Objective Structured Assessment Ultrasound Skill Scale for Hyomental distance competence – psychometric study

Sara Hora Gomes (saragomes@med.uminho.pt)
Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho
Marta Trindade
Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho
Cristina Petrisor
Anesthesia and Intensive Care II Department, "Iuliu Hatieganu" University of Medicine and Pharmacy Cluj-Napoca
Dinis Costa
Department of Anesthesia, Hospital de Braga
José M. Pêgo
Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho
Jorge Correia-Pinto
Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho
Patrício S. Costa
Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho

Research Article

Keywords: OSAUS, hyomental, distance, airway, ultrasound, assessment, psychometric properties

Posted Date: June 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1739820/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Medical Education on March 22nd, 2023. See the published version at https://doi.org/10.1186/s12909-023-04146-y.

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 for different medical fields. The 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. Fifteen volunteers were recruited and divided into three groups depending on participants expertise in airway ultrasound. Each participant performed ultrasonographic HMD evaluation, with the patient in head neutral and extended position, in 3 standardized patients. The performance was videorecord 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: 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±0.08 (mean±SD), the intermediate 14.3±0.33 and expert 13.5±0.68, with a significant difference between novice and expert groups (p=0.036). The time in seconds to complete the task was evaluated: novice (90.40±34.15) (mean±SD), intermediate (84.20±22.88) and experts (82.60±15.07), with no significant differences between groups. A strong correlation was observed between OSAUS and global rating scale (r=0.970, p<0.001).

Conclusions: The study demonstrated evidence of validity. Further studies are needed to move forward to implement OSAUS scale in the clinical setting for training and assessment of airway ultrasound competence.

1. Background

In the last decades, the use of ultrasonography expanded from the laboratory to the patient bedside evaluation. Nowadays, almost all clinical 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 cardiac, lung, gastric, abdominopelvic, regional anesthesia, vascular access and airway ultrasound evaluation. (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, 10, 11) However, there is no evidence-based guidelines for education or evaluation of airway ultrasound assessment.

Traditionally, skills competence was achieved after a subjective evaluation by 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, 13, 14) In the last two decades, many valid and reliable instruments were developed in order to improve the objectivity of assessment through the creation of checklists and global rating scales. (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 gastro-enterology) 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 in trauma (eFAST) (22, 23), lung (24), head and neck (25) ultrasound.

Global rating scales (GRS) have been globally used in medical education and in clinical training to assess competency in many technical skills. (15, 26, 27). In the present study, GRS was used as an "overall performance" scale based on 5-points of Likert.

This study explores the psychometric properties of OSAUS scale for hyomental ultrasound distance in neutral and extended position measurement.

2. Methods

2.1 Materials

2.1.1 Ethical approval

The study was carried out in accordance with 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 (Chairperson Prof. Dr Cecília Leão) on 15th November 2020 (CEICVS15/2020) (available Supplementary material Ethical file). Participation in this study was absolutely voluntary and all participants and assessors gave an oral and written informed.

In addition, all were aware that they could leave the study at any time without any problem.

2.1.2 Study dates

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

2.1.3 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 in head neutral and extended position.

Figure 1 represents the study design (Figure 1).

The study has developed in 4 steps. In the first step an educational moment was organized with a theoretical presentation of 1 hour. Two experienced airway ultrasound anesthesiologists presented the OSAUS scale (17) and its applicability to measuring ultrasound hyomental distance in neutral and extended positions. The protocol for HMD measurement in neutral and extended position was very well-defined and given to participants and was available for consult as a guide for practice. After this session, 3 hours of practical training of the HMD ultrasound measurements were done, tutorized by the same trainers.

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

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 study was done.

In step 4, assessors blindly rated participants videorecords 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). For further analysis, the time needed to complete participants' tasks was also collected. The assessors' evaluation was sent in an anonymous excel file.

2.1.4 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 up to six months experience in airway ultrasound, and an expert uses airway ultrasound for more than two years. The intermediate group enrolled participants within the two groups.

Standardized patients (SP) volunteers: Nine SP were recruited for the study and participated in both moments (training and validation) and provided informed consent before participating on the study.

2.1.5 Equipment and environment

Steps 1 and 2 were realized at School of Medicine, University of Minho, and the study equipment and environment was the same for the practical session and the assessment time. Ultrasound measurements were obtained using a SonoSite®, portable ultrasound machine (Fujifilm, SonoSite® Edge II and SonoSite® SII, Ultrasound System, Inc Bothell, WA, USA), using a curvilinear, multifrequency 3-8 MHz ultrasound transducer probe.

2.1.6 Procedure

Participant measurement of HMD in neutral and extended head 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 the all ultrasound machine. In all recordings, both the practitioner’s technique and the ultrasound were visible.

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

2.2 Methods

2.2.1 Psychometric study

The analysis of OSAUS’s psychometric properties was based on the Standards for Educational and Psychological Testing (American Educational Research Association - AERA, American Psychological Association & National Council on Measurement in Education, 2014)(28), following the category framework articulated by Messick (29) (content, internal structure, relation to other variables, response process and consequences).

The internal structure or construct validity was analyzed by internal consistency and interrater reliability.

The relation to other variables included the criterion-related analysis (concurrent validity), by comparing OSAUS with GRS, where raters were instructed to rate the overall participant´s performance and by comparing time to complete the task from different experience levels (expert, intermediate and novice).

The response process of OSAUS scale evaluation focused on data collection methods; rater instructions, training and performance; how scores were reported and summarized and in the methods responsible for the lack of bias in the process.(30)

2.2.2 Statistical analysis

The internal consistency was assessed through Cronbach's alpha for each item used (items 2, 3, 4 and 5). For each participant, we calculated the mean score of the 3 ultrasound measurements and the mean time in seconds to complete the task. Intraclass Correlation (ICC) estimates and their 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 5 raters (k=5), consistency, 2-way random effects model. (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 different levels of competence.

The statistical analysis was done using SPSS version 27 (IBM Corp, Armonk, NY) with P values below 0.05 were interpreted as statistical significance and the strength of agreement were interpreted according to *Portney* and co-workers (31) where values under 0.5 represent poor reliability, values between 0.5 to 0.69 considered moderate, values between 0.7 to 0.9 indicate strong and over 0.9 represent excellent reliability.

3. Results

3.2 General results

3.2.1 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±4.6, mean±SD, (min 25, max 39) years old.

3.2.2 Assessors

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

3.3 Psychometric Study

3.3.1 Content

The content analysis was not done since the development of OSAUS scale were done by Tolsgaard and co-workers (17). Due to the design of this study, the items: (1) Indication for the examination; (2) Documentation of the examination and (3) Medical decision making, were not adequate to investigate in the study. The item - Indication for the examination, is optional and in this study was an unvalued, since all participants were aware of the purpose of the study. The fifth item – Documentation of examination intends an image recording and a focused verbal or written documentation, so authors excluded this item since the participants could be identified by assessors and bias the results. The sixth item – Medical decision making is also optional and was out of the aim of the study.

3.3.1.1 Internal Consistency

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

3.3.1.2 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 was 0.720 (95%CI 0.408 to 0.893), with a significance level inferior to 0.001 ($p<0.001$).

3.3.2 Relation to other variables

To explore validity evidence for relations to 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).

3.3.2.1 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, $r^2=0.941$.

3.3.2.2 Group performance comparations

3.3.2.2.1 Score performance

The mean value of OSAUS-HMD scores for each group was evaluated during the analysis of the mean of 3 ultrasound measurement from each participant (15). The novice group scored 15.5±0.98, mean±SD

(95%CI 15.1 to 15.6), the intermediate 14.3±0.33, mean±SD (95%CI 13.0 to 14.90) and expert 13.5±0.68, mean±SD (95%CI 12.2 to 16.1). One-way ANOVA showed no significant differences between groups ($F(2,12)=4.42$, $p=0.036$). The novice group rated higher than the intermediate (mean difference of 1.14) and even higher compared with the expert group (mean difference of 1.84), with a significant result. The intermediate group rated a little higher than the expert group (mean difference - 0.7) also with no significant differences between groups. (Figure 2)

3.3.2.2.2 Time to complete the task

In the analysis of the mean time to complete the task for each participant (15), the novice, the intermediate, and the expert group spent in mean±SD (95%CI) respectively, 90.4±34.1 (95%CI 47.9 to 132.8), 84.2 ±22.88 (95%CI 55.78 to 112.61) and 82.6 ±15.07 (95%CI 63.88 to 101.31) seconds to complete the task. No significant differences were found between different competency levels ($F(2,12)=0.133$, $p=0.877$, one way ANOVA). Novice group spend more time to complete the task when compared with the intermediate group (mean difference of 6.2 seconds) and when compared with the expert group (mean difference of 7.8 sec), with no significant difference between groups. The same was observed for the intermediate group when compared with the experts (mean difference of 1.6 seconds) with no significant differences between groups. (Figure 3)

3.3.3 Response process

The participants performances were done individually, with no external interferences, and the video was recorded without participant’s faces or voices. With this approach we were able to blind the ratters. The videos were encoded by acquisition time.

The rater instructions and training consisted of an online meeting session with authors, where the OSAUS scale used for HMD measurement was presented, and an intermediate performance video was analyzed. All assessors agreed to use OSAUS to evaluate participants performance in ultrasound HMD measurement in neutral and extended head position.

4. Discussion

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

Documenting evidence of validity and reliability are essential for any new educational tool before its implementation, in order to have confidence in the collected data. (30) In this study, we used the content, internal structure, relation to other variables and response process domains of validity evidence described by Messick (29). Because no social implication or impact of participant’s scores was inherent to this study, the consequence domain was not investigated.

The content of OSAUS Scale was validated in the aforementioned study.(17) Three Delphi rounds were necessary to develop the scale, which represents a robust concern within its items.

The final version of the OSAUS scale has 7 elements, each with a 5-points Likert score. Due to the design of this study, items number 1, 6 and 7 were excluded from the analysis. The indication for the examination was clearly defined at the beginning of the study, so it was not adequate to evaluate this item. The examination of documentation was not done, since it could contribute to identifying participants, compromising the response process and the medical decision-making category is out of the aim of this study.

The evidence of reliability was documented through internal consistency and inter-rater reliability. The internal structure measures the “degree to which individual items fit the underlying construct of interest” (32). In our study its evaluation was based on each time an item was rated (45 videos x 5 assessors=225 evaluations)(33). A relevant internal validity was achieved (alfa-Cronbach 0.916) with no evidence of redundant items nor excessive scale length. Three elements from OSAUS were excluded from the original scale in our study: (1) “Indication for the examination”, since all participants knew the purpose of the study; (2) “documentation”, to ensure anonymity and (3) “medical decision making”, since the study was done in a simulated environment without any clinical implications. The internal consistency achieved in this study is undervalued, since not all items were used in the present study.

Inter-rater reliability was evaluated using interclass correlation coefficient – ICC (2,5) with a model 2-way model random, since we selected 5 consistent ratters from a larger possible population. (32,34) A good reliability 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 correlation and agreement between measurements.

According to the relation to other variables, convergent evidence was achieved by a strong correlation ($r=0.970$, $p<0.001$) with Global Rating Scale with 5-points Likert.

The OSAUS scale was able to discriminate between novice and experts. We expected that experts members had higher scores than the rest of the groups, but experts had the lowest values. These results can be explained by the fact that participants self-reported their level of experience in airway ultrasound based on experienced-time. A previous objective evaluation by an independent expert panel could distribute participants differently. The study was developed in a simulated environment away from the clinical reality, so experts were not familiar with the settings nor commitment with the study, which could have compromised their performance. (35) 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.(36,37)

The comparation of scores from different competency groups does not represent an essential validity argument.(38-40) Similarly with the Cook’s study (41) several methodological problems can explain the observed differences between groups not related with the construct of the scale: (1) lack of representativity of the population; (2) the novice group was the most homogenous and the group average score does not represent the individual performance and

Although no significant differences were found in the total time to complete the task, experts performed faster than intermediate and novice groups due to their familiar use of the ultrasound equipment.

Methods to obtaining evidence about the response process are difficult to develop. A meta-analysis published by Beckman et co-workers (32) and a literature review published by Padilla and co-workers (42) reported response process as one of the least represented sources of validity in clinical teaching and assessment tools.

In our study, a valid response process was guaranteed since: (1) all assessors participated in a training session; (2) assessors approved on the application of OSAUS scale to ultrasound HMD measurement after a training session; (3) participants performed individualized, with no external interferences; and (4) adequate quality and security control throughout the steps of the study blinding assessors and authors, reducing the risk for a halo effect.

4.1 Limitations

Our study has some limitations that need to be considered: (1) The different competence groups were done following participant's experience based on time of practice. Although this evaluation could be similar to the number of procedures done, an external expert panel that could rate participants by level of competency could be more adequate. (2) 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. (3) The consequence domain of the Messick framework was not explored in this study, nevertheless our results express that this tool could be used in the in-training program. (4) Since this is the earliest study using OSAUS scale for airway ultrasound competence, it was not possible to compare with the methodology of similar studies.

5. Conclusion

The study demonstrated significant results in internal structure, relation to other variables and response process. Further studies are needed to complement the Messick's framework for validity. The OSAUS scale should be used as assessment tool in other airway ultrasound parameters. Finally, we need to move forward and implement the use of OSAUS scale in the clinical setting for training and assessment of airway ultrasound competence.

6. Abbreviations

OSAUS – Objective Structured Assessment Ultrasound Skill

PoCUS – Point-of-care ultrasound

HMD – Hyomental distance

GRS – Global rating scale

SP – Standardized patient

7. Declarations

Ethical declaration

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) (Related files Document). The study was done in accordance with Helsinki declaration. All participants gave written informed consent to participate in the study.

Availability of data and materials

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.

Competing interests

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.

Funding

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

Author's contribution

All authors contributed substantially to the study. Conceived and design the experiments: SHG, MT, JMP, JCP, PC. 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 countable 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.

Acknowledgements

Authors would like to thank to 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, Romenia; 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. 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.

8. References

1. Li L., Yong R.J., Kaye A.D., Urman R.D. Perioperative Point of Care Ultrasound (POCUS) for Anesthesiologists: an Overview. Curr Pain Headache Rep. 2020, 21;24(5):20. doi: 10.1007/s11916-020-0847-0.
2. Gottlieb M., Holladay D., Burns K.M., Nakitende D., Bailitz J. Ultrasound for airway management: An evidence-based review for the emergency clinician. Am J Emerg Med. 2020, 38(5):1007–1013. doi: 10.1016/j.ajem.2019.12.019
3. Petrisor C., Szabo R., Constantinescu C., Prie A., Hagau N. Ultrasound-based assessment of hyomental distances in neutral, ramped, and maximum hyperextended 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):110–116. doi: 10.5603/AIT.2018.0017.
4. 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. doi: 10.12659/MSM.891037
5. Reddy P.B., Punetha P., Chalam K.S. Ultrasonography - A viable tool for airway assessment. Indian J Anaesth. 2016, 60(11):807–813. doi: 10.4103/0019-5049.193660
6. 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. doi: 10.1111/j.1553-2712.2011.01099.x.
7. 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. doi: 10.7863/ultra.15.11098
8. Gomes S.H., 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. doi: 10.3389/fmed.2021.671658.
9. Moore C.L., Copel J.A. Point-of-care ultrasonography. N Engl J Med. 2011, 24;364(8):749–57. doi: 10.1056/NEJMra0909487.
10. Stolz L.A., et al. Emergency Medicine Resident Assessment of the Emergency Ultrasound Milestones and Current Training Recommendations. Acad Emerg Med. 2017, 24(3):353–361. doi: 10.1111/acem.13113.

11. 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):143–148. doi: 10.1016/j.redar.2020.06.020.
12. Cazes N., Desmots F., Geffroy Y., Renard A., Leyral J,. Chaumoître K. Emergency ultrasound: a prospective study on sufficient adequate training for military doctors. Diagn Interv Imaging. 2013, 94(11):1109-15. doi: 10.1016/j.diii.2013.04.016. Epub 2013 Aug 6. PMID: 23928178
13. Greaves J.D,. Grant J.W. Watching anesthesiologists work: using the professional judgement of consultants to assess the developing clinical competence of trainees. Br J Anesth. 2000, 84:525–33.
14. 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. doi: 10.1186/1471-2474-13-35. PMID: 22410241; PMCID: PMC3414749.
15. Sultan S.F., 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. doi: 10.1111/j.1399-6576.2012.02673.x.
16. Chin K.J.., Tse C., Chan V, Tan J.S., Lupu C.M., 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. doi: 10.1097/AAP.0b013e31820d4305.
17. Tolsgaard M.G., et al. International multispecialty consensus on how to evaluate ultrasound competence: a Delphi consensus survey. PLoS One. 2013, 8(2):e57687. doi: 10.1371/journal.pone.0057687.
18. Ambroise Grandjean G., Berveiller P., Hossu G., Bertholdt C.., Judlin P, Morel O. Évaluation des compétences pour la pratique de la biométrie échographique fœtale: validation prospective du score OSAUS METHOD. Gynecol Obstet Fertil Senol. 2021, 49(4):275–281. doi: 10.1016/j.gofs.2021.01.010.
19. Ambroise Grandjean G., Gabriel P., Hossu G., Zuily S., Morel O., Berveiller P. Apprentissage de la biométrie échographique fœtale: évaluation prospective de la performance de l’Objective Structured Assessment of Ultrasound Skills (OSAUS). Gynecol Obstet Fertil Senol. 2020, 48(11):800–805. doi: 10.1016/j.gofs.2020.05.009.
20. Grandjean G.A,. 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. doi: 10.1016/j.jogoh.2021.102135.
21. Tolsgaard M.G., et al. Reliable and valid assessment of ultrasound operator competence in obstetrics and gynecology. Ultrasound Obstet Gynecol. 2014, 43(4):437–43. doi: 10.1002/uog.13198.
22. Pencil K. eFAST Simulation Training for Trauma Providers. J Trauma Nurs. 2017, 24(6):376–380. doi: 10.1097/JTN.0000000000000329.
23. Todsen T., et al. Reliable and valid assessment of point-of-care ultrasonography. Ann Surg. 2015, 261(2):309–15. doi: 10.1097/SLA.0000000000000552.

24. 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. doi: 10.1007/s40477-020-00454-x.

25. Todsen T., et al. Competency-based assessment in surgeon-performed head and neck ultrasonography: A validity study. Laryngoscope. 2018, 128(6):1346–1352. doi:10.1002/lary.26841.

26. Primdahl S.C., 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;17(5):440–445. doi: 10.5301/jva.5000581.

27. Wong D.M., 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. doi: 10.1097/AAP.0000000000000126.

28. American Educational Research Assocition and American Psychological Association and National Council on Measurement in Education. Standards for Educational and Psychological Testing. Report of a Joint Committee, Washington; 2014.

29. Messick S. "Validity". In: R. L. Linn, editor. Educational Measurement. 3rd edition. New York: MacMillan (1988). p. 1-165.

30. Leslie G. Portney. "Concepts of Measurement". In: Foundation of Clinical Research: applications to evidence-based practice. Fourth edition. Philadelphia:F.A. Davis Company (2020). p. 106–166.

31. Buijs M.M., et al. Intra and inter-observer agreement on polyp detection in colon capsule endoscopy evaluations. United European Gastroenterol J. 2018, 6(10):1563–1568. doi: 10.1177/2050640618798182.

32. Beckman T.J., Cook D.A., Mandrekar J.N. What is the validity evidence for assessments of clinical teaching? J Gen Intern Med. 2005, 20(12):1159–64. doi: 10.1111/j.1525-1497.2005.0258.x.

33. Tavakol M., Dennick R. Making sense of Cronbach's alpha. Int J Med Educ. 2011, 27;2:53–55. doi: 10.5116/ijme.4dfb.8dfd.

34. Koo T.K., Li M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016, 15(2):155–63. doi: 10.1016/j.jcm.2016.02.012.

35. Everett T.C., et al. Simulation-based assessment in anesthesia: an international multicentre validation study. Can J Anaesth. 2019, 66(12):1440–1449. English. doi: 10.1007/s12630-019-01488-4.

36. Alvarez-Lopez F., Maina M.F., Arango F., Saigí-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, 27;8(4):e19723. doi: 10.2196/19723.

37. Blanié A., Amorim M.A., Meffert A., Perrot C., Dondelli L., Benhamou D. Assessing validity evidence for a serious game dedicated to patient clinical deterioration and communication. Adv Simul. 2020, 27;5:4. doi: 10.1186/s41077-020-00123-3.

38. Borgersen N.J., et al. Gathering Validity Evidence for Surgical Simulation: A Systematic Review. Ann Surg. 2018, 267(6):1063–1068. doi: 10.1097/SLA.0000000000002652.

39. Zendejas B., Ruparel R.K.., Cook DA. Validity evidence for the Fundamentals of Laparoscopic Surgery (FLS) program as an assessment tool: a systematic review. Surg Endosc. 2016, 30(2):512–520. doi: 10.1007/s00464-015-4233-7.
40. Cook D.A., Brydges R., Zendejas B., Hamstra S.J., 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. doi: 10.1097/ACM.0b013e31828ffdcf.
41. Cook D.A. 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. doi: 10.1007/s10459-014-9551-3.
42. Padilla J.L., Benítez I. Validity evidence based on response processes. Psicothema. 20, 26(1):136–44. doi: 10.7334/psicothema2013.259.

Figures

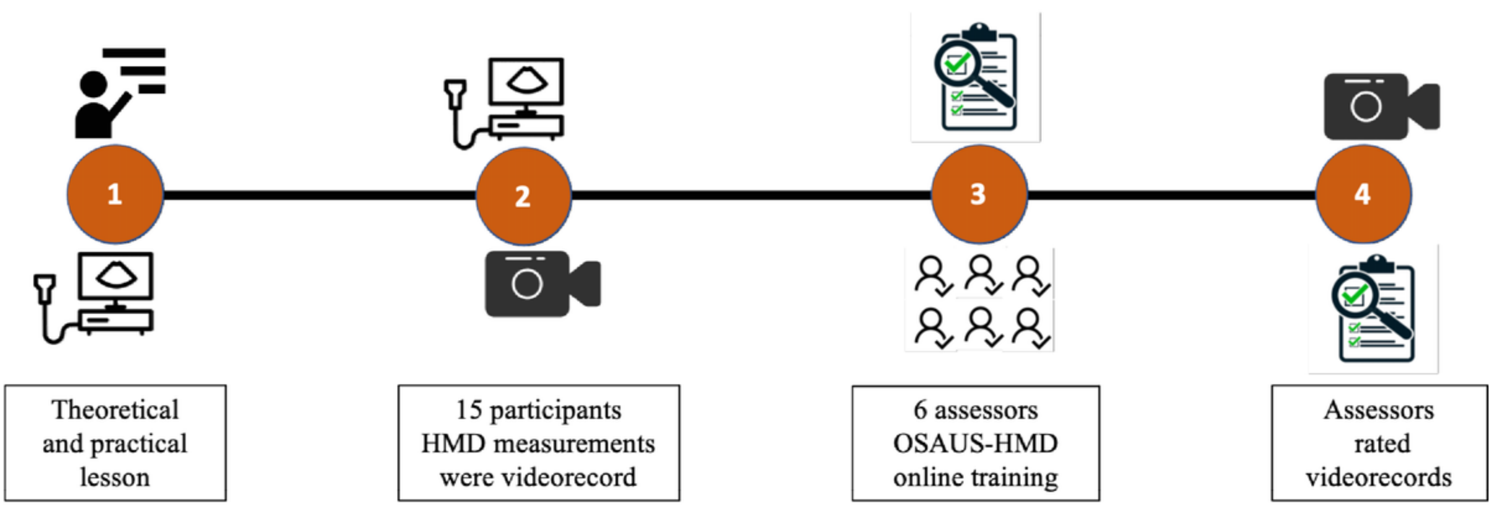

Figure 1

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.

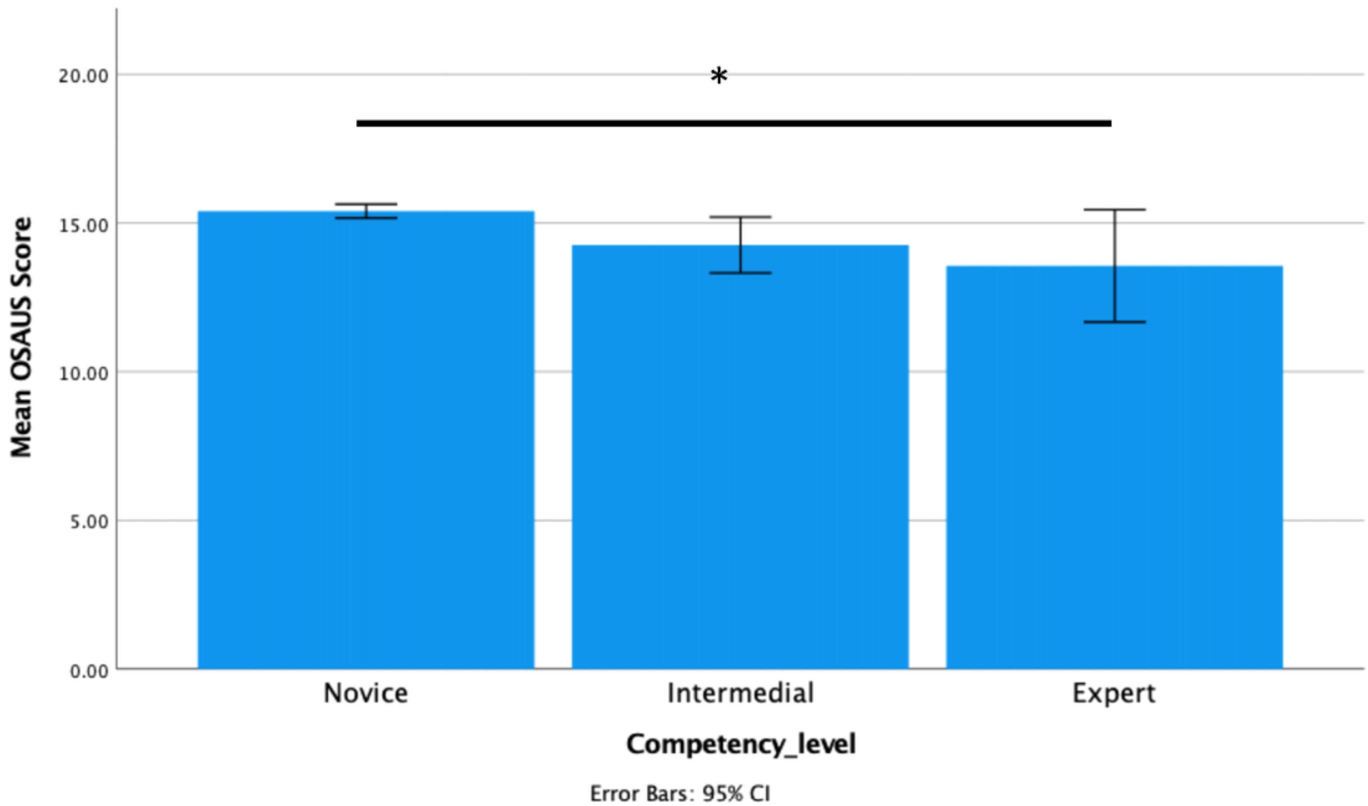

Figure 2

Mean of Scores in each group. Mean and error bars 95% CI for each competency group. * - p <0.05

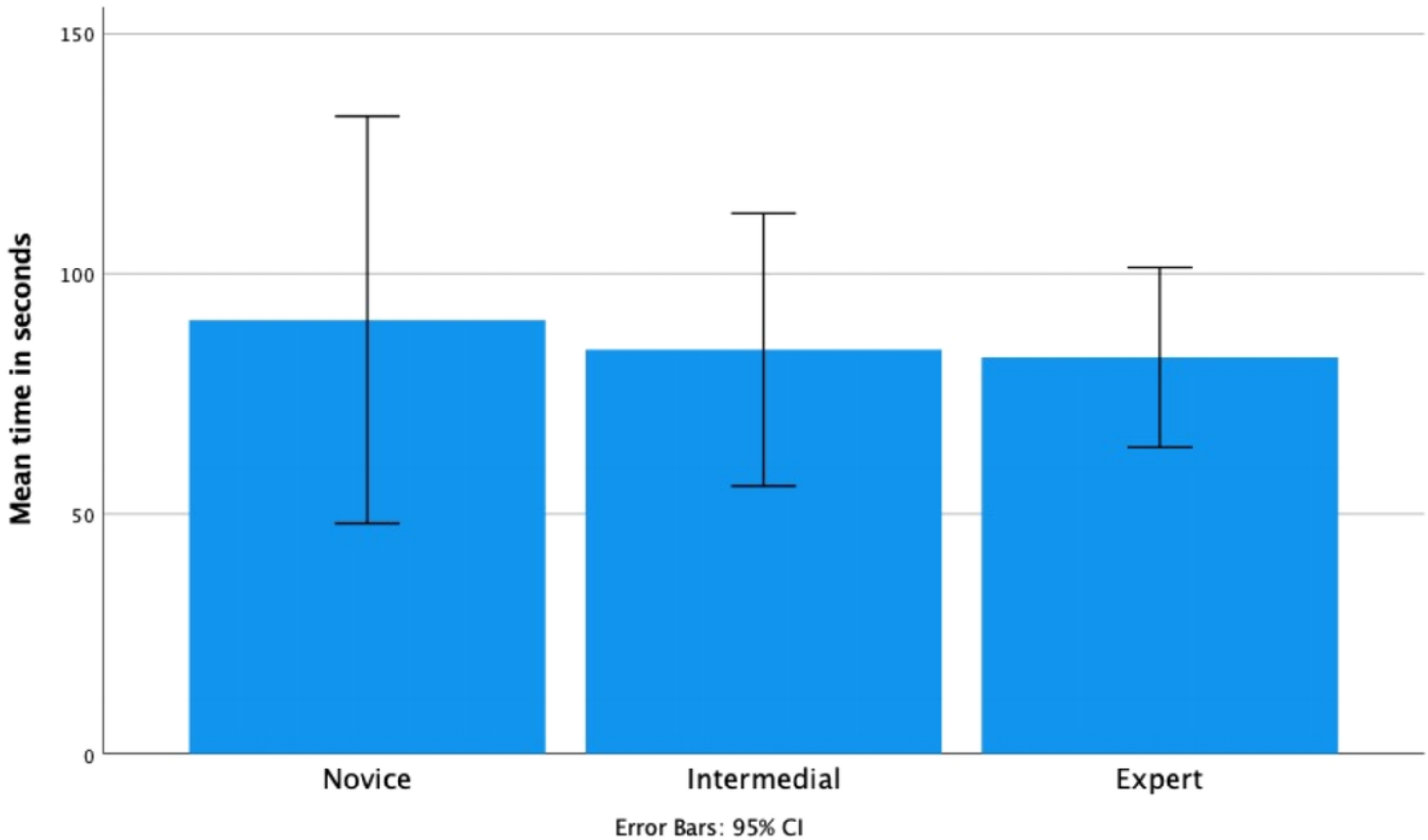

Figure 3

Mean total time (sec) to complete the task in each group. Mean and error bars 95% CI for each competency group.