Ultrasound in Med. & Biol., Vol. 48, No. 2, pp. 248–256, 2022

0301-5629/$ - see front matter

https://doi.org/10.1016/j.ultrasmedbio.2021.10.016

● *Original Contribution*

DEVELOPMENT OF AND GATHERING VALIDITY EVIDENCE FOR A THEORETICAL TEST IN CONTRAST-ENHANCED ULTRASOUND

Niels Jacobsen,[*,†,‡] Christian P. Nolsøe,[§,¶] Lars Konge,[¶] Ole Graumann,[‖,#] Christoph F. Dietrich,[**] Paul S. Sidhu,[††,‡‡] Odd H. Gilja,[§§,¶¶] Maria F. Meloni,[‖‖,##] Annalisa Berzigotti,[***] Chris J. Harvey,[†††] Annamaria Deganello,[††,‡‡] Francesco Prada,[‡‡‡,§§§,¶¶¶,‖‖‖] Markus H. Lerchbaumer,[###,***] and Christian B. Laursen[*,†,‡]

* Department of Respiratory Medicine, Odense University Hospital, Odense, Denmark; † Regional Center for Technical Simulation (TechSim), Odense University Hospital, Odense, Denmark; ‡ Odense Respiratory Research Unit (ODIN), Department of Clinical Research, University of Southern Denmark, Odense, Demark; § Center for Surgical Ultrasound, Department of Surgery, Zealand University Hospital, Køge, Denmark; ¶ Copenhagen Academy for Medical Education and Simulation (CAMES), Center for Human Resources and Education, The Capital Region of Denmark, Copenhagen, Denmark; ‖ Department of Radiology, Odense University Hospital, Odense, Denmark; # Research and Innovation Unit of Radiology, University of Southern Denmark, Odense, Denmark; ** Department of Internal Medicine, Hirslanden Clinic (Beau-Site, Salem-Spital, and Permanence), Bern, Switzerland; †† Department of Radiology, King's College Hospital, Denmark Hill, London, United Kingdom; ‡‡ School of Biomedical Engineering & Imaging Sciences, King's College London, United Kingdom; §§ National Centre for Ultrasound in Gastroenterology, Haukeland University Hospital, Bergen, Norway; ¶¶ Department of Clinical Medicine, University of Bergen, Bergen, Norway; ‖‖ Department of Interventional Ultrasound, IGEA S.p.A. Multispecialty Medical Clinic, Milan, Italy; ## Department of Radiology, University of Wisconsin, Madison, Wisconsin, USA; *** Department of Hepatology, University Clinic for Visceral Surgery and Medicine, University Hospital of Bern, University of Bern, Bern, Switzerland; ††† Department of Imaging, Imperial College NHS Healthcare Trust, Hammersmith Hospital, London, United Kingdom; ‡‡‡ Neurosurgery Unit, Department of Neuroscience, Alessandro Manzoni Hospital, Lecco, Italy; §§§ Acoustic Neuroimaging and Therapy Lab, Foundation IRCCS Carlo Besta Neurological Institute, Milan, Italy; ¶¶¶ Department of Neurological Surgery, University of Virginia Health Science Center, Charlottesville, Virginia, USA; ‖‖‖ Focused Ultrasound Foundation, Charlottesville, Virginia, USA; ### Charité University Hospital Berlin, Humboldt University of Berlin, Berlin, Germany; and *** Department of Radiology, Berlin Institute of Health, Berlin, Germany

(*Received* 6 *July* 2021; *revised* 18 *October* 2021; *in final form* 19 *October* 2021)

Abstract—Contrast-enhanced ultrasound (CEUS) is an imaging modality applied in a broad field of medical specialties for diagnostic uses, guidance during biopsy procedures and ablation therapies and sonoporation therapy. Appropriate training and assessment of theoretical and practical competencies are recommended before practicing CEUS, but no validated assessment tools exist. This study was aimed at developing a theoretical multiple-choice question-based test for core CEUS competencies and gathering validity evidence for the test. An expert team developed the test via a Delphi process. The test was administered to medical doctors with varying CEUS experience, and the results were used to evaluate test items, internal-consistency reliability, ability to distinguish between different proficiency levels and to establish a pass/fail score. Validity evidence was gathered according to Messick's framework. The final test with 47 test items could distinguish between operators with and without CEUS experience with acceptable reliability. The pass/fail score led to considerable risk of false positives and negatives. The test may be used as an entry test before learning practical CEUS competencies but is not recommended for certification purposes because of the risk of false positives and negatives. (E-mail: niels.jacobsen@rsyd.dk)

Key Words: **Contrast-enhanced ultrasound, Ultrasound contrast agent, Medical education, Theoretical test, Validity evidence, Multiple-choice questions.**

INTRODUCTION

Contrast-enhanced ultrasound

Contrast-enhanced ultrasound (CEUS) is an imaging modality in which an ultrasound contrast agent (UCA) is administered during the ultrasound (US) examination.

Address correspondence to: Niels Jacobsen, Kløvervænget 24A, 1.12, DK-5000 Odense C, Denmark. E-mail: niels.jacobsen@rsyd.dk

Currently used UCAs consist of micrometer-sized gas bubbles encapsulated in a stabilizing shell, typically made of phospholipids (Chong et al. 2018). The intravenously injected gas bubbles create a strong echo signal when exposed to low-power US beams. This signal can provide real-time information on tissue perfusion timing and pattern with high temporal and spatial resolution (Wilson et al. 2020). Additionally, UCAs are non-toxic and therefore safe to use (Piscaglia and Bolondi 2006).

Current uses and observer's competencies

CEUS is currently being applied in an expanding field of medical specialties as a diagnostic tool, for guidance during biopsy procedures, in the management of patients with liver and kidney tumors treated with ablation therapies and for sonoporation therapy (Dimcevski et al. 2016; Nolsoe et al. 2018; Sidhu et al. 2018; Dietrich et al. 2020). As with conventional US, proper, safe and effective use is highly operator dependent, and appropriate education and assurance of the competencies of CEUS operators are imperative (Quaia et al. 2010).

International US societies provide guidelines and recommendations for the clinical practice of hepatic and non-hepatic CEUS, but recommendations regarding training and competence assessment are less well established (European Federation of Societies for Ultrasound in Medicine and Biology [EFSUMB] 2010; Sidhu et al. 2018; Dietrich et al. 2020).

A useful model of the taxonomy of competencies is George Miller's pyramid (Miller 1990). This model describes four different layers of competencies and how these may be assessed. The levels dealing with cognitive (theoretical) knowledge are prerequisites for learning practical skills, and can be assessed adequately by written tests, such as multiple-choice question (MCQ) tests (Downing and Yudkowsky 2009).

Before any competence assessment test is implemented in an educational program, the test should be supported by sufficient validity evidence (Cook and Hatala 2016). Without this, it is not possible for educators to make qualified judgments of the trustworthiness of the test, thus risking misjudgment of learners, which ultimately may jeopardize patient safety. To our knowledge, no such standardized theoretical test for CEUS core competencies currently exists.

The aim of this study was to develop a theoretical MCQ-based test for core competencies in CEUS and gather validity evidence for the test.

METHODS

Setting and study design

This study was conducted at Odense University Hospital, Odense, Denmark, from July 2020 to April 2021 and comprised two consecutive phases: (1) development of a theoretical MCQ test in core CEUS competencies, (2) gathering validity evidence for the test.

Phase 1: Development of the test

For development of the initial test draft, we assembled a team of three clinical experts in CEUS (C.P.N., O. G., C.B.L.) (radiology [n = 2] and pulmonology [n = 1]), one professor in medical education (L.K.) and one junior resident (N.J.). Expertise in CEUS was defined as consultants with routine CEUS practice, frequent supervision of colleagues in CEUS procedures or substantial scientific research in CEUS, which are desired traits for the highest level of practice (EFSUMB 2010)

A previous published syllabus of core CEUS competencies was used as a framework (Jacobsen et al. 2020). MCQs (items) were developed for each section from the theoretical domain, ensuring that all relevant content from the syllabus was initially covered. Standard guidelines for writing MCQs were followed (Haladyna et al. 2002; Downing and Yudkowsky 2009) and were designed as one stem with three possible answers: one single best answer and two distractors (Fig. 1).

Phase 2: Gathering validity evidence for the test

Validity evidence was collected in accordance with Messick's framework, which comprises five sources of validity evidence: content, response process, internal structure, relationship to other variables and consequences (Downing and Yudkowsky 2009). Each source is described below, as is the process of recruiting test participants.

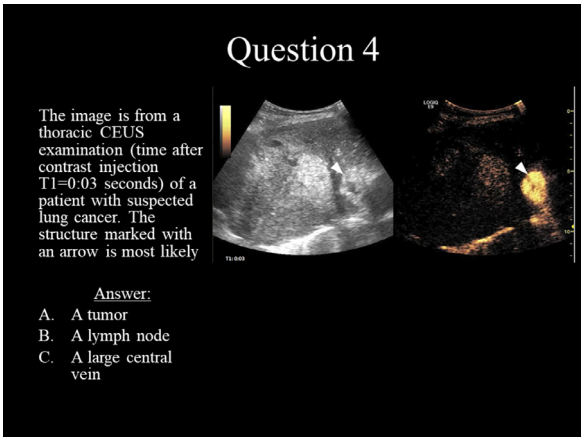

Fig. 1. Example of a final multiple-choice question. Designed with one single best answer (C) and two distractors. The learner must know the distribution phases of intravenously injected ultrasound contrast agents to correctly answer this question. Item difficulty and point biserial discriminatory index were 0.68 and 0.29, respectively. CEUS = contrast-enhanced ultrasound.

Recruitment of test participants

We recruited medical doctors from different countries and settings, with varying US and CEUS experience as test participants. They were categorized in three study groups based on their level of CEUS experience: "no CEUS experience," "limited CEUS experience" (limited exposure to CEUS examinations but no unsupervised CEUS examinations or participation in CEUS courses), and "experienced CEUS operators" (routine practice of independent CEUS examinations, frequent supervision of colleagues in clinical CEUS examinations, scientific research in CEUS or participation in CEUS courses).

We aimed to recruit more than 10 participants from each study group, as this sample size can be considered as normally distributed in medical educational research (Bloch and Norman 2012).

Content (relevance of the test content to the previously established syllabus)

Three clinical CEUS experts (C.F.D., P.S.S., O.H.G.) (hepato-gastroenterology [n = 2] and radiology [n = 1]) evaluated the initial test draft for consensus of selected items in a Delphi process. This method pools anonymous expert opinions through iterative rounds until a predefined level of consensus is achieved (De Villiers et al. 2005). The experts were asked to rate the initial items from 1 to 5 based on relevance (1 being completely irrelevant and 5 being very relevant); suggest new items; and assess whether the content for each section was sufficiently covered.

Items with a mean score ≤ 4 were then omitted, and the entire test developing team (N.J., C.P.N., L.K., O.G., C.F.D., P.S.S., O.H.G., C.B.L.,) was asked to reassess the remaining items for language, grammar and comprehensibility. These items were subsequently adjusted based on the experts' inputs in a Delphi process until final agreement for each item was achieved. This was defined as agreement by a minimum seven of the eight team members.

Response process (minimization of bias and standardization of the test process)

The selected items were uploaded to a free Web-based survey administration software (Google Forms) in randomized order, and corresponding item videos were uploaded to a free online video-sharing portal (YouTube). Access to the test was restricted to test participants only and was available via a link sent on acceptance of participation.

Baseline characteristics and information on competence level (reported as European Federation of Societies for Ultrasound in Medicine and Biology [EFSUMB] level of practice [EFSUMB 2006] and CEUS experience) were obtained prior to the test. The participants were informed to complete the test individually, in one take, with no support (*e.g.,* handbooks, web resources) and with no time limit. After the test, the participants could comment on the test for feedback purposes.

Internal structure (analysis of the quality of the developed test items and subsequent assessment of test reliability)

We performed an item analysis of item difficulty (proportion of test takers answering the item correctly) and point biserial item discrimination index (correlation between test takers' performance on an item and performance on the entire test) (Downing and Yudkowsky 2009). All items with a point biserial discrimination index >0.1 were included regardless of difficulty. Included items were used to measure internal consistency reliability via Cronbach's α. Additionally, we used generalizability theory and a decision study to explore the effect on internal consistency by altering the total number of test items.

Relationship to other variables (correlation of test scores to external variables such as level of CEUS experience)

No test for assessment of core CEUS knowledge currently exists to our knowledge.

We therefore assumed that different levels of experience with US and CEUS could be used as a proxy for core CEUS knowledge, because of its relative correlation with exposure to the procedure. Test variances between all study groups were compared using *post hoc* analyses of variance (ANOVAs) with subsequent independent sample tests (Levene's test for equality of variances and *t*-test for equality of means).

Consequences (outcome of establishing a pass/fail score)

We used the contrasting groups' standard setting method on the test score distribution of the lowest and highest proficiency groups to establish a pass/fail score (Jørgensen et al. 2018). We then explored the outcome of the pass/fail score and its consequences by calculating theoretical false positives (*i.e.,* medical doctors with no CEUS experience passing the test) and false negatives (*i.e.,* medical doctors with CEUS experience failing the test) (Jørgensen et al. 2018), using Fisher's exact test to test for statistical significance.

Statistical analyses

All statistical analyses were calculated using SPSS Version 25 (IBM, Armonk, NY, USA) and are described when relevant in each phase of the study. All statistics were considered significant at a 5% significance level.

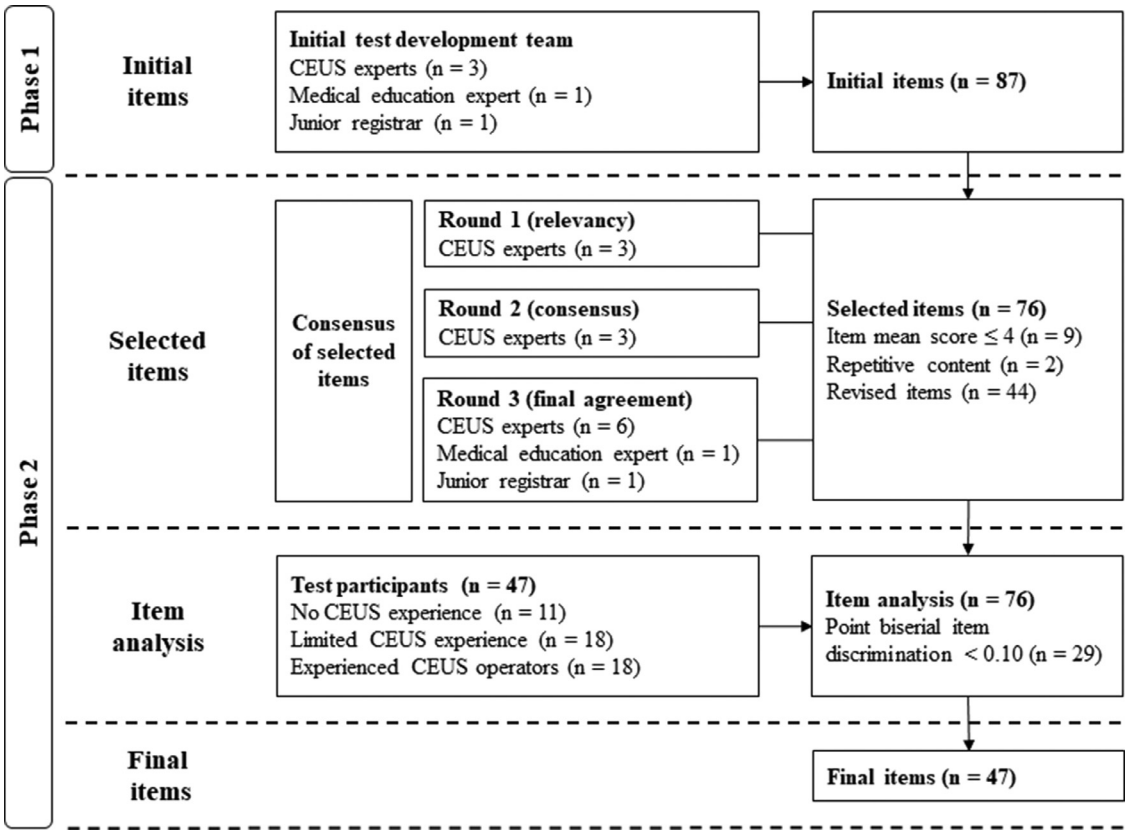

Fig. 2. Flowchart of test development. Initial items were developed in phase 1. Final items succeeded phase 2 after Delphi iterations and item analysis. CEUS = contrast-enhanced ultrasound.

RESULTS

Phase 1

A total of 87 MCQs were constructed for the initial draft. Figure 2 illustrates the development of the test.

Phase 2

Content. This phase comprised three Delphi rounds in which 76 items proceeded to the final test round. Forty-four items (57.9%) were improved in terms of grammar, clarity and minor corrections during the final Delphi round. No new items were suggested during the Delphi rounds (Fig. 2).

Response process. Baseline characteristics of the participants are summarized in Table 1. One participant completed the test twice, but only the first attempt was included for analysis.

Internal structure. Twenty-nine (38.2%) items had a point biserial discrimination index <0.1 and were removed. The average item difficulty and discriminatory indices of the final items (n = 47) were 0.8 (range: 0.36–0.98) and 0.27 (range: 0.11–0.53), respectively. The internal consistency reliability, measured as Cronbach's α, was 0.81 for the final items. The generalizability coefficients of 0.8 and 0.7 intersected at 45 and 26 total number of test items, respectively (Fig. 3).

Relationship to other variables. The total test scores of the groups are illustrated as boxplots in Figure 4.

Post hoc analysis of variance proved statistically significant differences between the groups ($p = 0.008$). Subsequent independent sample tests proved significant differences between the group with no CEUS experience and operators with CEUS experience ($p = 0.031$), and between the group with limited CEUS experience and operators with CEUS experience ($p = 0.007$), but not between the group with no CEUS experience and the group with limited CEUS experience ($p = 0.30$).

Consequences. The normal distribution curves of the mean test scores of the group with no CEUS experience and the operators with CEUS experience intersected at 36 points (77% correct), which is the pass/fail point according to the contrasting groups'

Table 1. Baseline characteristics of test participants and mean test scores of the groups*

	CEUS experience			
	None	Limited	Experienced	Total
No. of participants	11 (23)	18 (38)	18 (38)	47 (100)
Age, y	30.9 [27−38]	34.7 [30−54]	46.2 [30−72]	37.2 [27−72]
Male gender	6 (55)	10 (56) *	13 (72)	29 (62) *
US experience, n (%)				
EFSUMB level I	10 (91)	0	2 (11)	12 (26)
EFSUMB level II	1 (9)	15 (83)	2 (11)	18 (38)
EFSUMB level III	0	3 (17)	14 (78)	17 (36)
Specialty				
Radiology	5 (46)	10 (56)	11 (61)	26 (55)
Hepato-gastroenterology	2 (18)	2 (11)	4 (22)	8 (17)
Neurosurgery	1 (9)	2 (11)	1 (6)	4 (9)
Internal medicine	1 (9)	1 (6)	1 (6)	3 (6)
Pediatric radiology	0	0	1 (6)	1 (2)
Pulmonology	0	1 (6)	0	1 (2)
Nephrology/ICU	0	1 (6)	0	1 (2)
Otorhinolaryngology	1 (9)	0	0	1 (2)
General medicine	1 (9)	0	0	1 (2)
General surgery	0	1 (6)	0	1 (2)
Country				
United Kingdom	3 (27)	5 (28)	3 (17)	11 (23)
Italy	1 (9)	4 (22)	5 (28)	10 (21)
Germany	2 (18)	4 (22)	3 (17)	9 (19)
Switzerland	2 (18)	2 (11)	3 (17)	7 (15)
Denmark	0	1 (6)	1 (6)	2 (4)
Sweden	1 (9)	0	1 (6)	2 (4)
Belgium	0	2 (11)	0	2 (4)
France	1	0	0	1 (2)
Norway	0	0	1 (6)	1 (2)
United States	0	0	1 (6)	1 (2)
Not stated	1 (9)	0	0	1 (2)
Test score	34.7 ± 6.96 (74)	36.4 ± 5.01 (77)	40.6 ± 3.35 (86)	

CEUS = contrast-enhanced ultrasound; EFSUMB = European Federation of Societies for Ultrasound in Medicine and Biology; ICU = intensive care unit; US = ultrasound.

Values are expressed as the number (%), mean (range), mean ± standard deviation or mean ± standard deviation (% correct on average).

* One participant did not wish to specify gender.

method (Fig. 5). This pass/fail score yielded theoretical false-positive and false-negative rates of 40.4% and 10.9%, respectively, but these were statistically not significant (p = 0.107). The calculated false positives and false negatives from our sample was 45.5% and 11.1% in comparison to the theoretical estimates.

The final items, links to the uploaded cine clips and the answer key are available in the Supplementary Data.

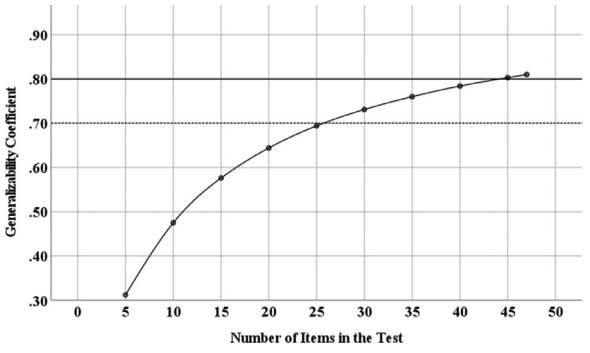

Fig. 3. Decision and generalizability study. Results from the decision study revealing the correlation of generalizability coefficient with an increasing number of test items. Generalizability coefficients of 0.7 and 0.8 intersected at 26 and 45 total test items, respectively.

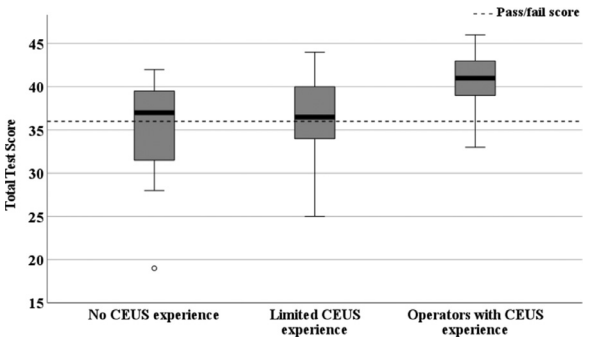

Fig. 4. Test scores. Boxplots with test scores for each study group as an independent variable, contrast-enhanced ultrasound (CEUS) experience as a dependent variable and the calculated pass/fail score as a *dotted line*.

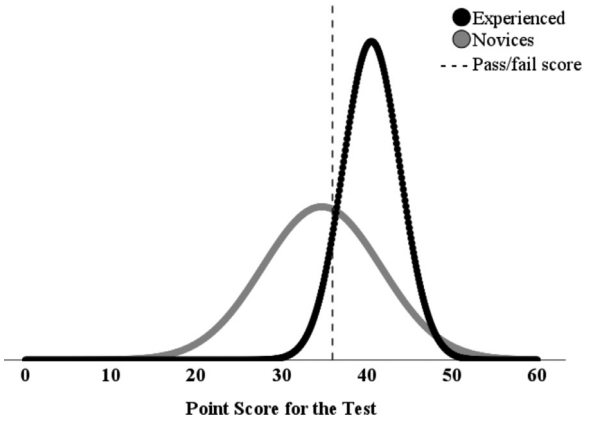

Fig. 5. The contrasting group's method. The normal distribution curves for mean test score of the group with no CEUS experience (*gray curve*) and operators with CEUS experience (*black curve*) are used to identify their intersection point (*dotted vertical line*) which is the pass/fail score (36 points). CEUS = contrast-enhanced ultrasound.

DISCUSSION

We developed a theoretical multiple-choice question-based test in core CEUS competencies and gathered multiple sources of validity evidence in accordance with Messick's framework. In 1999, this framework of validity evidence was accepted as standard in the field of medical education; as such, our study provides a theoretical test developed in accordance with best practice in educational research (American Educational Research Association et al. 1999).

Core competencies and competence assessment

Our test could distinguish experienced CEUS operators from operators with limited or no experience, but all groups had a high mean score, indicating that the difficulty of the test is low. This was somewhat anticipated because of the specific focus on assessment of core curriculum competencies rather than specialized and advanced content. Practical CEUS and interpretation are different for each organ, and learning objectives for CEUS operators are therefore not uniform for all medical specialties. It would be less feasible and with limited clinical relevance to require CEUS operators to master learning objectives for all CEUS applications within every specialty. More general CEUS features (*e.g.,* preparation and administration of UCAs) constitute core competencies and are essential to learn for all operators in the same manner as sterile principles are essential for all physicians, regardless of their subspecialty.

Assessing clinical competencies should not be limited to assessment of cognitive knowledge; assessment of performance skills is also required for a complete assessment of clinical competencies (Downing and Yudkowsky 2009).

This theoretical test is therefore not intended to stand alone but should be accompanied by a practical competence assessment test or serve as an entry before such test. Currently, no practical test exists, and until this has been developed and validated, our theoretical test can be used as an entry test prior to a practical CEUS skills course.

Internal reliability

The internal reliability of this test was sufficient for moderate stakes tests with a Cronbach α of 0.81, as recommended (Downing and Yudkowsky 2009). If the objective of the test is to provide feedback and serve as an entry before a practical test, a lower internal reliability may be acceptable ($\geq$0.7). We used generalizability theory to characterize the reliability of the scores from the pilot test, followed by a decision study to provide reliability estimates for different total numbers of test items. Twenty-six test items were estimated to achieve a generalizability coefficient of 0.7. Choosing this cutoff would allow for a shorter test, and the excess questions could be used for item rotation via an item bank to ensure test variety over time.

Content

We chose a MCQ-based test, as this format has proven useful for testing cognitive knowledge. It can broadly test large knowledge domains, can be scored truly objectively and is time efficient compared with other written test formats (Downing and Yudkowsky 2009). One concern may be that the MCQs excessively test trivial content (*i.e.,* recall knowledge) but this is hardly preventable when testing core content. A minimum of 35−40 items in total is advised for three-option MCQs, which we have accomplished with our test (Downing and Yudkowsky 2009).

A previously published international syllabus of core CEUS competences was used as a framework for the initial development of test items (Jacobsen et al. 2020). During the consensus phase for selected items, a Delphi method was used that reduces risk of cognitive biases by anonymizing response and feedback by individual experts. The developing group constituted both multispecialty and international experts in both CEUS and medical education. Consequentially, we consider the development phase of this theoretical test methodologically strong.

Only few items (11/87, 13%) were removed during the Delphi iterations, indicating an overall good agreement of content coverage of the initial test items. Additionally, the expert team suggested no new items, and the average score for relevancy was very high (4.6/5,

90%). However, more than half of the selected items (44/76, 58%) needed improvements before final agreement was achieved, indicating some discrepancy between experts, even for essential content.

Response process

For test administration, we chose a free online platform that allowed upload of pictures and clips and was simple to use and distribute. Because of the online administration, it was not possible to control if test participants used any supports for the test, which is a source for potential bias. Minimizing this type of bias would require more controlled administration of the test (*e.g.,* physically administration of the test in a classroom setting), but this would drastically reduce the feasibility of testing larger numbers of participants. Additionally, a strength in our study was the sampling of international test participants from countries with different US education systems based on teaching by local supervisors, which would have been impractical to achieve by any offline approach especially in the era of the Covid-19 pandemic.

Not all relevant medical specialties were represented by test participants (*e.g.,* pediatrics, cardiology etc.). While this was unintentional and may compromise the generalizability of the test for these specialties, we deem it less concerning because of the test's intention of assessing core competencies.

Internal structure

The distribution of item classes for the final test items was not optimal. Ideally, most items should be level I items (middle range of average item difficulty and with a high discriminatory index) followed by level II items (easier items with moderate to high discriminatory index) and level III items (difficult items with a moderate discriminatory index) (Downing and Yudkowsky 2009). In our study, only 38% of the items were level I−III items. However, the item classification system is considered more a guideline than a strict rule, and less optimal item classes can be considered if covering important content (Downing and Yudkowsky 2009). For the two major item quality parameters, the discriminatory index is considered more important than item difficulty (Downing and Yudkowsky 2009). We chose to include all items with acceptable discriminatory indices ($\geq$0.1) regardless of item difficulty. This yielded an average discriminatory index of 0.27, which is recommended for level I items, and an average item difficulty of 0.80, which is in range of level II items. The result is a test with easy test items on average, but with enhanced content-related validity, while preserving acceptable item discriminatory abilities.

Relationship to other variables

All test groups generally performed well in the test with a score for all groups of 37.2 points (79% correct) on average. A study group of CEUS naïve participants is not expected to perform that well in a test. Possible explanations are that the test content is not only testing CEUS knowledge, the test content is very easy, test participants are not truly CEUS naïve or any combination of these.

When assessing core and not specialized knowledge, a test will naturally cover more basic content, which may be simpler and thus easier to answer. Knowledge of US is a prerequisite for learning CEUS, and thus covering core CEUS content will inevitably also cover US content, which may be answered correctly even by CEUS-naïve participants.

The information on test participants' US and CEUS experience was self-reported during the online assessment, without interference by study facilitators. The risk of self-reporting bias is therefore present. In the online assessment form, brief written guidance on how to report on US and CEUS experience was provided, but test participants may have interpreted it differently.

In our study, all test participants had a relatively high level of US experience (equivalent to EFSUMB level I or higher), and thus our lowest proficiency group could be classified as "CEUS-naïve US intermediates." Several similar studies of theoretical test development within medical education compare test performance of medical students and specialists, which are the two extremes of proficiency level (Jørgensen et al. 2019; Pietersen et al. 2019). Naturally, when comparing very high and very low proficiency groups, the likelihood of observing a significant difference in test scores is higher, and the probability of false negatives and positives when establishing a pass/fail score is lower.

Ultrasound competencies equivalent to EFSUMB level II are recommended before beginning to learn the practice of CEUS (EFSUMB 2010). Thus, by comparing groups with baseline high US experience, our test results represent the educational context in which the test is intended to be used and is in line with current recommendations of pre-required competences before learning CEUS.

It should be noted that the clinical practice of CEUS differs between continents. Only one UCA (Sonazoid, GE Healthcare/Daiichi Sankyo, Oslo, Norway/Tokyo, Japan) is available in Norway, Korea and other Asian countries and is licensed for characterization of focal liver and breast lesions. Three UCAs are currently approved by the U.S. Food and Drug Administration (FDA) for intravenous use (Lumason [Bracco, Milan, Italy], Definity [Lantheus, North Billerica, MA, USA] and Optison [GE Healthcare, Oslo, Norway]). These

were originally approved for delineation of the endocardial border in suboptimal echocardiograms, but in 2016, Lumason was additionally licensed for characterization of focal liver lesions in adults and children. In Europe, the licensed applications are similar, and most other usages are "off-label" (Sidhu et al. 2018). However, off-label use is sanctioned by medical authorities if the prescriber (i) is satisfied that it will serve the patients' needs; (ii) is satisfied that sufficient evidence base and/or experience of using the UCA has demonstrated its safety and efficacy; and (iii) takes responsibility for prescribing the UCA and oversees the patient's care (Sidhu et al. 2012; General Medical Council 2021). The current guidelines of non-hepatic applications of CEUS by EFSUMB provide the evidence base to incorporate UCAs into clinical practice and encourage that this be done when meaningful (Sidhu et al. 2012, 2018). This is reflected by the current broader clinical use of UCAs in Europe compared with the United States.

Consequences

We calculated a pass/fail score using the contrasting groups' method with estimated theoretical false positives and false negatives of 40.4% and 10.9%, respectively.

We believe that the developed test does not suffice to certify CEUS trainees, but rather should be used as an entry test before taking a practical test in core CEUS competencies and for learning purposes via structured feedback. For this purpose, we deem the relatively large estimated theoretical false-positive and false-negative rates to be acceptable.

In summary, we recommend that the test may be used to assess medical doctors' knowledge in all aspects of core CEUS (preparation, performance and general interpretation principles of CEUS studies). A re-evaluation of the validity evidence is recommended before using the test in other setups (*e.g.,* a future competence assessment test in organ-specific CEUS).

CONCLUSIONS

We have developed the first theoretical test in core CEUS competencies and gathered multiple sources of validity evidence. The test may be used as an entry test before learning practical CEUS competencies and as an evaluation and feedback tool.

CONFLICT OF INTEREST DISCLOSURE

A.B. is an advisory board member for General Electrics, Inventiva and Boehringer Ingelheim? The remaining authors have nothing to disclose.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ultrasmedbio.2021.10.016.

REFERENCES

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; 1999.

Bloch R, Norman G. Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide No. 68. Med Teach 2012;34:960–992.

Chong WK, Papadopoulou V, Dayton PA. Imaging with ultrasound contrast agents: Current status and future. Abdom Radiol (NY) 2018;43:762–772.

Cook DA, Hatala R. Validation of educational assessments: A primer for simulation and beyond. Adv Simul (Lond) 2016;1:31.

de Villiers MR, de Villiers PJ, Kent AP. The Delphi technique in health sciences education research. Med Teach 2005;27:639–643.

Dietrich CF, Nolsøe CP, Barr RG, Berzigotti A, Burns PN, Cantisani V, Chammas MC, Chaubal N, Choi BI, Clevert DA, Cui X, Dong Y, D'Onofrio M, Fowlkes JB, Gilja OH, Huang P, Ignee A, Jenssen C, Kono Y, Kudo M, Lassau N, Lee WJ, Lee JY, Liang P, Lim A, Lyshchik A, Meloni MF, Correas JM, Minami Y, Moriyasu F, Nicolau C, Piscaglia F, Saftoiu A, Sidhu PS, Sporea I, Torzilli G, Xie X, Zheng R. Guidelines and Good Clinical Practice Recommendations for Contrast-Enhanced Ultrasound (CEUS) in the Liver—Update 2020 WFUMB in Cooperation with EFSUMB, AFSUMB, AIUM, and FLAUS. Ultrasound Med Biol 2020;46:2579–2604.

Dimcevski G, Kotopoulis S, Bjånes T, Hoem D, Schjøtt J, Gjertsen BT, Biermann M, Molven A, Sorbye H, McCormack E, Postema M, Gilja OH. A human clinical trial using ultrasound and microbubbles to enhance gemcitabine treatment of inoperable pancreatic cancer. J Control Release 2016;243:172–181.

Downing SM, Yudkowsky R. Assessment in health professions education. New York: Routledge; 2009.

European Federation of Societies for Ultrasound in Medicine and Biology [EFSUMB]. Minimum training recommendations for the practice of medical ultrasound. Ultraschall Med 2006;27:79–105.

European Federation of Societies for Ultrasound in Medicine and Biology [EFSUMB]. Minimum training requirements for the practice of medical ultrasound in Europe. Ultraschall Med 2010;31:426–427.

General Medical Council, United Kingdom (GMC). Good practice in prescribing medicine— Guidance for doctors. April 2021. Available at: https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/good-practice-in-prescribing-and-managing-medicines-and-devices#10.

Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. Appl Meas Educ 2002;15:309–334.

Jacobsen N, Nolsøe CP, Konge L, Graumann O, Dietrich CF, Sidhu PS, Piscaglia F, Gilja OH, Laursen CB. Contrast-enhanced ultrasound: Development of syllabus for core theoretical and practical competencies. Ultrasound Med Biol 2020;46:2287–2292.

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

Jørgensen M, Savran MM, Christakopoulos C, Bek T, Grauslund J, Toft PB, Ziemssen F, Konge L, Sørensen TL, Subhi Y. Development and validation of a multiple-choice questionnaire-based theoretical test in direct ophthalmoscopy. Acta Ophthalmol 2019;97:700–706.

Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990;65:S63–S67.

Nolsoe CP, Nolsoe AB, Klubien J, Pommergaard HC, Rosenberg J, Meloni MF, Lorentzen T. Use of ultrasound contrast agents in

relation to percutaneous interventional procedures: A systematic review and pictorial essay. J Ultrasound Med 2018;37:1305–1324.
Pietersen PI, Konge L, Madsen KR, Bendixen M, Maskell NA, Rahman N, Graumann O, Laursen CB. Development of and gathering validity evidence for a theoretical test in thoracic ultrasound. Respiration 2019;98:221–229.
Piscaglia F, Bolondi L. The safety of Sonovue in abdominal applications: Retrospective analysis of 23188 investigations. Ultrasound Med Biol 2006;32:1369–1375.
Quaia E, Alaimo V, Baratella E, Pizzolato R, Cester G, Medeot A, Cova MA. Effect of observer experience in the differentiation between benign and malignant liver tumors after ultrasound contrast agent injection. J Ultrasound Med 2010;29:25–36.
Sidhu PS, Choi BI, Nielsen MB. The EFSUMB Guidelines on the non-hepatic clinical applications of contrast enhanced ultrasound (CEUS): A new dawn for the escalating use of this ubiquitous technique. Ultraschall Med 2012;33:5–7.
Sidhu PS, Cantisani V, Dietrich CF, Gilja OH, Saftoiu A, Bartels E, Bertolotto M, Calliada F, Clevert DA, Cosgrove D, Deganello A, D'Onofrio M, Drudi FM, Freeman S, Harvey C, Jenssen C, Jung EM, Klauser AS, Lassau N, Meloni MF, Leen E, Nicolau C, Nolsoe C, Piscaglia F, Prada F, Prosch H, Radzina M, Savelli L, Weskott HP, Wijkstra H. The EFSUMB guidelines and recommendations for the clinical practice of contrast-enhanced ultrasound (CEUS) in non-hepatic applications: Update 2017 (long version). Ultraschall Med 2018;39:e2–e44.
Wilson SR, Burns PN, Kono Y. Contrast-enhanced ultrasound of focal liver masses: A success story. Ultrasound Med Biol 2020;46:1059–1070.