Objective Evaluation of Clock Face Suture Using the Objective Structured Assessment of Technical Skill (OSATS) Checklist

Lucille Treila, Nicole Neumann a, Nicolas Chanes d, Anne Lejay a, Tristan Bourciera, Jean Bismuth f, Jason T. Lee e, Malachi Sheahan d, Anne-Florence Rouby a, Nabil Chakfé a, * Collaborators

a GEPROMED, Strasbourg, France
b Department of Vascular Surgery and Kidney Transplantation, University Hospital of Strasbourg, Strasbourg, France
c UNISIMES (UNIté de SIMulation Européenne en Santé), Université de Strasbourg, Faculté de Médecine, Strasbourg, France
d Division of Vascular and Endovascular Surgery, Department of Surgery, Louisiana State University Health Sciences Centre, New Orleans, LA, USA
e Department of Ophthalmology, University Hospital of Strasbourg, Strasbourg, France
f Houston Methodist DeBakey Heart & Vascular Centre, Houston, TX, USA
g Division of Vascular Surgery, Stanford, CA, USA

Objective: Assessment of the quality of the final product (QFP) is critical in simulation training, such as the clock face suture (CFS) exercise that is used to assess trainees’ needle handling and suturing accuracy. Objective Structured Assessment of Technical Skill (OSATS) scores are the gold standard for the evaluation of trainees. The aim was to investigate variability in the use of OSATS checklists and to evaluate a semi-automatic method of suture analysis vs. OSATS scores.

Methods: Details of 287 CFSs performed by trainees during Fundamentals in Vascular Surgery examinations were collected. All were rated according to a seven item OSATS checklist, including QFP score and an overall score by one or two expert surgeons immediately after completion. Interassessor variability was assessed for the CFS that were assessed by two assessors.

In order to assess intra- and interassessor variability, 50 CFS pictures were chosen randomly and submitted to three expert surgeons to rate the QFP twice and to carry out a semi-automatic image analysis of each CFS and the estimated cumulative error (CE; mm) recorded. It was hypothesised that the CE correlates to OSATS checklist items or overall score. Variables were compared for correlation with OSATS results using a linear regression. A Pearson’s test was used to confirm the proposed hypothesis.

Results: Mean ± standard deviation overall score for the OSATS checklist was 20.61 ± 6.33. Inter- and intra-assessor correlation were statistically significant regarding OSATS checklist items. Both correlations presented a low coefficient of determination, indicating variability. The mean CE was 16.07 ± 4.84 mm, and the correlation between the QFP and CE was statistically significant, proving that CE is an objective metric by which to assess the QFP.

Conclusion: OSATS score demonstrated intra- and interassessor variability, although there was a significant correlation between scores. CE is an objective metric that is not subject to assessor subjectivity or interassessor variability and is correlated with the gold standard of evaluation.

© 2022 Published by Elsevier Ltd on behalf of European Society for Vascular Surgery. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Article history: Received 11 January 2022, Revised 30 September 2022, Accepted 6 October 2022,

Keywords: Education, Suture analysis, Vascular surgery

INTRODUCTION

Surgical training is changing significantly owing to restrictions on the number of working hours for trainees, especially in Europe.1 In parallel, open vascular surgery (OVS) is seeing a reduction in total procedure volume with the development of endovascular surgery.2 This raises the concern of their being fewer opportunities for trainees to see or perform some or all of an OVS procedure in recent years.3,4 Practical bench tests and simulation programs should help close the gap in trainee education and encourage self confidence in skill acquisition.5–8 It is important to consider simulation in surgical training not as a change in the expected competency of trainees, but as a new means of attaining competency.9

However, simulation is still met with scepticism, often with regard to the way it is implemented in training.
programmes and a lack of standardisation. These inconsistencies were addressed by the creation and development of two basic skills simulation programs in the USA: the Fundamentals of Vascular Surgery and the Fundamentals in Endovascular Surgery. These programs were developed by the Education Committee for the American Vascular Surgery Director, and have demonstrated an increased in proficiency in several open and endovascular surgical procedures. One of the exercises used during these simulation programs is the clock face suture exercise. It is a basic skill exercise to practice needle placement and suture precision.

The Objective Structured Assessment of Technical Skills (OSATS) has been validated to evaluate efficiently the acquisition of technical surgical skills. OSATS comprises seven task specific items, an overall score (OS), and a global rating score (GRS), which have both been found to be valid methods of assessment, although the GRS presents better reliability. This assessment tool remains limited in its application in surgical training programmes mainly due to expert assessor availability.

Since 2005, the European Board of Vascular Surgery has incorporated technical skills evaluation in bench models in their examination. Before obtaining certification in, for example, the European Board of Vascular Surgery Examination, trainee need regular detailed feedback to further their progress. It has been shown that simulation with expert feedback is more beneficial than without it. Therefore, there is a growing need for objective and less demanding ways to evaluate trainees. A semi-automatic method with which to evaluate suture precision using the clock face exercise is proposed.

The goal of this study was to evaluate variability in the evaluation associated with the OSATS checklist and propose a semi-automatic way to evaluate the clock face suture using the cumulative error (CE). The results of the suture evaluation were then correlated with the OSATS checklist score. It was hypothesised that objective evaluation of error correlates with the global OSATS rating score, the current gold standard in surgical evaluation, as well as with individualised items.

**MATERIAL AND METHODS**

**Materials**

Stamped on 10 cm wide × 10 cm long × 0.5 mm thick expanded polytetrafluoroethylene (ePTFE) patches (W.L. Gore & Associates, Newark, DE, USA), the clock face model consists of two concentric circles with diameters of 40 mm and 60 mm, respectively. The centre of these circles is crossed by six lines, which are evenly distributed circumferentially, providing 12 branches. For the suturing exercise, the patch is stretched with four clamps at a height of 13 cm on a 16.5 cm diameter plastic cylinder, allowing duplication of the restricted space in OVS. Standard surgical instruments, such as a needle holder and a forceps, and 4-0 monofilament Prolene (Ethicon, Bridgewater, NJ, USA) are used to perform the suture.

**Clock face suture exercise and associated Objective Structured Assessment of Technical Skills checklist score**

To complete a clock face suture exercise, the participant must start the suture by entering the needle and the thread at the intersection of the inner circle with the first branch and pulling out the needle at the intersection of the outer circle with the same branch. The motion continues by entering and pulling out the needle and the thread for each branch, one after the other, in a clockwise direction. This clock face suture is considered to be finished when the thread is tightened and the needle comes out at the intersection of the outer circle with the twelfth branch (Fig. 1A). Each participant was given 10 minutes to complete the clock face suture. If it was not completed within 10 minutes, or if the rules were not followed, the suture was classified as a failure (Fig. 1B).

Evaluation of each clock face suture was based on the OSATS checklist, including seven performance criteria, rated from one (lowest score) to five (highest score) by one or two independent senior vascular surgeon(s) in person on the day of suturing. The seven evaluation items were: respect for tissue; time and motion; instrument handling; knotting and suturing; use of assistant; procedural flow; and the quality of final product. An OS, rated on a scale of 7–35, was obtained by adding the score for these seven items. A GRS, scored from zero (lowest level) to four (highest level), was also given by each assessor. Even if the clock face suture was not completed or the rules not respected, the participant was still assessed.

**Study design**

Participants were asked to perform a clock face suture according to the pre-specified rules. Participants were recruited during the Louisiana State University Fundamentals of Vascular Surgery Symposium in New Orleans, USA. Some participants had attended this symposium several times and had thus performed several clock face sutures. As the study focused on the assessment of suture precision and not the evolution of suture precision, every repeated performance of a clock face suture by the same participant was included for analysis.
Between 2014 and 2019, 287 clock face sutures were performed by different vascular surgery residents, at different residency levels. Owing to some repeated participation of students, 258 participants performed 287 clock face sutures. Three different performance assessments were performed: the initial assessment at the time of completion of the exercise by each participant, the other two after completion of the exercise.

**Performance assessment**

Initial assessment was performed at the time of completion. The number of assessors depended on faculty availability. If more than two assessors were present, each of them rated the participant’s performance independently.

In order to extend the subjectivity analysis, 50 clock face sutures were given to three experts for independent review. These experts rated the quality of the final product (QFP) on the OSATS checklist. This specific item focuses on the final result of the exercise and therefore the precision of the performed suture. The experts evaluated each picture twice. Each review included all the images in a different, randomised order with an interval of one week between assessments. This additional review was designed to research the reliability of the OSATS checklist with the same assessor rating the same suture several times (i.e., intra-assessor correlation).

Evaluation of each clock face suture was then performed semi-automatically, based on a photo of the achieved clock face suture with a graduated rule (Fig. 1). After a calibration stage to compute the pixel to millimetre ratio, the CE was computed by successively adding the difference between the optimal entry point (intersection of the circle and the branch) and the real entry point of the thread. Based on the photo of the clock face suture, the CE of each suture was computed. The CE studied here was the sum of the total error on the inner circle, where the needle and thread entered the ePTFE patch, and the total error on the outer circle, where the needle and thread were pulled out from the patch. The error was defined as the difference between the optimal entry and exit point (intersection of the circle and the branch) and the real entry and exit point of the thread.

**Statistical analysis**

The OSATS score and CE were expressed as median (interquartile range) for each data set. Statistical analysis was performed using the statistical functions of the scipy.stats module of Python (Python Software Foundation, Wilmington, DE, USA). Statistical correlation tests were performed with Pearson’s test to analyse the linear correlation between two measures. For intra-assessor correlation, the QFP score from each review was used. The QFP score from the initial evaluation was used for analysis of correlation coupled with the corresponding CE. Results were considered to be statistically significant when the \( p \) value was \(< .050\).

**RESULTS**

Between 2014 and 2019, 287 clock face sutures were performed and assessed. Seventy-six per cent (\( n = 217 \)) of participants successfully completed the exercise and 24% (\( n = 70 \)) failed. Of these 70 participants, 60 did not complete the exercise on time (i.e., did not finish their suture within 10 minutes) and 10 did not follow the instructions (e.g., by suturing counter clockwise or by skipping a branch). On the day of suturing, 47% (\( n = 134 \)) of participants were evaluated by only one assessor; the other 53% (\( n = 153 \)) were evaluated by two assessors. Mean \( \pm \) standard deviation OS for the OSATS checklist was 20.61 \( \pm \) 6.33. Scores for each item were also collected independently, resulting in a mean score for each item: respect of tissue 3.02 \( \pm \) 0.91; time and motion 2.83 \( \pm \) 1.04; instrument handling 2.82 \( \pm \) 1.04; knotting and suturing 2.93 \( \pm \) 1.00; use of assistant 2.97 \( \pm \) 1.03; procedural flow 3.05 \( \pm \) 0.95; QFP 3.00 \( \pm \) 1.01.

**Interassessor correlation**

The interassessor analysis was performed on the 153 clock face suture that were evaluated by two assessors on the day of suturing. Fig. 2(A) provides the score distribution between both assessors for each of the seven items of the OSATS checklist. According to the colour map, one can visualise how many times assessor 1 gave the score X and assessor 2 the score Y. The more the pixel becomes dark blue, the more often the pair of grades (X,Y) was given, and the more it becomes light green, the less often the same pair of grades was given. There was a linear correlation between the grades of assessors 1 and 2. The same graphic has been plotted in Fig. 2(B) to assess OS.

The assessors’ evaluations of each of the seven items of the OSATS checklist was highly correlated (\( p < .001 \)). For individual checklist items, the linear regression slope ranged from 0.429 to 0.612, and \( r^2 \) ranged from 0.179 to 0.364. For OS, association was significant (\( p < .001 \)) with a linear regression slope of 0.597 and \( r^2 \) of 0.387.

**Intra-assessor variability**

Based on the acquired pictures, three senior vascular surgeons rated the QFP of 50 randomly chosen clock face sutures twice, with at least two calendar days between assessments. Fig. 3(A) provides the score distribution for this double assessment achieved by three assessors. The colouring of these matrices suggests a linear correlation between both assessments for each assessor and applying the Pearson test confirms the linear correlation between two successive assessments of the QFP by a single assessor (\( p < .001 \)), with the linear regression slope ranging from 0.727 to 1.02, and \( r^2 \) ranging from 0.519 to 0.621.

Fig. 3(B) shows the score distribution between the initial assessment on the day of suturing and the later assessment of each of the vascular surgeons. Among the 50 clock face sutures that were assessed later by three senior vascular
surgeons (assessors 1—3) looking only at the photo of the achieved suture awarded, on average, 0.40, 0.68, and 0.72 points out of 5, respectively, more than the initial on site assessor for the QFP.

Quantitative analysis
The CE was given for 277 participants; the results of the 10 participants who did not respect the suturing rules were not included. The mean CE was 6.80 ± 1.88 mm on the inside circle vs. 9.28 ± 3.66 mm on the outside circle. The mean total CE was 16.07 ± 4.84 mm.

The CE on the inside circle (see blue curve in Fig. 4A) was lower than the CE on the outside circle (see green curve in Fig. 4A).

The CE is a quantitative and, above all, objective metric that can be correlated with the QFP given by the initial assessor (see Fig. 4B). According to the Pearson and the Spearman tests, these variables were highly correlated (p < .001), whether a linear or a non-linear model is considered, indicating that the CE is a reliable metric with which to assess the QFP result.

DISCUSSION
The results indicate that evaluation using an OSATS checklist is significantly correlated between assessors and for the same assessor, but it still introduces variability in student assessments as there is an intrinsic subjectivity to this type of evaluation. This brings forward one of the main limitations of the current gold standard for evaluating performances in a simulation exercise like the clock face suture. The CE proposed here is a new objective metric that is significantly correlated to the current gold standard of evaluation, the QFP score, making it a reliable means of estimating suture precision.

Winckel et al. demonstrated that the Structured Assessment for Technical Skills is reliable; however, they did not use the coefficient of determination to test for variability between assessors. Indeed, the interassessor correlation appears to be linear in Fig. 1. However, the slope computed from a simple linear regression is always far from 1, proving that there is an important variability between assessors. The coefficient of determination, \( r^2 \), is low for all the items, which also shows that the linear model was not adapted to the collected data. Regarding intra-assessor correlation, which should be perfect as it is the same expert assessing the same picture, the coefficient of determination, \( r^2 \), evaluating the choice of the linear model, was, again, far from 1. Today, with technological advances it is necessary to develop objective measures to evaluate technical skills.

Rater bias is one of the main limitations of the OSATS, which has already been raised by Martin et al. When looking at interassessor correlation of the QFP score between the initial evaluation and the expert review, it was shown that the assessor might be influenced by watching the procedure instead of focusing solely on the final product. This is a new bias that has not been studied before and merits further exploration. The assessor might give a lower score on every item if one is poorly acquired, or unconsciously inflate every score if one item is well executed.

Finally, a correlation was found between the CE and the QFP of the OSATS checklist. No significant correlation between the CE and the other checklist items was found, indicating that the objective measure is correlated with one specific item of the OSATS checklist. It is logical that one measure does not represent the whole evaluation process. It shows the need for other objective measures correlated with items of the checklist to provide students with a well-rounded evaluation. In this study, the sample size was a limitation in finding a correlation between the CE and QFP, as the coefficient of determination remains low, which is why the decision was made to include all completed clock face sutures in the analysis, even if they were performed by a student who had already participated in the symposium.

Figure 2. (A) Score distribution of two assessors for the seven items of the OSATS checklist and the global rating score (GRS). (B) Score distribution of both assessors for the GRS.
In OVS simulation, objective metrics have been validated as reliable and able to differentiate between senior surgeons and residents using electromagnetic motion capture or, more recently, optical motion capture. Anastomosis quality has been assessed by objective metrics such as the total amount of leakage when infused with saline, this time with non-correlated results. However, they have not been validated against the gold standard for most final exam validations, which are the OSATS type checklist. The decision was made to start by comparing the metric to an OSATS type checklist; studies of its sensitivity in discriminating between experts and novices are to come. Moreover, it remains a costly investment with lot of single use materials. Consequently, it is not possible for

![Figure 3](image_url)

Figure 3. (A) Score distribution for the double assessment of 50 randomly chosen clock face sutures by three senior vascular surgeons. (B) Score distribution of quality of final product between the initial assessment and a later assessment by three senior vascular surgeons.

![Figure 4](image_url)

Figure 4. (A) Total cumulative error (CE) and CE for the inner and outer circle for all the participants. (B) Total CE as a function of the quality of final product score given by the initial assessor.
every surgery training programme to invest in those technologies. The aim is to develop a metric that is easy to obtain, inexpensive, reproducible, and correlated with the OSATS checklist.

CONCLUSION

The QFP score given by the OSATS checklist entails significant intra- and interassessor variability, even when employing assessors involved in education on a regular basis. Moreover, this specific item can be influenced by the score given to previous items. This underlines the current need for objective metrics in simulation settings. In the case of the clock face suture exercise studied in this paper, the objective measure of the Cumulated Error is correlated with the QFP score and is a promising metric with which to reflect suture precision.

ACKNOWLEDGEMENTS

Our studies are partially supported by the Eurometropole de Strasbourg and the région Grand’Est without conflict of interest with industry.

APPENDIX

Collaborators:
John Eid1, Yannick Georg2, Erica Leith Mitchell3, David Rigberg4, Murray Shames5, Fabien Thaveau2,6, Claudie Sheahan6
1. Texas Vascular Associate, Houston, USA
2. Department of Vascular Surgery and Kidney Transplantation University Hospital of Strasbourg, Strasbourg, France
3. Department of Vascular and Endovascular Surgery, University of Tennessee Health Science, Memphis, USA
4. Department of Vascular Surgery, University of California Los Angeles, USA
5. Department of Vascular Surgery, Tampa General Hospital, Tampa, USA
6. GEPROVAS (Groupe Européen de Recherche sur les Prothèses Appliquées à la Chirurgie Vasculaire), Strasbourg, France
7. Division of Vascular and Endovascular Surgery, Department of Surgery, Louisiana State University Health Sciences Center, New Orleans, LA, USA

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ejvsvf.2022.10.001.

REFERENCES

1. Pickersgill T. The European working time directive for doctors in training. BMJ 2001;323:1266.
2. Schanzer A, Steppacher R, Eslami M, Arous E, Messina L, Belkin M. Vascular surgery training trends from 2001—2007: a substantial increase in total procedure volume is driven by escalating endovascular procedure volume and stable open procedure volume. J Vasc Surg 2009;49:1339—44.
3. Cronenwett JL. Vascular surgery training: is there enough case material? Semin Vasc Surg 2006;19:187—90.
4. Dua A, Koprowski S, Uphchur G, Lee CJ, Desai SS. Progressive shortfall in open aneurysm experience for vascular surgery trainees with the impact of fenestrated and branched endovascular technology. J Vasc Surg 2017;65:257—61.
5. Bath J, Lawrence P. Why we need open simulation to train surgeons in an era of work-hour restrictions. Vascular 2011;19:175—7.
6. Sidhu RS, Park J, Brydges R, MacRae HM, Dubrowski A. Laboratory-based vascular anastomosis training: a randomized controlled trial evaluating the effects of bench model fidelity and level of training on skill acquisition. J Vasc Surg 2007;45:343—9.
7. Egle JP, Malladi SVS, Gopinath N, Mittal VK. Simulation training improves resident performance in hand-sown vascular and bowel anastomoses. J Surg Educ 2015;72:291—6.
8. Tavlasoglu M, Durukan AB, Gurbuz HB, Jahollari A, Guler A. Skill acquisition process in vascular anastomosis procedures: a simulation-based study. Eur J Cardiothorac Surg 2015;47:812—8.
9. Boyle JR, Ghysen A, d’Orío V, Vega de Ceniga M, Chalké F. Time to embrace simulation in vascular surgical training: little things matter. Eur J Vasc Endovasc Surg 2019;58:161—2.
10. Duran C, Estrada S, O’Malley M, Sheahan MG, Shames ML, Lee JT, et al. The model for Fundamentals of Endovascular Surgery (FEVS) successfully defines the competent endovascular surgeon. J Vasc Surg 2015;62:1660—1666.e3.
11. Sheahan MG, Duran C, Bismuth J. National simulation-based training of fellows: the vascular surgery example. Surg Clin North Am 2015;95:781—90.
12. Martin JA, Regehr G, Reznick R, Macrae H, Murnaghan J, Hutchison C, et al. Objective Structured Assessment of Technical Skill (OSATS) for surgical residents. Br J Surg 1997;84:273—8.
13. Nayahangan LJ, Lawaetz J, Strøm M, de la Motte L, Rørdam P, Gottschalksen BC, et al. Ensuring competency in open aortic aneurysm repair — development and validation of a new assessment tool. Eur J Vasc Endovasc Surg 2020;59:767—74.
14. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998;73:993—7.
15. Moorthy K. Objective assessment of technical skills in surgery. BMJ 2003;327:1032—7.
16. Pandey VA, Wolfe JHN, Lindahl AK, Rauwerda JA, Bergqvist D. Validity of an exam assessment in surgical skill: EBSQ-VASC pilot study. Eur J Vasc Endovasc Surg 2004;27:341—8.
17. Sidhu RS, Grober ED, Musselman LJ, Reznick RK. Assessing competency in surgery: where to begin? Surgery 2004;135:6—20.
18. Rogers DA, Regehr G, Howdieshell TR, Yeh KA, Palm E. The impact of external feedback on computer-assisted learning for surgical technical skill training. Am J Surg 2000;179:341—3.
19. Robinson WP, Baril DT, Taha O, Schanzer A, Larkin AC, Bismuth J, et al. Simulation-based training to teach open abdominal aortic aneurysm repair to surgical residents requires dedicated faculty instruction. J Vasc Surg 2013;58:247—253.e2.
20. Winckel CP, Reznick RK, Cohen R, Taylor B. Reliability, and construct validity of a structured technical skills assessment form. Am J Surg 1994;167:423—7.
21 Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J Am Coll Surg* 2001;193:479–85.

22 Bann SD, Khan MS, Darzi AW. Measurement of surgical dexterity using motion analysis of simple bench tasks. *World J Surg* 2003;27:390–4.

23 Sun X, Byrns S, Cheng I, Zheng B, Basu A. Smart sensor-based motion detection system for hand movement training in open surgery. *J Med Syst* 2016;41:24.

24 Jensen AR, Milner R, Achild O, Gaughan J, Wilhite DB, Grewal H. Effective instruction of vascular anastomosis in the surgical skills laboratory. *Am J Surg* 2008;195:189–94.