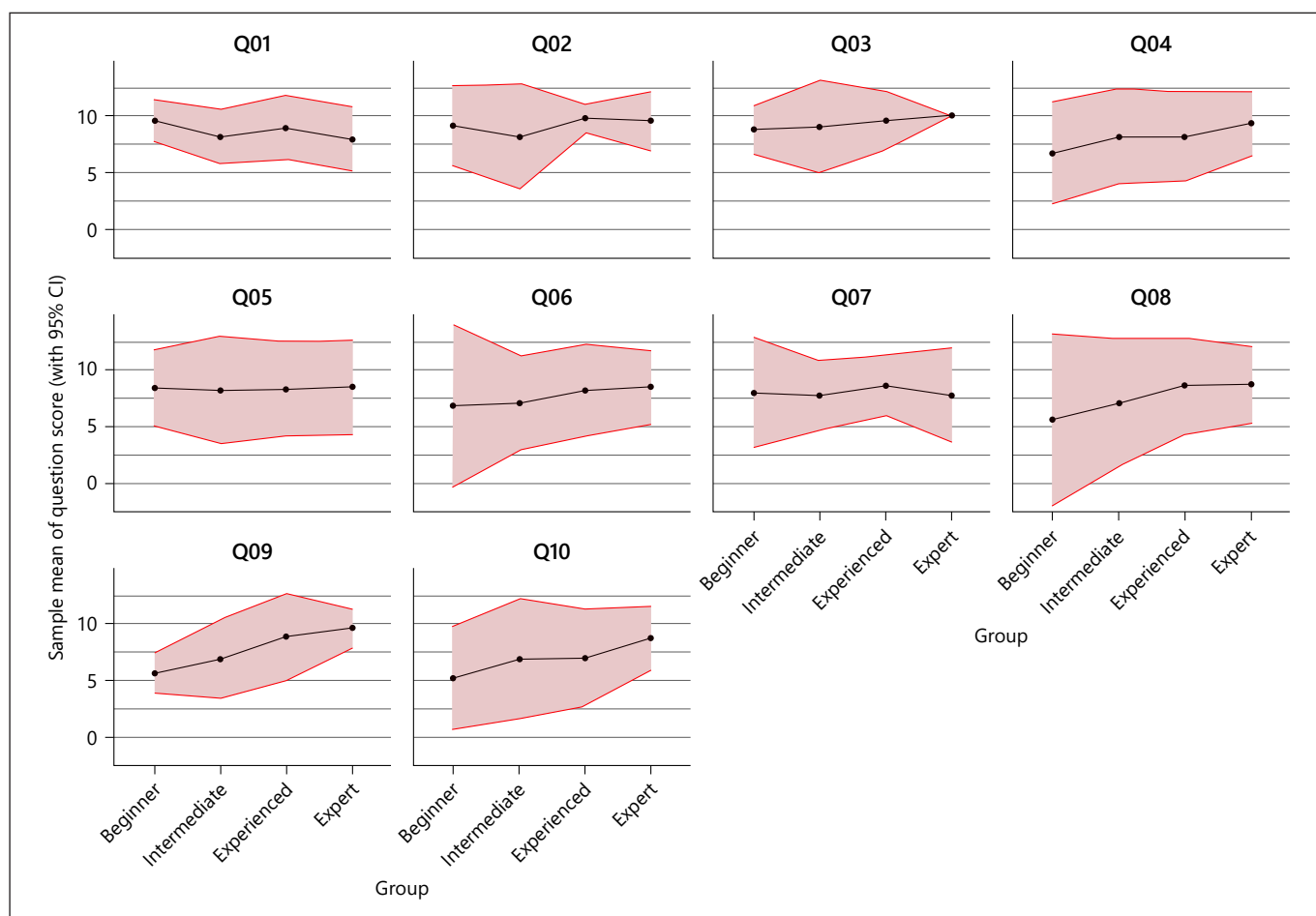
Fig. 4. Sample means of the 10 questions in BRadSTAT by group.

Table 3. Estimated cut points, sensitivity, specificity, and accuracy for BRadSTAT and BRadSTAT-CT

Score	Comparison	B versus I + Ed + Ex	B + I versus Ed + Ex	B + I + Ed versus Ex
BRadSTAT	Cut point	≤67	≤80	≤88
	Sensitivity	40%	50%	92%
	Specificity	89%	83%	25%
	Accuracy	81%	69%	75%
BRadSTAT-CT	Cut point	≤50	≤69	≤82
	Sensitivity	20%	43%	67%
	Specificity	100%	89%	89%
	Accuracy	88%	70%	73%

B, Beginner; I, Intermediate; Ed, Experienced; Ex, Expert. Sensitivity: probability of correct classification in the lower ability group. Specificity: probability of correct classification in the higher ability group. Accuracy: probability of correct classification overall.

Yap/Vandal/Williamson/Nguyen/Colt

Cut Point Analysis

Table 3 presents the cut points for the BRadSTAT and BRadSTAT-CT scores and the associated cross-validated sensitivity, specificity, and accuracy. Both the BRadSTAT and BRadSTAT-CT cut points have lower sensitivity than specificity. The accuracy of the BRadSTAT cut points (≤67 for Beginner level, 68–80 for Intermediate, 81–88 for Experienced, and ≥89 for Expert) ranged between 69% and 81%. The accuracy of the BRadSTAT-CT cut points (≤50 for Beginner level, 51–69 for Intermediate, 70–82 for Experienced, and ≥83 for Expert) ranged between 70% and 88%. The BRadSTAT-CT cut points have better specificity, accuracy, and lower sensitivity compared to BRadSTAT cut points.

Discussion

BRadSTAT objectively and systematically tests a user's combined knowledge of radiologic interpretation and ability to maneuver a bronchoscope to a desired peripheral location. Technological advances in airway imaging and peripheral bronchoscopy make these skills increasingly necessary, particularly for nodule management in the era of CT screening for lung cancer. Precise knowledge of bronchopulmonary segments is also essential for accurate bronchoscopic lavage in patients with patchy interstitial lung disease.

The importance of accurate endobronchial path selection to maximize diagnostic yield was highlighted by Dolina et al. [7], who noted a wide range of accuracy across experience levels. Variability in yields might also result from varying bronchoscopic dexterity and handling [8] and inexperience [9]. While both can be measured using existing bronchoscopy assessment tools [1–3], it is noteworthy that none apply to the combined use of radiologic imaging and technical skill. The BRadSTAT was designed to fill this void.

In this study, we demonstrated the BRadSTAT's construct validity; it was able to discriminate between participants at different levels of experience, from Beginner to Expert, and between thoracic trainees and consultants. The time to complete the assessment was also reduced from Beginner to Expert, reflecting difficulty to complete the assessment based on level of experience. Time should be included in assessments because proficiency is characterized by good results as well as task efficiency. Although we did not test the longitudinal aspects of BRadSTAT, we propose the repeated use of this assessment tool in training programs would be consistent with the strategy of using formative and barrier assessments such as the BSTAT [1], EBUS-STAT [10], and UG-STAT [11], employed by TSANZ to assess competency in flexible bronchoscopy [12], EBUS bronchoscopy [13], and pleural ultrasound [14]. Deliberate practice and improving specific tasks under supervision with constructive feedback helps achieve mastery learning in other medical fields [15–17].

The secondary aim of this study was to analyze our data to potentially refine the BRadSTAT. Based on high discriminative power, low misclassification, and a clinically relevant weighting scheme, we identified and studied a BRadSTAT-CT scoring system which consisted of seven CT-related questions. That BRadSTAT-CT had superior discriminative power compared with BRadSTAT suggests that an assessment tool based on CT-related questions alone could better discriminate between participants of varying levels of experience, i.e., had better construct validity compared with BRadSTAT. This may be related to the possibility that CXR interpretation has become a fundamental skill for all physicians; therefore, few differences in chest interpretation skill might be expected between Beginners and more advanced trainees or consultants. It is also possible that a tool using only three CXR-related questions is insufficient to detect statistically significant differences. Regardless, in an era when CT is the imaging modality of choice for patients requiring bronchoscopy for peripheral airway pathology, a CXR-related bronchoscopy assessment tool might rapidly become obsolete.

BRadSTAT-CT shows how an assessment tool may be more discriminating when different weights are allotted to questions based on their complexity. This was also the case for the RIGID-TASC [18], an instrument that objectively measures rigid bronchoscopy-related skills. In BRadSTAT-CT, combined radiologic interpretation and technical skill questions 9 and 10 required multiple steps to complete and thus were weighted more than the remaining five questions that required only one step. This approach is in line with Miller's [19] suggestion that assessment should not only capture important components of a complex task but also ensure that test elements are weighted appropriately.

Cut points, also referred to as cut scores, provide a precise way of establishing a standard of performance for a test [20]. They are often used by educators and policy makers to help define whether a particular test score is sufficient for some purpose [21] and are valuable for setting levels of proficiency or competency [22, 23]. If used longitudinally, cut points can also enable assessors to determine a trainee's progress along the learning curve.

Those of the BRadSTAT and BRadSTAT-CT were identified via cross-validation. They allowed stratification into Expert, Experienced, Intermediate, and Beginner groups (Table 3). Both sets of cut points had high accuracy rates for predicting these groups, further strengthened by the results of the Jonckheere-Terpstra test which showed a stronger predictive power of the BRadSTAT-CT cut points compared with those of the BRadSTAT.

The BRadSTAT-CT's cut points also make clinical sense. For instance, a participant who scored perfectly in CT interpretation (questions 3, 4, 5, 6, 8) but made technical errors (questions 9 and 10) could score 50 points, i.e., Beginner level, which could be consistent with a subjective assessment of trainees with CT interpretation skills but not yet able to complete the more complex task of combining radiologic interpretation with bronchoscopic navigation. Similarly, to perform at an Expert level (cut point ≥83), one might expect a perfect score in CT interpretation and certainly no more than three mistakes in either technical question, which each contained five images of 5 points each for a total possible score of 25 per question. It is strangely coincidental that our proposed BRadSTAT-CT cut points closely resemble the grading system used in New Zealand and Australia's tertiary education systems, where a grade A is conferred for scores ≥80% (≥85% in Australia), B for scores ≥65% (≥70% in Australia), and C for scores ≥50% [24, 25].

Limitations

Because the lead investigator (E.Y.) also helped administer and score BRadSTAT, one limitation of this study is possible investigator bias. A similar possibility for bias is found in studies of other assessment tools [1, 10]. Overall, the objective nature of BRadSTAT's scoring system makes investigator bias unlikely. A second limitation of this study is that trainees recruited from centers with a strong focus on bronchoscopy education may have been better trained, thereby reducing BRadSTAT's ability to discriminate between groups. A third limitation relates to small sample size and the possibility of introducing a priori group classification error. Each of these limitations can be addressed through a larger scale study. A fourth limitation relates to our decision to test the technical skills components of questions 9 and 10 using the ORSIM bronchoscopy simulator rather than patients. While this device is validated and provides a realistic procedural experience [4, 5], it is not widely available. Future studies can incorporate assessments in the clinical setting.

Finally, while BRadSTAT and BRadSTAT-CT are important steps toward the development of an objective, fully validated assessment tool, the BRadSTAT-CT was derived from differential scoring and not tested independently. Future clinical studies employing subjects of varying experience are needed to determine its repeatability, reproducibility, and longitudinal applicability. We believe that the objective nature of its questions and BRadSTAT-CT's weighted scoring system make instability over time or between different assessors unlikely. How cut points might be used to help define competency, however, and how often BRadSTAT or BRadSTAT-CT should be performed during training is less clear.

Conclusions

Combining accurate radiologic interpretation, precise airway navigation, and dexterous technical skill is essential to competent bronchoscopic access to peripheral airways. BRadSTAT is the first validated assessment tool that objectively measures these skills and effectively discriminates between bronchoscopists with different levels of experience. Refining BRadSTAT and strengthening its discriminative power by using differential weighting and CT-related questions only produced the BRadSTAT-CT. Studies are warranted to demonstrate the reproducibility and repeatability of both these tools and to assess their usefulness in competency-based bronchoscopy training programs.

Author Contributions

Elaine L.C. Yap: conceptualization: lead, formal analysis: supporting, investigation: equal, methodology: lead, writing – original draft: lead, and writing – review and editing: lead. Alain C. Vandal: formal analysis: lead, methodology: equal, writing – original draft: equal, and writing – review and editing: equal. Jonathan P. Williamson: conceptualization: supporting, investigation: equal, methodology: supporting, and writing – review and editing: equal. Phan Nguyen: investigation: equal, methodology: supporting, and writing – review and editing: equal. Henri Colt: conceptualization: lead, formal analysis: supporting, methodology: lead, writing – original draft: lead, and writing – review and editing: lead.

Data Availability Statement

All data generated or analyzed during this study are included in this article and its online supplementary material. Further inquiries can be directed to the corresponding author.

References

1 Davoudi M, Osann K, Colt HG. Validation of two instruments to assess technical bronchoscopic skill using virtual reality simulation. Respiration. 2008;76(1):92–101.

2 Voduc N, Dudek N, Parker CM, Sharma KB, Wood TJ. Development and validation of a bronchoscopy competence assessment tool in a clinical setting. Ann Am Thorac Soc. 2016; 13(4):495–501.

3 Konge L, Larsen KR, Clementsen P, Arendrup H, von Buchwald C, Ringsted C. Reliable and valid assessment of clinical bronchoscopy performance. Respiration. 2012;83(1):53–60.

4 Baker PA, Weller JM, Baker MJ, Hounsell GL, Scott J, Gardiner PJ, et al. Evaluating the OR-SIM simulator for assessment of anaesthetists' skills in flexible bronchoscopy aspects of validity and reliability. Br J Anaesth. 2016; 117(Suppl 1):i87–91.

5 Wong DT, Mehta A, Singh KP, Leong SM, Ooi A, Niazi A, et al. The effect of virtual reality bronchoscopy simulator training on performance of bronchoscopic-guided intubation in patients: a randomised controlled trial. Eur J Anaesthesiol. 2019;36(3):227–33.

6 Jonckheere AR. A distribution-free k-sample test against ordered alternatives. Biometrika. 1954;41(1/2):133–45.

7 Dolina MY, Cornish DC, Merritt SA, Rai L, Mahraj R, Higgins WE, et al. Interbronchoscopist variability in endobronchial path selection. A simulation study. Chest. 2008;133(4): 897–905.

8 Minami H, Ando Y, Nomura F, Sakai S, Shimokata K. Interbronchoscopist variability in the diagnosis of lung cancer by flexible bronchoscopy. Chest. 1994;105(6):1658–62.

9 Ouellette DR. The safety of bronchoscopy in a pulmonary fellowship program. Chest. 2006;130(4):1185–90.

10 Davoudi M, Colt HG, Osann KE, Lamb CR, Mullon JJ. 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(8):773–9.

11 Salamonsen M, Mcgrath D, Steiler G, Ware R, Colt H, Fielding D. A new instrument to assess physician skill at thoracic ultrasound, including pleural effusion markup. Chest. 2013; 144(3):930–4.

12 Thoracic Society Australia & New Zealand. Recognition of competency in flexible bronchoscopy for adults. 2019 Apr [cited July 2021]. Available from: https://www.thoracic.org.au/documents/item/1630.

13 Thoracic Society Australia & New Zealand. Recognition of competency in endobronchial ultrasound transbronchial needle aspiration (EBUS-TBNA). 2019 Apr [cited July 2021]. Available from: https://www.thoracic.org.au/documents/item/1629.

14 Thoracic Society Australia & New Zealand. Recognition of competency in thoracic ultrasound. 2016 Nov [cited July 2021]. Available from: https://www.thoracic.org.au/documents/item/1626.

15 Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med. 2004;79(10 Suppl l):S70–81.

16 Wayne DB, Butter J, Siddall VJ, Fudala MJ, Wade LD, Feinglass J, et al. Mastery learning of advanced cardiac life support skills by internal medicine residents using simulation technology and deliberate practice. J Gen Intern Med. 2006;21(3):251–6.

17 Zendejas B, Cook DA, Bingener J, Huebner M, Dunn WF, Sarr MG, et al. Simulation-based mastery learning improves patient outcomes in laparoscopic inguinal hernia repair: a randomized controlled trial. Ann Surg. 2011;254(3):502–9; discussion 509–11.

18 Mahmood K, Wahidi MM, Osann KE, Coles K, Shofer SL, Volker EE, et al. Development of a tool to assess basic competency in the performance of rigid bronchoscopy. Ann Am Thorac Soc. 2016;13(4):502–11.

19 Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990; 65(9 Suppl):S63–7.

20 Fey B. SAGE Encyclopedia of educational research, Measurement and evaluation [cited July 2021]. Available from: https://methods.sagepub.com/reference/the-sage-encyclopedia-of-educational-research-measurement-and-evaluation/i6685.xml.

21 A primer on setting cut scores on tests of educational achievement [cited July 2021]. Available from: https://www.ets.org/Media/Research/pdf/Cut_Scores_Primer.pdf.

22 Konge L, Clementsen P, Larsen KR, Arendrup H, Buchwald C, Ringsted C. Establishing pass/fail criteria for bronchoscopy performance. Respiration. 2012;83(2):140–6.

23 Colt HG, Davoudi M, Quadrelli S, Zamanian Rohani N. Use of competency-based metrics to determine effectiveness of a post-graduate thoracoscopy course. Respiration. 2010; 80(6):553–9.

24 Academic grading in New Zealand [cited July 2021]. Available from: https://www.newzealandeducation.info/education-system/academic-grading-in-new-zealand.html.

25 Grading system in Australia [cited July 2021]. Available from: https://www.australiaeducation.info/education-system/grading-system.html.