

RESEARCH

Open Access



# Objective structured assessment ultrasound skill scale for hyomental distance competence – psychometric study

Sara Hora Gomes<sup>1,2\*</sup>, Marta Trindade<sup>1</sup>, Cristina Petrisor<sup>3</sup>, Dinis Costa<sup>4</sup>, Jorge Correia-Pinto<sup>1,2,5</sup>, Patrício S. Costa<sup>1,2</sup> and José M. Pêgo<sup>1,2,6</sup>

## Abstract

**Background** Ultrasound assessment of the airway recently integrates the point-of-care approach to patient evaluation since ultrasound measurements can predict a difficult laryngoscopy and tracheal intubation. Because ultrasonography is performer-dependent, a proper training and assessment tool is needed to increase diagnostic accuracy. An objective, structured assessment ultrasound skill (OSAUS) scale was recently developed to guide training and assess competence. This work aims to study the psychometric properties of OSAUS Scale when used to evaluate competence in ultrasound hyomental distance (HMD) measurement. **Methods:** Prospective and experimental study. Volunteers were recruited and enrolled in groups with different expertise. Each participant performed three ultrasonographic HMD evaluation. The performance was videorecorded and anonymized. Five assessors blindly rated participants' performance using OSAUS scale and a Global Rating Scale (GRS). A psychometric study of OSAUS scale as assessment tool for ultrasound HMD competence was done. **Results:** Fifteen voluntaries participated on the study. Psychometric analysis of OSAUS showed strong internal consistency (Cronbach's alpha 0.916) and inter-rater reliability (ICC 0.720;  $p < 0.001$ ). The novice group scored  $15.4 \pm 0.18$  (mean  $\pm$  SD), the intermediate  $14.3 \pm 0.75$  and expert  $13.6 \pm 0.1.25$ , with a significant difference between novice and expert groups ( $p = 0.036$ ). The time in seconds to complete the task was evaluated: novice ( $90 \pm 34$ ) (mean  $\pm$  SD), intermediate ( $84 \pm 23$ ) and experts ( $83 \pm 15$ ), with no significant differences between groups. A strong correlation was observed between OSAUS and global rating scale ( $r = 0.970$ ,  $p < 0.001$ ).

**Conclusion** The study demonstrated evidence of validity and reliability. Further studies are needed to implement OSAUS scale in the clinical setting for training and assessment of airway ultrasound competence.

**Keywords** OSAUS, Hyomental, Distance, Airway, Ultrasound, Assessment, Psychometric properties

\*Correspondence:

Sara Hora Gomes  
saragomes@med.uminho.pt

<sup>1</sup>Life and Health Sciences Research Institute (ICVS), School of Medicine,  
University of Minho, Campus Gualtar, Braga 4710-057, Portugal

<sup>2</sup>ICVS/3B's-PT Government Associate Laboratory,  
Braga/Guimarães 4710-057, Portugal

<sup>3</sup>Anesthesia and Intensive Care II Department, Pharmacy Cluj-Napoca and  
Anesthesia and Intensive Care Department, "Iuliu Hatieganu" University of  
Medicine, Clinical Emergency County Hospital, Cluj-Napoca  
400347, Romania

<sup>4</sup>Department of Anesthesia, Hospital de Braga, Braga 4710-243, Portugal

<sup>5</sup>Department of Pediatric Surgery, Hospital de Braga, Braga  
4710-243, Portugal

<sup>6</sup>iCognitus4ALL – IT Solutions, Braga 4470-057, Portugal



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

In the last decades, the use of ultrasonography expanded from the imaging laboratory to patient bedside evaluation. Nowadays, almost all medical specialties have developed a Point-of-Care Ultrasonography (PoCUS) approach to enhance patients' primary assessment, diagnostics and treatment. In anesthesiology training curriculum and daily practice several PoCUS approaches to patient evaluation have been incorporated, namely for cardiac, lung, gastric and airway evaluation and to guide regional anesthesia and vascular access [1].

Ultrasound can be applied to multiple aspects of airway management, such as tube size positioning, predicting successful extubation, guiding cricothyrotomy and predict difficult airway [2]. Several ultrasound parameters have been studied as predictors of difficult laryngoscopy, specifically hyomental distance in neutral and extended position [3], distance from skin to hyoid bone [4], to vocal cords [5], to epiglottis [6], tongue cross-sectional area and volume [7] and many others. Nevertheless, in a recent systematic review and meta-analysis, the most consistent predictor was hyomental distance (HMD) in a neutral position [8].

Since ultrasound is an operator-dependent technique there is a need for structured training and standardized assessment to certify clinician's skills and competence [9–11]. However, there is no evidence-based guidelines for education or assessment of airway ultrasound.

Traditionally, skill competence was achieved after tutorized clinical training where residents were progressively trusted to practice autonomously by their tutors. The process was complex and was supported by knowledge assessment, gather information from third parties, structured supervision and a direct practical observation of trainees' performance [12–14]. In the last two decades, many valid and reliable instruments were developed to improve the objectivity of assessment by creating of scales and checklists [15, 16].

In 2013, Tolsgaard and co-workers [17] led an international multispecialty consensus on the content of a generic ultrasound rating scale using a Delphi technique. A total of 60 international ultrasound experts from different medical specialties (radiology, emergency medicine, obstetrics, surgery, urology, rheumatology and gastroenterology) were invited to participate in three Delphi rounds [17]. Since then, several authors have used this tool to train and assess proficiency for clinical ultrasound in a variety of fields, namely obstetric and gynecology [18–21], abdominal trauma (eFAST) [22, 23], lung [24], head and neck [25] ultrasound.

The Global rating scale (GRS) has been widely used in medical education and in clinical practice as a training and assessment tool. It is usually used as an independent tool or as a complement to support checklists

of technical, communication skills or other professional tasks [15, 26, 27]. The present study used GRS as an "overall performance scale" based on 5-points of Likert scale.

We developed an opportunity to train ultrasound hyomental distance measurement for different expertise groups in a simulated environment with standardized patients. The study aims to explore the psychometric properties of the OSAUS scale when used to assess hyomental ultrasound competence.

## Methods

### Materials

#### *Ethical approval*

The study was carried out following relevant guidelines and regulations. The study was conducted after institutional review committee approval from Ethical Committee for Institute of Life and Health Sciences (CEICVS) of School of Medicine, University of Braga, Braga, Portugal on 15th November 2020 (CEICVS15/2020). Participation in this study was voluntary, and all participants and assessors gave verbal and written informed consent.

#### *Study dates*

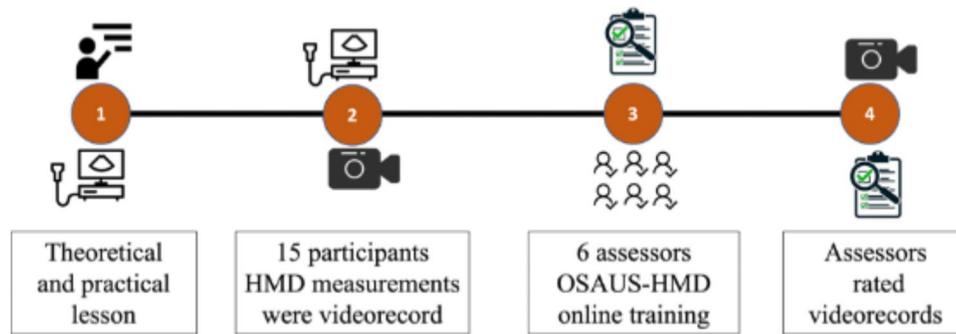
The study was conducted at the School of Medicine, University of Minho, Braga, Portugal, from November 2020 to June 2021.

#### *Study design*

This is a prospective experimental, rater and principal investigator double-blinded study to determine OSAUS's psychometric properties when the scale is used for the ultrasound measurement of hyomental distance with the head in neutral position.

The study has developed in 4 steps. (Fig. 1) In the first step an educational moment was organized with an one-hour theoretical presentation. Two experienced airway ultrasound anesthesiologists presented the OSAUS scale [17] and its applicability to measuring ultrasound hyomental distance in a neutral position. The protocol for HMD measurement in neutral position was very well-defined and available to consult during the training moment. After this session, a 3-hours of practical training was done with a one-to-one feedback by the experienced trainers. The participants evaluated 9 standardized patients.

Two weeks later, in the second step, participants completed the ultrasound measurement of ultrasound HMD in neutral head position and the performance was video recorded. Each participant evaluated ultrasound HMD in the same three standardized patients, generating 45 videos.



**Fig. 1** Study design. Step 1 - theoretical presentation and practical session; Step 2 - participants evaluation of HMD in standardized patients, videorecord; Step 3 - assessors' preparation to use OSAUS scale; Step 4 - evaluation of videorecord participants performance using OSAUS scale.

In step 3, six assessors were recruited and received online guidance on the OSAUS scale and its applicability for assessing ultrasound HMD measurement. A concrete preparation analyzing a pilot video was done.

In step 4, assessors blindly rated participants performance according to OSAUS and a global rating scale (GRS) with 5-Likert points (1 point - unacceptable; 2 points - weak; 3 points - acceptable; 4 points - very good; 5 points - excellent performance). The time needed to complete participants' task was also collected for further analysis. The assessors' evaluation was sent in an anonymous excel file.

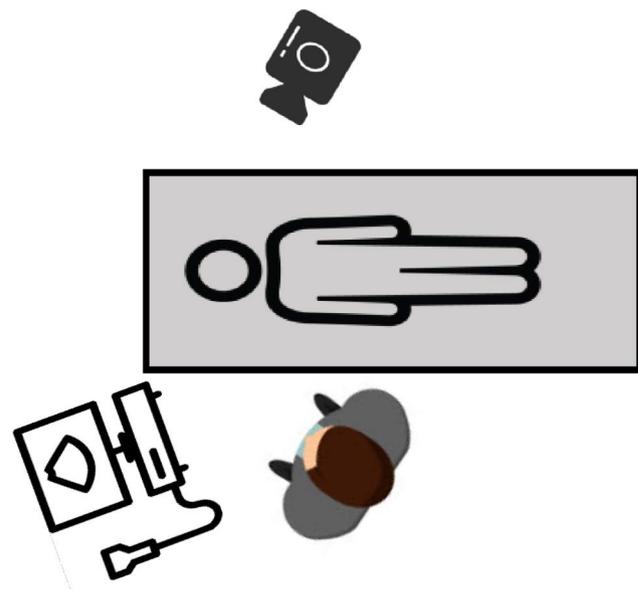
### Participants

Volunteer participants provided informed consent and self-reported their experience with airway ultrasound before their enrolment. According to participants experience, three categories were created: novices, intermediates and experts. A novice participant had no experience to up to six months of practice in airway ultrasound and expert has used airway ultrasound for more than two years. The intermediate group enrolled participants with experience within six months to two years.

Standardized patients (SP) volunteers: Nine SP were recruited for the study. All SP participated in the training moment and only three of the nine collaborate in the assessment time. All SP were women, between 20 and 25 years old, with healthy weight, a normal to low BMI, and with no previous history of difficult airway or neck deformation or scars. All provided informed consent before participating on the study.

### Equipment and environment

Steps 1 and 2 were performed at the School of Medicine, University of Minho, and the study equipment and environment were the same for the practical session and for the assessment time. Ultrasound measurements were obtained using a SonoSite<sup>®</sup>, portable ultrasound machine (Fujifilm, SonoSite<sup>®</sup> Edge II and SonoSite<sup>®</sup> SII, Ultrasound



**Fig. 2** Setting of the ultrasound workstation for training and videorecording. The camera was able to videorecord simultaneously participant's hands, standardized patient face and neck and the ultrasound screen

System, Inc Bothell, WA, USA), using a curvilinear, multifrequency 3–8 MHz ultrasound transducer probe.

### Procedure

Participant measurement of HMD in a neutral position was video recorded. The angle of video records provided a global overview of the technique, including the face of the SP; both hands of the sonographer and all ultrasound machine. (Fig. 2).

Participants were anonymized by not recording their faces or voices. Once the procedure was finalized, the film clips were stored and referenced by order of collection. An anonymous link to a folder with the videos was sent independently to each assessor.

## Methods

### Psychometric study

The analysis of OSAUS's psychometric properties was based on the category framework articulated by Messick [28, 29] and according to three domains (internal structure, relation to other variables and response process). The internal structure was analyzed by internal consistency and interrater reliability. The study of the relation to other variables domain included the criterion-related analysis, by comparing [1] OSAUS with GRS and [2] the performance score and [3] time to complete the task from novice-intermediate-expert groups. The response process domain of OSAUS scale evaluation focused on data collection methods; rater instructions, training and performance; how scores were stored and on strategies adopted for the lack of bias in the process [30].

### Statistical analysis

The internal consistency was assessed through Cronbach's alpha for each item used (items 2 – Applied knowledge of ultrasound equipment, 3 – Image optimization, 4 – Systematic examination and 5 – Interpretation of images). For each participant, we calculated the mean score of the three ultrasound measurements and the mean time in seconds to complete the task. Intra-class Correlation (ICC) estimates and its 95% confidence interval were calculated. The ICC two-way-random effect model was used to evaluate consistency between raters based on the mean value of the OSAUS score from raters ( $k=n$ ) [31]. An analysis of variance (ANOVA) for repeated measurements was also done. Convergent validity was assessed by a Pearson's correlation between OSAUS and Global Rating Scale.

One-way ANOVA explored differences in OSAUS rating scores and differences in time to complete the task between groups with different levels of expertise.

The statistical analysis was done using IBM SPSS Statistics version 29.0.0.0 (IBM Corp, Armonk, NY) with  $P$  values below 0.05 were interpreted as the statistical significance and the strength of agreement were interpreted according to Portney [32] where values under 0.5 represent poor reliability, values between 0.5 and 0.69 considered moderate, values between 0.7 and 0.9 indicate strong and over 0.9 represent excellent reliability.

## Results

### General results

#### Participants

Fifteen participants were enrolled on the study, 10 (66.6%) were female, and 5 (33.3%) were male. The mean age of participants was  $30\pm 4.6$ , mean $\pm$ SD, (min 25, max 39) years old. Each groups received 5 participants.

### Assessors

A total of 6 assessors were recruited for the study. One assessor was excluded due to an incomplete assessment.

### Psychometric study

#### Internal structure

**Internal consistency** The scale's internal consistency was achieved by analysing the use of each item of OSAUS scale ( $225=45$  videos \* 5 assessors). The internal consistency of OSAUS scale for ultrasound HMD measurement was evaluated with a Cronbach's alpha of 0.916.

**Inter-rater reliability** The inter-rater reliability was assessed using the Interclass Correlation Coefficient (model 2,5) when analysed the mean score of the three ultrasound measurements from each participant. The ICC was 0.720 (95%CI 0.408 to 0.893), with a significance level inferior to 0.001 ( $p<0.001$ ).

#### Relation to other variables

To explore evidence of validity concerning other variables, we compared OSAUS scale with GRS, and compared the OSAUS scores and time to complete the task across different experience levels (novice, intermediate and expert).

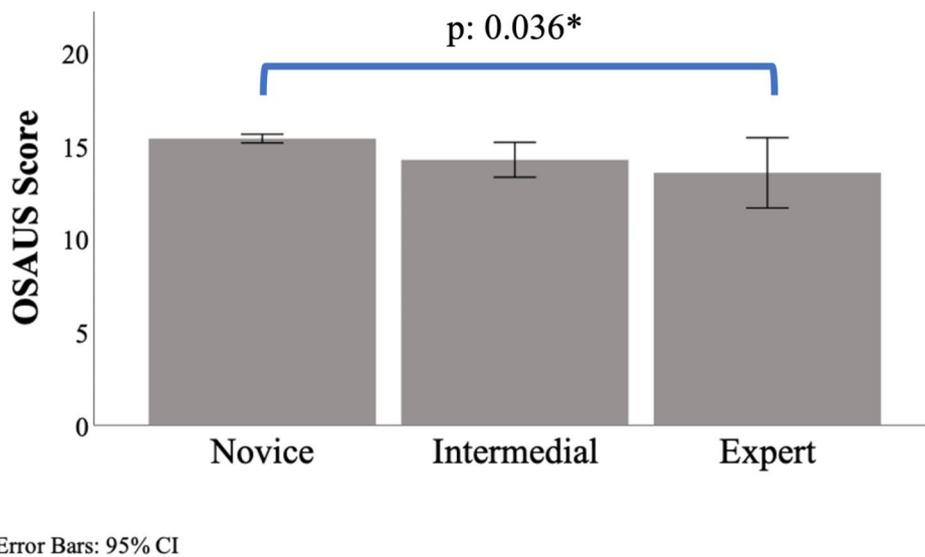
**OSAUS compared with GRS** OSAUS for HMD measurement was compared with a 5 points-Likert Global Rating Scale (1- poor performance; 5 – perfect performance). The correlation between OSAUS and GRS was studied using the mean score of the three ultrasound measurements done by each participant.

The correlation between OSAUS-HMD and GRS was  $r=0.970$  ( $p<0.001$ ; with 94% of shared variance,  $r^2=0.941$ ).

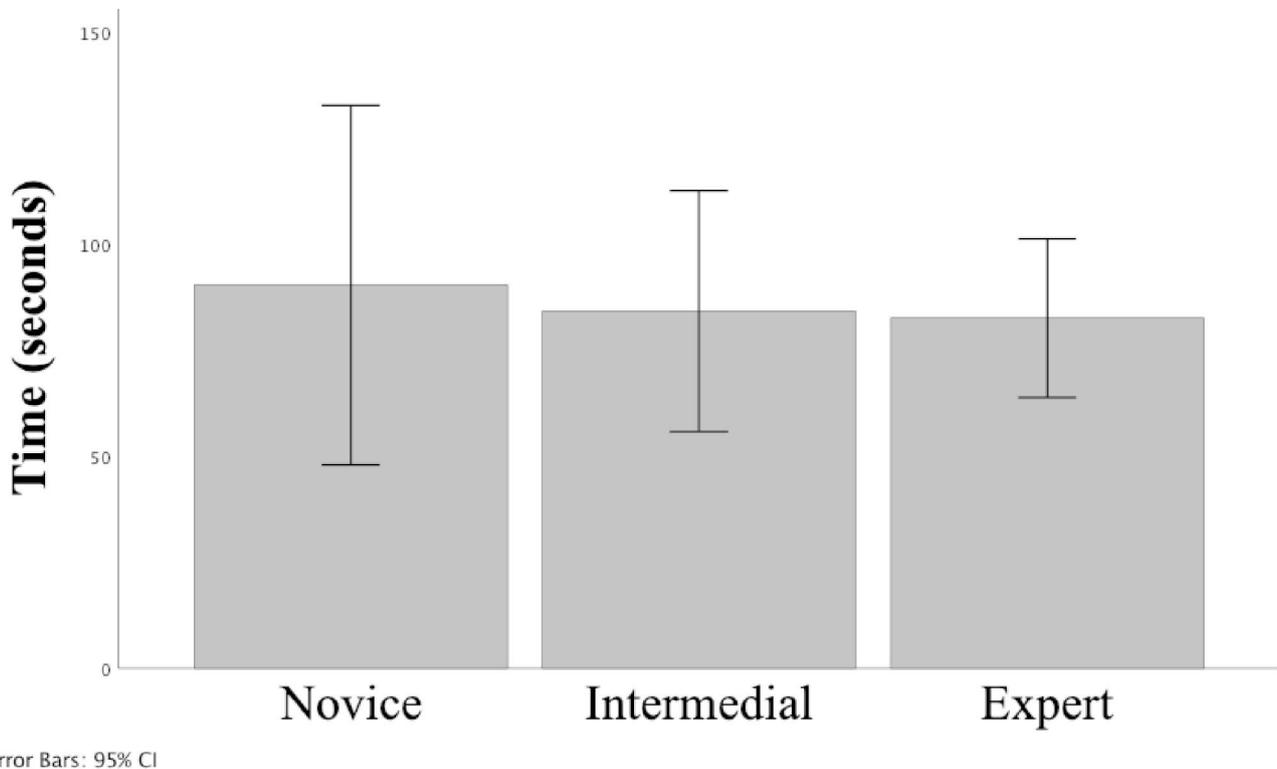
#### Group performance comparisons OSAUS score

The mean value of OSAUS-HMD scores for each group was evaluated considering the mean score of the three measurements from each participant. The novice group scored  $15.4\pm 0.18$ , mean $\pm$ SD (95%CI 15.1 to 15.6), the intermediate  $14.3\pm 0.75$ , mean $\pm$ SD (95%CI 13.0 to 14.90) and expert  $13.6\pm 0.125$ , mean $\pm$ SD (95%CI 12.2 to 16.1) (Fig. 3). One-way ANOVA showed significant differences between groups ( $F(2, 12)=4.422$ ,  $p=0.036$ ). The novice rated higher than the expert group (mean difference of 1.8), with a significant result ( $p=0.037$ , post hoc test, Bonferroni), and higher than intermediate (mean difference of 1.1), with no significant differences ( $p=0.279$ , post hoc teste, Bonferroni). The intermediate group rated a little higher than the expert group (mean difference of 0.7), with no significant differences between groups ( $p=0.853$ , post hoc teste, Bonferroni).

**Time to complete the task**



**Fig. 3** Mean OSAUS scores in each competency group. Values are presented as mean and error bars 95% CI. \*One-way ANOVA



Error Bars: 95% CI

**Fig. 4** Mean total time in seconds to complete the task in each competency group. Mean and error bars 95% CI for each competency group

In the analysis of the mean time to complete the task for each participant, the novice, the intermediate, and the expert group spent in mean±SD (95%CI) respectively, 90.4±34.2 (95%CI 47.9 to 132.8), 84.2 ±22.8 (95%CI 55.8 to 112.6) and 82.6 ±15.0 (95%CI 63.8 to 101.3) seconds. No significant differences were found between different

competency levels (F(2, 12)=0.133, p=0.877, one way ANOVA). (Fig. 4)

**Response process**

The participants performances were done individually, with no external interferences, and the video was

recorded without participant's faces or voices. The videos were encoded by acquisition time, which contributed to anonymity. With this approach, we were able to blind the raters.

The rater's instructions and training consisted of one-hour online session with authors, where the OSAUS scale was presented, and an intermediate performance video was analyzed and rated for training. All assessors agreed to use OSAUS to evaluate participants performance.

## Discussion

The global change for a competency-based medical education extended the need for good assessment tools that could accurately evaluate clinical performance [33–35]. Those tools need to be submitted to a rigorous process of validation before its implementation, so trainees and assessors thrust on the results [36–38]. The present study supports evidence of validity of OSAUS scale, when used to assessed ultrasound HMD, with a strong internal structure, adequate relation to other variables and consistent evidence in the response process domain.

The internal structure measures the “degree to which individual items fit the underlying construct of interest”. In our study its evaluation was based on each time an item was rated (45 videos \* 5 assessors=225 evaluations) [39]. A relevantly excellent internal-consistency was achieved (alfa-Cronbach 0.916) with no evidence of redundant items nor excessive scale length. As the authors excluded three items from OSAUS the internal consistency achieved in this study could be undervalued.

Inter-rater reliability (IRR) was evaluated using inter-class correlation coefficient – ICC [2, 5] with a model 2-way model random, since we selected five consistent raters from a larger possible population [31]. IRR was achieved with an ICC of 0.720 (95%CI 0.408 to 0.893), with a significance level ( $p < 0.001$ ), which reflects a strong agreement between evaluators.

According to the relation to other variables, we compared scores from OSAUS scale with GRS and the ability of the scale to discriminate performance between different groups. We achieved an excellent correlation ( $r = 0.970$ ,  $p < 0.001$ ) with Global Rating Scale. OSAUS scale could discriminate between novice and experts. We expected inverted results with experts gathering the highest scores compared with other groups. These results can be secondary to how participants reported their level of competence, and experts were more confident about their level of expertise. Second, the study was developed in a simulated environment away from the clinical reality, where experts were not so familiar with the settings. Nevertheless, they were voluntaries, experts were very difficult to recruit due to their full clinical work agenda and seemed less committed to the study. Those conditions could have compromised their performance.

Similar results were obtained from Alvarez-Lopez and co-workers, when used a low-cost 3D portable virtual simulator for skill training in minimally invasive surgery [40]. Simultaneously, the novice participants were at a relatively early stage in their residency program, so that their appetency to learn new techniques and increase their knowledge can explain the higher scores achieved by this group [40]. Although operator-dependent, airway ultrasound skills are easy to acquire, and even the less experienced can achieve performance accuracy with appropriated training. Two recent studies highlighted novice doctors' ability to quickly learn new technical skills related to ultrasound. Pratheb et al. [41] reported no significant differences in airway ultrasound when comparing novice anesthesia residents with experienced anesthesiologists. Oliveira et al. [42] demonstrated that the learning curve for novices to identify the cricothyroid membrane was relatively short even after a short training of 2 h, with a need of less than 20 scans to achieve competence. The comparison of scores from different competency groups does not represent an essential validity argument [41, 42]. Similarly with the Cook's study [43] several methodological problems can explain the observed differences between groups not related with the construct of the scale: i) lack of representativity of the population; ii) the novice group was the most homogeneous and iii) the group average score does not represent the individual performance.

Methods to obtain evidence about the response process are difficult to develop. A meta-analysis published by Beckman et co-workers [34] and a literature review published by Padilla and co-workers [30] reported the response process as one of the least represented sources of validity in clinical teaching and assessment tools. In our study, a consistent response process was guaranteed by i) assessor's participation in a training session; ii) assessors's approvement of the use of OSAUS scale to ultrasound HMD measurement; iii) participants performed individualized, with no external interferences; and iv) adequate quality and security control throughout the steps of the study blinding assessors and authors and reducing the risk for a halo effect.

The authors used the previously developed OSAUS scale to assess ultrasound hyomental distance measurement competence. This approach allowed us to compare our results with previous work and contribute to a boarder application of the scale.

OSAUS scale has been used as an assessment tool for head and neck ultrasonography (HNUS) by otolaryngologists. Todsen et co-workers [25] studied the diagnostic accuracy of surgeon who performed HNUS and stablished the evidence of validity of the OSAUS scale. Similarly with our study, the work recruited a small sample size, used 5 of the 7 items of OSAUS scale. Participants

were distributed in two different competence levels (otolaryngologists and interns with no experience in HNUS); and only two raters were invited. This study enrolled patients instead of healthy standardized patients. This approach allowed the authors to establish a correlation between the scale and the diagnostic accuracy. The study presented a good internal structure and discriminated different competence levels.

OSAUS scale was also used to assess competence of ultrasound fetal biometry [18–21]. All studies enrolled participants in two groups (novice and intermedial) according to the number of fetal US. The scale could discriminate the level of competence and the consequence domain was explored with a pass/fail score.

OSAUS scale was used as an assessment tool during the implementation of a simulation training in emergent Focused Abdominal Sonography in Trauma (eFAST) [22] and in transvaginal ultrasound [44]. The scale was also able to discriminate competence group with a good reliability in the assessment of point-of-care ultrasonography.

### Limitations

The study has some limitations that need to be considered the study didn't explore the content nor the consequences domains of the validity framework from Messick. The content of OSAUS Scale was validated in previous studies [17, 23]. The scale was developed after an international panel of multidisciplinary experts. Three Delphi rounds were necessary to develop the scale, representing a robust concern within its items. The consequence domain of Messick's framework includes the impact on examinee performance, examinee effects (such as anxiety, stress) and the definition of pass/fail standard [29, 33]. This source of evidence was not explored since the study was done in a simulated environment, far from the workplace and without any professional consequences from a pass/fail evaluation.

The final version of the OSAUS scale has seven elements, with two elements facultative (indication for the examination and medical decision making). Due to the design of this study, we excluded the facultative and the documentation item. The first item - Indication for the examination, is optional and in this study was not evaluated, since all participants were aware of the purpose of the study. The sixth item – Documentation of examination, intends an image recording and a focused verbal or written documentation. It was inappropriate since it could identify participants and consequently bias the results. The seventh item – Medical decision making is also optional and was out of the aim of the study.

Selection of level of competence groups. In our study, participants self-reported their competence level, which might introduce a bias. An external expert panel that

could rank participants by level of competency would be more appropriate. The study reported time-of-practice instead of the number-of-procedures criteria to define the three expertise groups. We think that competence in technical skills depends on much more than the number of procedures done, namely on the quality of formative feedback [45]. Nevertheless, experts were voluntary and participated in both moments, their recruitment was challenging, and they seemed to be less engaged with the study when compared with novices and the intermedial group.

Additionally, a portfolio report is not mandatory for consultants, so it would be very difficult to quantify the experience based on numbers instead of self-reported experience. The small sample size is also a limitation.

We use GRS with a single item, as an overall performance scale. Although it was used in order to simplify the process of assessment and to decrease the time spent by assessors to complete the task it could influence the results.

All participants did the assessment moment evaluating a SP already known from the training opportunity. Although the moments were separated by fifteen days, it might improve the scores of all participants, since they were familiar with SP sono-anatomy. Simultaneously, all SP had a thin neck with no dysmorphias. A way to solve this limitation was to consider SP more heterogenous in relation to neck morphology and weight, to assure that participants could do the measurement in both “easy and difficult” necks. This approach could contribute to establishing a pass/fail score and to the process of implementation of the scale in the workplace-based assessment.

### Conclusion

The study demonstrated a strong evidence of validity supporting OSAUS scale to assess HMD competence. The use of OSAUS scale should be integrated into the clinical setting for training and assessment of airway ultrasound competence.

### List of Abbreviations

OSAUS	Objective Structured Assessment Ultrasound Skill
PoCUS	Point-of-care ultrasound
HMD	Hyomental distance
GRS	Global rating scale
SP	Standardized patient
IRR	Inter-rater- reliability
ICC	Intraclass correlation Coefficient
US	Ultrasound

### Acknowledgements

Authors would like to thank all assessors: Dr. Robert Szabo, Dr Catalin Constantinescu and Dr Robert Simon from Anesthesia and Intensive Care 1 Department, Clinical Emergency County Hospital Cluj-Napoca, Romania; Prof. Dr Javier Onrubia from University Hospital Dr. Preset València, Spain; Dr. Patrícia Santos from Department of Anesthesiology, Hospital and University Centre S. João, Portugal and to Dra Ana Isabel Pereira from Department of Anesthesiology, Hospital Vila Nova de Gaia/Espinho, Portugal; Prof Dr. António Manuel Melo from School of Medicine, University of Minho. All Colleagues

from the Department of Anesthesiology, Hospital of Braga, Portugal, specially to Dr Joana Veiga for the support with the experiments. Thanks to FujiFim (Sonosite®) for the support with the ultrasound and probes.

#### Author Contribution

All authors contributed substantially to the study. Conceived and design the experiments: SHG, MT, JMP, JCP, PSC. Performed the experiments: SHG, MT, DC. Analyzed the data: SHG, JMP, JCP, PSC. In depth revising the manuscript critically for important intellectual content: SHG, CP, JMP, JCP, PSC. Final approval of the version to be published and all agree to be accountable for all aspects of the work thereby ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: SHG, MT, CP, DC, JMP, JCP, PSC.

#### Funding

This work was supported by National funds, through the Foundation for Science and Technology (FCT) – project UIDB/50026/2020 and UDP/50026/2020.

#### Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to the anonymization of the data, but are available from the corresponding author on reasonable request.

#### Declarations

##### Ethical approval

was obtained from the Ethical Committee for Institute of Life and Health Sciences (CEICVS) of School of Medicine, University of Braga, Braga, Portugal (Chairperson Prof. Dr Cecília Leão) on 15th November 2020 (ethical approval number - CEICVS15/2020). The study was done in accordance with Helsinki declaration. All participants gave written informed consent to participate in the study.

##### Consent for publication

Not applicable.

##### Conflict of Interest

The authors declare that the study was conducted in absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 9 June 2022 / Accepted: 9 March 2023

Published online: 22 March 2023

#### References

- Li L, Yong RJ, Kaye AD, Urman RD. Perioperative Point of Care Ultrasound (POCUS) for Anesthesiologists: an Overview. *Curr Pain Headache Rep.* 2020; 21;24(5):20. doi: <https://doi.org/10.1007/s11916-020-0847-0>.
- Gottlieb M, Holladay D, Burns KM, Nakitende D, Bailitz J. Ultrasound for airway management: an evidence-based review for the emergency clinician. *Am J Emerg Med.* 2020;38(5):1007–13. <https://doi.org/10.1016/j.ajem.2019.12.019>.
- Petrisor C, Szabo R, Constantinescu C, Prie A, Hagau N. Ultrasound-based assessment of hyomental distances in neutral, ramped, and maximum hyper-extended positions, and derived ratios, for the prediction of difficult airway in the obese population: a pilot diagnostic accuracy study. *Anaesthesiol Intensive Ther.* 2018;50(2):110–6. <https://doi.org/10.5603/AIT.2018.0017>.
- Wu J, Dong J, Ding Y, Zheng J. Role of anterior neck soft tissue quantifications by ultrasound in predicting difficult laryngoscopy. *Med Sci Monit.* 2014;18(20):2343–50. <https://doi.org/10.12659/MSM.891037>.
- Reddy PB, Punetha P, Chalam KS. Ultrasonography - A viable tool for airway assessment. *Indian J Anaesth.* 2016;60(11):807–13. <https://doi.org/10.4103/0019-5049.193660>.
- Adhikari S, et al. Pilot study to determine the utility of point-of-care ultrasound in the assessment of difficult laryngoscopy. *Acad Emerg Med.* 2011;18(7):754–8. <https://doi.org/10.1111/j.1553-2712.2011.01099.x>.
- Andruszkiewicz P, Wojtczak J, Sobczyk D, Stach O, Kowalik I. Effectiveness and validity of Sonographic Upper Airway evaluation to predict difficult Laryngoscopy. *J Ultrasound Med.* 2016;35(10):2243–52. <https://doi.org/10.7863/ultra.15.11098>.
- Gomes SH, et al. Useful Ultrasonographic Parameters to predict difficult laryngoscopy and difficult tracheal Intubation-A systematic review and Meta-analysis. *Front Med.* 2021;28:8:671658. <https://doi.org/10.3389/fmed.2021.671658>.
- Moore CL, Copel JA. Point-of-care ultrasonography. *N Engl J Med.* 2011;24(8):749–57. <https://doi.org/10.1056/NEJMr0909487>.
- Stolz LA, et al. Emergency Medicine Resident Assessment of the emergency Ultrasound Milestones and current training recommendations. *Acad Emerg Med.* 2017;24(3):353–61. <https://doi.org/10.1111/acem.13113>.
- Vives M, et al. Consensus document of the Spanish Society of Anesthesia (SEDAR), Spanish Society of Internal Medicine (SEMI) and Spanish Society of Emergency Medicine (SEMES). *Rev Esp Anestesiol Reanim.* 2021;68(3):143–8. <https://doi.org/10.1016/j.redar.2020.06.020>.
- Cazes N, Desmots F, Geffroy Y, Renard A, Leyral J, Chaumoitre K. Emergency ultrasound: a prospective study on sufficient adequate training for military doctors. *Diagn Interv Imaging.* 2013, 94(11):1109–15. doi: <https://doi.org/10.1016/j.diii.2013.04.016>. Epub 2013 Aug 6. PMID: 23928178
- Greaves JD, Grant JW. Watching anesthesiologists work: using the professional judgement of consultants to assess the developing clinical competence of trainees. *Br J Anesth.* 2000;84:525–33.
- Ellegaard K, et al. Feasibility of a standardized ultrasound examination in patients with rheumatoid arthritis: a quality improvement among rheumatologists' cohort. *BMC Musculoskelet Disord.* 2012;12:13:35. <https://doi.org/10.1186/1471-2474-13-35>. PMID: 22410241; PMCID: PMC3414749.
- Sultan SF, Iohom G, Saunders J, Shorten G. A clinical assessment tool for ultrasound-guided axillary brachial plexus block. *Acta Anaesthesiol Scand.* 2012;56(5):616–23. <https://doi.org/10.1111/j.1399-6576.2012.02673.x>.
- Chin KJ, Tse C, Chan V, Tan JS, Lupu CM, Hayter M. Hand motion analysis using the imperial college surgical assessment device: validation of a novel and objective performance measure in ultrasound-guided peripheral nerve blockade. *Reg Anesth Pain Med.* 2011;36(3):213–9. <https://doi.org/10.1097/AAP.0b013e31820d4305>.
- Tolsgaard MG, et al. International multispecialty consensus on how to evaluate ultrasound competence: a Delphi consensus survey. *PLoS ONE.* 2013;8(2):e57687. <https://doi.org/10.1371/journal.pone.0057687>.
- Ambroise Grandjean G, Berveiller P, Hossu G, Bertholdt C, Judlin P, Morel O. Évaluation des compétences pour la pratique de la biométrie échographique foetale: validation prospective du score OSAUS METHOD. *Gynecol Obstet Fertil Senol.* 2021;49(4):275–81. <https://doi.org/10.1016/j.gofs.2021.01.010>.
- Ambroise Grandjean G, Gabriel P, Hossu G, Zuily S, Morel O, Berveiller P. Apprentissage de la biométrie échographique foetale: évaluation prospective. *Gynecol Obstet Fertil Senol.* 2020;48(11):800–5. <https://doi.org/10.1016/j.gofs.2020.05.009>. de la performance de l'Objective Structured Assessment of Ultrasound Skills (OSAUS).
- Grandjean GA, et al. Fetal biometry in ultrasound: a new approach to assess the long-term impact of simulation on learning patterns. *J Gynecol Obstet Hum Reprod.* 2021;50(8):102135. <https://doi.org/10.1016/j.jogoh.2021.102135>.
- Tolsgaard MG, et al. Reliable and valid assessment of ultrasound operator competence in obstetrics and gynecology. *Ultrasound Obstet Gynecol.* 2014;43(4):437–43. <https://doi.org/10.1002/uog.13198>.
- Pencil K. eFAST Simulation Training for Trauma Providers. *J Trauma Nurs.* 2017;24(6):376–80. <https://doi.org/10.1097/JTN.0000000000000329>.
- Todsen T, et al. Reliable and valid assessment of point-of-care ultrasonography. *Ann Surg.* 2015;261(2):309–15. <https://doi.org/10.1097/SLA.0000000000000552>.
- Di Pietro S, et al. Lung-ultrasound objective structured assessment of technical skills (LUS-OSAUS): utility in the assessment of lung-ultrasound trained medical undergraduates. *J Ultrasound.* 2021;24(1):57–65. <https://doi.org/10.1007/s40477-020-00454-x>.
- Todsen T, et al. Competency-based assessment in surgeon-performed head and neck ultrasonography: a validity study. *Laryngoscope.* 2018;128(6):1346–52. <https://doi.org/10.1002/lary.26841>.
- Primdahl SC, Todsen T, Clemmesen L, Knudsen L, Weile J. Rating scale for the assessment of competence in ultrasound-guided peripheral vascular access - a Delphi Consensus Study. *J Vasc Access.* 2016;21(5):440–5. <https://doi.org/10.5301/jva.5000581>.
- Wong DM, et al. Evaluation of a task-specific checklist and global rating scale for ultrasound-guided regional anesthesia. *Reg Anesth Pain Med.* 2014;39(5):399–408. <https://doi.org/10.1097/AAP.0000000000000126>.

28. American Educational Research Association and American Psychological Association and National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington: Report of a Joint Committee; 2014.
29. Messick S. "Validity". In: R. L. Linn, editor. *Educational Measurement*. 3rd edition. New York: MacMillan (1988). p. 1-165.
30. Padilla JL, Benítez I. Validity evidence based on response processes. *Psicothema*. 2014;26(1):136–44. <https://doi.org/10.7334/psicothema2013.259>.
31. Koo TK, Li MY. A Guideline of selecting and reporting Intraclass correlation coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155–63. <https://doi.org/10.1016/j.jcm.2016.02.012>.
32. Leslie G, Portney. "Concepts of Measurement". In: *Foundation of Clinical Research: applications to evidence-based practice*. Fourth edition. Philadelphia:FA. Davis Company (2020). p. 106–166.
33. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul*. 2016;1:31. <https://doi.org/10.1186/s41077-016-0033-y>.
34. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med*. 2005;20(12):1159–64. <https://doi.org/10.1111/j.1525-1497.2005.0258.x>.
35. Everett TC, et al. Simulation-based assessment in anesthesia: an international multicentre validation study. *Can J Anaesth*. 2019;66(12):1440–9. <https://doi.org/10.1007/s12630-019-01488-4>. English. doi:
36. Borgersen NJ, et al. Gathering validity evidence for Surgical Simulation: a systematic review. *Ann Surg*. 2018;267(6):1063–8. <https://doi.org/10.1097/SLA.0000000000002652>.
37. Zendejas B, Ruparel RK, Cook DA. Validity evidence for the Fundamentals of laparoscopic surgery (FLS) program as an assessment tool: a systematic review. *Surg Endosc*. 2016;30(2):512–20. <https://doi.org/10.1007/s00464-015-4233-7>.
38. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Acad Med*. 2013;88(6):872–83. <https://doi.org/10.1097/ACM.0b013e31828ffdcf>.
39. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011;27:2:53–5. <https://doi.org/10.5116/ijme.4dfb8dfd>.
40. Alvarez-Lopez F, Maina MF, Arango F, Saigi-Rubió F. Use of a low-cost portable 3D virtual reality Simulator for Psychomotor Skill Training in minimally invasive surgery: Task Metrics and score validity. *JMIR Serious Games*. 2019;8(427). <https://doi.org/10.2196/19723>.
41. Prathep S, Sikaretrakul P, Nimmaanrat S, Juthasantikul W, Karnjanawanichkul O, Kovitwanawong N. Improvement in Validity and Reliability of Airway Assessment Using Ultrasonography by Novice Personnel: A Learning Cohort Study. *Advances in medical education and practice*. 2022; 13: 1103–1111. <https://doi.org/10.2147/AMEPS377403>
42. Oliveira KF, Arzola C, Ye XY, Clivatti J, Siddiqui N, You-Ten KE. Determining the amount of training needed for competency of anesthesia trainees in ultrasonographic identification of the cricothyroid membrane. *BMC Anesthesiol*. 2017;17(1):74. <https://doi.org/10.1186/s12871-017-0366-7>.
43. Cook DA. Much ado about differences: why expert-novice comparisons add little to the validity argument. *Adv Health Sci Educ Theory Pract*. 2015;20(3):829–34. <https://doi.org/10.1007/s10459-014-9551-3>.
44. Byford S, Janssens S, Cook R. Implementing the transvaginal ultrasound simulation training (TRUSST) programme for obstetric registrars. *Adv Simul*. 2021;6(1):1. <https://doi.org/10.1186/s41077-020-00152-y>.
45. Schott CK, Kode KR, Mader MJ. Teaching vs learning: impact of deliberate practice and formative feedback on developing point of care ultrasound skills. *J Clin Ultrasound*. 2020;48(8):437–42. <https://doi.org/10.1002/jcu.22878>.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.