Assessment of Basic Thoracic Ultrasound Skills in Immersive Virtual Reality

Gathering Validity Evidence

Nielsen, Anders Bo; Dragsbæk, Jonas; Jacobsen, Niels; Laursen, Christian B; Farr, Amy; Slavicky, Marek; Konge, Lars; Pietersen, Pia Iben

Published in:
Ultrasound in Medicine & Biology

DOI:
10.1016/j.ultrasmedbio.2023.12.002

Publication date:
2024

Document version:
Final published version

Ultrasound in Medicine & Biology 50 (2024) 467−473

Contents lists available at ScienceDirect

Ultrasound in Medicine & Biology

journal homepage: www.elsevier.com/locate/ultrasmedbio

Original Contribution

Assessment of Basic Thoracic Ultrasound Skills in Immersive Virtual Reality: Gathering Validity Evidence

Anders Bo Nielsen [a,b], Jonas Dragsbæk [c], Niels Jacobsen [d], Christian B. Laursen [e,f], Amy Farr [g], Marek Slavicky [g], Lars Konge [h], Pia Iben Pietersen [i,j,*]

[a] Department of Anesthesiology and Intensive Care, Odense University Hospital, Svendborg, Denmark
[b] SimC Simulation Centre, Odense University Hospital, Odense, Denmark
[c] Department of Radiology, Odense University Hospital, Odense, Denmark
[d] Department of Respiratory Medicine, Sygehus Lillebælt−Vejle, Vejle, Denmark
[e] Department of Respiratory Medicine, Odense University Hospital, Odense, Denmark
[f] Odense Respiratory Research Unit (ODIN), Department of Clinical Research, University of Southern Denmark, Odense, Denmark
[g] Educational Activities, European Respiratory Society, Lausanne, Switzerland
[h] Copenhagen Academy for Medical Education and Simulation (CAMES), Department of Radiology, Odense University Hospital, Odense, Denmark
[i] Department of Radiology, Odense University Hospital, Svendborg, Denmark
[j] UNIFY−Research and Innovation Unit of Radiology, Department of Clinical Research, University of Southern Denmark, Odense, Denmark

A R T I C L E I N F O

Keywords:
Medical education
Virtual reality
Thoracic ultrasound
Ultrasound
Skill assessment
Competency

Objective: Operator skills are essential for thoracic ultrasound (TUS) to ensure diagnostic accuracy. Immersive virtual reality (IVR) has shown potential within medical education but never for assessment of TUS skills. This study was aimed at developing an IVR test for assessing TUS skills, gathering validity evidence and establishing a pass/fail score.

Methods: An expert panel developed a test based on the TUS protocol by the European Respiratory Society (ERS), including a tutorial and two clinical cases (pleural effusion and interstitial syndrome), using an IVR platform (VitaSim, Odense, Denmark). Four anterior, four lateral and six posterior zones were available for examination and decision of diagnosis. Each correct examination equaled one point. The contrasting groups' method was used to set a pass/fail score.

Results: Data were collected during the 2022 ERS Congress. We included 13 novices (N, experience: 0 TUS), 22 intermediates (I, 1−50 TUS) and 11 experienced clinicians (E, >50 TUS). Cronbach's α was 0.86. The total mean point scores in case 1 (C1) were (N) 5.0 ± 2.7, (I) 7.3 ± 2.4 and (E) 8.7 ± 1.3, and the scores in case 2 (C2) were (N) 4.5 ± 1.8, (I) 6.7 ± 2.3 and (E) 8.5 ± 2.1. Significant differences were found between N and I for C1 ($p = 0.007$) and C2 ($p = 0.02$), I and E for C1 ($p = 0.04$) and C2 ($p = 0.019$) and N and E for C1 ($p < 0.001$) and C2 ($p < 0.001$). The pass/fail score was 7 points in each case.

Conclusion: We established an IVR test that can distinguish between operators with different TUS skills. This enables a standardized, objective and evidence-based approach to assessment of TUS skills.

Introduction

Thoracic ultrasound is increasingly being used because of its high sensitivity and specificity for many pulmonary and pleural diseases [1−3]. Technical development has made ultrasound available in most institutions worldwide, and this development has many advantages that can benefit patients with a faster investigation process, faster diagnosis and the possibility of closer monitoring [4,5]. Unlike most other imaging modalities, ultrasound examinations are dynamic and highly user dependent. Sufficient education and training are needed to obtain both high diagnostic accuracy and correct integration of findings with the clinical context.

Simulation makes it possible to practice technical skills and competencies in a calm and uniform environment without influencing factors such as worsened patient status, critically ill patients that need acute management and treatment or limited allowed time in the patient room [6].

Currently, simulation-based education has been restricted to ultrasound training on simulated patients, hence without sonographic pathologies [7] or virtual reality solutions requiring a large physical component, which is often expensive and not portable [8].

Immersive virtual reality is a new simulation-based modality that does not require a physical phantom but uses a head-mounted display and a 3-D interactive environment. The setup is relatively cheap, compared with physical simulators, and immersive virtual reality has shown

* Corresponding author. Kløvervænget 10, Entrance 112, 2nd Floor, 5000 Odense C, Denmark.
E-mail address: pia.iben.pietersen3@rsyd.dk (P.I. Pietersen).

https://doi.org/10.1016/j.ultrasmedbio.2023.12.002
Received 3 July 2023; Revised 30 November 2023; Accepted 1 December 2023

great potential within medical education in general, but also for ultrasound skill acquisition including thoracic ultrasound skills [9−11].

Immersive virtual reality holds the possibility of practicing ultrasound examinations with real, dynamic ultrasound images with feedback and assessment of competencies before real-life performance. This concept meets the principles of mastery learning, the educational framework that works toward anyone being able to learn a competency but with individual learning paces and need for supervision [12]. Thus, it remains unclear to which level of competency the immersive virtual reality training is needed.

This study aimed to develop an immersive virtual reality test for assessing thoracic ultrasound competencies, to gather validity evidence for the test and to establish a credible pass/fail score.

Methods

Setting

The study was designed as a validation study and carried out as training sessions at the Clinical Skills Zone during the European Respiratory Society (ERS) International Congress, September 2022, in Barcelona, Spain. The study and data collection were done in accordance with the principle of the Helsinki Declaration, and all included participants provided written consent.

Validity evidence

The internationally recommended principles for gathering validity evidence for a test using Messick's framework were used [13]. The framework included five sources of evidence to be explored: evidence of validity for content, response process, internal structure, relationship to other variables and consequences of testing.

Immersive virtual reality test (validity evidence for content)

The content was based on the thoracic ultrasound curriculum presented in the previous official ERS publication and the ERS Monograph [14,15]. Two experienced clinicians in thoracic ultrasound and two experts in medical education reached a consensus on a test including a tutorial followed by two clinical cases in English. The tutorial contained an introduction to immersive virtual reality, handling, adjustments and submission of answers.

The first case was a 56-y-old man with increasing dyspnea, cough and increased respiration rate. The thoracic ultrasound findings were a heterogenic pleural effusion with septations in zone 3R and 5R. The lung tissue was consolidated with air bronchograms suggesting pneumonia with parapneumonic pleural effusion. The diagnosis was pneumonia and complex, right-sided pleural effusion (Video S1, online only).

The second case was a 67-y-old man with increasing dyspnea and an increased respiration rate during the last hours. The thoracic ultrasound findings were multiple B-lines in zones 1R, 1L, 4R and 4L. These findings fulfilled the criteria of the interstitial syndrome.

The content was converted to immersive virtual reality cases by a software developer employed by the Danish virtual reality design company VitaSim (Odense, Denmark). The company has specialized in virtual reality for medical education and training. The test was designed using software provided by VitaSim (VitaSim Author version 0.8.5-rc3) and was installed on Meta Quest 2 (Meta Platforms Inc., Menlo Park, CA, USA) head-mounted displays.

The immersive world and patients were combined with real-life ultrasound recordings for each scanning zone shown on the virtual ultrasound machine when the respective zones were scanned. Adjustments of gain, depth and focus were possible during the scans.

A pilot test was conducted at SimC Simulation Centre, Odense University Hospital, Denmark, in August 2022. Two novices, an intermediate and an experienced physician in thoracic ultrasound, satisfactorily completed the program without improvement proposals. None of the participants from the pilot test participated in the final test and data collection.

Participants

Three study groups were created and pre-defined by the project group. The three groups were based on experience in thoracic ultrasound and covered (i) novices who were clinicians with no unsupervised thoracic ultrasound examinations, (ii) intermediates who were clinicians who had performed 1−50 thoracic ultrasound examinations in a clinical setting and, finally, (iii) an experienced group including clinicians who had performed more than 50 thoracic ultrasound examinations.

Congress delegates visiting the Clinical Skill Zone were eligible for inclusion. The Clinical Skills Zone is an area of the congress where congress delegates can drop by and experiment with simulation of clinical skills and procedures. All included participants filled out a questionnaire on demographics and thoracic ultrasound experience and were subsequently categorized into one of the three groups. Potential participants without any clinically relevant position (not physicians) who were applying at the Clinical Skills Zone were allowed to try the VR test but were excluded from enrollment in the study.

A minimum of 10 participants in each group were needed to meet the assumption of normally distributed data in medical education research [16].

Completion of the test and test collection (validity evidence for response process)

Each participant received a unique identification number to preserve anonymity and for logging performance. All instructions were programmed into the software, and a virtual guide ensured a standardized, uniform and objective collection of data in the testing situation and thereby ensured the validity evidence toward the response process. The participants had to go through all steps of the tutorial to become acquainted with the immersive virtual reality setting and the ultrasound examination setup. The participants were virtually placed in a hospital-like examination room with an ultrasound machine and a patient during both cases (Fig. 1).

Written instructions in front of the participant introduced the medical case history and when understood, the participant could move on to the thoracic ultrasound examination as illustrated in Figure 1. It was possible to scan four anterior zones, four lateral zones and six posterior zones at the virtual patient. A virtual board with eight optional diagnoses appeared when the participants pushed a button in the virtual room indicating the examination was finished. The optional diagnoses were normal examination (no pathology), pneumothorax dexter, pneumothorax sinister, simple pleural effusion dexter, simple pleural effusion sinister, complex pleural effusion dexter, complex pleural effusion sinister, pneumonia dexter, pneumonia sinister and interstitial syndrome.

The same setup was used for both cases and the program automatically moved on when answers were submitted. The software was programmed to register the choice of probe (linear, curved or phased array), pre-setting (abdominal, lung or heart) and scanning. Additionally, the participants were asked to register the diagnosis for each case.

A completed scan of each of the anterior and lateral zones equaled one point with a total of eight points per case. Additionally, each correct diagnosis equaled one point. The maximum point score in each case was 10 points. The data registrations and point scores were uploaded to a database provided by VitaSim.

Statistical analyses

All statistical analyses were carried out in SPSS statistics version 28 (IBM, Armonk, NY, USA).

Figure 1. Virtual reality. (A) Presentation of the ultrasound room in the immersive virtual reality world. The physician is scanning the patient with real ultrasound images on the virtual reality ultrasound machine; in this patient, there is a pleural effusion on the patient's right side. The board behind the ultrasound operator contains patient information with clinical history and vital parameters. (B) Following the ultrasound scan, the physician must push the button bottom of the diagnosis/diagnoses.

The descriptive statistics are expressed as means, standard deviations and ranges for some results.

The internal structure defined as the internal consistency reliability measured by Cronbach's α was used to measure the consistency of the test across test items. A Cronbach's α between 0.70 and 0.95 indicates high consistency across items. Item analysis was performed to explore the correlation and difficulty indices of the items. An item difficulty between 0.45 and 0.75 was preferred. An item difficulty between 0.25 and 0.44 was categorized as difficult, and an interval between 0.76 and 0.91 as easy. Item difficulties <0.24 and >0.91 were categorized as too difficult and too easy, respectively.

Pearson's r was used to determine the correlation between the two cases.

Independent *t*-tests were used to compare scores between the groups according to the relationship to other variables. The consequences of testing were explored using the contrasting groups' method to calculate a pass/fail standard [17]. A 5% significance level was considered significant.

Results

A total of 46 participants (22 women, 47.8%) including 21 nationalities took part in the immersive virtual reality test. Thirteen were categorized as novices, 22 as intermediates and 11 as experienced. Eighty-two cases (case 1: 46, case 2: 36) were completed during the test period. The baseline characteristics of the participants are seen in Table 1.

Validity evidence for internal structure

The internal consistency reliabilities for the outcomes are outlined in Table 2. Cronbach's α was 0.86 for the entire test. The item analysis revealed a low difficulty in choosing the right probe of 0.81 and a presetting of 1.00. None of the items was found too difficult; however, zones 2R (0.92), 3R (0.97) and 4R (0.95) in the first case were categorized as too easy. Zone 3R was also too easy to scan in case 2 (0.92).

The correlation between the two cases in the test had a Pearson r linear value of 0.50. This indicates a relatively low test−retest reliability between the two cases.

Validity evidence for the relations to other variables

The total mean scores in case 1 were 5.0 ± 2.7 points for the novices (range: 0−8), 7.3 ± 2.4 points for the intermediates (range: 6−9) and 8.7 ± 1.3 points for the experienced group (range: 6−10) (Fig. 2). The total mean scores in case 2 were 4.5 ± 1.8 points for the novices (range: 2−6), 6.7 ± 2.3 points for the intermediates (range: 1−10) and 8.5 ± 2.1 points for the experienced group (range: 5−10) (Fig. 3). Independent *t*-tests

Table 1
Participant demographics

Characteristic	Novices	Intermediates	Experienced
Total, n	13	22	11
Women, n (%)	9 (69%)	9 (41%)	4 (36%)
Mean (range) age, y	34 (22−60)	33 (26−48)	33 (23−42)
Level of education, n (%)			
Specialized physician	4 (31%)	5 (23%)	4 (36%)
Physicians under specialization	3 (23 %)	13 (59 %)	5 (45%)
Physicians not under specialization	3 (23 %)	4 (18%)	2 (19%)
Other	3 (23 %)		
Specialty, n (%)			
Respiratory medicine	7 (54%)	16 (72%)	10 (91%)
Other	6 (46%)	6 (28%)	1 (9%)
Country, n	Serbia 3	Spain 5	Denmark 3
	Ukraine 3	United Kingdom 3	United Kingdom 2
	Netherlands 2	Poland 2	Albania 1
	Austria 1	Portugal 2	Germany 1
	Denmark 1	Belgium 1	Italy 1
	Italy 1	Croatia 1	Singapore 1
	Saudi Arabia 1	Denmark 1	Spain 1
	Spain 1	France 1	Thailand 1
		Italy 1	
		Kenya 1	
		Mexico 1	
		Morocco 1	
		Saudi Arabia 1	
		Serbia 1	

Table 2
Item analysis

Case	Item	Item difficulty index	Item discrimination index
Intro	Probe	0.81	0.14
	Pre-set	1.00	0.00
Case 1	Zone 1R	0.81	0.55
	Zone 2R	0.89	0.56
	Zone 3R	0.86	0.21
	Zone 4R	0.92	0.09
	Zone 1L	0.38	0.61
	Zone 2L	0.97	0.64
	Zone 3L	0.35	0.71
	Zone 4L	0.95	0.66
	Diagnosis	0.68	0.15
Case 2	Zone 1R	0.70	0.62
	Zone 2R	0.86	0.51
	Zone 3R	0.89	0.21
	Zone 4R	0.92	0.40
	Zone 1L	0.38	0.48
	Zone 2L	0.92	0.45
	Zone 3L	0.38	0.74
	Zone 4L	0.89	0.74
	Diagnosis	0.32	0.43

revealed statistical significance between the novices and intermediates for case 1 ($p = 0.007$) and case 2 ($p = 0.02$). Significant differences were also found between the intermediates and the experienced group for case 1 ($p = 0.04$) and case 2 ($p = 0.019$) and for case 1 ($p < 0.001$) and between the novices and experienced group for case 2 ($p < 0.001$).

Validity evidence for consequences of testing

The contrasting groups' standard-setting method was used to establish a credible pass/fail score of the whole test combining both cases. The pass/fail scores were based on the mean performance of the novice and experienced groups (Figs. 4 and 5). Both cases needed to be passed to pass the test.

The pass/fail scores in both cases 1 and 2 were seven points each (Figs. 2 and 3). One participant from the experienced group failed both cases, and none of the novices passed both cases.

In case 1, three novices passed and one experienced clinician failed. In case 2 none of the novices passed but three experienced ultrasound operators failed the test. This resulted in a rate of 9% false negatives (failing experienced) in case 1 and 30% in case 2. The rates of false positives (passing novices) were 23% in case 1 and 0% in case 2.

The theoretical rates of false negatives were 8.5% in case 1 and 17.7% in case 2, while the theoretical rates of false positives were 24.7% in case 1 and 13.7% in case 2.

Discussion

We aimed to develop and gather solid validity evidence for an immersive virtual reality test of basic thoracic ultrasound competences. The test could significantly distinguish between self-reported novices and self-reported experienced thoracic ultrasound operators with few false negatives.

Previous studies in thoracic ultrasound education and competence assessment have explored various methods, including educational approaches, assessment sheets for clinical use [18,19], and simulation-based features [7,18,20,21]. To our knowledge, only one study has presented a head-mounted display immersive virtual reality setup as an opportunity for learning thoracic ultrasound [9]. The study found no significant difference between trainees who used virtual reality and those from a simulation-based study. Additionally, the study explored the educational impact of gamified versus non-gamified virtual reality training and found no difference between the two training approaches. Immersive virtual reality proved useful for practical training outside a clinical setting, and it can engage trainees for practical training, for example, during the COVID-19 lockdown periods or during other epidemics where trainees are restricted to limited clinical access.

It is, however, crucial to determine the optimal level of competence for virtual reality training and when to transition to other educational methods in the clinical setting for continuous competence development [22]. Assessing competence in virtual reality creates a unique opportunity to standardize and streamline the assessment and eliminate external influences, such as patient discomfort and interruptions. Another important factor in the assessment of competence is the examiner−examinee relationship, which can affect and bias the test [23]. Video-recorded assessments have been suggested for the complete objective evaluation, but technological setup and an examiner to watch the video and fill out assessment sheets are required. Therefore, the trainee must also wait for the results of the test, whereas in virtual reality, immediate feedback can be integrated and/or results of summative assessments can be provided.

Several disadvantages of head-mounted display virtual reality also exist. Haptic feedback and the dynamic movements of the transducer are currently not in the same way possible to simulate [9]. In the test's two cases, it is possible to move the transducer to one point on the

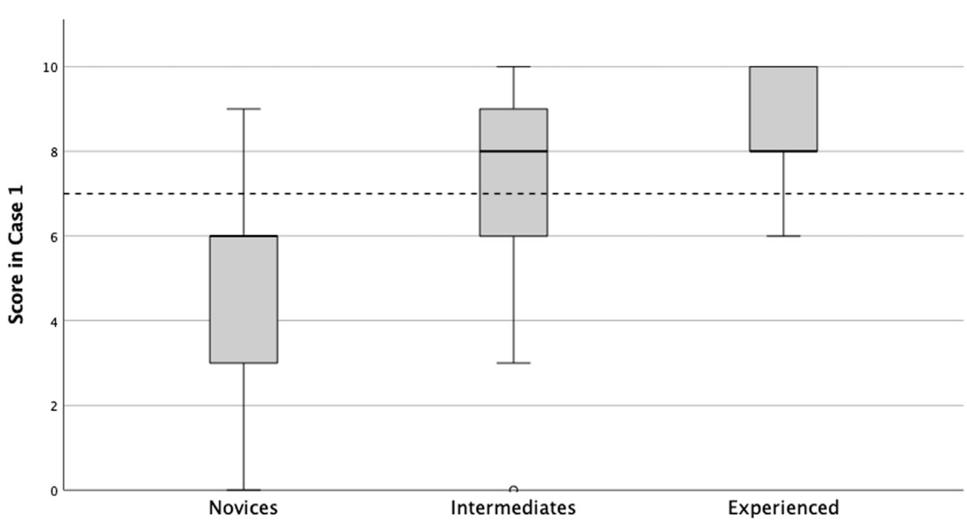

Figure 2. Case scores for case 1 divided by study group. The box-and-whisker plot illustrates the total point score for case 1 for the three study groups. The median, maximum and minimum scores are depicted. The *dashed horizontal line* represents the pass/fail score.

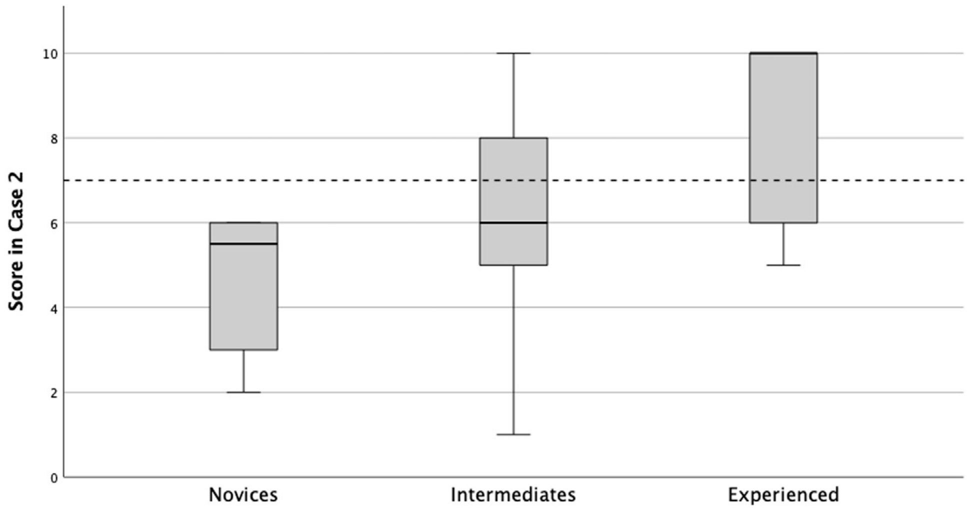

Figure 3. Case scores for case 2 divided by study group. The box-and-whisker plot illustrates the total point score for case 1 for the three study groups. The median, maximum and minimum scores are depicted. The *dashed horizontal line* represents the pass/fail score.

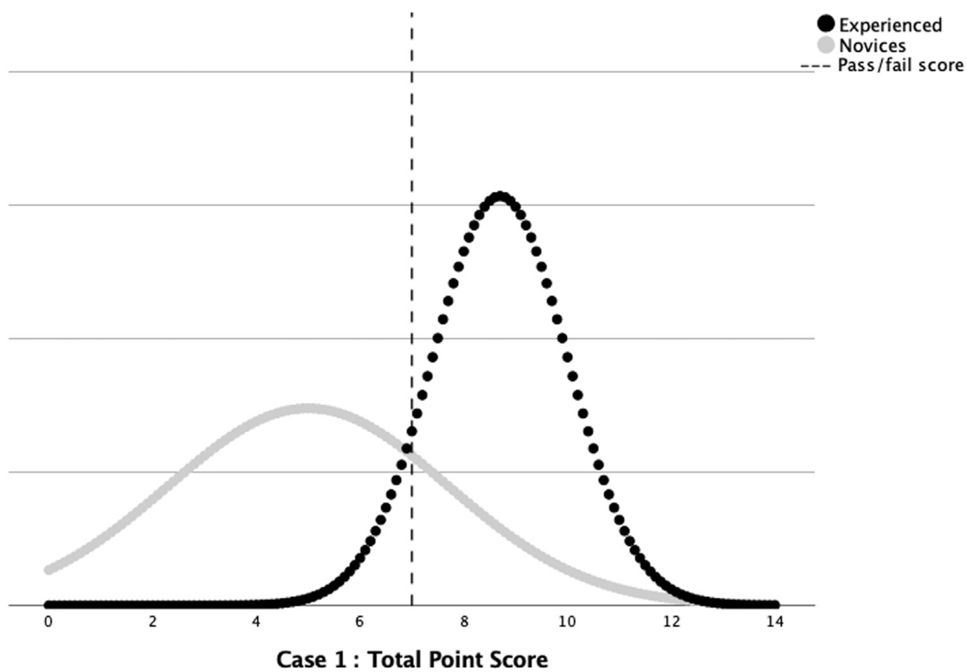

Figure 4. Pass/fail score using the contrasting groups' method. The pass/fail score using the contrasting groups' method for case 1 is represented by the *dashed vertical line*.

simulated patient's thorax, but the trainee is not able to slide the transducer back and forth or angle it as you would do when performing a real ultrasound examination or, for example, when using a phantom-based simulator like the US Mentor (Simbionix, 3D Systems, Circle Rock Hill, SC, USA) [20]. That said, no one has explored whether this possibility of dynamic movements and haptic feedback is important for the transfer of skills into a clinical setting and real ultrasound. Additionally, results from previous studies suggest that thoracic ultrasound is not a technically difficult ultrasound examination and does not require the same focus on angling in the proper direction as cardiac or abdominal ultrasound [7,24].

Virtual reality, alongside simulation-based training, cannot replace clinical education and supervision. Instead, it can enhance trainees' competence for their initial patient ultrasound examination. Converting theoretical and practical ultrasound training into virtual reality reduces instructional hours at the beginning of the learning curve, allowing physical instructors to focus on more complex aspects later on. Augmented feedback in simulation can guide trainees' probe position, while experienced supervisors can provide intricate feedback for advanced learned as trainees progress beyond the initial stages [25].

We acknowledge that the study suffers from several limitations. First is the creation of the study groups on the basis of arbitrary measures. We relied on previously published papers on education and training studies in lung ultrasound and reshaped the groups to fit the VR and international setting [20,26]. The groups were changed so that novices were not allowed to have performed any unsupervised ultrasound examinations, whereas in the studies in the literature they were allowed to have performed up to 20 examinations—supervised or unsupervised. This change could potentially lead to a lower mean score and standard deviation and thereby a lower pass/fail score. VR cannot replace all parts of a training program and does not entail experienced operators, but it increases the level of competence by the first patient-performed ultrasound examination, for which supervision is still needed. Therefore, we do not believe that the redefinition of the study groups would compromise patient safety or the validity of the test. Second, the demographics and data on previous experience in thoracic ultrasound were obtained from a written questionnaire completed by the

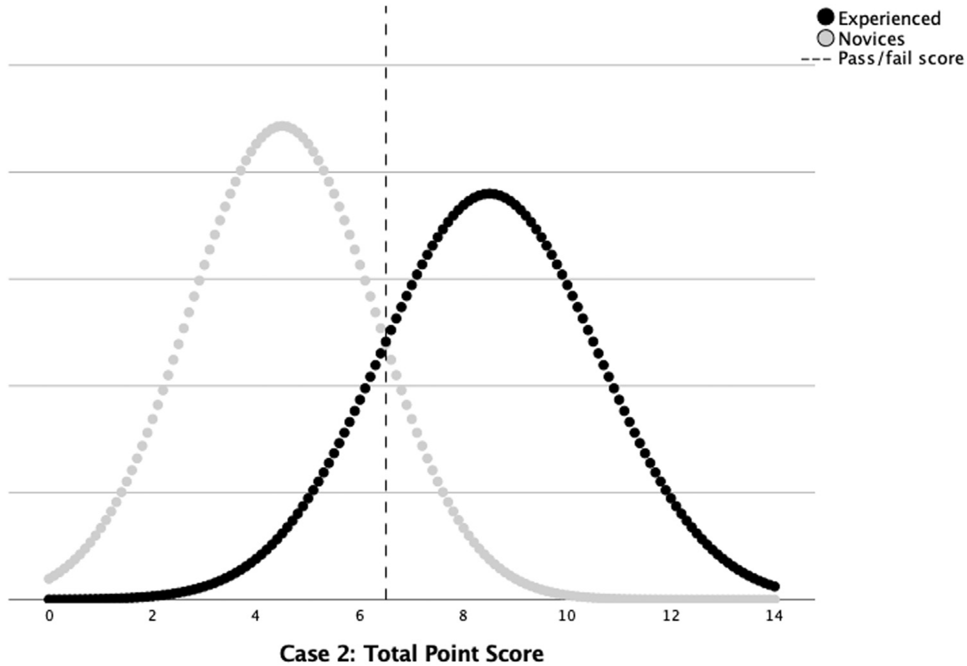

Figure 5. Pass/fail score using the contrasting groups' method. The pass/fail score using the contrasting groups' method for case 2 is represented by the *dashed vertical line*.

participants. The use and protocols of thoracic ultrasound vary worldwide; in some institutions, thoracic ultrasound includes only looking for pleural effusion, and in other institutions, it contains looking for signs of pneumothorax, pulmonary edema (interstitial syndrome) and pneumonia, among other things. Therefore, the participants could have stated that they were familiar with and experienced in thoracic ultrasound but only in parts of the content, which might explain the failing participants in the self-reported experienced group. On the other hand, compared with many other studies gathering validity evidence of tests in medical education, the participants are multinational and not from a specifically selected group, which increases the generalizability of our results.

Future perspectives in using immersive virtual reality for training thoracic ultrasound could comprise creation of a broad variety of educational tools that are structured and evidence based. Depending on the setting, the institution or participant could then pick out the most convenient and efficient educational tool. The needs and setup for educational methods most likely differ between institutions with access to a simulation center, institutions with many thoracic ultrasound operators and low-income country institutions with a limited budget and no access to face-to-face sparring and supervision.

With this study, there has been added a new tool for the training and assessment of thoracic ultrasound competencies, which could be used in several settings and educational programs.

Conclusion

We developed a simulation-based test of thoracic ultrasound skills in immersive virtual reality. We gathered solid validity evidence, determined that the test can discriminate between experienced and inexperienced ultrasound operators and established a credible pass/fail score. This standardized portable test could be a prerequisite in a structured simulation-based mastery learning program.

Conflict of interest

J.D. is employed at VitaSim. C.B.L. received speakers' honoraria for lectures at educational events/symposia/courses organized by AstraZeneca. P.I.P. received a sponsorship from Boehringer Ingelheim to participate in a thoracic imaging course (ESTI winter course, 2022). A.B.N., N. J., A.F., M.S. and L.K. declare no competing interests.

Acknowledgments

The authors thank SimC, the Simulation Centre at Odense University Hospital, for the loan of the virtual reality equipment.

Data availability statement

Data are available on request to the corresponding author.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ultrasmedbio.2023.12.002.

References

[1] Maw AM, Hassanin A, Ho PM, McInnes DF, Moss A, Juarez-Colunga E, et al. Diagnostic accuracy of point-of-care lung ultrasonography and chest radiography in adults with symptoms suggestive of acute decompensated heart failure: a systematic review and meta-analysis. JAMA Netw Open 2019;2:e190703.
[2] Orso D, Guglielmo N, Copetti R. Lung ultrasound in diagnosing pneumonia in the emergency department: a systematic review and meta-analysis. Eur J Emerg Med 2018;25:312–21.
[3] Staub LJ, Biscaro RRM, Kaszubowski E, Maurici R. Chest ultrasonography for the emergency diagnosis of traumatic pneumothorax and haemothorax: a systematic review and meta-analysis. Injury 2018;49:457–66.
[4] Riishede M, Lassen AT, Baatrup G, Pietersen PI, Jacobsen N, Jeschke KN, et al. Point-of-care ultrasound of the heart and lungs in patients with respiratory failure: a pragmatic randomized controlled multicenter trial. Scand J Trauma Resusc Emerg Med 2021;29:60.
[5] Araiza-Garaygordobil D, Gopar-Nieto R, Martinez-Amezcua P, Cabello-López A, Alanis-Estrada G, Luna-Herbert A, et al. A randomized controlled trial of lung ultrasound-guided therapy in heart failure (CLUSTER-HF study). Am Heart J 2020;227:31–9.
[6] McGaghie WC. Mastery learning: it is time for medical education to join the 21st century. Acad Med 2015;90:1438–41.
[7] Pietersen PI, Jorgensen R, Graumann O, Konge L, Skaarup SH, Lawaetz Schultz HH, et al. Training thoracic ultrasound skills: a randomized controlled trial of simulation-based training versus training on healthy volunteers. Respiration 2021;100:34–43.
[8] Østergaard ML, Konge L, Kahr N, Albrecht-Beste E, Bachmann Nielsen M, Nielsen KR. Four virtual-reality simulators for diagnostic abdominal ultrasound training in radiology. Diagnostics (Basel) 2019;9:50.
[9] Larsen JD, Jensen RO, Pietersen PI, Jacobsen N, Falster C, Nielsen AB, et al. Education in focused lung ultrasound using gamified immersive virtual reality: a randomized controlled study. Ultrasound Med Biol 2023;49:P841–52.

[10] Hu KC, Salcedo D, Kang YN, Lin CW, Hsu CW, Cheng CY, et al. Impact of virtual reality anatomy training on ultrasound competency development: a randomized controlled trial. PLoS One 2020;15:e0242731.

[11] Turso-Finnich T, Jensen RO, Jensen LX, Konge L, Thinggard E. Virtual reality head-mounted displays in medical education—a systematic review. Simul Healthc 2023;18:42–50.

[12] McGaghie WC, Barsuk JH, Wayne DB, editors. Comprehensive healthcare simulation: mastery learning in health professions education. Cham: Springer; 2020.

[13] American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME). Standards for educational and psychological testing. Washington, DC: AERA; 2014.

[14] Laursen CB, Clive A, Hallifax R, Iben P, Asciak R, Rømhild Davidsen J, et al. European Respiratory Society statement on thoracic ultrasound. Eur Respir J 2021;57(3).

[15] Laursen CB, Rahman NM, Volpicelli G. Thoracic ultrasound. European Respiratory Society Monograph 2018. doi: 10.1183/2312508X.erm7918.

[16] Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. Med Teach 2012;34:960–92.

[17] Jorgensen M, Konge L, Subhi Y. Contrasting groups' standard setting for consequences analysis in validity studies: reporting considerations. Adv Simul (Lond) 2018;3:5.

[18] Skaarup SH, Laursen CB, Bjerrum AS, Hilberg O. Objective and structured assessment of lung ultrasound competence: a multispecialty Delphi consensus and construct validity study. Ann Am Thorac Soc 2017;14:555–60.

[19] Millington SJ, Arntfield RT, Guo RJ, Koenig S, Kory P, Noble V, et al. The Assessment of Competency in Thoracic Sonography (ACTS) scale: validation of a tool for point-of-care ultrasound. Crit Ultrasound J 2017;9:25.

[20] Pietersen PI, Konge L, Graumann O, Nielsen BU, Laurse CB. Developing and gathering validity evidence for a simulation-based test of competencies in lung ultrasound. Respiration 2019;97:329–36.

[21] Pietersen PI, Konge L, Madsen KR, Bendixen M, Maskell NA, Rahman N, et al. Development of and gathering validity evidence for a theoretical test in thoracic ultrasound. Respiration 2019;98:221–9.

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

[23] Yudkowsky R, Downing SM. Assessment in health professions education. 2nd ed. New York: Routledge; 2020.

[24] Russell FM, Ferre R, Ehrman RR, Nobel V, Gargani L, Collins SP, et al. What are the minimum requirements to establish proficiency in lung ultrasound training for quantifying B-lines? ESC Heart Fail 2020;7:2941–7.

[25] Hatala R, Cook DA, Zendejas B, Hamstra SJ, Brydges R. Feedback for simulation-based procedural skills training: a meta-analysis and critical narrative synthesis. Adv Health Sci Educ 2014;19:251–72.

[26] Pietersen PI, Bhatnagar R, Andreasen F, Konge L, Borbjerg Laursen C, Rahman N, et al. Objective structured clinical examination in basic thoracic ultrasound: a European study of validity evidence. BMC Pulm Med 2023;23:15.