Validity and Reliability of 'Objective Structured Assessment of Technical Skills' (OSATS) for Vaginal Delivery

Dr. Jaweria Faisal

Abstract

Objective: To assess validity and reliability metrics of objective structured assessment of technical skills (OSATS) for assessing the competence of technical skills of vaginal delivery

Materials and Methods: Co-relational reliability study, the validation study was conducted at the Department of Obstetrics and Gynaecology of Akhtar Saeed Medical College and Farooq Hospital Rawalpindi from April to December 2023. The study comprised nine candidates at different training levels (LHV, staff, postgraduate trainees & senior registrars), who were evaluated on OSATS by 8 evaluators over 54 NVD observations. Cronbach's alpha and intraclass correlation coefficient (ICC) were used to estimate reliability metrics of OSATS scores. The construct validity; that OSATS can measure and reveal the underlying construct, was determined by three shreds of evidence; evidence of content, response process and internal structure, and evidence of improvement in performance scores as the procedure-specific experience accrues with OSATS using linear mixed model and evidence of the discriminatory power of the tool between various levels of the training program, with one-way ANOVA.

Results: The OSATS tool demonstrated good reliability metrics with Cronbach's alpha of 0.897 and good inter-rater reliability calculated by intraclass correlation coefficient as 0.801. OSATS was validated as examinee scores improved showing gradual progression on performance curves with succeeding OSATS encounters. Validation also yielded that the tool can differentiate among different training levels with a p-value of 0.019

Conclusion: The OSATS tool was found to be valid and reliable for assessing the competence of technical skills of vaginal delivery.

Keywords: Cronbach's alpha, Validity, OSATS, Reliability, Workplace-based assessment.

How to cite: Faisal J. Validity and Reliability of 'Objective Structured Assessment of Technical Skills' (OSATS) for Vaginal Delivery. Esculapio - JSIMS 2024;20(02):166-170

DOI: https://doi.org/10.51273/esc24.25132024

Introduction

Acquisition of technical procedural capabilities must ensue parallel to the attainment of knowledge and professionalism. Assessment drives learning.[1] In competency-based education, workplace-based assessment (WPBA) has gained importance as it demonstrates that professionals are making appropriate progress through a structured program of specialty training.[2,3] Objective Structured Assessment of Technical Skills (OSATS) is a WPBA, where series of standardized procedural tasks are measured.[4]

The OSATS was first introduced by surgical educators at the University of Toronto for formative assessment.[5] Since then OSATS, has demonstrated high validity and reliability metrics in technical skills assessment.[6,7,8] OSATS is being used, by the Royal College of Obstetrics and Gynaecology for various assessments but OSATS for normal vaginal delivery is not available anywhere. One study has been done for OSATS for vertex delivery at the University Hospital of Nice but that form hasn't

1. *Department of Obstetrics & Gynaecology, Akhtar Saeed Medical College & Farooq Hospital Rawalpindi*

Correspondence:
Dr. Jaweria Faisal, *Professor/ HOD, Department of Obstetrics & Gynaecology, Akhtar Saeed Medical College & Farooq Hospital Rawalpindi.*
E-mail: Jaweria_faisal@hotmail.com

Submission Date: 01-03-2024
1st Revision Date: 27-03-2024
Acceptance Date: 11-05-2024

been acquired by accreditation bodies.[9]

Mostly, normal vaginal delivery (NVD) is learned by the master-apprenticeship model, by observation and assisting, gradually becoming competent to perform it independently; and assessed subjectively by looking in Logbooks. Surgical and allied Disciplines have historically lacked objective assessment for technical skills of learners.[4] Whenever assessment is largely subjective and base heavily on preceptor ratings, it carries low reliability.[2] To ensure the technical proficiency of trainees on NVD, we crucially need an assessment tool that is capable of providing objective scores to evaluate this predefined competency and ratify its proper training, before doctors are allowed to perform it unsupervised.[9] OSATS permits objective structured assessment of technical skills on a global rating scale, but teachers are also not sensitized with this new techniques.[3,10]

In my research, modified Delphi's method was used to design OSATS tool, a set of 20 items of six responses with descriptors that could measure the underlying construct, the educational achievement, that candidates are competent to perform safe normal vaginal delivery. The rationale of this study was to establish its scores validity and reliability metrics so that a standardized objective tool would be available for enhancement of training. It may be incorporated in curriculum of midwives, and doctors for formative assessment and revalidation of NVD skill to improve patient care

Patients and Methods

After taking approval from Akhtar Saeed Medical College and Farooq Hospital Rawalpindi Institutional Review Board Ref. No. RAC&IRB-05/04/2023, the study was conducted at the Department of Obstetrics and Gynaecology. Inclusion criteria: All candidates; LHV, staff, postgraduate trainees (PG) & senior registrars (SR), who were already conducting NVD in low risk mothers, were briefed about the study proposal so that they could give an informed consent to participate. A total number of 54 patients, who had signed the inpatient consent form, were included in the study. Sampling techniques was non probability convenience sampling. Any delivery in high-risk woman with comorbidities was excluded. All candidates were pre trained on NVD skill by using OSATS that was freely available in labour ward. The candidates selected the time of WPBA with OSATS with assessors provided they both agreed. For data collection, while conducting the delivery, each candidates was observed by two assessors simultaneously. They both rated the performance of candidates on NVD skill with OSATS form on rating scale, independent of each other. All assessors were voluntary pre-trained consultants. This study was not duration limited, it finished when all candidates had taken six OSATS. The immediate feedback covered strength and suggestions for improvement. Detailed psychometric analysis was done to assess OSATS construct validity and reliability. Cronbach's alpha, was calculated from average (Mode) of the 6 response of each candidate's OSATS score. The standard error of measurement to create confidence bands around observed scores was used to indicate the precision of measurement. For estimating inter-rater reliability, Intraclass correlation coefficient was used. To establish the caseload, all consecutive OSATS of each candidate were numbered. OSATS scores has capacity to reveal and prove the underlying construct, the construct validity; was determined by three evidences. Evidence of content, response process and internal structure. Evidence of improvement in performance scores as the NVD experience increased with OSATS. To see progression, learning curves for each candidate was mapped by plotting her OSATS scores against the total procedure-specific caseload, during total time till all OSATS are signed off, using Linear mixed model. Evidence of the discriminatory power of the tool between various levels of the training program, using one-way ANOVA with level of training program as the independent variable and the examination scores, performance, as the dependent variables. All scores were entered in Statistical Package for Social Sciences (SPSS 26) datasheet for analysis.

Results

Nine candidates were evaluated by 8 evaluators over 54 NVD observations with OSATS. Cronbach's alpha based on standardized Items (n= 20) was calculated as 0.897. Intraclass Correlation Coefficient came as 0.801 (0.737-0.861) as the random error in the responses are very low, the 95% confidence bands of each item shows a very close difference. All items correlated positively with each other. The mean ratings on each item ranged from 3.09 to 3.74 with 'documentation' getting the least and 'positions the patient' being awarded the highest mean score.**(Table-I)** As validity evidence, performance curves showed gradual progression as candidate's accrued experience with OSATS on specific skill. Candidates learned, removed their mistakes, gradually performed better and scored highest in their last attempt. (Figure 1). The result showed that senior regis-

trars (SR) took a very high OSATS score as compared to the others with a p value 0.019. (Table II). Although the average OSATS score of all candidates differ slightly, p value 0.126, but their pair to pair comparison showed that senior registrars were way better than staff, LHV and post graduate trainee with significant P value 0.003

Table 1: *Descriptive item analysis of OSATS questionnaire*

No.	Items	Mean	Confidence interval	
			Lower	Upper
1	Patient doctor communication	3.50 ± 1.34	3.13	3.87
2	Positions the patient	3.74 ± 1.26	3.40	4.09
3	Infection control	3.31 ± 1.40	2.93	3.70
4	Bladder emptying	3.30 ± 1.27	2.95	3.64
5	Fetal heart monitoring	3.46 ± 1.28	3.11	3.81
6	Lignocaine perineal infiltratic	3.43 ± 1.42	3.04	3.81
7	Medio-lateral episiotomy	3.46 ± 1.41	3.08	3.85
8	Perineal phase management	3.61 ± 1.32	3.25	3.97
9	Positions hands on head	3.57 ± 1.16	3.26	3.89
10	Control extension	3.26 ± 1.31	2.90	3.62
11	Uterotonic	3.52 ± 1.30	3.16	3.87
12	Cord cutting	3.57 ± 1.42	3.19	3.96
13	Placental separation	3.61 ± 1.37	3.24	3.98
14	Controlled cord traction	3.13 ± 1.29	2.78	3.48
15	Placental inspection	3.46 ±1.31	3.10	3.82
16	Trauma evaluation & repair	3.37 ± 1.09	3.07	3.67
17	Handling of instruments/ knots	3.11 ± 1.28	2.76	3.46
18	Hemostasis and tissue respect	3.28 ± 1.12	2.97	3.58
19	Count completed	3.26 ± 1.14	2.95	3.57
20	Documentation	3.09 ± 1.19	2.77	3.42

Table 2: *Comparison of OSATS score with respect of candidate and training program*

Characteristic	Category	Mean ± Standard deviation	P value
Candidate	LHV	56.50 ± 18.13	0.126
	SR	82.50 ± 29.08	
	SR	83.33 ± 22.55	
	PG 3	74.50 ± 26.49	
	PG 3	74.33 ± 23.95	
	PG 2	59.50 ± 18.36	
	PG 2	63.67 ± 23.69	
	Staff	49.33 ± 16.00	
	LHV	68.83 ± 16.46	
Training program	Staff	49.33 ± 16.00	0.019
	LHV	62.67 ± 17.72	
	PG	68.00 ± 22.75	
	SR	82.92 ± 24.81	

Table 3: *Post ANOVA comparison of OSATS score with respect of training programs*

Candidate 1	Candidate 2	Difference of mean between candidate 1 and candidate 2	P value
Senior Registrar (SR)	Staff	33.58	0.003*
	LHV	20.25	0.026*
	PG	14.92	0.057
Post-Graduate	Staff	18.67	0.065
	LHV	5.33	0.489
LHV	Staff	13.33	0.224

(Table III). This shows that instrument is valid and can discriminate well between candidates of varying learning levels.

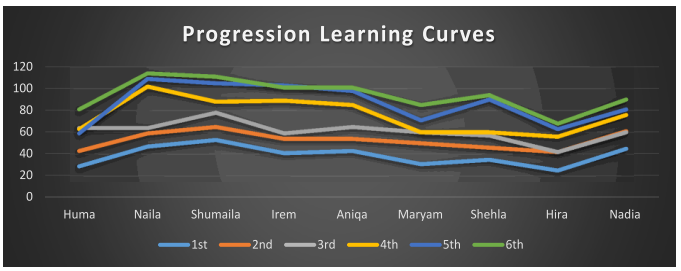

Figure 1: *Progression learning curve.*

Discussion

An assessment tool is a set of items that can measure the underlying construct.[11] Reliability is prerequisite of validity and refers to true variance in the ability of the students. The goal of psychometric analysis is to estimate and minimize, the error variance (E), so that the observed score X is a good measure of the true score.[12] Reliability coefficients, Cronbach's alpha estimate measurement error in assessment data, '0' indicate no reliability and '1' indicate no measurement error with perfect reliability.[12, 13] The Cronbach's alpha of our questionnaire was sufficiently high (0.897). Likewise Carsuzaa F et al., reported Cronbach's alpha 0.9 in a validation study of French version of OSATS.[14] If assessment scores are not reproducible with great level of certainty, the accurate interpretation of the test scores becomes questionable and validity evidence to support or refute the assessment is compromised. A study by Fouillen KJ et al., for validation of OSATS, established its construct validity with good internal consistency and inter-rater reliability, its finding are comparable to my findings.[15] Reliability also depends on number of items, sample size, observations, rating scale, objectivity and standardization.[16] In my study, for standardi-

zation and objectivity, every candidate was tested on the same skill of NVD as WPBA, with the same items, OSATS, and by the pre-trained examiner according to the same criteria of rating scale with descriptors and scoring rubric.

The largest threat to the reproducibility is rater inconsistency. Intraclass correlation coefficient (ICC) determines if items can be rated reliably by different raters.[17] The ICC '0' indicates no reliability among raters and '1' means perfect reliability. My study showed good rater consistency with ICC score (0.801- 0.988). This compares well to a research carried out by Maybodi FR et al., on applicability of OSATS (ICC = 0.99), where no significant difference was found between the checklist scores of the two raters.[18] In a validation study of OSATS by Schmidt MW et al., excellent inter-rater reliability by ICC was calculated as (0.923–0.924, $p < 0.001$).[6]

Construct validity that tool can measure which it intends to measure, is the whole validity, it requires different sources of evidence; content, response process and internal structure. Content validity was ensured by obtaining assessors agreement for OSATS items. This tool is representative of our demographic needs and represent curricula of doctors and staff, as the skill of normal vaginal delivery is taught to them by going through the four training levels; observation, to independent performance. The internal structure of the OSATS demonstrated good evidence of homogeneity in the items with a positive correlation. This OSATS validation study showed good discrimination between performances of candidate on NVD skill at different stages of training. Highest scores were taken by senior most candidates, indicating that OSATS based decisions were accurate. Likewise, Ramazani F et al., argued that increasing scores on OSATS correlate well with level of training.[19] Results also depicted a significant improvement in every candidate's performance on OSATS with increasing procedural experience giving another evidence of construct validity. This is consistent with reports from other OSATS researches stating that an increased experience and training, leads to better competency.[19,20]. Validity evidence of this study correlate well with many international studies where OSATS is considered the best tool.[21] Navrazhina K et al., reported significant correlation between trainee's scores and year of training. They also reported increasing scores on rating scales ($P = .003$) and checklist ($P = .04$) with more advanced procedural training with OSATS.[22] In another study by Chavescampos ME et al., for the construct validation, OSATS scores showed the capacity to differentiate the performance between experts and beginners as in my research.[23] Response process: For, fit for purpose assessment tool construction, its domain, purpose, and construct was defined. For OSATS instrument generation modified Delphi's model was used. With given reliability and validity evidences, OSATS, seems to be an effective instrument that generated objective scores for assessing the competence of technical skill of vaginal delivery as in international studies.[24] This study is also comparable to local study conducted in Services Institute of Medical Sciences, where OSATS was validated for formative assessment in Gynaecology.[25] Timely and constructive feedback is essence of workplace based assessments (OSATS) which promote reflective practice.[26] This study was limited to single hospital which may be reflected as strength as it limits extraneous elements which may affect the evaluator and candidate's performance assessment on the other side, limited number of observed candidates barred a comprehensive analysis.

Conclusion

With these evidences, OSATS was found to be valid and reliable for assessing the competence of technical skills of vaginal delivery in WPBA but for its acceptance as summative assessment further validation and time is required.

Conflict of Interest: None

Funding Source: None

References

1. Strandbygaard J, Scheele F, Sørensen JL. Twelve tips for assessing surgical performance and use of technical assessment scales. Med Teach 2017 Jan 2; 39(1):32-7
2. Khaliq T. Reliability of Results Produced Through Objectively Structured Assessment of Technical Skills (OSATS) for Endotracheal Intubation (ETI). JCPSP 2013; 23(1): 51-5.
3. Norcini JJ. Work based assessment ABC of learning and teaching in medicine. BMJ 2003; 326:753-5.
4. Zoller A, Holle T, Wepler M, Radermacher P, Nussbaum BL. Development of a novel global rating scale for objective structured assessment of technical skills in an emergency medical simulation training. BMC Medical Education 2021 Dec; 21:1-8.
5. Martin JA, Regehr G, Reznick R, Macrae H, Murnaghan J, Hutchison C, Brow M. Objective structured assessment of technical skill (OSATS) for surgical residents. British Journal of Surgery 1997 Feb; 84(2):273-8.

6. Schmidt MW, Haney CM, Kowalewski KF, Bintintan VV, Abu Hilal M, Arezzo A, et al. Development and validity evidence of an objective structured assessment of technical skills score for minimally invasive linear-stapled, hand-sewn intestinal anastomoses: the A-OSATS score. Surgical endoscopy 2022 Jun; 1:1-3.
7. Asif H, McInnis C, Dang F, Ajzenberg H, Wang PL, Mosa A, et al. Objective Structured Assessment of technical skill (OSATS) in the Surgical Skills and Technology Elective Program (SSTEP): comparison of peer and expert raters. The American Journal of Surgery 2022 Feb 1; 223(2):276-9.
8. Hatala R, Cook DA, Brydges R, Hawkins R. Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): a systematic review of validity evidence. Adv in Health Sci Educ 2015; 20:1149–75.
9. Antomarchi J, Delotte J, Jordan A, Tran A, Bongai A. Development and validation of an objective structured assessment of technical skill tool for the practice of vertex presentation delivery. Arch Gynecol Obstet 2014; 290:243–7.
10. Treil L, Neumann N, Chanes N, Lejay A, Bourcier T, Bismuth J, et al. Objective Evaluation of Clock Face Suture Using the Objective Structured Assessment of Technical Skill (OSATS) Checklist. In EJVES Vascular Forum 2022 Jan 1; 57: 5-11.
11. Kyriazos TA, Stalikas A. Applied Psychometrics: The Steps of Scale Development and Standardization Process. Psychology 2018; 9: 2531-60.
12. Steger D, Jankowsky K, Schroeders U, Wilhelm O. The road to hell is paved with good intentions: How common practices in scale construction hurt validity. Assessment 2023 Sep; 30(6):1811-24.
13. Amirrudin M, Nasution K, Supahar S. Effect of variability on Cronbach alpha reliability in research practice. Jurnal Matematika, Statistika dan Komputasi 2021; 17(2):223-30.
14. Carsuzaa F, Payen C, Gallet P, Favier V. French translation and validation of the OSATS tool for the assessment of surgical skill. Journal of Visceral Surgery 2023 Dec 1; 160(6):402-6.
15. Fouillen KJ, Duncan HF, de Vries P, Chevalier V. Development and preliminary validation of an Objective Structured Assessment of Technical Skills (OSATS) for a partial pulpotomy procedure. International Endodontic Journal 2023 May 25; 56(8):1011-21.
16. Sürücü L, Maslakci A. Validity and reliability in quantitative research. Business & Management Studies: An International Journal. 2020 Sep 25; 8(3):2694-726.
17. Xue C, Yuan J, Lo GG, Chang AT, Poon DM, Wong OL, et al. Radiomics feature reliability assessed by intraclass correlation coefficient: a systematic review. Quantitative Imaging in Medicine and Surgery. 2021 Oct; 11 (10):4431.
18. Maybodi FR, Keshmiri F, Kazemipoor M, Owlia F. Assessment of suturing and scaling skills of periodontology and oral medicine residents by OSATS method: a pilot study. BMC Medical Education 2023 Nov 21; 23(1):889.
19. Ramazani F, Wright ED, Randall DR, Lin JR, Jeffery CC. Developing an Objective Structured Assessment of Technical Skills (OSATS) for Microlaryngoscopy. The Laryngoscope 2023 Feb 23.
20. Jakubowski L, Leader BA, Ishman SL, Chun R, Tarima SS, Parikh S, Skinner M, Thorne M, Weatherly R, Wiet G, Brown DJ. Validation of an objective assessment tool for tonsillectomy in otolaryngology resident training. The Laryngoscope. 2021 Feb; 131(2):E359-66.
21. Vaidya A, Aydin A, Ridgley J, Raison N, Dasgupta P, Ahmed K. Current status of technical skills assessment tools in surgery: a systematic review. Journal of Surgical Research. 2020 Feb 1; 246:342-78.
22. Navrazhina K, Alam M, Tung R, Cressey BD, Decker A, Surprenant D, et al. A blinded, multirater and multi-institutional study evaluating the Objective Structured Assessment of Technical Skills (OSATS) tool in dermatology education. JAAD 2021 Nov 1; 85(5):1346-8.
23. Chavescampos ME, de Oliveira MM, de Assis LB, Reis AB, Gonçalves FB. Validation of the objective structured assessment of technical skill in brasil. Revista da Associação Médica Brasileira 2020; 66(3):328-33.
24. Benmansour M, Malti A, Jannin P. Deep neural network architecture for automated soft surgical skills evaluation using objective structured assessment of technical skills criteria. International Journal of Computer Assisted Radiology and Surgery. 2023 Jan 25; 18:1-9.
25. Yunus S, Naqi A. OSAT (Objective Structured Assessment of Technical Skill) as a Formative Assessment Tool in Gynecology. Esculapio - JSIMS 2023; 19(04): 430-4 https://doi.org/10.51273/esc23.251319411
26. Rehman PM, Qureshi MA, Latif MZ. Experience of Dops (Direct Observation Practical Skills) By Postgraduate General Surgical Residents of Lahore. Esculapio - JSIMS 2023; 19(01):91-5 https:// doi.org/ 10. 51273/ esc23.2519119

Authors Contribution

JF: Conceptualization of Project
JF: Data Collection
JF: Literature Search
JF: Statistical Analysis
JF: Drafting, Revision
JF: Writing of Manuscript