Accepted Manuscript online: 2023-07-04 Article published online: 2023-10-06

Validity evidence for simulator-based obstetric ultrasound competency assessment tool: a multi-center study

Validitätsnachweis für ein simulatorbasiertes Tool zur Beurteilung der Ultraschallkompetenz in der Geburtshilfe: Eine multizentrische Studie

Authors
Yongfeng Zhao[1], Ping Zhou[1], Wenhui Zhu[1], Jidong Xiao[1], Wengang Liu[1], Yingchun Luo[2], Junhui Zhang[2], Lan Yang[3], Yan Xu[3], Xiaohong Tang[4]

Affiliations
1 Ultrasound, Central South University Third Xiangya Hospital, Changsha, China
2 Ultrasound, Hunan Provincial Maternal and Child Health Care Hospital, Changsha, China
3 Ultrasound, Nanjing Drum Tower Hospital, Nanjing, China
4 Clinical Skill Center, Central South University Third Xiangya Hospital, Changsha, China

Key words
obstetric ultrasound, competency, assessment, simulation, validity

received 07.10.2022
accepted after revision 04.07.2023
accepted manuscript online 04.07.2023
published online 06.10.2023

Bibliography
Ultraschall in Med 2024; 45: 168–175
DOI 10.1055/a-2122-6746
ISSN 0172-4614
© 2023. Thieme. All rights reserved.
Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Correspondence
Prof. Xiaohong Tang
Clinical Skill Center, Central South University Third Xiangya Hospital, Tongzipo Road 138, 410013 Changsha, China
tangxh007007@163.com

Additional material is available at https://doi.org/10.1055/a-2122-6746.

ABSTRACT

Purpose To collect validity evidence for a simulator-based obstetric ultrasound competency assessment tool (OUCAT).

Methods 89 sonographers from three centers (XY, MC, DT), including novices (n = 21), experienced trainees (n = 44), and experts (n = 24), participated in the competency assessment. Validity evidence of OUCAT was collected according to Standards for Educational and Psychological Testing. Content validity was ensured by reviewing guidelines and reaching expert consensus. The response process was ensured via training raters. Internal structure was explored through internal consistency, inter-rater reliability, and test-retest reliability. Relations to other variables were explored by comparing OUCAT scores of sonographers with different experience. Evidence for consequences was collected by determining the pass/fail level.

Results OUCAT included 123 items, of which 117 items could effectively distinguish novices from experts (P< 0.05). The internal consistency was represented by the Cronbach's α coefficient (0.978). The inter-rater reliability was high, with XY being 0.868, MC being 0.877, and DT being 0.937 (P< 0.001). Test-retest reliability was 0.732 (P = 0.001). The performance of experts was significantly better than experienced trainees, and the performance of experienced trainees was significantly better than novices (70.3 ± 10.7 vs. 39.8 ± 15.0 vs. 20.5 ± 10.6, P< 0.001). The pass/fail level determined by contrast group method was 45 points. The passing rate of novices, experienced trainees and experts was 0 % (0/21), 31.8 % (14/44), and 100 % (24/24), respectively.

Conclusion Simulator-based OUCAT exhibits good reliability and validity in assessing obstetric ultrasound skills.

ZUSAMMENFASSUNG

Ziel Erhebung von Validitätsnachweisen für ein simulatorbasiertes Tool zur Beurteilung der Ultraschallkompetenz in der Geburtshilfe (OUCAT).

Material und Methoden 89 Sonografen aus 3 Zentren (XY, MC, DT), darunter Anfänger (n = 21), erfahrene Auszubildende (n = 44) und Experten (n = 24), nahmen an der Kompetenzbewertung teil. Der Validitätsnachweis für OUCAT wurde gemäß den Standards für pädagogische und psychologische Tests ermittelt. Die inhaltliche Validität wurde durch die Überprüfung von Richtlinien und die Erzielung eines Expertenkonsenses sichergestellt. Der Antwortprozess wurde durch die Schulung von Bewertern sichergestellt. Die interne Struktur wurde durch interne Konsistenz, Inter-Rater-Reliabilität und Test-Retest-Reliabilität untersucht. Die Relationen zu anderen Variablen wurden durch den Vergleich der OUCAT-Scores der

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Sonografen mit der unterschiedlichen Erfahrung untersucht. Der Nachweis von Konsequenzen wurde erfasst, indem das Pass/Fail-Level bestimmt wurde.

Ergebnisse OUCAT umfasste 123 Elemente, von denen 117 effektiv zwischen Anfängern und Experten unterscheiden konnten (P< 0,05). Die interne Konsistenz wurde durch den Cronbachs-α-Koeffizienten (0,978) dargestellt. Die Inter-Rater-Reliabilität war hoch: XY lag bei 0,868, MC bei 0,877 und DT bei 0,937 (p < 0,001). Die Test-Retest-Reliabilität betrug 0,732 (p = 0,001). Die Leistung von Experten war signifikant besser als die von erfahrenen Auszubildenden, und die Leistung von erfahrenen Auszubildenden war signifikant besser als die von Anfängern (70,3 ± 10,7 vs. 39,8 ± 15,0 vs. 20,5 ± 10,6; p < 0,001). Das durch die Kontrastgruppenmethode ermittelte Pass/Fail-Level betrug 45 Punkte. Die Quote für Bestehen betrug bei Anfängern 0 % (0/21), bei erfahrenen Auszubildenden 31,8 % (14/44) und bei Experten 100 % (24/24).

Schlussfolgerung Das simulatorbasierte OUCAT zeigt eine gute Reliabilität und Validität bei der Bewertung von Ultraschallfähigkeiten in der Geburtshilfe.

Introduction

Ultrasound, free of ionizing radiation, is frequently used for prenatal diagnosis. Routine second-trimester ultrasound examination is of great value for assessing fetal growth and detecting fetal anomalies [1]. Obstetric ultrasound relies on the sonographer's operational skills. Sonographers need to adjust the probe to obtain appropriate planes, which requires hand-eye coordination. Incompetent scanning can lead to a missed diagnosis and misdiagnosis. Therefore, sonographers need training to meet a minimum skill requirement [2]. The International Society for Ultrasound in Obstetrics and Gynecology (ISUOG) proposed that at least 100 hours of supervised scanning and 100 obstetric ultrasound scans should be completed [3]. However, it may not be accurate to use the number of completed examinations as an assessment of clinical competency. Different training courses have different learning curves. Individual differences also affect training effect. Simply completing a certain number of examinations does not mean that all trainees will have mastered the skills [4].

Competency-based medical education calls for direct observation and objective assessment of a trainee's competencies [5]. To ensure that trainees who have completed training are qualified for performing obstetric ultrasound, it is necessary to conduct a composite assessment [6]. Previously, trainees were assessed when examining pregnant volunteers. This form of assessment often has several limitations: 1) comprehensive assessment is time-consuming, and causes discomfort for the pregnant woman, making it difficult to recruit volunteers; 2) extra examination time increases unnecessary exposure of the fetus to ultrasound, which does not conform to the principle of ALARA (as low as reasonably achievable) [7]; and 3) due to the influence of gestational week, amniotic fluid volume, and maternal obesity, the difficulty of obtaining ultrasound planes varies greatly.

Ultrasound simulators are highly realistic and cost-effective [8]. When trainees move the probe, the simulator perceives 3 D position data of the probe via an electromagnetic tracking system and displays 2 D image of the virtual fetus. A few scholars have tried to use simulators to assess trainees' skills [9]. Tolsgaard designed the Objective Structured Assessment of Ultrasound Skills (OSAUS), which evaluates trainees' abilities with respect to 5 aspects, including image optimization, systematic examination, and interpretation of images [10]. OSAUS has good reliability and validity and can effectively distinguish between novices and experts [10, 11]. Chalouhi developed an image quality score-based criterion. Scores achieved in 6 morphological planes and 3 biometric measurements were selected to evaluate trainees' abilities [12]. These representative views can reflect the level of a trainee's skill.

However, the use of simulators for obstetric ultrasound skill assessment is still in the exploratory stage. Evidence and a mature assessment tool for comprehensive assessment are lacking. This study intends to develop a simulator-based obstetric ultrasound competency assessment tool (OUCAT) to determine whether trainees' skill performance meets the requirements for second-trimester obstetric ultrasound examination [1, 6], and collect its validity evidence.

Materials and Methods

Settings

This prospective study was conducted in the Third Xiangya Hospital of Central South University (XY), Hunan Provincial Maternal and Child Health Care Hospital (MC) and Nanjing Drum Tower hospital (DT). The ethics committee of the Third Xiangya Hospital of Central South University approved this study. All procedures were conducted in accordance with the Helsinki Declaration. Validity evidence for the OUCAT was collected with reference to the framework proposed in Standards for Education and Psychological Testing [13]: content evidence, response process, internal structure, relationship to other variables, and consequences.

Participants

89 volunteers were included as research subjects, including 30 from XY, 30 from MC, and 29 from DT. They were divided into 3 groups: Novices (N = 21) were first-year residents without obstetric ultrasound training experience. Experienced trainees (N = 44) had more than one year of training but had not been independently engaged in obstetric ultrasound examination. Experts (N = 24) were sonographers who have completed residency training and independently engaged in obstetric ultrasound examination. Since the medical education system in China is different from other countries, the educational background of residents varies. Residents joining the post-graduate program and residency train-

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

ing at the same time will be granted a master's degree, while other residents choosing residency training alone will have a bachelor's degree.

6 consultants from 3 hospitals were selected as raters. The inclusion criteria were as follows: raters must have independently conducted obstetric ultrasound examinations for more than 5 years and have completed at least 500 second-trimester obstetric ultrasound examinations per year.

Equipment

Assessments were performed on Simbionix U/S Mentor simulators (Tel Aviv, Israel). All sonographers chose the same normal fetal module for assessment. During the assessment, the 3 D anatomical schematic was hidden.

Content evidence

A consultant designed the OUCAT checklist and scoring rules based on second-trimester fetal ultrasound scan guidelines and competency assessment consensus [1, 6]. Participants received a pass or fail score for each item. Then, the draft OUCAT was emailed to 15 consultants to solicit their opinions on whether the settings were reasonable. According to the feedback, consultants were organized for expert group discussion to revise the draft. Finally, all consultants reached an agreement on the items and scoring rules. A total of 123 items in 11 sections have been developed (**supplementary file 1**).

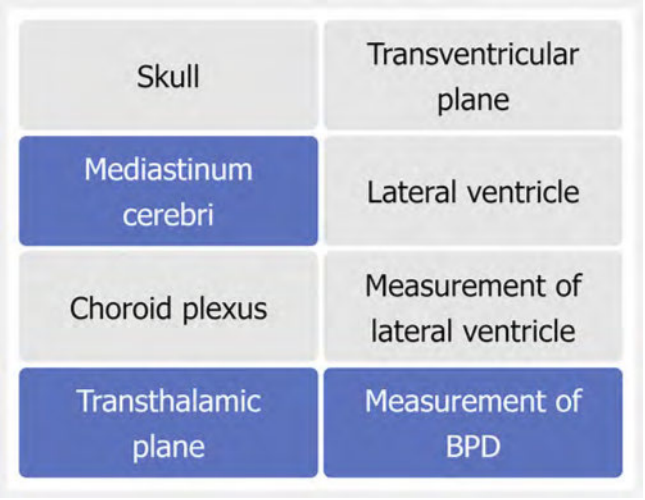
Response process

Participants learned how to operate the simulator for 15 minutes before assessment, but they did not receive any skill training. Then the participants watched a demonstration video. They were told to follow the examination procedures for a real pregnant woman. Participants were supposed to complete the examination independently within 30 minutes (the period of 30 minutes was based on the time required for a skilled sonographer in China to complete a routine second-trimester obstetric ultrasound examination). A high-definition video was recorded during the assessment. After the assessment, the video was anonymously numbered. Each video was distributed to two raters for independent scoring, and the final score was averaged.

The OUCAT scoring was performed online (▶ **Fig. 1**). Before scoring, the scoring rules and the link to the score book were distributed to raters so they could familiarize themselves. During the training of raters, two assessment videos were distributed to all raters, and the raters graded the videos separately. Then the designer summarized the inconsistencies of raters and the problems encountered during scoring. All raters attended an online discussion, and the designer explained the scoring rules until consensus was reached among the raters.

Internal structure

Item difficulty was calculated by average/full score of each item. Differences in scores for items between novices and experts were compared. The Cronbach's alpha coefficient was calculated to assess internal consistency. Inter-rater reliability was assessed



▶ **Fig. 1** Diagram of online OUCAT score book. The rater could press the button on the touch screen to complete the scoring.

by analyzing the intraclass correlation coefficient (ICC) of the scores given by two raters. Experts took a second-round assessment one week after the first round, and the test-retest reliability was explored by calculating the ICC of scores achieved in two rounds.

Relations to other variables

Differences in OUCAT scores among participants with different experience/educational background were compared. The relationship between participants' learning time spent on obstetric ultrasound and the OUCAT scores was analyzed. In addition, raters made two extra assessments: using a Likert scale to rate the performance of participants on a scale of 1–5 (**supplementary file 2**) and giving a pass/fail expert evaluation (**supplementary file 3**). Relationships between extra assessments and OUCAT scores were explored.

Consequences

Pass/fail level was set up via contrast group method to determine the criteria for qualified sonographers competent to perform second-trimester obstetric ultrasound scans. The minimum passing score was set as the optimal cut-off value to minimize false positives/false negatives. The mean score of the experts was set as the mastery learning level. The OUCAT pass rate and mastery learning pass rate were compared to the pass/fail rate in the rater's extra assessment.

Statistical analysis

Statistical analysis was conducted by SPSS 21. Measurement data were compared using analysis of variance. Pairwise comparisons were performed with the LSD-t test. Non-normally distributed measurement data were compared using Wilcoxon signed-rank test. Comparisons of enumeration data were performed using Fisher's exact test. The Mann-Whitney U test was used to compare scores of items between novices and experts. Cronbach's alpha coefficient was calculated to evaluate internal consistency.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

ICC was analyzed to explore inter-rater reliability and test-retest reliability. Spearman correlation analysis was used to analyze the association between OUCAT score and learning time spent on obstetric ultrasound, and the relation between OUCAT scores and Likert scores in the extra assessment. The McNemar test was used to compare the pass rate based on OUCAT with the pass rate based on the expert evaluation. $P<0.05$ was considered to be statistically significant.

Result

The demographic characteristics of the 89 participants are shown in ▶ **Table 1**. There were significant differences among the three groups in terms of educational background and learning time spent on obstetric ultrasound (all $P<0.001$).

Internal Structure

The median item difficulty was 0.42 (IQR 0.30–0.55) (**Supplementary file 1**). The difficulty of different sections ranged from 0.21 to 0.66, with the least difficult part being section 1 (Preparation) and the most difficult part being section 13 (Conclusion) (▶ **Table 2**). Among 123 items, 117 items could effectively distinguish experts from novices (both $P<0.05$) (**supplementary file 1**). The internal consistency of OUCAT was high, with a Cronbach's alpha coefficient of 0.978. The inter-rater reliability was 0.868 in XY, 0.877 in MC, and 0.937 in DT (all $P<0.001$). The test-retest reliability was 0.732 ($P=0.001$).

Relations to other variables

The score of novices, experienced trainees, and experts was 20.5 ± 10.6, 39.8 ± 15.0, and 70.3 ± 10.7, respectively (▶ **Table 3**). The performance of experienced trainees was significantly better than that of novices ($P<0.001$), and the performance of experts was significantly better than that of experienced trainees ($P<0.001$). OUCAT scores between participants with different educational backgrounds also had a significant statistical difference ($P=0.007$) (▶ **Table 3**). The performance of those with PhD and master's degrees was significantly better than that of those with a bachelor's degree (both $P<0.05$). There were no differences between those with PhD and master's degrees ($P=0.241$). OUCAT scores were positively correlated with learning time spent on obstetric ultrasound, with correlation coefficient $r_s=0.635$, $P<0.001$. There was a positive correlation between OUCAT scores and Likert scores in the extra assessments ($r_s=0.852$, $P<0.001$). The score of failed participants judged by experts in the extra assessment was 36.6 ± 17.8, and the score of passed participants was 75.0 ± 10.1. The difference between these two groups was statistically significant ($P<0.001$).

Consequences

45 out of 100 was set as the minimum passing score for qualified sonographers (▶ **Fig. 2**). Based on this standard, the pass rate was 0 % (0/21) for novices, 31.8 % (14/44) for experienced trainees, and 100 % (24/24) for experts. 70 out of 100 was set as the mastery learning score, which was the mean score of experts. Neither novices nor experienced trainees reached the mastery learning level. The pass rate based on the subjective evaluation of experts in the extra assessment was lower than that of the OUCAT pass/fail level ($P<0.001$). The pass rate based on the expert evaluation in the extra assessment was 0 % (0/21) for novices, 2.3 % (1/44) for experienced trainees, and 62.5 % (15/24) for experts. No significant statistical difference was found between the pass rates based on the OUCAT mastery learning level and the expert subjective evaluation in the extra assessment ($P=0.250$) (▶ **Table 4**).

Discussion

Ultrasound simulators based on virtual reality technology provide a standardized training environment for trainees with good usability and fidelity [14]. In recent years, a few scholars have tried to use simulators to assess the skill level of trainees and found that applying simulators for assessment was feasible [15, 16, 17]. This study is the first to comprehensively assess sonographers using simulators. The results show that a simulator-based OUCAT assessment scheme has good feasibility and validity. It can distinguish

▶ **Table 1** Demographic characteristics of participants.

Category		Novices	Experienced trainees	Experts	P
Number	Capitals	21	44	24	
Gender	Male	1	1	5	0.025*
	Female	20	43	19	
Age	Years	25.2 ± 2.5	30.4 ± 5.6	36.0 ± 4.2	< 0.001**
Educational background	Bachelor	15	34	7	< 0.001*
	Master	6	9	12	
	PhD	0	1	5	
Learning time spent on obstetric ultrasound	Years	0.0 ± 0.0	4.4 ± 4.0	9.5 ± 4.6	< 0.001**

*Fisher's exact test, **Wilcoxon signed-rank test

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

▶ **Table 2** OUCAT scores achieved by participants in different sections.

Section	Overall score	Difficulty	Novices		Experienced trainees		Experts	
			Mean score	Score rate	Mean score	Score rate	Mean score	Score rate
Preparation	5	0.66	3.1	0.62	3.2	0.63	3.8	0.76
General information	4	0.42	1.1	0.28	1.6	0.41	2.4	0.60
Skull	19	0.51	5.3	0.28	9.2	0.48	15.1	0.80
Face	8	0.56	2.5	0.32	4.2	0.52	7.0	0.88
Neck	3	0.50	0.9	0.28	1.7	0.56	1.8	0.60
Chest	7	0.28	0.4	0.06	1.7	0.24	4.1	0.58
Heart	32	0.32	4.2	0.13	7.8	0.24	20.0	0.63
Abdomen	14	0.51	3.7	0.26	7.2	0.52	10.3	0.73
Spine	5	0.41	1.0	0.21	1.7	0.34	3.6	0.72
Limbs	14	0.49	3.8	0.27	5.7	0.41	11.8	0.84
Placenta, umbilical cord and amniotic fluid	6	0.41	1.8	0.29	2.1	0.36	3.8	0.63
Maternal uterus/adnexa	5	0.30	0.7	0.15	1.5	0.30	2.3	0.46
Conclusions	1	0.21	0.1	0.11	0.1	0.13	0.4	0.44

▶ **Table 3** OUCAT scores of participants with different experience and educational backgrounds.

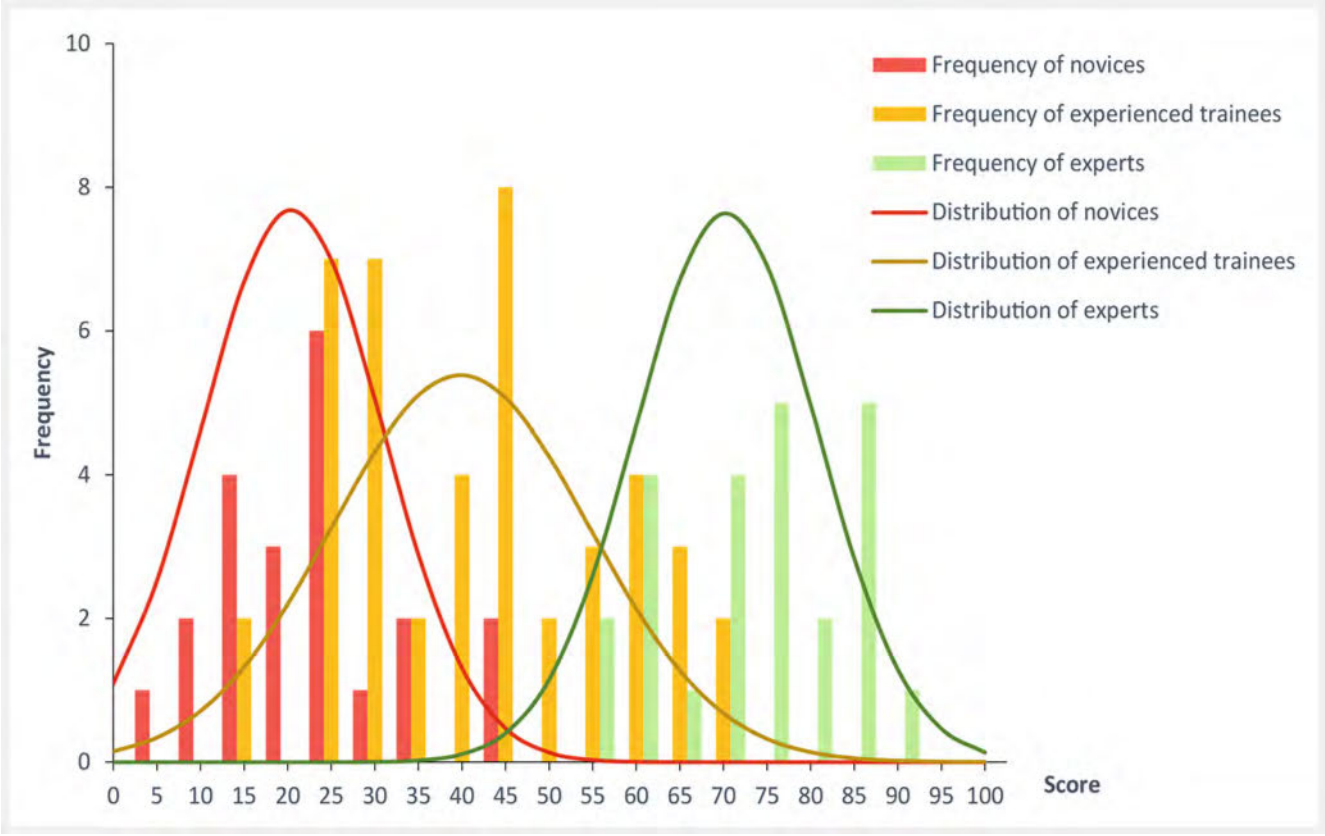
		Mean	Standard deviation	Standard error	Min.	Max.	P
Group	Novices	20.5	10.6	2.3	4.1	43.5	< 0.001*
	Experienced trainees	39.8	15.0	2.3	11.0	69.9	
	Experts	70.3	10.7	2.2	50.8	87.0	
Educational background	Bachelor	38.3	20.1	2.7	4.1	76.0	0.007*
	Master	50.3	23.5	4.5	6.5	87.0	
	PhD	61.6	21.1	8.6	26.8	84.2	

*Analysis of variance

between sonographers at different skill levels to determine which sonographers are qualified to perform second-trimester obstetric ultrasound scans.

Evidence of content validity could be obtained by analyzing the relationship between the content of the test and the structure it was intended to measure. In other studies, scholars used metrics such as dexterity scores, OSAUS scores, image quality scores, fetal biometric accuracy scores, and built-in parameters in simulators to assess student skill levels. Most of those metrics reflect the skill level of sonographers [10, 12, 18, 19, 20]. However, establishing reliable metrics to assess obstetric ultrasound skills is still challenging [9]. At present, a widely recognized method is the consensus proposed by ISUOG and American Institute of Ultrasound in Medicine [6]. OUCAT was designed based on the widely acknowledged guidelines and consensus. It aimed to assess candidates' ability with respect to the following aspects: 1) obtaining appropriate ultrasound planes; 2) accurately identifying and checking the anatomical structures; and 3) accurately conducting fetal biometry measurements, all of which are required in clinical work. Similarly, in Rosen's study, experts also evaluated candidates' performance based on quality of images, acquisition of required landmarks, and accuracy of landmarks [18]. For the definition of scoring rules, reference was made to Salomon's ultrasound image quality control standards and guidelines [1, 6, 21]. After OUCAT was for-

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.



▶ **Fig. 2** Distribution curve of OUCAT scores in each group.

▶ **Table 4** Comparison of pass rates between OUCAT passing score/mastery learning score and subjective expert judgement in the extra assessment.

		Expert judgement					Expert judgement		
		Fail	**Pass**	**Sum**			**Fail**	**Pass**	**Sum**
Passing score = 45	Fail	50	1	51	Mastery learning score = 70	Fail	73	3	76
	Pass	23	15	38		Pass	0	13	13
	Sum	73	16	89		Sum	73	16	89
*McNemar test		P < 0.001*					P = 0.250*		

mulated, expert opinions were collected by a survey following an expert group discussion. Expert consensus was eventually reached, thereby ensuring content validity.

Validity in terms of response process came from the management of assessment tools and the collection of data to control errors. In this study, candidates were properly informed about the operation of the simulator, as well as the assessment form. During the assessment, the candidates completed the operation independently with no intervention. The raters were strictly screened and well trained. They were familiar with the content and image quality control standards and had the same understanding of the scoring rules. The online scoring made data collection easier and lightened the raters' workload. Raters could rewatch the high-definition video of candidates' operations to reduce omissions.

Validity regarding internal structure is related to the psychometric characteristics of the assessment tool. This study collected evidence from several perspectives. The item difficulty of OUCAT was moderate. The median item difficulty was 0.42. Novices and experienced trainees received low scores in the sections regarding the chest, heart, and maternal uterus/adnexa, as well as the drawing of conclusions. The above aspects may be challenging for trainees, or they may have been neglected in previous training. Among the 123 items, 117 items can effectively distinguish between novices and experts, and the other six items include check-

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

ing of the patient's name, age, etc. Although these items cannot effectively distinguish novices from experts, they are all indispensable in the examination procedure. Thus, they were not excluded from the final score. A high Cronbach's alpha coefficient of 0.978 revealed that OUCAT had good internal consistency. Reliability reflects the reproducibility of the assessment. In most tests, a reliability of 0.7–0.8 is acceptable, while for high-stakes tests, the reliability is required to be greater than 0.9 [22] . The inter-rater reliability of three hospitals in this study was all greater than 0.8, indicating that the OUCAT assessment had good inter-rater reliability. The inter-rater reliability of DT was greater than 0.9, which was sufficient to meet the demand of admission assessment for reliability. In this study, the test-retest reliability was 0.732, which was acceptable. From the above, we found that OUCAT exhibited good internal structure validity evidence. In some other studies, Dyre found that the internal consistency of simulator assessment was high, while the test-retest reliability was as low as 0.62 [16]. In another study, Chalouhi also confirmed that the inter-rater consistency of simulator-based skill assessment was good. The correlation coefficient between raters was 0.888, which was comparable to that of the assessment of pregnant women [12].

By setting OUCAT pass/fail criteria, the researchers collected validity evidence of consequences. With a passing score of 45/100, the pass rate was 100 % for experts and 0 for novices. It means that there were no false-negative or false-positive results, and OUCAT shows validity in identifying qualified sonographers. In this study, an extra assessment was attempted, which allowed raters to judge from their own professional perspective and experience. It was found that there was a significant difference between the pass rate determined by OUCAT and the pass rate assessed by the raters' subjective evaluation, with the latter being more stringent. Interestingly, the pass rate was not significantly different from that of the raters' subjective evaluation when the passing standard was set as the mastery learning score. It means that if the subjective evaluation is given by raters, they may make inappropriate decisions based on their own experience. Thus, the pass/fail criteria based on OUCAT may be more objective and reasonable.

The advantage of the simulator-based OUCAT is that it is a comprehensive assessment tool, and the test content and procedure resemble the daily routine. The items and scoring rules are based on guidelines that are easy for raters to understand, and satisfactory inter-rater consistency can be obtained after simple training. Based on the assessment, we can understand the difficulties regarding the second-trimester obstetric ultrasound scan, identify deficiencies of individuals with respect to completing specific tasks, provide feedback to trainees, and optimize the training course. Its feasibility has also been verified by a relatively large sample size in multiple centers. There are still limitations to this study: 1) The method to reach expert consensus regarding the OUCAT scorebook uses a questionnaire and expert discussion, whereas consensus obtained via the Delphi survey may be more objective; 2) OUCAT requires a long period of concentration on the part of the raters because it has many scoring items; 3) OUCAT was designed for competency assessment of single routine second-trimester obstetric ultrasound scan skills, not for first-trimester scans; and 4) performance in the simulated environment does not fully represent the skills required to examine a real pregnant woman. When performing an examination on a real pregnant woman, the candidates will face more pressure and the environment will be more complicated. Previous studies have proved that candidates get higher test scores using the simulator than examining real pregnant women [12].

Conclusion

This study constructed a feasible OUCAT for comprehensive assessment of second-trimester obstetric ultrasound skills, which met the requirements of the obstetric ultrasound examination guidelines and competency assessment consensus. Evidence collected in the multi-center verification revealed that OUCAT had good validity and reliability and could effectively identify qualified sonographers. The results supported that obstetric ultrasound skills could be objectively assessed by simulator-based competency assessment.

Funding

Project of the 14th five year plan of Educational Science in Hunan Province (XJK21AGD002)

Acknowledgement

We acknowledge support from the China National Medical Examination Center.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Salomon LJ, Alfirevic Z, Berghella V et al. Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. Ultrasound Obstet Gynecol 2011; 37: 116–126

[2] Chalouhi GE, Bernardi V, Ville Y. Ultrasound simulators in obstetrics and gynecology: state of the art. Ultrasound Obstet Gynecol 2015; 46: 255–263

[3] ISUOG Education Committee recommendations for basic training in obstetric and gynecological ultrasound. Ultrasound Obstet Gynecol 2014; 43: 113–116

[4] Damewood SC, Leo M, Bailitz J et al. Tools for Measuring Clinical Ultrasound Competency: Recommendations From the Ultrasound Competency Work Group. AEM Educ Train 2020; 4: S106–S112

[5] Lockyer J, Carraccio C, Chan MK et al. Core principles of assessment in competency-based medical education. Med Teach 2017; 39: 609–616

[6] Abuhamad A, Minton KK, Benson CB et al. Obstetric and gynecologic ultrasound curriculum and competency assessment in residency training programs: consensus report. Ultrasound Obstet Gynecol 2018; 51: 150–155

[7] Gill RW. The Physics and Technology of Diagnostic Ultrasound: A Practitioner's Guide. Sydney: High Frequency Publishing; 2012

[8] Taksoe-Vester C, Dyre L, Schroll J et al. Simulation-Based Ultrasound Training in Obstetrics and Gynecology: A Systematic Review and Meta-Analysis. Ultraschall in Med 2021; 42: e42–e54

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

[9] Le Lous M, Klein M, Tesson C et al. Metrics used to evaluate obstetric ultrasound skills on simulators: A systematic review. Eur J Obstet Gynecol Reprod Biol 2021; 258: 16–22

[10] Tolsgaard MG, Ringsted C, Dreisler E et al. Reliable and valid assessment of ultrasound operator competence in obstetrics and gynecology. Ultrasound Obstet Gynecol 2014; 43: 437–443

[11] Tolsgaard MG, Ringsted C, Dreisler E et al. Sustained effect of simulation-based ultrasound training on clinical performance: a randomized trial. Ultrasound Obstet Gynecol 2015; 46: 312–318

[12] Chalouhi GE, Bernardi V, Gueneuc A et al. Evaluation of trainees' ability to perform obstetrical ultrasound using simulation: challenges and opportunities. Am J Obstet Gynecol 2016; 214: 525.e1–525.e8

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

[14] Ostergaard ML, Konge L, Kahr N et al. Four Virtual-Reality Simulators for Diagnostic Abdominal Ultrasound Training in Radiology. Diagnostics (Basel) 2019; 9: 50

[15] Le Lous M, Tsatsaris V, Tesniere A et al. Improving students' ability to perform a standardized foetal biometry plane using ultrasound simulators. J Gynecol Obstet Hum Reprod 2017; 46: 439–443

[16] Dyre L, Norgaard LN, Tabor A et al. Collecting Validity Evidence for the Assessment of Mastery Learning in Simulation-Based Ultrasound Training. Ultraschall in Med 2016; 37: 386–392

[17] Dromey BP, Ahmed S, Vasconcelos F et al. Dimensionless squared jerk: An objective differential to assess experienced and novice probe movement in obstetric ultrasound. Prenat Diagn 2021; 41: 271–277

[18] Rosen H, Windrim R, Lee YM et al. Simulator Based Obstetric Ultrasound Training: A Prospective, Randomized Single-Blinded Study. J Obstet Gynaecol Can 2017; 39: 166–173

[19] Burden C, Preshaw J, White P et al. Usability of virtual-reality simulation training in obstetric ultrasonography: a prospective cohort study. Ultrasound Obstet Gynecol 2013; 42: 213–217

[20] Madsen ME, Norgaard LN, Tabor A et al. The Predictive Value of Ultrasound Learning Curves Across Simulated and Clinical Settings. J Ultrasound Med 2017; 36: 201–208

[21] Salomon LJ, Winer N, Bernard JP et al. A score-based method for quality control of fetal images at routine second-trimester ultrasound examination. Prenat Diagn 2008; 28: 822–827

[22] Thomas PA, Kern DE, Hughes MT et al. Curriculum development for medical education: a six-step approach. 4th edition. Baltimore: Johns Hopkins University Press; 2022

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.