



Development of and Gathering Validity Evidence for a Theoretical Test in Thoracic Ultrasound

Pietersen, Pia Iben; Konge, Lars; Madsen, Kristian Rørbæk; Bendixen, Morten; Maskell, Nick A.; Rahman, Najib; Graumann, Ole; Laursen, Christian B.

Published in:
Respiration

DOI:
[10.1159/000500146](https://doi.org/10.1159/000500146)

Publication date:
2019

Document version
Early version, also known as pre-print

Citation for published version (APA):
Pietersen, P. I., Konge, L., Madsen, K. R., Bendixen, M., Maskell, N. A., Rahman, N., Graumann, O., & Laursen, C. B. (2019). Development of and Gathering Validity Evidence for a Theoretical Test in Thoracic Ultrasound. *Respiration*, 98(3), 221-229. <https://doi.org/10.1159/000500146>

Development and gathering validity evidence for a theoretical test in thoracic ultrasound.

Short title: Test of knowledge in thoracic ultrasound

Pia Iben Pietersen^{abc}, Lars Konge^d, Kristian Rørbæk Madsen^e, Morten Bendixen^f, Nick A. Maskell^g, Najib Rahman^h, Ole Graumann^{ci}, Christian B. Laursen^{ac}

^aDepartment of Respiratory Medicine, Odense University Hospital, Denmark

^bRegional Center for Technical Simulation, Odense University Hospital, Denmark,

^cInstitute for Clinical Research, University of Southern Denmark, Denmark,

^dCopenhagen Academy for Medical Education and Simulation, Capital Region of Denmark, Denmark,

^eDepartment of Anaesthesiology and Intensive Care Medicine, Odense University Hospital, Denmark,

^fDepartment of Cardiothoracic and Vascular Surgery, Aarhus University Hospital,

^gUniversity of Bristol, England

^hOxford Pleural Unit and NIHR Biomedical Research Unit, Oxford University Hospital, England

ⁱDepartment of Radiology, Odense University Hospital, Denmark

Corresponding author:

Pia Iben Pietersen

Department of Respiratory Medicine

Odense University Hospital

Kranvej 71

5000 Odense C, Denmark

Telephone: +45 2060 5630

Mail: pia.iben.pietersen3@rsyd.dk

Keywords:

- Ultrasound technique
- Education
- Medical training
- Thoracic imaging

Abstract:

Background: Thoracic ultrasound (TUS) has a high diagnostic accuracy for many common pulmonary diseases, but theoretic knowledge in sonographic physics, thoracic anatomy and physiology, and sonopathologic patterns is required to develop competence.

Objectives: The aims of the study were to develop and gather validity evidence for a theoretical test in TUS and to establish a pass/fail standard.

Methods: Content was provided based on expert interviews, leading to creation of 113 initial multiple-choice (MCQ) items. Consensus was reached on 92 proceeding items through a Delphi-process, and items were presented to physicians with different knowledge and experience in TUS. Answers were used for item statistics in order to select the items with the most optimal item discrimination and difficulty (i.e. Level 1 items) to be included in the final test. Mean scores of the novice-, intermediate- and experienced group were compared, and a pass/fail score was established using the contrasting groups' standard setting method.

Results: Item statistic revealed 38 Level 1 items, of which 30 were selected to be included in the final test. The internal consistency was high (Cronbach's alpha =0.88). Differences in mean scores were 8.6 points ($p<0.001$), 6.3 points ($p=0.01$), and 14.0 points ($p<0.001$) between novices and intermediates, intermediates and experienced, and novices and experienced, respectively. A pass/fail standard of 20 points was established.

Conclusion: The established MCQ-test can distinguish between physicians with different level of competence in TUS, and enables an objective, evidence-based approach for assessing the theoretical knowledge for trainees undergoing an educational program in TUS.

Introduction

Thoracic ultrasound (TUS) is a bedside, rapid and low-cost diagnostic tool for clinical assessment of patients with suspected disease in the chest [1-3]. In contrast to other thoracic imaging techniques like chest x-ray, computed tomography (CT), and high resolution computed tomography (HRCT), TUS is a dynamic examination performed real-time by a physician in a clinical setting.

Several requirements and competencies are needed to perform a sufficient ultrasound examination [4]. The physician must possess the capability to consider differential diagnoses, perform the examination, and put the findings in context to the history taken, vital parameters, objective assessment, and results of other tests in order to use TUS for clinical decision making [5]. Incorrect treatment or decisions can be made if the physician misinterprets sonopathologic patterns. Therefore, basic knowledge in anatomy, physiology, pathophysiology, and sonographic physics is necessary to develop the complex competence of an ultrasound examination [6].

Recommendations for training and assessment in TUS from the international societies and federations include lists of theoretical knowledge needed prior to certification or completion of a course [7,8]. The lists include: physics and technology, ultrasound techniques and sonographic anatomy like diaphragm and chest wall, but do not elaborate on *how* this knowledge should be obtained. Assessment and tests have been widely used in order to ensure adequate level of theoretical knowledge, which is a prerequisite for developing technical skills [6,9,10].

Validity evidence should be gathered prior to implementation when incorporating a theoretic test into an educational programme [11]. This to confirm that the test actually measures what it is meant to measure, so that no incompetent trainee passes and performs insufficient examinations.

If the adequate level is not complied, it would provide risk to the patients due to potential misinterpretation of TUS findings, and poor clinical decision-making. A vast majority of the published literature in TUS education include theoretical assessment, but do not present validity evidence for the tests, which make the assessment data futile [12-15].

To our knowledge no previous studies have gathered validity evidence for a theoretical test in clinical TUS. The aims of this study were 1) to develop and establish validity evidence of a theoretical MCQ-test in TUS, and 2) to establish a pass/fail score.

Material and methods

Setting

The study took place at the Regional Center for Technical Simulation in Region of Southern Denmark, Odense University Hospital, and at Copenhagen Academy for Medical Education and Simulation (CAMES), the Capital Region of Denmark, from November 2017 to June 2018.

The study consisted of four phases 1) development of initial questions, 2) consensus of proceeding questions, 3) selection of final questions, and 4) gathering validity evidence for the whole test (Figure 1). Gathering validity evidence for the test was done according to the framework by Messick. Table 1 summarize the framework and approaches in the study [11].

Phase 1 - Development of initial items

Five experts in TUS from four specialities (pulmonary medicine (n=2), intensive care medicine (n=1), thoracic surgery (n=1), and radiology (n=1)), were invited to take part of the study. The experts represented four different centres in two European countries (Denmark, England). In order to assign content for the curriculum, the five experts were asked to answer the following questions qualitatively;

- 1) Which technical knowledge/functions are prerequisite when making a thoracic ultrasound?*
- 2) Which sonoanatomic structures are mandatory to know about and recognize when performing thoracic ultrasound?*
- 3) Which pathological sonographic patterns/artefacts are essential to recognize, to make a sufficient and valid thoracic ultrasound examination?*

Based on the expert opinions, initial questions were constructed according to guidelines of writing MCQs [16-18]. The selected MCQ format was single-best-answer with three possible answers; one correct answer and two less correct options, which is proven to be an effective testing approach in MCQ testing [17]. Figure 2, presents one of the questions included in the final test.

Phase 2 - Consensus of proceeding items (Content validity)

According to the Delphi method [19], the experts were individually asked to rate the questions from 1-5 in relation to relevance (1 being completely irrelevant, 5 being completely relevant), and to comment on the items if there were issues in relation to language, grammar, and understanding of items. Ratings were gathered after each round, and the mean score of each question was calculated. Questions with a mean score less than 4 were excluded. Additionally, language and grammar issues were corrected.

Phase 3 - Selection of final items

Three study groups were created based on level of experience in TUS:

- Novices: Medical students or physicians with no or very little ultrasound experience. No attendances on ultrasound courses or educational programs in TUS. Basic course in general ultrasound was allowed. No or ≤ 20 TUS examinations performed.
- Intermediates: First to fifth year residents in respiratory medicine, or residents in other specialities that use TUS on regularly basis (i.e. intensive care medicine, emergency medicine, thoracic surgery, cardiology, or radiology). Moderate experience in TUS from attendances in ultrasound courses. Performed ≥ 20 TUS examinations. No research or providing teaching in TUS.
- Experienced: Physicians with a long experience in TUS, comparable to Royal College of Radiologists (RCR) or European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB) competency level 3 in TUS, doing TUS almost daily, and provide teaching or research within this field [7,8].

Participants from 7 different hospitals and two universities (medical students) were invited to take part of the study by mail or educational groups. They were informed about the study, and no salary or compensation was received. The included participants were given an anonymous study ID in order to match their performances with the information about level of competencies, number of courses and number of performed examinations prior to the test. No support (e.g. handbooks, internet access) was allowed, and no time limit was complied.

After testing the participants, answers were used to assess item quality measured by item difficulty and item discrimination. Items were divided into one of four categories (I, II, III, and IV), where level I items are considered as the best items for a theoretical test [18]. Characteristics of level I items are middle range of average item difficulty (0.45 - 0.75), and with a high item discrimination (≥ 0.20).

Phase 4 - Gathering validity evidence

After the exclusion of Level II-IV items, answers from remaining Level I items were used to explore whether the test could distinguish between novices, intermediates, and experienced. The groups' mean scores were compared using ANOVA with Bonferroni's correction for multiplicity, and Cronbach's alpha was calculated as a measure of intraclass correlation coefficient (ICC) for the final test.

We established a pass/fail score with the contrasting groups' standard setting method based on the responses of the novices and experienced, and the consequences in terms of number of false-positives and false-negatives were explored [20].

Statistical analysis

Statistical analyses were performed in SPSS version 21 (IBM, Armonk, New York). Statistical analyses are described under each phases of the study, all statistics were considered significant at a 5% significance level.

Ethics

The Regional Committee on Health Research Ethics for Southern Denmark was asked for permission, the project was categorized as a quality assurance study, and approved with the case number: S-20172000-44. All data were anonymized, and did not include personal information.

Results

Identification of experts in TUS was done in the pre study phase, and was done based on previous published literature within TUS, and among physicians known to use the procedure on almost daily basis, and who have been teaching TUS. All 5 invited experts accepted the invitation and participated.

Phase 1

The interview resulted in a list of 38 topics from the experts, which formed the curriculum of the items created. A total of 113 items were constructed (Figure 1).

Phase 2

Phase 2 comprised three Delphi-rounds in which 92 of the initial 113 items reached a mean score above 4, and thereby proceeded to be tested by 31 participants (in Phase 3).

Phase 3

The 92 proceeding items were subsequently presented to 31 participants who took part in the study. The answers were used for item statistics that identified 38 level I items (proper difficulty), 29 level II items (difficult, but usable), 7 level III items (very easy or very difficult), and 18 level IV (very easy or very difficult) items. All Level II, III and IV items were excluded (n=54) and three authors reviewed the remaining 38 items. Eight items were excluded in order to balance the content between sonographic physics and clinical items and avoid repetitive items. The final test comprised 30 level I items. Cronbach's alpha was 0.88.

Phase 4

Mean score of the novices was 11.0 (95% CI 8.8 - 13.4) with a minimum of 5 points and a maximum of 14 points. The intermediates' mean score was 19.7 (95% CI 17.1 - 22.3) and a minimum of 14 and a maximum of 27. The experienced had a significantly higher mean score of 26 points (95% CI 24.2 - 27.8), minimum was 23 and maximum was 29 points. Figure 3 presents a boxplot illustrating the test scores of each group including median scores. The post hoc ANOVA with Bonferroni correction for multiplicity, proved statistically significant differences between all three groups (Table 3).

Using the contrasting groups' standard setting method, a pass/fail standard of 20 was calculated based on the distribution test score from the novices and the experienced (Figure 4). The consequences were that no novices passed (no false-positives), and no experienced failed (no false-negatives).

Discussion

We created a theoretical test in TUS, consisting of 30 items with appropriate item difficulty and good item discrimination according to the recommendations [18]. The test could convincingly distinguish between novices, intermediates, and experienced, respectively. We established a pass/fail standard according to the contrasting groups' standard method with no false negative and no false positive [18].

The *content* of the items created was based on correspondences with five international experts. We included experts from different specialities in order to establish a curriculum suitable for

educational programmes in a broad variation of specialities and thereby increase the generalizability. TUS, as a diagnostic tool, could benefit practitioners other than pulmonologists, e.g. thoracic surgeons, paediatricians, anaesthesiologists, intensivists, emergency medicine physicians, and radiologists because of the high diagnostic accuracy [1,3,21,22]. Furthermore, TUS can also be used for evaluation of potential postprocedural complications like pneumothorax after placement of central venous catheter, oesophageal or bronchial intubation in emergency settings. Therefore the use of TUS can be expected to increase even more in the future [2,23]. This multidisciplinary approach could contribute to a more cost effective educational programme, where the focus is on the specific procedure, rather than an educational programme for physicians in a specific speciality. To our knowledge this study is the first theoretical assessment tool with proven validity evidence in a thoracic procedures [24,25].

Hopefully, this multidisciplinary approach and generalizability will contribute to a more uniform procedure among countries and specialities, which today is a major issue in TUS, because of the different protocols and terms [26-28].

A priority was that the test included both technical items regarding sonographic physics, knowledge regarding the practical use of an ultrasound machine, as well as clinical items. Previous studies have described the use of a theoretical test in the form of MCQs for assessment in TUS [12,14,15,29]. These studies however used theoretical assessment as a measure of skills obtained in TUS after a course or educational programme.

The current test was not intended as a pass/fail test of overall competencies in TUS. Theoretical tests are incomplete to assess practical skills, assess the ability to integrate the procedure into a clinical setting, and incorporate the findings in the overall assessment of the patients [6]. Three very important aspects of a technical procedure, but regardless if the basic knowledge is not understood. In these higher and more complex layers of a technical procedure, the trainee should be assessed for their practical skills on e.g. healthy figurants or as simulation-based training [30], and later on, assessment in a clinical setting performing the examination on patients [6,31].

Several assessment tools for the clinical setting have been developed, with solid validity evidence for assessment of competences in a clinical setting [32,33]. In these assessment tools, all aspects are being evaluated. Hence, there is a demand for moving away from the volume thresholds or theoretical assessments to the competence-based assessments with appropriate tests for the different levels of the competence.

In order to minimize sources of errors or bias associated to the test administration, *response process*, all participants, except three, filled out the answer sheet in written form with the items administered in a PowerPoint presentation on a computer disconnected from the Internet. Three participants were not able to fulfil the answer sheet while being observed, either because of geographic location or being interrupted during the completion of the test. Because all three participants were allocated in the same group (experienced) this could potentially have influenced the final results and increased the mean of the respective group.

Calculating item difficulty and item discrimination assessed the internal structure. All included items were level I items, which meets the recommendations [18]. Compared to other studies gathering validity evidence for a theoretical test in a technical procedure, not all fulfils these recommendations, and most often both level I and II items are included, like seen in Jensen et al., that presents a theoretical test in non-anaesthesiologist-administered propofol sedation for gastrointestinal endoscopy [34].

The concept of a mastery learning approach is that all trainees can learn a topic or a skill to mastery, but some need more time or more guidance than others to reach the level of competence [35]. The test between each level should then be manageable and effective, otherwise the risk of discontinuation of trainees in the educational programme, is considerable. Thus, thirty items were included in the final test covering most aspects of the curriculum with clear objectives, and because the test is administered online or by a Power-Point presentation, it can easily be applied into any setting or any educational programme. In practice, the intention with the test was that physicians should obtain theoretical knowledge based on e-learning or online material, and thereby initiate self-studying ahead of hands-on training. Subsequently, a hands-on training session for 2-3 hours, and then supervised examinations on patients with possible pathology until passed clinical assessment. This set-up requires experienced supervisors that can guide and help the trainee and make clinical assessment. However, it leads to an increased outcome for the trainees than a theoretical course or training among patients in the beginning of the learning curve, which can be stressful. The clinical, experienced supervisors are first involved in the training programme at the very end, where the supervisor can contribute to the complexity and perspective of the ultrasound findings.

To our knowledge, no validated and evidence-based mastery-learning programme in TUS has been published, but in e.g. endobronchial ultrasound (EBUS) a training programme based on the mastery-learning concept has been developed with solid validity [36]. It manufactures intended and

clear demands to the level of knowledge better than the present guidelines, and describes what should follow the theoretical module.

When evaluating the *relationship to other variables*, Messick refers to the relationship of the test scores to variables external to the test [11]. In this study the characteristics expected to correlate with the test scores were knowledge and previous experience with TUS. The results proved a conformation of this assumption. The test could discriminate between the three groups with no false-positive and no false-negatives (*consequences*), which assumes convergent and discriminant validity evidence, but in order to sustain true correlation, multiple measures of the same trait or subgroup analysis are required in additional studies [18].

A potential variable in this study design is the inter-procedural transfer of skills. Even though a trainee is novice within TUS, based on the answers from the questionnaire, the trainee could possess skills in other technical procedures that directly or indirectly imply knowledge used in TUS. No studies are identified describing transfer of skills within TUS from other procedures. Seen in a research perspective, the transfer of skills from one procedure to another, can influence the results in an educational intervention or validation study, and provide points to a trainee that was not a result of an education or directly related to, in this study, TUS. But that does not make the specific part of the assessment less valuable, and makes it possible to allocate the resources to the pitfalls and where the trainee needs more guidance.

Limitations

No gold standard for defining experienced or experts in validation studies was found in the literature. Therefore, the research group invited experts to be a part of the study based on their previous work, research, and based on the levels presented by RCR and EFSUMB [7,8]. We acknowledge this limitation and clarify that the selection of participants and division into study groups could potentially have led to selection bias, because a great amount of published literature in TUS does not necessarily ensure competencies in all aspects of TUS. The potential selection bias is probably limited since all invited experts had the highest competency level according to current recommendations and the multidisciplinary approach ensured that the expert group had a wide range of competencies within TUS.

Even though we aimed to include more than 10 participants in each group, we only managed to include 9 in the experienced group. This is a limitation to the study but was a compromise of time

and resource use in the study. Looking at other validation studies, the amount of participants vary significantly [24,32,37], and none of the studies include a power- or sample size calculations, but rely on the theory from Downing [18]. The generalizability would improve if the study-groups were larger and included international participants.

Conclusion

We succeeded in developing and gathering validity evidence for an MCQ test in TUS. The developed test can discriminate between trainees with different level of competencies, and has a pass/fail standard with no false-negative or false-positive results. As the first evidence-based theoretical assessment tool in TUS, we recommend that the test should be used as part of a TUS mastery-learning program prior to hands-on training. The test is assumed to be a more concrete approach to assess theoretical knowledge than the current guidelines.

1. Grimberg A, Shigueoka DC, Atallah AN et al. Diagnostic accuracy of sonography for pleural effusion: systematic review. *Sao Paulo Med J* 2010; 128: 90-95
2. Alrajab S, Youssef AM, Akkus NI et al. Pleural ultrasonography versus chest radiography for the diagnosis of pneumothorax: review of the literature and meta-analysis. *Crit Care* 2013; 17: R208
3. Chavez MA, Shams N, Ellington LE et al. Lung ultrasound for the diagnosis of pneumonia in adults: a systematic review and meta-analysis. *Respir Res* 2014; 15: 50
4. Nicholls D, Sweet L, Hyett J. Psychomotor skills in medical ultrasound imaging: an analysis of the core skill set. *J Ultrasound Med* 2014; 33: 1349-1352
5. Lohman D. Spatial Ability and g. In, *Human Abilities: Their Nature and Measurements*: Lawrence Erlbaum Associates, Inc. ; 1998: 97-116
6. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65: S63-67
7. EFSUMB. European Federation of Societies for Ultrasound in Medicine and Biology. Appendix 11: Thoracic Ultrasound and Assessment sheet. Available: <http://www.efsumb.org/guidelines/2009-04-14apx11.pdf>;
8. Radiologists TRCo. Ultrasound training and recommendations for medical and surgical specialities. In. 3 ed; 2017: 5-10, 58-62
9. Green ML, Moeller JJ, Spak JM. Test-enhanced learning in health professions education: A systematic review: BEME Guide No. 48. *Med Teach* 2018; 40: 337-350
10. Roediger HL, 3rd, Karpicke JD. The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspect Psychol Sci* 2006; 1: 181-210
11. Messick S. Validity. In, *Educational Measurement*. 3 ed. New York; 1988
12. Bhat SR, Johnson DA, Pierog JE et al. Prehospital Evaluation of Effusion, Pneumothorax, and Standstill (PEEPS): Point-of-care Ultrasound in Emergency Medical Services. *West J Emerg Med* 2015; 16: 503-509
13. Breitzkreutz R, Dutine M, Scheiermann P et al. Thorax, trachea, and lung ultrasonography in emergency and critical care medicine: assessment of an objective structured training concept. *Emerg Med Int* 2013; 2013: 312758
14. Noble VE, Lamhaut L, Capp R et al. Evaluation of a thoracic ultrasound training module for the detection of pneumothorax and pulmonary edema by prehospital physician care providers. *BMC Med Educ* 2009; 9: 3
15. Cuca C, Scheiermann P, Hempel D et al. Assessment of a new e-learning system on thorax, trachea, and lung ultrasound. *Emerg Med Int* 2013; 2013: 145361
16. Case SM SD. *Constructing Written Test Questions For the Basic and Clinical Sciences*. Philadelphia, PA, USA: National Board of Medical Examiners; 2001
17. Haladyna TH DS, Rodriguez MC. A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Measurement in Education* 2002; 15: 309-334
18. Downing SM. *Assessment in Health Professions Education* 1ed: Routledge; 2009
19. Harold A Linstone MT. *The Delphi Method - Techniques and Applications* In; 2002: 7-118
20. Jorgensen M, Konge L, Subhi Y. Contrasting groups' standard setting for consequences analysis in validity studies: reporting considerations. *Adv Simul (Lond)* 2018; 3: 5

21. Volpicelli G, Elbarbary M, Blaivas M et al. International evidence-based recommendations for point-of-care lung ultrasound. *Intensive Care Med* 2012; 38: 577-591
22. Lichtenstein DA. Lung ultrasound in the critically ill. *Ann Intensive Care* 2014; 4: 1
23. Das SK, Choupoo NS, Haldar R et al. Transtracheal ultrasound for verification of endotracheal tube placement: a systematic review and meta-analysis. *Can J Anaesth* 2015; 62: 413-423
24. Savran MM, Hansen HJ, Petersen RH et al. Development and validation of a theoretical test of proficiency for video-assisted thoracoscopic surgery (VATS) lobectomy. *Surg Endosc* 2015; 29: 2598-2604
25. 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
26. Lichtenstein DA. BLUE-protocol and FALLS-protocol: two applications of lung ultrasound in the critically ill. *Chest* 2015; 147: 1659-1670
27. Enghard P, Rademacher S, Nee J et al. Simplified lung ultrasound protocol shows excellent prediction of extravascular lung water in ventilated intensive care patients. *Crit Care* 2015; 19: 36
28. Gargani L. Lung ultrasound: a new tool for the cardiologist. *Cardiovasc Ultrasound* 2011; 9: 6
29. Pietersen PI, Madsen KR, Graumann O et al. Lung ultrasound training: a systematic review of published literature in clinical lung ultrasound training. *Crit Ultrasound J* 2018; 10: 23
30. Pietersen PI, Konge L, Graumann O et al. Developing and Gathering Validity Evidence for a Simulation-Based Test of Competencies in Lung Ultrasound. *Respiration* 2018; DOI: 10.1159/000493758: 1-8
31. Al-Eraky M, Marei H. A fresh look at Miller's pyramid: assessment at the 'Is' and 'Do' levels. *Med Educ* 2016; 50: 1253-1257
32. Skaarup SH, Laursen CB, Bjerrum AS et al. Objective and Structured Assessment of Lung Ultrasound Competence. A Multispecialty Delphi Consensus and Construct Validity Study. *Ann Am Thorac Soc* 2017; 14: 555-560
33. Salamonsen M, McGrath D, Steiler G et al. A new instrument to assess physician skill at thoracic ultrasound, including pleural effusion markup. *Chest* 2013; 144: 930-934
34. Jensen JT, Savran MM, Moller AM et al. Development and validation of a theoretical test in non-anaesthesiologist-administered propofol sedation for gastrointestinal endoscopy. *Scand J Gastroenterol* 2016; 51: 872-879
35. McGaghie WC. Mastery learning: it is time for medical education to join the 21st century. *Acad Med* 2015; 90: 1438-1441
36. Farr A, Clementsen P, Herth F et al. Endobronchial ultrasound: launch of an ERS structured training programme. *Breathe (Sheff)* 2016; 12: 217-220
37. Strandbygaard J, Maagaard M, Larsen CR et al. Development and validation of a theoretical test in basic laparoscopy. *Surg Endosc* 2013; 27: 1353-1359

Figure legends:

Figure 1. Study phases in the development and validation process of the test.

Figure 2. Example of one of the questions in the final test. The question type is single-best-answer with 3 answer options. The correct answer to this question is: A - hydropneumothorax. When the questions are presented to the trainees in a power point presentation or online, this question comprises a ultrasound clip, not a still picture.

Figure 3. Box-plot of results with level of competencies as dependent variable, and number of correct answers as independent variable. Dotted line indicates the calculated pass/fail score.

Figure 4. Contrasting groups' methods for standard pass/fail score

Figure 1

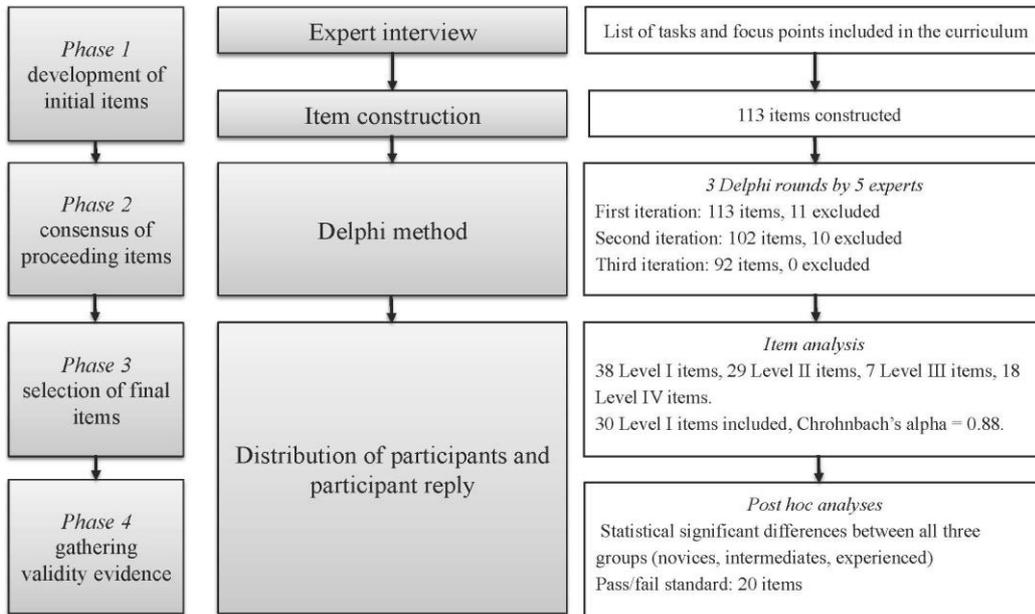
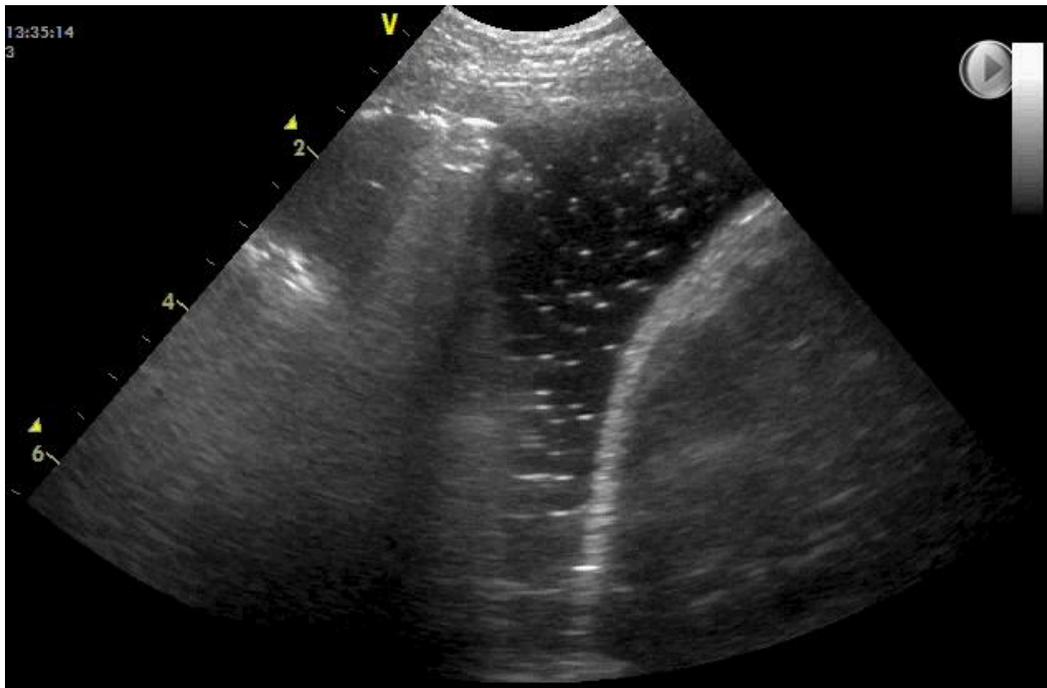


Figure 2



FLUS clip from a patient in which thoracocentesis had just been performed. What is seen in the clip?

1. Hydropneumothorax
2. Pneumothorax
3. Pleural empyema

7

Figure 3

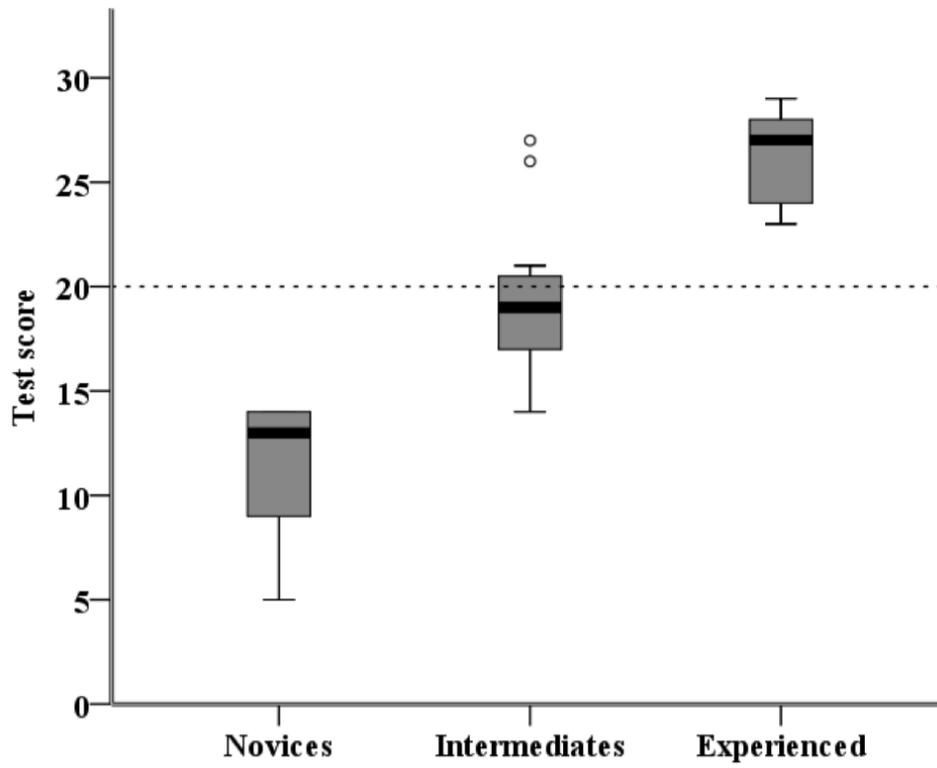


Figure 4

