Reviewer A:

This study evaluated whether or not the SP, DC, and AIT are the indicators on the competence of bronchoscopy by using a bronchoscopy model. The authors demonstrated these items were correlated well with the experience of bronchoscopy.
In this study, the bronchoscopy skill of operators, "novice," "intermediate," and "expert," was classified by using the number of procedures performed, thus the number of procedures seems to be still a gold standard for assessing the competence of bronchoscopists. However, from the results of this study, SP, DC and AIT is likely to be able to be used as another assessment tool for competence. This is a well written paper.

We thank the reviewer for his/her positive words and kind feedback.

Comment 1: *Although the results of this study by itself seem to provide new knowledge to readers not so much, they will become a base for making an assessment tool on competence of bronchoscopy in future.*
Reply 1: We agree with the reviewer that SP will become a basis for making an assessment tool on competence in bronchoscopy. It adds new nuances to the assessment of bronchoscopy performance by highlighting the importance of a systematic progression. Future studies should compare our automatic and objective assessment tool with that of expert raters, which is now included in the manuscript.

Changes in the text:

Line 261, "Based on this study we aim in the future to develop a computerized feedback system that is feasible when training on phantoms to ensure a higher learning outcome. This could also work as an automatic assessment tool, which will be able to set an unbiased pass/fail criterion that is not susceptible to rater bias. Future studies should compare our automatic, objective scores with existing more subjective scoring systems from expert raters [1]."
Line 278, "This study identified SP as a useful outcome measure to assess competency in bronchoscopy. SP can be used alongside existing measures of competence (DC and AIT) [2-4], as it add new useful nuances to the assessment of bronchoscopy performance that are not captured by existing measures – it was the only of the three outcome measures that was able to differentiate between intermediates and experts."

Reviewer B:
The authors present a study of bronchoscopy competency using a new checklist tool - structured progress - that tracks the ability of the bronchoscopist to move from airway to airway in structured sequential order. 34 participants were recruited from 11 centers with varying levels of experience.

The participants were evaluated on a standard model (broncho boy) for a complete airway inspection with 3 methods: structured progress, diagnostic completeness, and average intersegmental time. There were differences between novices and experts in diagnostic completeness and intersegmental time. There were differences between novice, intermedius and experts in structured progress.

Pros:
1. Well-developed study with a relatively large cohort of subjects that should give a sufficent number for validity
2. Standardized model was used
3. STROBE statement used

We thank the reviewer for his/her valuable comments and suggestions.

Comment 1: *I have trouble with the fundamental use of the structured progress. First it is rated from 0-18 and marks are given when there is correct 'transition' from one airway to the next. Since there are only 18 standard anatomic airways, it should be out of 17(17 transitions).*
+ The method of SP evaluation needs to be better described in the manuscript.
Reply 1: We thank the reviewer for pointing this out. He/she is totally correct that it is the transition from one segment to the succeeding segment that scores a point. However, one point is also awarded by transitioning from the carina to segment RB1. This is explained in the revised manuscript along with a more comprehensive description of the SP evaluation that now reads:

<u>Changes in the text:</u>

Line 155, "SP indicates the number of times the operator progressed from one segment to the immediate succeeding segment. Passing from segment #1 to #2 gives one point but going from segment #2 to #1 gives zero point. SP ranges from 0-18 as the potential first point is awarded by passing from the carina to segment #1 either on the right or left side depending on which lung is inspected first. Progression from the carina to segment #1 to segment #2 to segment #3 gives 3 out of three points, while progressing from the carina to segment #1 to segment #3 to segment #2 gives 1 out of three points."

Comment 2: *The SP scoring is not shown and I worry that it would be very arbitrary. For instance, is it correct to move from RB1 to RB2 to RB3 or RB1 to RB3 to RB2? The former may grant 2 points and the latter may grant 0 points, but all 3 airways are examined.*
The arbitrary nature is highlighted by the range of scores. Experts (with a diagnostic completeness of 0.94) has a mean score of 11/18 with a broad interquartile range of 5. This either shows that there is broad variability between experts (and they shouldn't be called experts) or it is a flawed test for competency.
Since SP is a fundamental component of the manuscript, I believe the entire basis would have problems.

Reply 2: We agree with the reviewer when he/she points out the difference between examining <u>all segments</u> (i.e. RB1 to RB3 to RB2 in the example: diagnostic completeness=100%, structured progression (SP) score=1 point out of three) and examining the segments <u>in the correct order</u> (i.e. RB1 to RB2 to RB3 in the example: diagnostic completeness=100%, SP score=3 points out of three). As the reviewer points out this difference is evident throughout the results. However, we do not consider this trait of the SP scoring to be "very arbitrary" – on the contrary, we have shown that the SP score adds nuances to the assessment of bronchoscopy performance that is not captured by diagnostic completeness. As reported by the reviewer, SP was the only outcome measure (SP, DC and AIT) that could differentiate between intermediates and experts. SP would therefore be the best assessment tool to differentiate between experience levels.

SP correlated positively with DC and negatively with AIT both across groups and within groups (aspect 3: internal structure). This indicates that those expert who proceeds from RB1 to RB2 to RB3 inspect more segments and have a lower procedure time. Therefore, SP is an extra measure when evaluating bronchoscopy performance with a clear message for new operators: The segments should be inspected in an orderly ascending order as those experts who do inspect the most segments and have the lowest procedure time.

These points are now further elaborated in the manuscript:

Changes in the text:

Line 278, "This study identified SP as a useful outcome measure to assess competency in bronchoscopy. SP can be used alongside existing measures of competence (DC and AIT) [2-4], adding new useful nuances to the assessment of bronchoscopy performance that are not captured by existing measures – it was the only of the three outcome measures that was able to differentiate between intermediates and experts."

Comment 3: *There are other evaluation tools available for bronchoscopy and the study should compare their new scoring method against an already validated method.*

Reply 3: We agree with the reviewer that assessment tools with validity evidence already exist for flexible bronchoscopy. The following paper was published in the Journal of Thoracic Diseases in 2019 and highlights some of these: Pietersen PI, Laursen CB, Petersen RH, Konge L. Structured and evidence-based training of technical skills in respiratory medicine and thoracic surgery. J Thorac Dis 2019. doi: 10.21037/jtd.2019.02.39. This paper is now referenced in the revised manuscript. However, the existing tools require the availability of expert rater(s) and are subject to bias. We thank the reviewer for his/her remark and totally agree that future studies should aim at comparing our automatic, objective scores with existing more subjective scoring systems.

Changes in the text:

Line 261, "Based on this study we aim in the future to develop a computerized feedback system that is feasible when training on phantoms to ensure a higher learning outcome. This could also work as an automatic assessment tool, which will be able to set an unbiased pass/fail criterion that is not

susceptible to rater bias. Future studies should compare our automatic, objective scores with existing more subjective scoring systems from expert raters [1]."

Reviewer C:

This is an interesting study in the field of procedural education, a field that is sadly neglected in current research. I do feel that a few clarifications in the discussion would be helpful.

We thank the reviewer for his/her positive feedback.

Comment 1: *At line 205, you state "Several findings in our study indicated this correlation to be a causal relation." What follows, however, is an explanation of why better standards for procedural competency are needed. This breaks up the flow of ideas and leads to confusion when reading the manuscript. I would reorganize the discussion such that this comment immediately precedes the discussion of findings that indicate causal relationships. This would help with readability.*
Reply 1: We agree with the reviewer that the sentence breaks up the flow of ideas. It has now been moved to the following section:

Changes in the text:

Line 220, "Our study showed a positive correlation between DC and SP (Pearson correlation, *r*=.62, *P*<.001). Several findings in our study indicated this correlation to be a causal relation. It seems obvious that operators should visualize as much of the bronchial tree as possible; to ensure this, a bronchoscopy should optimally be completed in a systematic approach. This finding holds an obvious confounder, as the groups were divided by experience level."

Comment 2: *Please clarify whether systemic progression should be used alongside your comparator measures or instead of them.*
Reply 2: SP should be used alongside the comparator measures. This point is now further elaborated in the manuscript:

Changes in the text:

Line 278, "This study identified SP as a useful outcome measure to assess competency in bronchoscopy. SP can be used alongside existing measures of competence (DC and AIT) [2-4], adding new useful nuances to the assessment of bronchoscopy performance that are not captured by existing measures – it was the only of the three outcome measures that was able to differentiate between intermediates and experts."

Comment 3: *Please comment on whether presence of a known abnormality should alter the systematic progression and in what way. You use the correlation of staging EBUS in your argument in favor of*

systematic progression, but the way EBUS is performed is different depending on the laterality of the lesion in question. Many bronchoscopists, myself among them, advocate starting bronchoscopic inspection on the contralateral side to the lesion in order to ensure complete airway inspection.

Reply 3: We agree with the reviewer that the presence of a known abnormality would alter the systematic progression of the procedure. In the case of a known abnormality in the right lung, the first point for SP should be awarded by progressing from the vocal cords to LB1 and vice versa. The manuscript now reads:

<u>Changes in the text:</u>

Line 157, "SP ranges from 0-18 as the potential first point is awarded by passing from the carina to segment #1 either on the right or left side depending on which lung is inspected first."

Line 242, "Therefore, a systematic approach is important when performing flexible bronchoscopy to ensure no segments are missed, and our measure of SP can be used as a valid assessment tool for user competence. The scoring system for SP is also adaptable for example in the case of a known abnormality where bronchoscopic inspection should start in the contralateral side of the lesion. One point is awarded for each systematic progression such that if the inspection starts in the left lung, the first point is given when progressing from carina to segment #1 on the left side."