



# Artificial intelligence for automatic and objective assessment of competencies in flexible bronchoscopy

Kristoffer Mazanti Cold<sup>1^</sup>, Kaladerhan Agbontaen<sup>2</sup>, Anne Orholm Nielsen<sup>1,3</sup>, Christian Skjoldvang Andersen<sup>4</sup>, Suveer Singh<sup>5,6</sup>, Lars Konge<sup>1</sup>

<sup>1</sup>Copenhagen Academy for Medical Education and Simulation (CAMES), Rigshospitalet, University of Copenhagen, the Capital Region of Denmark, Copenhagen, Denmark; <sup>2</sup>Department of Intensive Care Unit, Chelsea and Westminster Hospital, Chelsea, London, UK; <sup>3</sup>Bispebjerg Hospital, Department of Pulmonary Medicine, Capital Region of Denmark, Copenhagen, Denmark; <sup>4</sup>Department of Pulmonary Medicine, North Zealand Hospital, Capital Region of Denmark, Hilleroed, Denmark; <sup>5</sup>Department of Intensive Care Unit, Royal Brompton Hospital, Chelsea, London, UK; <sup>6</sup>Faculty of Medicine, Imperial College London, Chelsea, London, UK

**Contributions:** (I) Conception and design: All authors; (II) Administrative support: All authors; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: KM Cold, K Agbontaen; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

**Correspondence to:** Dr. Kristoffer Mazanti Cold, MD. Copenhagen Academy for Medical Education and Simulation (CAMES), Rigshospitalet, University of Copenhagen, Ryesgade 53B, the Capital Region of Denmark, 2100 Copenhagen, Denmark. Email: Kristoffer.mazanti.cold.01@regionh.dk.

**Background:** Bronchoscopy is a challenging technical procedure, and assessment of competence currently relies on expert raters. Human rating is time consuming and prone to rater bias. The aim of this study was to evaluate if a bronchial segment identification system based on artificial intelligence (AI) could automatically, instantly, and objectively assess competencies in flexible bronchoscopy in a valid way.

**Methods:** Participants were recruited at the Clinical Skills Zone of the European Respiratory Society Annual Conference in Milan, 9<sup>th</sup>–13<sup>th</sup> September 2023. The participants performed one full diagnostic bronchoscopy in a simulated setting and were rated immediately by the AI according to its four outcome measures: diagnostic completeness (DC), structured progress (SP), procedure time (PT), and mean intersegmental time (MIT). The procedures were video-recorded and rated after the conference by two blinded, expert raters using a previously validated assessment tool with nine items regarding anatomy and dexterity.

**Results:** Fifty-two participants from six different continents were included. All four outcome measures of the AI correlated significantly with the experts' anatomy-ratings (Pearson's correlation coefficient, P value): DC ( $r=0.47$ ,  $P<0.001$ ), SP ( $r=0.57$ ,  $P<0.001$ ), PT ( $r=-0.32$ ,  $P=0.02$ ), and MIT ( $r=-0.55$ ,  $P<0.001$ ) and also with the experts' dexterity-ratings: DC ( $r=0.38$ ,  $P=0.006$ ), SP ( $r=0.53$ ,  $P<0.001$ ), PT ( $r=-0.34$ ,  $P=0.014$ ), and MIT ( $r=-0.47$ ,  $P<0.001$ ).

**Conclusions:** The study provides initial validity evidence for AI-based immediate and automatic assessment of anatomical and navigational competencies in flexible bronchoscopy. SP provided stronger correlations with human experts' ratings than the traditional DC.

**Keywords:** Artificial intelligence (AI); flexible bronchoscopy; simulation; medical education; competence assessment

Submitted May 21, 2024. Accepted for publication Jul 12, 2024. Published online Sep 06, 2024.

doi: 10.21037/jtd-24-841

View this article at: <https://dx.doi.org/10.21037/jtd-24-841>

<sup>^</sup> ORCID: 0000-0001-7715-4073.

## Introduction

Flexible bronchoscopy is a challenging technical procedure that requires extensive training for new bronchoscopists. Accurate navigation through the bronchial tree is crucial for identifying the correct segments, which is essential for proper diagnosis and treatment (1). Novice bronchoscopists exhibit lower yields of positive biopsy samples, higher complication rates, and increased patient discomfort during the early stages of their learning curve (2-4). Trainees should meet certain competency requirements before being allowed to perform unsupervised procedures on patients. The traditional approach to certification in medical education has relied on the completion of a certain (arbitrary) number of procedures, as recommended by established organizations such as The American College of Chest Physicians (5) and the European Respiratory Society (6). However, this volume-based approach is not evidence-based, given that trainees learn at different paces and procedural experience does not guarantee competence (7,8). An expert panel in flexible bronchoscopy recommended a shift towards assessing trainee competency through skill acquisition (9). Several tools have been developed for the assessment of bronchoscopy performance, but unfortunately, they have several limitations. Those based on expert raters (10-12), are susceptible to rater bias and rely on experts' limited time. Those relying on expensive virtual reality (VR) simulators (13-15) have limited discriminatory abilities. A novel artificial intelligence (AI)-system (Ambu BPS, Prototype version AmbuBPStrainingGUIDEv.0.0.1, Ambu) for bronchoscopy

has been developed and improve the performance of novel bronchoscopists by providing feedback doing training (16). However, it needs to be established whether it can be used as an automatic, instant, and unbiased assessment tool for bronchoscopy performance (17,18).

### Study aim

This study aims to determine whether an automatic tool based on AI—the AI bronchoscopy assessment (AIBA)—assess competencies in flexible bronchoscopy in a valid way. We present this article in accordance with the STROBE reporting checklist (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-841/rc>).

### Methods

AIBA was evaluated according to Messick's framework of validity evidence from five sources: content, response process, internal structure, relationship to other variables, and consequences (*Table 1*) (19).

### Development of the assessment tool

The content of AIBA was established through consensus by the entire author group, encompassing three pulmonary consultants (A.O.N., C.S.A., and S.S.) with a combined more than 40-year experience in bronchoscopy, a thoracic surgeon and professor of medical education with a PhD in assessment of bronchoscopy (L.K.), and two doctors with research experience within bronchoscopy (K.M.C. and K.A.). The group decided that the AIBA should automatically register a competently performed diagnostic bronchoscopy that examines all segments in a structured order in an efficient manner using a limited amount of time. A previous validation study using “hand-held” data collection by a research assistant could differentiate performance by relying on the following four outcome measures: diagnostic completeness (DC): inspected segments, range 0–18. Structured progress (SP): structured progressions, range 0–18, defined as progressions following a chronological order (RB1→RB2 = 1 point, RB2→RB1 = 0 points) (20). Procedure time (PT): seconds, range 0–∞, defined as time spent from visualizing the carina to extraction of the scope (14,15). Mean intersegmental time (MIT), range 0–∞, defined as mean time spent navigating between segments as calculated by DC divided by PT (20).

### Highlight box

#### Key findings

- Artificial intelligence (AI) can automatically, instantly, and objectively assess bronchoscopy performance.

#### What is known and what is new?

- Assessment of bronchoscopy performance currently relies on expert raters that is susceptible to rater bias and rely on experts' limited time.
- This is the first AI system in flexible bronchoscopy which can replace expert's raters of flexible bronchoscopy maneuvering skills.

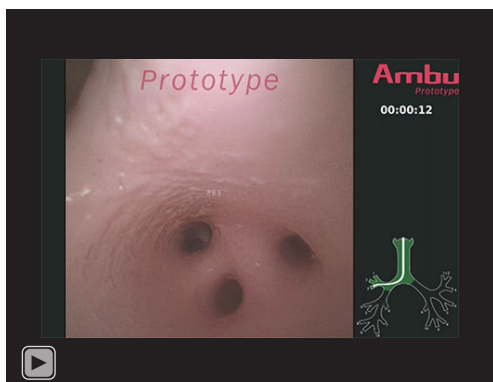
#### What is the implication, and what should change now?

- Basic navigational skills in flexible bronchoscopy can be assessed automatically, instantly, and objectively by the AI, before the trainee advanced to perform more advanced procedures.

**Table 1** Different sources of validity evidence based on Messick's validity framework (19)

Source of evidence for validity and description	Validity evidence for AIBA
Content: the test content should measure what it is supposed to measure	Three pulmonary consultants (A.O.N., C.S.A. and S.S.) with more than 40 years' combined experience in bronchoscopy hypothesized along with a thoracic surgeon and professor of medical education, who has a PhD in bronchoscopy assessment (L.K.), and two doctors with research experience in bronchoscopy education (K.M.C. and K.A.) that simulated bronchoscopy performance can be assessed automatically using an AIBA, relying on the following outcome measures: DC, SP, PT, and MIT
Response process: integrity of data should always be maintained. Test administration should be controlled or standardized at a maximum level possible	All procedures were performed in a controlled, simulated environment, making the tests comparable as the participants were using the same scope, monitor and phantom. For data integrity and to avoid bias, all recordings were automatically rated by AIBA. All videos were additionally rated in a blinded fashion by two expert bronchoscopists (A.O.N. and C.S.A.) using an established rating tool (11)
Internal structure: this refers to the reliability of the test results. The outcome measures should correlate with one another	DC correlated significantly with SP (Pearson's $r=0.75$ , $P<0.001$ ). DC did not correlate significantly to PT ( $r=0.22$ , $P=0.11$ ), neither did SP with PT ( $r=-0.09$ , $P=0.51$ ). DC correlated significantly with MIT ( $r=-0.55$ , $P<0.001$ ), as did SP with MIT ( $r=-0.51$ , $P<0.001$ ).
Relationship to other variables: assessment scores should correlate with known measures of competence—AIBA should correlate with the experts' ratings	All the four outcome measures of AIBA correlated significantly with the experts' anatomy rating: DC (Pearson's $r=0.47$ , $P<0.001$ ), SP ( $r=0.57$ , $P<0.001$ ), PT ( $r=-0.32$ , $P=0.02$ ) and MIT ( $r=-0.55$ , $P<0.001$ ), and with the experts' dexterity rating: DC ( $r=0.38$ , $P=0.006$ ), SP ( $r=0.53$ , $P<0.001$ ), PT ( $r=-0.34$ , $P=0.01$ ) and MIT ( $r=-0.47$ , $P<0.001$ )
Consequences: consequences of testing relates to the pass/fail standard that is set	The pass-/fail criterion of 8 points in anatomy-rating made 30 participants fail and 22 pass. The participants that passed performed significantly better on all four outcome measures: DC ( $P=0.01$ ), SP ( $P=0.004$ ), PT ( $P=0.03$ ), MIT ( $P<0.001$ )

AIBA, artificial intelligence bronchoscopy assessment; DC, diagnostic completeness; SP, structured progress; PT, procedure time; MIT, mean intersegmental time.



**Video 1** Live recording of how the AI can assess bronchoscopy performance and work as a feedback tool. The bronchoscopists did not receive the AI's rating or guidance. AI, artificial intelligence.

### *Participants, and collection of videos for assessment*

The study design was a prospective study. Participants included were doctors participating in the European Respiratory Society (ERS) annual conference in Milan, 9<sup>th</sup>–13<sup>th</sup> September 2023. Participants for this study performed one complete bronchoscopy procedure on a Koken phantom (Bronchoscopy Training Model LM-092, Koken Co., Lth., Tokyo, Japan) without any warm-up, familiarization, or demonstration of the AI-system (*Video 1*). The timer started automatically when the main carina was visualized and ended with extraction of the scope. All the bronchoscopy recordings were automatically assessed by the AI and were also video-recorded for expert rating after the conference. This study was conducted in accordance with

**Table 2** Participants' demographic and outcome measures

Demographic and outcomes	Values
Participants' demographic (n=52)	
Male	39 (75.0)
Age, years	36.5±9.1
Bronchoscopies performed, number	373±1,415
Bronchoscopies performed within last 6 months, number	33±49
Expert-rating (points)	
Total-rating (range, 0–18)	9.8±4.8
Anatomy-rating (range, 0–12)	6.8±3.5
Dexterity-rating (range, 0–6)	3.1±1.7
Outcome measures AIBA	
DC, segments (range, 0–18)	8.4±4.7
SP, progressions (range, 0–18)	3.2±3.2
PT, seconds	264±145
MIT, seconds	39±27

Data are presented as mean ± SD or number (percentage). DC, diagnostic completeness; SP, structured progress; PT, procedure time; MIT, mean intersegmental time; SD, standard deviation; AIBA, artificial intelligence bronchoscopy assessment.

the Declaration of Helsinki (as revised in 2013). Ethical approval is not required for this study in accordance with local or national guidelines, as the study was conducted in a simulated setting not involving patients, but voluntary participants at the ERS 2023 conference signed an informed consent form.

### Expert rating

Two expert raters (A.O.N. and C.S.A.) rated all recordings according to an established and validated rating tool for operator competence (11)—referred to as the expert rating tool. The tool consisted of 9 items each ranging from 0–2 points (18 points maximum, [Figure S1](#)). The first six items report on anatomical knowledge and navigational skills (right upper lobe, right middle lobe, right lower lobe, left upper division segments 1+2 and 3, lingula, and left lower lobe). The rating tool was slightly modified, as the participants were not asked to freeze the image at the six different locations as in the original validation study (11). Two points were awarded when navigating securely and

structured to the location, 1 point when navigating to the location but not in a securely and structured manner, and 0 point was awarded if the location was not visualized. Ratings of the first six items were referred to as anatomy-ratings.

The last three items report on scope maneuvering skills (wall collisions, red-out, and scope centering). These items were not modified from the original rating tool. Ratings of the last three items were referred to as dexterity-ratings.

To ensure consistent interpretation of the rating tool, the raters were invited to live rate three bronchoscopy videos (not included in the study) and having their ratings discussed afterwards with the authors K.M.C. and L.K. Furthermore, they were sent a guideline on how to use the rating tool with examples ([Appendix 1](#)). To ensure individualized rating, the raters were sent the recordings and not allowed to discuss their ratings.

A former study established a pass/fail of 8 point in anatomy-rating for the used rating tool (12).

### Statistical analysis

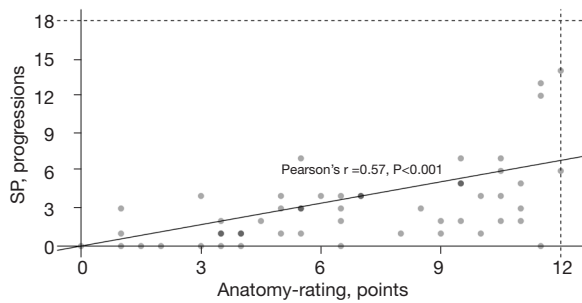
Means and standard deviation were used for normally distributed data. Statistical testing was completed in Statistical Package for the Social Sciences (SPSS) version 29 (PASW v29.0; SPSS Inc., Chicago, IL, USA). To confirm accurate use of the expert rating tool, interrater reliability between the two raters was tested using Pearson's correlation and internal consistency reliability was tested using Cronbach's alpha.

Internal structure: Correlations between the individual outcome measures of AIBA (DC, SP, PT, and MIT), were tested using Pearson's r. Relationship to other variables: Correlations between anatomy-rating, dexterity-rating, and the outcome measures of AIBA, were tested using Pearson's r. Consequences: A pass-/fail criterion of 8 points in anatomy-rating was set based on a previous study (12). The AIBA outcome measures of the failing group and the passing group were compared using independent samples *t*-tests.

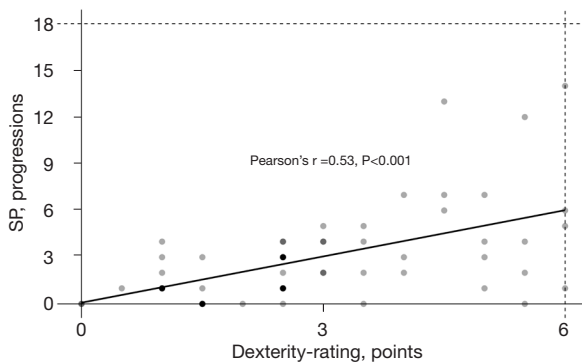
### Results

Fifty-two participants were included from 20 different countries spanning six continents ([Table 2](#)).

The experts' ratings had a very high level of agreement, as the interrater reliability was high (Pearson's  $r=0.92$ ,  $P<0.001$ ) and the internal consistency reliability was also



**Figure 1** Correlation between SP by AIBA and anatomy-rating by expert raters. Grey datapoints indicate one participant. Dark-grey datapoints indicate two participants with identical scores. SP, structured progress; AIBA, artificial intelligence bronchoscopy assessment.



**Figure 2** Correlation between SP by AIBA and dexterity-rating by expert raters. Grey datapoints indicate one participant, dark-grey datapoints indicate two participants, and black datapoints indicate three participants with identical scores. SP, structured progress; AIBA, artificial intelligence bronchoscopy assessment.

high (Cronbach's alpha: 0.84).

### Internal structure

DC correlated significantly with SP (Pearson's  $r=0.75$ ,  $P<0.001$ ). DC did not correlate significantly with PT ( $r=0.22$ ,  $P=0.11$ ), neither did SP with PT ( $r=-0.09$ ,  $P=0.51$ ). DC correlated significantly with MIT ( $r=-0.55$ ,  $P<0.001$ ), as did SP with MIT ( $r=-0.51$ ,  $P<0.001$ ).

### Relationship to other variables

All four outcome measures of AIBA correlated significantly with the expert's anatomy rating: DC ( $r=0.47$ ,  $P<0.001$ ), SP ( $r=0.57$ ,  $P<0.001$ , *Figure 1*), PT ( $r=-0.32$ ,  $P=0.02$ ), and MIT ( $r=-0.55$ ,  $P<0.001$ ) and for the experts dexterity rating: DC ( $r=0.38$ ,  $P=0.006$ ), SP ( $r=0.53$ ,  $P<0.001$ , *Figure 2*), PT ( $r=-0.34$ ,  $P=0.01$ ), and MIT ( $r=-0.47$ ,  $P<0.001$ ) (*Table 3*).

### Consequences

The pass-/fail criterion of 8 points in anatomy-rating made 30 participants fail and 22 pass. The participants who were failed by the expert raters had significantly lower AIBA scores than participants who passed: DC (mean difference: 3.7 segments,  $P=0.01$ ), SP (2.8 progressions,  $P=0.004$ ), PT ( $-84.3$  seconds,  $P=0.03$ ), MIT ( $-23.1$  seconds,  $P<0.001$ ) (*Table 1*).

## Discussion

This is the first study to establish validity evidence for

**Table 3** Validity evidence towards relationship to other variables: correlations between former, subjective rating tool (expert-rating and dexterity-rating) and new, automatic, and objective rating tool (AIBA)

AIBA/expert rating	Anatomy-rating		Dexterity-rating	
	Pearson's r	P	Pearson's r	P
DC, segments	0.47	<0.001	0.38	0.006
SP, progressions	0.57	<0.001	0.53	<0.001
PT, seconds	-0.32	0.02	-0.34	0.01
MIT, seconds	-0.55	<0.001	-0.47	<0.001

AIBA, artificial intelligence bronchoscopy assessment; DC, diagnostic completeness; SP, structured progress; PT, procedure time; MIT, mean intersegmental time.

automatic, instant, and unbiased assessment of bronchoscopy performance using AI. The expert rating tool could be used as a gold standard, as it has previously documented evidence of validity (11) and our study confirmed a high interrater reliability and high internal consistency reliability.

A bronchoscopist should have a high level of anatomy knowledge to ensure inspection of all bronchial segments. Therefore, DC is the first and most widely used metric to assess bronchoscopy competence (15,21). DC correlated significantly with anatomy rating ( $r=0.47$ ,  $P<0.001$ ), however only providing a moderate correlation. By repeatedly entering all segments that are randomly visualized during a procedure, novice bronchoscopists can achieve a high DC score without knowing where they are. This makes it relatively easy to get a DC score close to maximum introducing a ceiling effect and limiting its discriminatory abilities. This is in accordance with previous studies that found DC to have no value in differentiating skill level (14,22). Colt *et al.* (15) even found that novice bronchoscopists with one day of training on a VR simulator had a higher DC than clinically experienced bronchoscopists when tested on a phantom. Therefore, DC indicates a thorough but not necessarily structured examination and cannot assess anatomy knowledge on its own.

SP had the best correlation to anatomy-rating (Pearson's correlation coefficient: 0.57,  $P<0.001$ , *Figure 1*). This is in accordance with former studies, showing SP to be a more discriminatory measure for competence than DC (16,20,23). Therefore, to provide a valid interpretation of DC, this study suggests evaluating DC in the light of SP, as a measure of anatomical knowledge indicated by a measure of systematically inspection of the segments. Novices are able to obtain a median DC of 14.5 while only having an SP score of 3 (16) indicating an unstructured inspection of the bronchial tree where all segments in sight are visited (and re-visited) in a random fashion (24). Novices trained using the AI however obtain a median DC of 18 and SP of 16.5 (16). This finding is in accordance with our study, showing SP to have a higher discriminatory ability than DC, as it shows a stronger correlation to anatomy-rating. SP is an outcome measure with a high level of validity evidence, and the most recent systematic review in the effectiveness of bronchoscopy training suggests the measure's incorporation into future assessment tools as done for this study (21).

The interpretation of SP should be done with caution, as some experts might follow their own inspection order (RB3→RB2→RB1) resulting in a low SP. We believe that

experts should be free to develop and follow their own preferred procedural strategies but to ensure consistent performance of trainees, novice bronchoscopists should be trained in a structured order, e.g., based on SP. AIBA was not designed to assess experts but it allows automatic and instant assessment of trainees without relying on the busy schedule of expert raters.

Besides anatomy-knowledge, the bronchoscopists should have a high level of dexterity to perform an efficient and careful procedure. AIBA does not present a direct measure of dexterity like wall collisions or centering. However, DC correlated significantly with dexterity-rating ( $r=0.38$ ,  $P=0.006$ ), and SP showed a stronger correlation to dexterity-rating. This is in accordance with a previous study, as bronchoscopists with good anatomical knowledge (SP-score) also showed a higher dexterity level (23). This is the first simulator to provide an automatic measure of dexterity, since VR-simulator metrics such as wall collisions and % time midlumen are not able to distinguish dexterity (14,25-27), and external systems relying on extensive setups are needed to assess dexterity (25). One pretest-posttest study using a VR-simulator actually observed increased level of wall-contacts and less % time in midlumen after a training intervention (27). Since the first study using a VR-simulator in bronchoscopy by Ost *et al.* in 2001 (13), ten studies have shown that training on a simulator in general improves performance on a simulator (21) even though simulators' performance metrics are severely limited and cannot assess dexterity performance. Previously, performance could not be automatically assessed using phantoms, which is unfortunate as phantoms have several strengths. They provide haptic feedback and use the same setup as in a clinical setting. Experienced bronchoscopists rated an inexpensive 3D-printed phantom as more realistic than VR-simulators (28) and phantoms might be superior to VR-simulators for bronchoscopy training (29). This is the first study to provide automatic, instant, and reliable ratings that are correlated with dexterity, but further development of the AI algorithm is needed to directly measure scope centering and wall collisions.

Both time metrics had a significant inverse correlation to anatomy-rating: PT ( $r=-0.32$ ,  $P=0.02$ ), MIT ( $r=-0.55$ ,  $P<0.001$ ) and to dexterity-rating: PT ( $r=-0.34$ ,  $P=0.01$ ), MIT ( $r=-0.47$ ,  $P<0.001$ ). PT does not hold a self-explanatory value towards how well a bronchoscopy was conducted and cannot be used as the only measure of bronchoscopy quality. A bronchoscopist can insert the scope and extract it after a few seconds resulting in a PT of only

a few seconds, however not investigating any segments. PT should therefore always be interpreted along with DC and SP of the procedure. To have a self-explanatory measure of procedure efficiency, MIT can be used, which several studies have done (13,20,26,30-32). To evaluate procedure efficiency, MIT should be used instead of PT, as it correlated significantly to both SP and DC (aspect 3, internal structure, *Table 1*). Even though a systematic review in simulation-based bronchoscopy concluded that when gathering validity evidence, the outcome measure that consistently differentiated performance was PT (33), we suggest future validation studies using PT could include MIT for internal consistency analysis.

In 2015, a CHEST expert panel suggested a move away from volume-based certification to competency-based certification with the use of simulation-based training (9). A pass-fail score of 8 points in anatomy-rating would pass 22 and fail 30 of the participants, and the passing group performed significantly better for all AIBA's outcome measure than the failing group. This finding indicates that the outcome measures constituting AIBA can be used to set proficiency training criteria, enabling mastery learning, which is a training modality where trainees practice until these proficiency targets are met (34). No training studies in simulation based bronchoscopy training has used mastery learning (21), even though it has been recommended (33). We encourage fellow researchers to do a mastery learning training study using AIBA, to examine how much training time with feedback from the AI is needed to reach a pre-defined mastery learning level (16).

This is the first and largest validation study in bronchoscopy to assess performance automatically, instantly and without bias. However, our study holds several limitations. The study was conducted in a simulated setting to provide a standardized and safe environment, therefore AIBA cannot be directly applicable to clinical bronchoscopies. The participants only conducted one bronchoscopy due to the busy nature of a conference and our wish to capture a wide range of performances. Their performance might have improved and even plateaued with consecutive trials allowing them to practice the procedure and get used to the equipment. However, it is the biggest validation study regarding assessment of competence in flexible bronchoscopy and the only one with participants from six continents providing a strong level of external validity. We chose not to divide the participants into groups based on experience. When collecting validity evidence,

it is erroneous to make experienced-novice comparisons for aspect four of Messick's validity framework: i.e., relationship to other variables (35-37). We therefore chose to gather validity evidence for this aspect by correlating to a gold standard for performance by a validated assessment tool. Bronchoscopy competence was therefore assessed by expert raters, rather than experience level following current guidelines for competency rather than volume-based assessment (9). The AI holds the potential to replace expert raters, if being further developed to provide the bronchoscopists with direct dexterity measures. This is the first AI navigational system tested in bronchoscopy, and development of the AI should entail direct dexterity measures as those assessed by the expert raters. In the future, the AI could be implemented in everyday clinical practice to help ensure competent performance before allowing trainees to begin supervised practice on patients.

## Conclusions

AI can provide immediate and automatic assessment of anatomical and navigational competencies in flexible bronchoscopy. SP provided stronger correlations with human experts' ratings than the traditional DC.

## Acknowledgments

*Funding:* None.

## Footnote

*Reporting Checklist:* The authors have completed the STROBE reporting checklist. Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-841/rc>

*Data Sharing Statement:* Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-841/dss>

*Peer Review File:* Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-841/prf>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-841/coif>). K.M.C. received funding from Ambu regarding The CoRS-feedback study in colonoscopy: NCT04862793. S.S. has received funding from Ambu A/S. L.K. has annotated

clinical bronchoscopy videos for Ambu's development of the AI system. The other authors have no conflicts of interest to disclose.

*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). Ethical approval is not required for this study in accordance with local or national guidelines, as the study was conducted in a simulated setting not involving patients, but voluntary participants at the ERS 2023 conference signed an informed consent form.

*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

- Andolfi M, Potenza R, Capozzi R, et al. The role of bronchoscopy in the diagnosis of early lung cancer: a review. *J Thorac Dis* 2016;8:3329-37.
- Hsu LH, Liu CC, Ko JS. Education and experience improve the performance of transbronchial needle aspiration: a learning curve at a cancer center. *Chest* 2004;125:532-40.
- Ouellette DR. The safety of bronchoscopy in a pulmonary fellowship program. *Chest* 2006;130:1185-90.
- Stather DR, MacEachern P, Chee A, et al. Trainee impact on procedural complications: an analysis of 967 consecutive flexible bronchoscopy procedures in an interventional pulmonology practice. *Respiration* 2013;85:422-8.
- Ernst A, Silvestri GA, Johnstone D, et al. Interventional pulmonary procedures: Guidelines from the American College of Chest Physicians. *Chest* 2003;123:1693-717.
- Bolliger CT, Mathur PN, Beamis JF, et al. ERS/ATS statement on interventional pulmonology. *European Respiratory Society/American Thoracic Society. Eur Respir J* 2002;19:356-73.
- Barsuk JH, Cohen ER, Feinglass J, et al. Residents' Procedural Experience Does Not Ensure Competence: A Research Synthesis. *J Grad Med Educ* 2017;9:201-8.
- Leong S, Shaipanich T, Lam S, et al. Diagnostic bronchoscopy--current and future perspectives. *J Thorac Dis* 2013;5 Suppl 5:S498-510.
- Ernst A, Wahidi MM, Read CA, et al. Adult Bronchoscopy Training: Current State and Suggestions for the Future: CHEST Expert Panel Report. *Chest* 2015;148:321-32.
- Davoudi M, Osann K, Colt HG. Validation of two instruments to assess technical bronchoscopic skill using virtual reality simulation. *Respiration* 2008;76:92-101.
- Konge L, Larsen KR, Clementsen P, et al. Reliable and valid assessment of clinical bronchoscopy performance. *Respiration* 2012;83:53-60.
- Konge L, Clementsen P, Larsen KR, et al. Establishing pass/fail criteria for bronchoscopy performance. *Respiration* 2012;83:140-6.
- Ost D, DeRosiers A, Britt EJ, et al. Assessment of a bronchoscopy simulator. *Am J Respir Crit Care Med* 2001;164:2248-55.
- Konge L, Arendrup H, von Buchwald C, et al. Using performance in multiple simulated scenarios to assess bronchoscopy skills. *Respiration* 2011;81:483-90.
- Colt HG, Crawford SW, Galbraith O 3rd. Virtual reality bronchoscopy simulation: a revolution in procedural training. *Chest* 2001;120:1333-9.
- Cold KM, Xie S, Nielsen AO, et al. Artificial Intelligence Improves Novices' Bronchoscopy Performance: A Randomized Controlled Trial in a Simulated Setting. *Chest* 2024;165:405-13.
- Huang J, Lin J, Lin ZL, Shiyue., Changhao. Z. Old wine in a new bottle or true innovation. *CHEST*. 2023;In Press.
- Cold KMK, Lars. Response Letter to the Editor. *Chest*. 2023;In Press.
- Messick S. Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-104). New York, NY: American Council on education and Macmillan.; 1989.
- Cold KM, Svendsen MBS, Bodtger U, et al. Using structured progress to measure competence in flexible bronchoscopy. *J Thorac Dis* 2020;12:6797-805.
- Gerretsen ECF, Chen A, Annema JT, et al. Effectiveness of Flexible Bronchoscopy Simulation-Based Training: A Systematic Review. *Chest* 2023;164:952-62.
- Moorthy K, Smith S, Brown T, et al. Evaluation of virtual reality bronchoscopy as a learning and assessment tool. *Respiration* 2003;70:195-9.
- Cold KM, Svendsen MBS, Bodtger U, et al. Automatic and Objective Assessment of Motor Skills Performance in



- Flexible Bronchoscopy. *Respiration* 2021;100:347-55.
24. Cold KM, Konge L. Simulation-Based Training in Flexible Bronchoscopy: Best Practices and Future Directions. *Chest* 2023;164:820-1.
  25. Colella S, Søndergaard Svendsen MB, Konge L, et al. Assessment of competence in simulated flexible bronchoscopy using motion analysis. *Respiration* 2015;89:155-61.
  26. Veaudor M, Gérinière L, Souquet PJ, et al. High-fidelity simulation self-training enables novice bronchoscopists to acquire basic bronchoscopy skills comparable to their moderately and highly experienced counterparts. *BMC Med Educ* 2018;18:191.
  27. Schertel A, Geiser T, Hautz WE. Man or machine? Impact of tutor-guided versus simulator-guided short-time bronchoscopy training on students learning outcomes. *BMC Med Educ* 2021;21:123.
  28. Pedersen TH, Gysin J, Wegmann A, et al. A randomised, controlled trial evaluating a low cost, 3D-printed bronchoscopy simulator. *Anaesthesia* 2017;72:1005-9.
  29. Kennedy CC, Maldonado F, Cook DA. Simulation-based bronchoscopy training: systematic review and meta-analysis. *Chest* 2013;144:183-92.
  30. Bjerrum AS, Eika B, Charles P, et al. Dyad practice is efficient practice: a randomised bronchoscopy simulation study. *Med Educ* 2014;48:705-12.
  31. Bjerrum AS, Eika B, Charles P, et al. Distributed practice. The more the merrier? A randomised bronchoscopy simulation study. *Med Educ Online* 2016;21:30517.
  32. Bjerrum AS, Hilberg O, van Gog T, et al. Effects of modelling examples in complex procedural skills training: a randomised study. *Med Educ* 2013;47:888-98.
  33. Naur TMH, Nilsson PM, Pietersen PI, et al. Simulation-Based Training in Flexible Bronchoscopy and Endobronchial Ultrasound-Guided Transbronchial Needle Aspiration (EBUS-TBNA): A Systematic Review. *Respiration* 2017;93:355-62.
  34. Siddaiah-Subramanya M, Smith S, Lonie J. Mastery learning: how is it helpful? An analytical review. *Adv Med Educ Pract* 2017;8:269-75.
  35. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)* 2016;1:31.
  36. Cook DA. Much ado about differences: why expert-novice comparisons add little to the validity argument. *Adv Health Sci Educ Theory Pract* 2015;20:829-34.
  37. Cook DA, West CP. Perspective: Reconsidering the focus on "outcomes research" in medical education: a cautionary note. *Acad Med* 2013;88:162-7.

**Cite this article as:** Cold KM, Agbontaen K, Nielsen AO, Andersen CS, Singh S, Konge L. Artificial intelligence for automatic and objective assessment of competencies in flexible bronchoscopy. *J Thorac Dis* 2024;16(9):5718-5726. doi: 10.21037/jtd-24-841