Eur J Vasc Endovasc Surg (2020) 60, 933—941

ESVS CERTIFICATION

European Society for Vascular Surgery (ESVS) Certification of Theoretical and Practical Competences in Basic Vascular Ultrasound: Validity Investigation of the Assessment Tools

Leizl J. Nayahangan [a,*], Ramon Vila [b,c], Jonathan Lawaetz [a,d], José I. Leal Lorenzo [e], Kim K. Bredahl [c,d], Martin Lawaetz [d], Gaspar Mestres [f], Lise Westerlin [d], Steven K. Rogers [g], Filipe Fernandes [h,i], Elena I. Ortega [b,c], Lars Konge [a,j], Jonas Eiberg [a,c,d,j]

[a] Copenhagen Academy for Medical Education and Simulation, The Capital Region of Denmark, Copenhagen, Denmark
[b] Department of Vascular Surgery, Hospital Universitari de Bellvitge, Barcelona, Spain
[c] European Society for Vascular Surgery (ESVS), France
[d] Department of Vascular Surgery, Rigshospitalet, Copenhagen, Denmark
[e] Department of Vascular Surgery, Clinica Universidad de Navarra, Madrid, Pamplona, Spain
[f] Servicio de Cirugía Cardiovascular, Hospital Clínic de Barcelona, Barcelona, Spain
[g] Academic Surgery Unit, Division of Cardiovascular Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, and The Vascular Studies Unit, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Manchester, UK
[h] Vascular Laboratory, Medical Engineering and Physics Department, King's College NHS Foundation Trust, London, UK
[i] Faculty of Life Sciences and Medicine, King's College London, UK
[j] Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

WHAT THE PAPER ADDS
This paper presents evidence of validity of the comprehensive assessment tool, the Assessment of basic Vascular Ultrasound Expertise, designed to assess competences during the European Society for Vascular Surgery Academy's certification programme in basic vascular ultrasound. A pass/fail score was defined for each component, ensuring that operators achieve the competences in all aspects to be certified.

Objective: The aim of this study was to gather validity evidence for the Assessment of basic Vascular Ultrasound Expertise (AVAUSE) tool, and to establish a pass/fail score for each component, to support decisions for certification.
Methods: A cross sectional validation study performed during the European Society for Vascular Surgery's annual meeting. Validity evidence was sought for the theoretical test and two practical tests based on Messick's framework. The participants were vascular surgeons, vascular surgical trainees, sonographers, and nurses with varying experience levels. Five vascular ultrasound experts developed the theoretical and two practical test components of the AVAUSE tool for each test component. Two stations were set up for carotid examinations and two for superficial venous incompetence (SVI) examinations. Eight raters were assigned in pairs to each station. Three methods were used to set pass/fail scores: contrasting groups' method; rater consensus; and extended Angoff.
Results: Nineteen participants were enrolled. Acceptable internal consistency reliability (Cronbach's alpha) for the AVAUSE theoretical (0.93), carotid (0.84), and SVI (0.65) practical test were shown. In the carotid examination, inter-rater reliability (IRR) for the two rater pairs was good: 0.68 and 0.78, respectively. The carotid scores correlated significantly with years of experience (Pearson's $r = 0.56$, $p = .013$) but not with number of examinations in the last five years. For SVI, IRR was excellent at 0.81 and 0.87. SVI performance scores did not correlate with years of experience and number of examinations. The pass/fail score set by the contrasting groups' method was 29 points out of 50. The rater set pass/fail scores were 3.0 points for both carotid and SVI examinations and were used to determine successful participants. Ten of 19 participants passed the tests and were certified.
Conclusion: Validity evidence was sought and established for the AVAUSE comprehensive tool, including pass/fail standards. AVAUSE can be used to assess competences in basic vascular ultrasound, allowing operators to progress towards independent practice.

Keywords: Assessment, Certification, Education, Messick, Validity, Vascular ultrasound
Article history: Received 23 March 2020, Accepted 24 July 2020, Available online 6 September 2020

* Corresponding author. CAMES Rigshospitalet, Blegdamsvej 9, Copenhagen, Denmark.
E-mail address: leizl.joy.nayahangan@regionh.dk (Leizl J. Nayahangan).
1078-5884/© 2020 European Society for Vascular Surgery. Published by Elsevier B.V. All rights reserved.
https://doi.org/10.1016/j.ejvs.2020.07.075

INTRODUCTION

Vascular ultrasound (US) has become a cornerstone in vascular surgery as an essential part of screening programmes, pre-operative work up, intra-operative quality control, and post-operative follow up.[1–4] Vascular surgeons are embracing this non-invasive diagnostic modality in the outpatient clinic, operating theatre, and when on call.[5,6] The importance of mastering vascular US at a basic level was demonstrated in studies investigating the training needs of residents in Denmark and across Europe.[7,8] The major challenge is to ensure that the operator has achieved a defined level of competence in knowledge and technical skills, ensuring high sonographic quality, image optimisation, and interpretation.[5,9]

In an era of competency based medical education where learning is based on objectives rather than time, new education models such as mastery learning have been introduced to equip health professionals with knowledge and skills throughout the training continuum.[10] Timely and meaningful assessment of competence is critical to ensure continued progression from the training milieu to the clinical setting.

The European Society for Vascular Surgery Academy Committee (ESVS Academy) provides accredited workshops in basic vascular US to unify the education and training of vascular health professionals in Europe. Consequently, a certification programme was initiated in 2018 to assess basic vascular US competences and determine the readiness of an operator to progress towards independent practice (www.esvs.org/ccvus-certification). The ambition was that an ESVS certified sonographer should be able to demonstrate satisfactory knowledge and technical skills to scan without direct supervision and achieve decision making skills to call on expert supervisors when needed. For this

Table 1. Sources of validity evidence for the Assessment of Vascular Ultrasound Expertise (AVAUSE) theoretical and practical tests based on Messick's framework of validity

Source of validity evidence	Definition	Validity evidence for the AVAUSE tests
1. Content	The content of the test should measure what it intends to measure, which in this study was competence in basic vascular ultrasound	The theoretical and two practical tests were developed by experts in vascular US (vascular surgeons and scientists/sonographers) based on the most important aspects when performing basic US examinations
2. Internal structure	This refers to the internal consistency reliability of the test items, where similar scores should be achieved when measuring the same construct Cronbach's alpha ≥.90 is sound for high stakes examination	Theoretical test: Cronbach's alpha = .93 Carotid US examination: • Internal consistency reliability = Cronbach's alpha = .84 • IRR: rater 1 and rater 2: Pearson's $r = .68$, $p = .062$, while rater 3 and rater 4: $r = .78$, $p = .005$ SVI US examination: • Internal consistency reliability = Cronbach's alpha = .65 • IRR: rater 1 and rater 2: $r = .81$, $p = .005$, while rater 3 and rater 4: $r = .87$, $p = .002$ No significant correlation with number of examinations in the last five years ($r = .23$, $p = .34$)
3. Response process	Quality and integrity of the test process is observed and maintained at all times	The test setting was standardised, with the same two facilitators managing the process. The raters underwent rater training prior to the certification programme, ensuring understanding of the test items and description of the rating scale
4. Relationships to other variables	The test scores should correlate with known measures of competence, in this case levels of experience	Theoretical test: the novice group as defined by the medical students ($n = 14$) had a mean ± SD test score of 20.7 ± 4.6, while the experienced group as defined by the participants had a mean test score of 40.1 ± 6.6 Carotid US examination: total performance scores correlated significantly with years of experience with Pearson $r = .56$, $p = .013$, no significant correlation with number of examinations in the last five years ($r = .23$, $p = .34$) SVI US examination: total performance scores did not correlate with years of experience ($r = .28$, $p = .25$), as well as number of examinations in the last five years
5. Consequences	This refers to standard setting of a pass/fail score to support decisions regarding test results	Theoretical test: pass/fail score = 29 points Carotid US examination: • rater set pass/fail score = 3 • Angoff set pass/fail score = 4.1 SVI US examination: • rater set pass/fail score = 3 • Angoff set pass/fail score = 3.7

US = ultrasound; IRR = inter-rater reliability; SVI = superficial venous incompetence; SD = standard deviation.

purpose, the Assessment of basic Vascular Ultrasound Expertise (AVAUSE) tool was developed, which includes both theoretical and practical tests covering the most important US examinations within the vascular curriculum. International certification is high stake, with potential consequences for patients; hence, solid validity evidence is needed to support the proposed use and ensure that the interpretation of scores and decisions based on the results are trustworthy, defensible, and repeatable over time.[11] In this study, the aim was to investigate both the theoretical and practical tests for evidence of validity, as well as establish credible pass/fail scores to explore the consequences of testing.

MATERIALS AND METHODS

This cross sectional validation study was conducted during the certification programme at the 2019 ESVS annual meeting in Hamburg, Germany.

Validity was investigated using the unified, contemporary framework proposed by Messick,[11] which gathers validity evidence from five sources: content; response process; internal structure; relations to other variables; and consequences (Table 1).[12]

Participants

Participants included vascular surgeons, vascular surgical trainees, sonographers, and nurses with varying experience levels (Table 2). To qualify for certification, participants must have a minimum of 100 vascular US scans across different modalities (e.g., carotid, lower limb venous, peripheral arteries, and abdominal aortic aneurysm), in addition to a letter of recommendation from their departmental clinical lead (Table 3).

Table 2. Characteristics of the participants who enrolled and have taken the European Society for Vascular Surgery certification programme in basic vascular ultrasound

Profession	Participants (n = 19)
Vascular surgeon	9 (47)
Vascular surgical trainee	6 (32)
Nurse	1 (5)
Sonographer	2 (11)
Other	1 (5)
Median (range) time of experience – y	5 (2–22)

Data are presented as n (%) unless stated otherwise.

Theoretical test

Phase 1: development of the test

Step 1 (content). A theoretical test for basic vascular US was developed in 2018 by vascular surgeons and vascular scientists/sonographers (RV, JE, FF, SR, and JIL) with extensive expertise in performing and teaching vascular US. The test consisted of 53 multiple choice questions (MCQs) with one correct answer from five choices. This first version was administered during the ESVS annual meeting in 2018 in Valencia, Spain (Fig. 1).

Step 2 (modification of content). To improve the content of the theoretical test and establish validity evidence, the 2018 test results were used to evaluate the quality of the MCQs. Item analysis was performed by calculating the proportion of the examinees who answered the MCQ correctly and categorised them according to difficulty based on the Item Classification guide by Haladyna and colleagues (described in the data analysis section).[13,14] Poor quality MCQs were eliminated. The test developers were asked to contribute 5–10 new MCQs to the remaining pool.

Phase 2: validity investigation for the test

Step 3 (content validity). The MCQs were sent back in random order to the test developers to appraise and rate according to relevance, where one was considered “completely irrelevant” and five considered “extremely relevant”. They were also asked to revise the wording to ensure clarity and accuracy.

Step 4 (gathering validity evidence). The MCQ test was administered to the participants during the certification programme in 2019. The test was developed in Moodle and was accessed using individual computers (Moodle, Perth, Australia).

Standard setting for the theoretical test. The contrasting groups’ method was used to set a pass/fail score for the theoretical test by involving 14 Danish medical students without US experience as novices and the participants as the experienced group.[15] The pass/fail score was the intercept between the normalised distribution of performance scores between the two groups.

Practical tests

Owing to time constraints, two of four US examinations were chosen for the practical test: a unilateral carotid US

Table 3. Number of ultrasound examinations performed in the last five years in 19 participants of the examination for the European Society for Vascular Surgery certification program in basic vascular ultrasound

Vascular ultrasound experiences	Number of examinations				
	0–50	51–100	101–500	501–1000	1 000+
Carotid ultrasound	2	4	8	5	0
Arterial ultrasound	2	5	5	6	1
Venous ultrasound	2	1	7	5	4
Aortic ultrasound	6	1	8	4	0

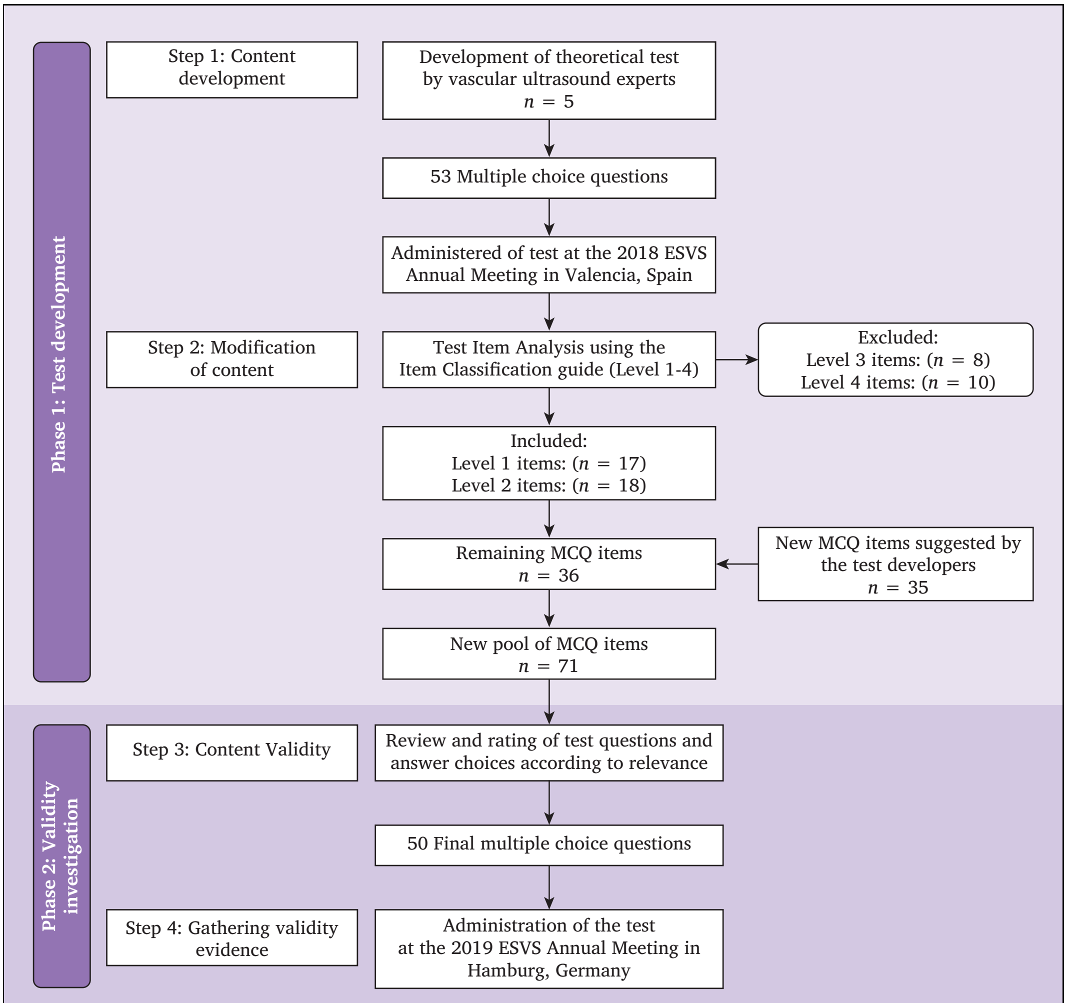

Figure 1. Flow diagram showing the development and validation of the multiple choice questionnaire (MCQ) that comprises the theoretical part of the European Society for Vascular Surgery (ESVS) certification programme in basic vascular ultrasound.

examination and a unilateral superficial venous incompetence (SVI) examination.

Content development. Practical assessment tools in carotid and SVI US examinations were developed by the same five experts. These included the most important aspects when performing basic US examinations, such as positioning of patient, choosing the correct transducer and US application, and adjusting settings (gain, time gain compensation, focus, depth, and pulse repetition frequency). For the carotid examination, the participants needed to make appropriate velocity measurements, and differentiate between the external and internal carotid arteries correctly, as recommended in recent guidelines.[16,17] For SVI examination, participants needed to identify the common femoral vein and test for reflux, as well as the great and small saphenous veins. These specific tasks were expressed as individual test items using a five point rating scale, where 1 indicated “very poor/needs help”; 3 indicated “acceptable”; and 5 indicated “very good”.

Raters and rater training. Eight expert vascular US operators (RV, JE, FF, SR, EIO, KB, LW, and ML [vascular surgeons, vascular nurses, and vascular scientists/sonographers]) assessed the participants during the practical tests. Rater training was initiated prior to the certification to ensure consistent use of the assessment tools. For this purpose, four pilot videos were sent to the raters to individually assess a carotid US examination performed by a novice nurse trainee and by an experienced vascular nurse (>1 000 carotid scans); and an SVI examination performed by a novice nurse trainee and by an experienced vascular nurse (>1 000 venous scans). Instructions were provided, including an explanation of the test items and scale. Results of the pilot rating were presented to the raters during a face to face meeting to discuss scoring differences until

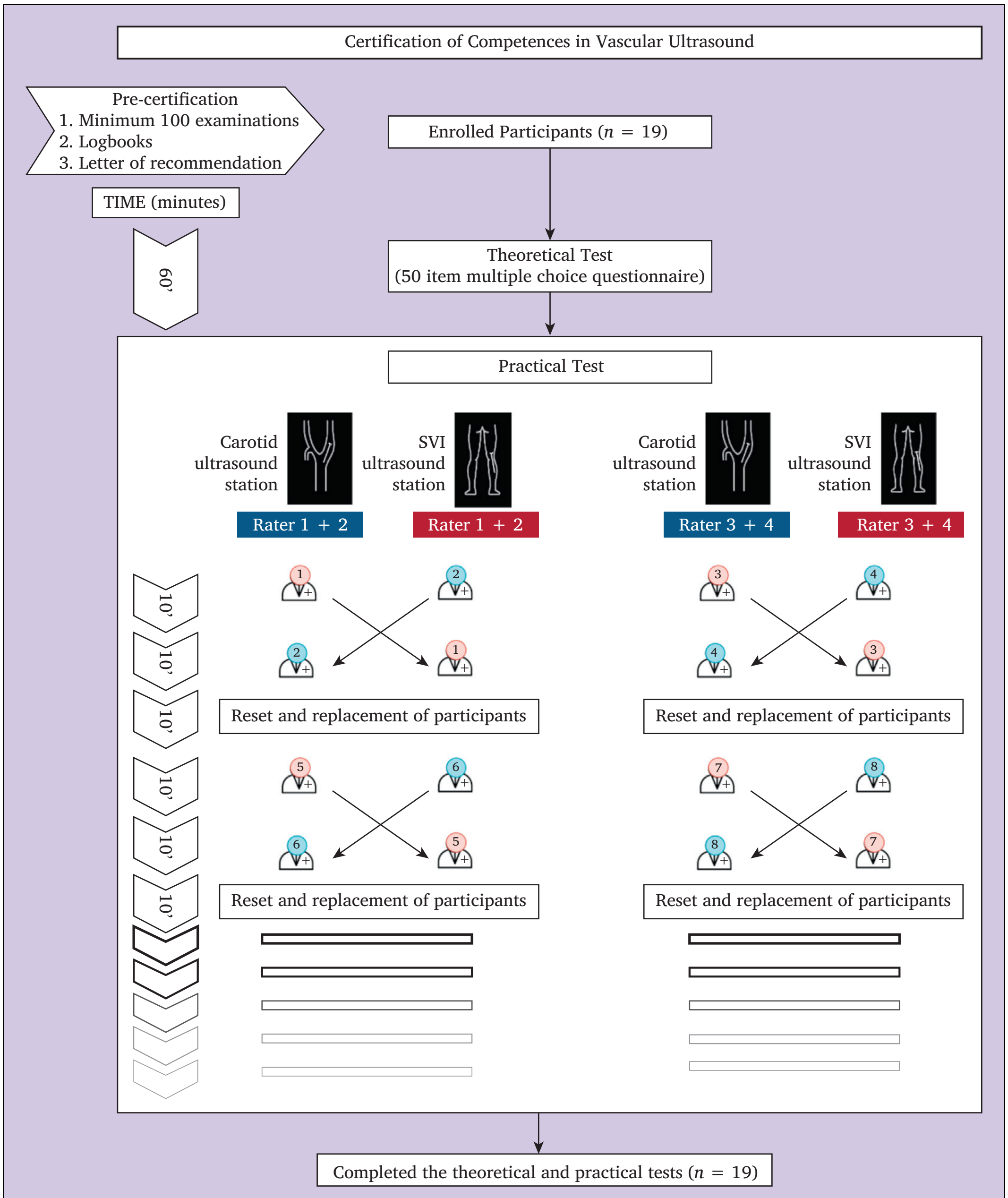

Figure 2. Flow diagram showing the set up during the practical part of the European Society for Vascular Surgery certification programme in basic vascular ultrasound. The set up included two carotid ultrasound (US) stations and two superficial venous incompetence (SVI) US stations, with two raters assigned to each station to assess the participants.

consensus was reached. This also allowed for rewording of the test items and include a global rating scale (GRS) to assess overall performance, rated across a scale of 1 indicating "incompetent"; 3 indicating "borderline"; and 5 indicating "competent".

The carotid assessment tool consisted of 14 items plus the GRS, while the SVI assessment tool consisted of 10 items plus the GRS.

Test developers and raters. The five experts who developed the theoretical test, as well as the eight raters in the practical tests, are ESVS members and were affiliated with the ESVS Academy as convenors during workshops/courses. They were recommended by the Academy leadership for their expertise and interest in clinical vascular US and training. Additionally, the experts were distributed across different countries and professions.

Test process and data collection. Plenary introduction was provided prior to the test, succeeded by the 60 min theoretical test, after which the platform automatically closes down. The practical test followed immediately and consisted of two stations designated for carotid and two for SVI testing; all stations were equipped with a high end US system, relevant transducers, scanning bed, US model (healthy medical student), and two raters (Fig. 2). Participants were randomly tested in both carotid and SVI examinations, with a maximum of 10 min per station. Assessment was by direct observation, using eight tablets equipped with a specially developed online assessment platform through the Research Electronic Data Capture Software (REDCap). Two facilitators (LJN and JL) guided the participants to assigned stations, solved technical issues, and ensured the completeness of rating.

Standard setting, practical test. Two methods were used to set pass/fail scores for the practical tests. Firstly, the eight expert raters were asked prior to certification to estimate and discuss how well participants should perform in each item before being allowed to work independently in the outpatient clinic. An acceptable pass/fail score was set after several discussions and when consensus was reached. Secondly, the extended Angoff method was used to collect expert judgement regarding pass/fail scores.[18,19] Vascular US experts ($n = 10$ for carotid; $n = 8$ for SVI)) from the UK, Spain, Serbia, and Denmark were identified and invited to determine an appropriate passing mark based on the minimum practical skill level that is demanded of a sonographer before being allowed to work independently in the expert clinic.

Certification. Participants must pass all three components (theoretical test and two practical tests) to be an ESVS certified operator in basic vascular US.

Data and statistical analysis

Item analysis of the theoretical test results from 2018 was performed to explore question quality. The item classification guide was used, where level 1 is considered the best fit with middle difficulty (item difficulty = 0.45–0.75); level 2 questions are considered easy questions (item difficulty = 0.76–0.91); level 3 questions are considered difficult (item difficulty = 0.25–0.44); level 4 questions are either very easy or very difficult, and are considered poor quality questions (item difficulty <0.24 or >0.91).[14] Levels 3 and 4 were considered as poor quality questions and were eliminated (Fig. 1, step 2). During the review and rating of questions in step 3, mean scores per questions were calculated. Questions that scored <4 were eliminated.

Internal consistency reliability of the AVAUSE theoretical and practical tools was explored using Cronbach's alpha to ensure that the individual items appropriately measure the same construct. For the practical tests, the total performance score was calculated as the mean between the multiple item rating scale and the GRS scores. The Pearson correlation coefficient was used to explore the relationship between the item based score and the GRS; the inter-rater reliability (IRR) between the paired raters for each station; the correlation between the carotid performance scores and SVI performance scores; and the relationship between total performance scores and years of experience. Spearman rho's correlation coefficient (r) was used to explore the correlation between total performance scores and number of examinations in the last five years.

Data analysis was done using SPSS Statistics version 25.0 (IBM, Armonk, NY, USA).

Ethics

No patient data were used in this study. Informed consent was provided by the participants, indicating their agreement to participate and with the understanding that their test results would be used for research purposes.

RESULTS

Nineteen participants were enrolled. Ten participants passed all three components and were certified; nine failed.

Theoretical test

The initial theoretical test consisted of 53 MCQs, of which 18 were eliminated after item analysis (Fig. 1, step 2). A new set of 36 MCQs was added to the remaining pool. These 71 MCQs were sent to test developers to be rated; 21 were eliminated, and after further discussions and revisions, a 50 item MCQ test was generated, covering US physics, vascular haemodynamics, abdominal aorta, SVI, carotid, and supragenicular arteries (Fig. 1, step 3).

The internal consistency reliability of the MCQ test was high, with a Cronbach's alpha of 0.93. The novice group, as defined by the medical students ($n = 14$), had a mean $\pm$ standard deviation test score of 20.7 $\pm$ 4.6, while the experienced group, as defined by the participants, had a mean test score of 40.1 $\pm$ 6.6.

Practical tests

Assessment of carotid US skills. Investigating the internal consistency reliability of the practical test resulted in a Cronbach's alpha of 0.84. A significant correlation was found between the total item scores and the global scores with a Pearson's r of .79 ($p < .001$). The IRR for raters 1 and 2 was acceptable but not significant ($r = .68$, $p = .062$), while the IRR for raters 3 and 4 was good ($r = .78$, $p = .005$). The carotid total performance scores correlated significantly with years of experience ($r = .56$ $p = .013$; Fig. 3); however, there was no significant correlation with number of examinations in the last five years ($r = .23$, $p = .34$).

Assessment of SVI US skills. The internal consistency for the test items was acceptable, with a Cronbach's alpha of .65. A significant correlation was found between the total item scores and the global scores ($r = .75$, $p < .001$). The IRR for raters 1 and 2 was excellent ($r = .81$, $p = .005$); for raters 3 and 4 Pearson's r was .87 ($p = .002$). The SVI performance scores did not correlate with years of experience ($r = .28$, $p = .25$), as well as number of examinations in the last five years ($r = .28$, $p = .25$).

Pass/fail scores

Theoretical test. The pass/fail score set by the contrasting groups' method was 29 points out of 50. Seventeen of the 19 experienced participants passed, while none of the novice medical students passed.

Practical test. The mean pass/fail score set by the raters for the carotid US examination was 3, while according to the Angoff method the mean pass/fail score was 4.1. For the SVI examination, the rater set mean pass/fail score was 3, while according to the Angoff method the mean score was 3.7. For the purpose of this certification, the rater set pass/fail score was used, where 12 and 17 participants passed the carotid and SVI examination, respectively. There was no significant correlation between the total performance scores in carotid examinations and the total performance score for SVI examinations ($r = -.5$, $p = .86$; Fig. 4.).

DISCUSSION

Reliable and timely assessment of the competence of healthcare professionals provides an impression of actual performance, identifies gaps, targets future learning, and improves patient safety.[20] In this study, two domains of competence were focused on, theoretical and technical skills in basic vascular US, and a newly developed comprehensive assessment tool (AVAUSE) for validity evidence was investigated based on five sources outlined by the Messick framework.

The content of the theoretical and practical tests was developed by vascular US experts from different professions and countries. This strategy ensured the inclusion of a wider range of topics and important competences.[21] The test developers went through a systematic approach to appraise the test questions for relevance and quality, ensuring that the multiple items in the practical test were specific and evaluative of the targeted knowledge and competences. The inclusion of a GRS provided an opportunity for raters to objectively assess a participant who may have performed most steps correctly (and potentially pass the test) but have missed a critical step that could compromise patient safety. As an example, one participant performed an SVI examination on a supine patient and therefore failed despite having performed all other steps correctly (reflux cannot be assessed without either being erect or having the patient in a steep reverse Trendelenburg position). The correlation between the multiple item rating scale scores and the GRS was high, suggesting that both scales measure the same construct and quality. This is consistent with previous studies exploring the use of the Objective Structured Assessment of Technical Skills (OSATS) and the Objective Structured Clinical Examination (OSCE), where the GRS was shown to have discriminatory ability to detect different levels of expertise.[22–24] Additionally, multiple item rating scales are preferred when assessing complex performances or experienced participants.[25]

The internal consistency reliability of the AVAUSE practical tools were acceptable, confirming that the test items measure relevant competence in the two US scans. Reliability coefficients of 0.80–0.89 are acceptable for tests such as end of year examinations, while coefficients <0.70 is useful as a component of an overall composite score.[26] For SVI US examination, the internal consistency was 0.65; however, it is still a useful component of the overall AVAUSE score. One way to increase reliability is to increase the number of test items.[26] IRR between the paired raters was significant except for raters 1 and 2 in one of the carotid stations. Rater 2 was not able to attend the rater training meeting but received instructions; nevertheless, the agreement between the two was variable. This underlines the importance of rater training prior to assessment to increase IRR and accuracy.[27] Standardised instructions and processes were observed to minimise bias and maintain assessment security. The raters were familiarised with the tests and were trained on the online ratings. Test facilitators ensured efficient test flow and confidentiality of ratings and scores between raters.

Interestingly, the AVAUSE practical tool for carotid US examination showed significant ability to discriminate based on years of experience ($r = .56$, $p = .01$) but SVI total performance scores did not. Furthermore, no significant correlation was found in this dataset between the number of examinations and total performance scores. This emphasises that absolute volume of procedures does not necessarily correlate with competence.[28] These numbers were also self reported and based on recall, and therefore may not be reliable. A systematic review of accurate self assessment found weak or no associations between self assessment of competence *vs.* external objective observations.[29] Other measures such as years of experience cannot

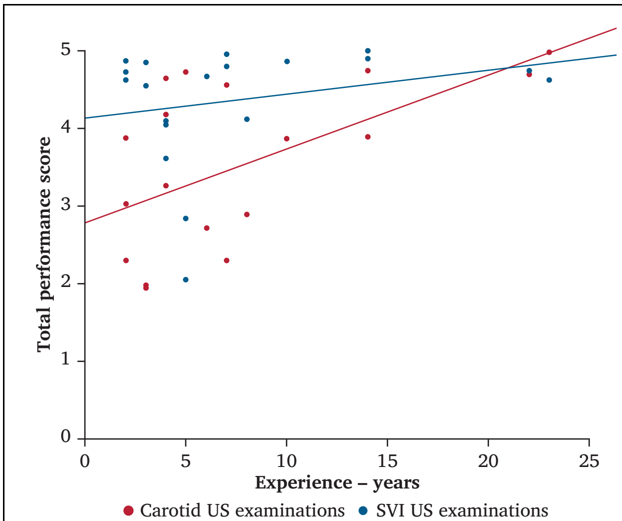

Figure 3. Correlation between years of experience and total performance score of the 19 participants in carotid ultrasound (US) examination ($r = .56$; $p = .013$); and years of experience and total performance score in superficial venous incompetence (SVI) examination ($r = .23$; $p = .34$) for the European Society for Vascular Surgery (ESVS) certification programme.

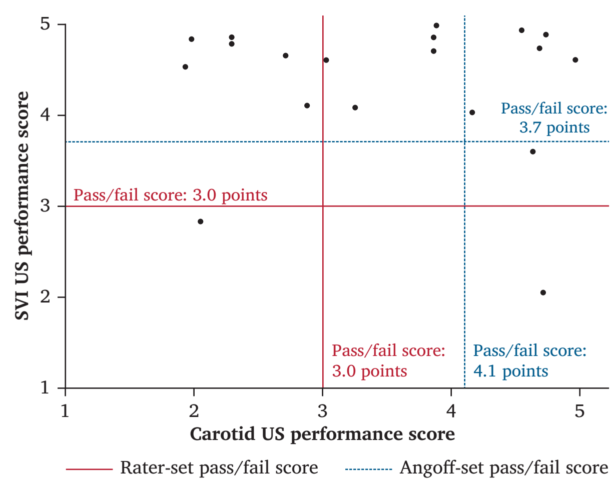

Figure 4. Correlation between total performance score of the 19 participants in carotid ultrasound (US) examination and the total performance score in superficial venous incompetence (SVI) US examination for the European Society for Vascular Surgery (ESVS) certification program plotted against the rater set (red solid line) and Angoff (blue dashed line) pass/fail scores.

predict competence nor acceptable performance in a valid way.[30] To identify proficient vascular US operators, objective assessment of performance, combined with experience, must be considered.[31] No significant correlation was found between the carotid and the SVI performance scores, indicating that competence in one examination does not reflect competence in another, as the skill sets are entirely different.

An imperative asset of this study is the determination of credible and defensible pass/fail scores for each part of the AVAUSE tool, all of which must be passed in order to be certified. This non-compensatory approach ensures that the participants reach a certain level of competence before they are certified in independent practice. The pass/fail score that was set by the raters was found to be a credible and reasonable benchmark. Given that this was decided after pilot rating, they had a shared mental model of how the borderline operator would perform. In contrast, the pass/fail score set by the extended Angoff method was very high. The experts invited to take part in this process have not seen any actual participant performance and have therefore influenced the unrealistically high scores. Expert practitioners tend to have (too) high expectations of their trainees.[32] Extensive instructions when setting standards, as well as a clear definition of minimum performance, are needed.[32] A rigorous approach to setting standards especially for high stakes examinations such as certification is essential. If the pass/fail score is set too low, trainees will not be well prepared for clinical practice, which might put patients in harm's way. If the standard is too high, trainees risk wasting valuable time and focusing their efforts on the test rather than educational goals.[31] However, maintaining a high standard for procedural performance decreases adverse events and promotes patient safety. An alternative standard setting method driven by patient safety was proposed where essential and non-essential test items are identified and a separate passing percentage for each category is determined. Essential items are those that when neglected can compromise patient safety, comfort, and procedure outcome.[33]

The small number of participants is a limitation as it could have affected the test scores and the correlations between performances. It would be interesting to explore further evidence between the relationship of performance scores and other variables (i.e., different professional groups and years of experience) in a larger cohort once the certification is fully implemented. Direct observation during the practical test may have introduced negative or positive bias but was minimised by allocating participants to unfamiliar raters. The use of healthy models was also a limitation. One of the challenges of performing assessment in a simulated setting is its accuracy to predict how a participant performs a procedure independently in the clinical environment.[34] It is proposed that this ESVS certification programme is considered an initial step and that it entrust the final decision for independent practice to the local clinical leads.

The diverse professional backgrounds of vascular US operators across Europe makes it challenging to establish professional level certification. However, pursuant of the goal to standardise training and practice of vascular US skills, the next steps may be towards certification of advanced skills, including an "ESVS ultrasound residency" in selected departments, followed by "ESVS workshops in advanced vascular ultrasound".

Conclusion

The comprehensive AVAUSE tool has established arguments for validity, supporting the scores and consequence decisions. The AVAUSE will be used to evaluate theoretical and technical competences in vascular US during the annual ESVS certification programme. It is currently certifying basic vascular US, but, in the future, it is also expected to include advanced vascular US. This is an important initial step that informs an operator's readiness to progress towards independence in the clinical environment.

CONFLICTS OF INTEREST

None.

FUNDING

None.

ACKNOWLEDGEMENTS

The authors would like to thank Anishan Vamadevan, who helped and facilitated the recruitment of novice participants, as well as the 14 medical students who participated in the theoretical test.

REFERENCES

1 Moll FL, Powell JT, Fraedrich G, Verzini F, Haulon S, Waltham M, et al. Management of abdominal aortic aneurysms clinical practice

guidelines of the European Society for Vascular Surgery. *Eur J Vasc Endovasc Surg* 2011;**41**(suppl. 1):1–58.
2 Ricco JB, Schneider F, Illuminati G, Samson RH. The role of completion imaging following carotid artery endarterectomy. *J Vasc Surg* 2013;**57**:1432–8.
3 Rockman CB, Halm EA. Intraoperative imaging: does it really improve perioperative outcomes of carotid endarterectomy? *Semin Vasc Surg* 2007;**20**:236–43.
4 Bredahl K, Mestre XM, Coll RV, Ghulam QM, Sillesen H, Eiberg J. Contrast-enhanced ultrasound in vascular surgery: review and update. *Ann Vasc Surg* 2017;**45**:287–93.
5 Eiberg J, Konge L, Vila R, Sillesen H. How vascular surgeons can learn ultrasound. *Semin Vasc Surg* 2019;**32**:33–40.
6 Eiberg JP, Hansen MA, Gronvall Rasmussen JB, Schroeder TV. Minimum training requirement in ultrasound imaging of peripheral arterial disease. *Eur J Vasc Endovasc Surg* 2008;**36**:325–30.
7 Nayahangan LJ, Konge L, Schroeder TV, Paltved C, Lindorff-Larsen KG, Nielsen BU, et al. A national needs assessment to identify technical procedures in vascular surgery for simulation based training. *Eur J Vasc Endovasc Surg* 2017;**53**:591–9.
8 Nayahangan LJ, Van Herzeele I, Konge L, Koncar I, Cieri E, Mansilha A, et al. Achieving consensus to define curricular content for simulation based education in vascular surgery: a Europe wide needs assessment initiative. *Eur J Vasc Endovasc Surg* 2019;**58**:284–91.
9 Nicholls D, Sweet L, Hyett J. Psychomotor skills in medical ultrasound imaging: an analysis of the core skill set. *J Ultrasound Med* 2014;**33**:1349–52.
10 McGaghie WC. Mastery learning: it is time for medical education to join the 21st century. *Acad Med* 2015;**90**:1438–41.
11 Messick S. *Foundations of Validity: Meaning and Consequences in Psychological Assessment*. ETS Research Report Series; 1993.
12 Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med* 2006;**119**:166.
13 Haladyna TM, Downing SM. A taxonomy of multiple-choice item-writing rules. *Appl Measure Educ* 1989;**2**:37–50.
14 Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. *Appl Measure Educ* 2002;**15**:309–33.
15 Jorgensen M, Konge L, Subhi Y. Contrasting groups' standard setting for consequences analysis in validity studies: reporting considerations. *Adv Simul (Lond)* 2018;**3**:5.
16 Naylor AR, Ricco JB, de Borst GJ, Debus S, de Haro J, Halliday A, et al. Editor's choice – Management of atherosclerotic carotid and vertebral artery disease: 2017 clinical practice guidelines of the European Society for Vascular Surgery (ESVS). *Eur J Vasc Endovasc Surg* 2018;**55**:3–81.
17 Oates CP, Naylor AR, Hartshorne T, Charles SM, Fail T, Humphries K, et al. Joint recommendations for reporting carotid ultrasound investigations in the United Kingdom. *Eur J Vasc Endovasc Surg* 2009;**37**:251–61.
18 Norcini JJ. Setting standards on educational tests. *Med Educ* 2003;**37**:464–9.
19 Angoff WH. Scales, norms, and equivalent scores. In: Thorndike RL, editor. *Educational measurment*. 2nd ed. Washington, DC: American Council on Education; 1971. p. 508–600.
20 Epstein RM. Assessment in medical education. *N Engl J Med* 2007;**356**:387–96.
21 Streiner DL, Norman GR, Cairney J. *Devising items. Health measurement scales: a practical guide to their development and use*. 5th ed. Oxford: Oxford University Press; 2015.
22 Martin J, Regehr G, Reznick R, Macrae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997;**84**:273–8.
23 Hodges B, McIlroy JH. Analytic global OSCE ratings are sensitive to level of training. *Med Educ* 2003;**37**:1012–6.
24 Ilgen JS, Ma IW, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ* 2015;**49**:161–73.
25 Yudkowsky R. Performance test. In: Yudkowsky R, Park YS, Downing SM, editors. *Assessment in health professions education*. 2nd ed. New York: Routledge; 2020.
26 Park YS. Reliability. In: Yudkowsky R, Park YS, Downing SM, editors. *Assessment in health professions education*. 2nd ed. New York: Routledge; 2020.
27 Feldman M, Lazzara EH, Vanderbilt AA, Diaz Granados D. Rater training to support high-stakes simulation-based assessments. *J Contin Educ Health Prof* 2012;**32**:279–86.
28 Reznick RK, MacRae H. Teaching surgical skills—changes in the wind. *N Engl J Med* 2006;**355**:2664–9.
29 Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* 2006;**296**:1094–102.
30 Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. *Ann Intern Med* 2005;**142**:260–73.
31 Yudkowsky R, Park YS, Lineberry M, Knox A, Ritter EM. Setting mastery learning standards. *Acad Med* 2015;**90**:1495–500.
32 Yudkowsky R, Downing SM, Tekian A. Standard setting. I. In: Yudkowsky R, Park YS, Downing SM, editors. *Assessment in health professions education*. 2nd ed. New York: Routledge; 2020.
33 Yudkowsky R, Tumuluru S, Casey P, Herlich N, Ledonne C. A patient safety approach to setting pass/fail standards for basic procedural skills checklists. *Simul Healthc* 2014;**9**:277–82.
34 Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;**65**:S63–7.