A novel simulator-based checklist for evaluating residents' competence in cerebral angiography in China

Xuxia Yi[1], Gang Wang[2], Nai Zhang[3], Weixin Si[4]* and Jianping Lv[1]*

[1]Department of Neurosurgery, Guangzhou First People's Hospital, Guangzhou, China, [2]National Center for Mental Health, China, Beijing, China, [3]Department of Neurosurgery, Tianjin Medical University General Hospital, Tianjin, China, [4]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Background: Nowadays, with the fast-increasing demand for neuro-endovascular therapy, surgeons in this field are in urgent need. Unfortunately, there is still no formal skill assessment in neuro-endovascular therapy in China.

Methods: We used a Delphi method to design a newly objective checklist for standards of cerebrovascular angiography in China and evaluated its validity and reliability. A total of 19 neuro-residents with no interventional experience and 19 neuro-endovascular surgeons from two centers (Guangzhou and Tianjin) were recruited; they were divided into two groups: residents and surgeons. Residents completed a simulation-based cerebrovascular angiography operation training before assessment. Assessments were under live and video record forms with two tools: the existing global rating scale (GRS) of endovascular performance and the new checklist.

Results: The average scores of residents were significantly increased after training in two centers ($p < 0.05$). There is good consistency between GRS and the checklist ($p = 0.856$). Intra-rater reliability (Spearman's rho) of the checklist was >0.9, and the same result was also observed in raters between different centers and different assessment forms ($p < 0.001$, rho > 0.9). The reliability of the checklist was higher than that of the GRS (Kendall's harmonious coefficient is 0.849, while GRS is 0.684).

Conclusion: The newly developed checklist appears reliable and valid for evaluating the technical performance of cerebral angiography and differentiating between trained and untrained trainees' performance well. For its efficiency, our method has been proven to be a feasible tool for resident angiography examination in certification nationwide.

KEYWORDS

checklist, reliability, GRS of endovascular performance, cerebral angiography, simulation

1. Introduction

In most vascular surgery centers, endovascular treatments have now superseded open surgical repair in the management of a large number of pathologies (1). Neuro-interventional surgery has been carried out in more than 700 hospitals in China, nearly 3,500 physicians were engaged in this field by 2018, and 124,100 operations were performed nationwide in 2019 (2). The increase in therapeutic endovascular treatment options has also led to a need for training in endovascular skills for the practitioners of the future. With the development of standardized residency training in China since 2014, most of the fresh interventional surgeons were trained under a residency program. Limited by several factors, qualification evaluation for

neuro-interventional surgeons still remains like written mold, although government authorities in China demand that doctors who prepare to be neuro-interventional surgeons be trained and qualified (3).

These factors include the uneven distribution and development of medical resources and competencies of physicians across different regions in China (4), work-hour restrictions and changes in medicine (5), and the complexity of clinical assessment, involved ethics and fairness. These traditional examination methods only assess basic theoretical knowledge and do not evaluate technical performance (6), which cannot ensure a licensed and certified surgeon is capable of delivering quality care. In addition to the above difficult situation, the assessment of methodical problems and a blank assessment tool also block the implementation of the qualifying policy.

Simulation-based assessment (SBA) is a great idea to solve assessment methodical problems, which has been incorporated in a number of high-stakes certification and licensure examinations (7–9). At present, many intervention centers or simulation centers have developed their assessment tools (9–12) for training and assessment or even certification and licensure examinations. The two most commonly used are checklists and GRS, and the most widely used is the GRS of endovascular performance, which is adapted from a previously validated scoring system (13). There were two common limitations among these assessment tools, which are as follows: (1) not unique to diagnostic neurovascular intervention, they can be applied to all interventional procedures; and (2) rater-dependent, GRS is subjective, which allows a rater to evaluate the degree (on a 1 to 5 scale) of a learner performing all steps in a given assessment exercise (13). A Likert rating scale (0–4; 4 being best) was used for checklist grading (9), which means they were semi-subjective. Another issue is that the objective is to assess the examinees' competence to complete the surgery while ignoring the standardization of basic surgical steps.

As mentioned earlier, the development of neuro-interventional training in different centers has obvious regional differences since regional representative hospitals have their training protocols and assessment methods. There is no standard method existing for skill evaluation. In the face of important assessments such as the national examination for readiness and competence for certification, it is difficult to guarantee the fairness of the results. Moreover, large-scale examiner training needs to consume a lot of time, manpower, and economic costs. These existing assessment tools are difficult to implement for such an important examination in China. To overcome subjectivity, an assessment tool should be designed so that there is little room for assessor interpretation, such as with a uniform rubric.

Therefore, with the support of the Chinese National Medical Examination Center, we conducted the following research. The newly designed objective checklist should reflect and distinguish the different levers of examinees' competency reliably. Different from the existing GRS of endovascular performance and checklists, the evaluation indicators of the scale are combined with the objective automatic metrics of the simulator, which means the checklist is more objective and has less rater bias. First, the score-based decisions must be validated and demonstrated to be reliable before using this scale to evaluate the ability of the examinee.

2. Materials and methods

This is a multi-stepped approach with mixed methods to assess the validity, reliability, and applicability of our new objective checklist for cerebrovascular angiography. The study was approved by the Committee of Guangzhou First People's Hospital, Guangzhou, China.

2.1. Development of the assessment tool

The checklist was developed by a team of nine experts from three hospitals: six key members of the Chinese Society for Neurointervention, with over 2,000 neuro-endovascular cases' experiences, two persons with experience in both interventional and open surgery fields of at least 20 years, and one expert in measurement.

We began with the literature "Chinese expert Consensus on the Operation Specification of Cerebrovascular Angiography" (14) to select initial 40 items of surgical parameters in the main procedures of cerebrovascular angiography with the addition of automatic metrics from virtual reality simulator (total fluoroscope time, total time, amount of contrast used, and handling events), which can adequately help interventionalists in grading intervention difficulty (15). Performance metrics used a Delphi method through three rounds of survey (16) to develop the consensus opinion on kept/omitted items, appropriateness of each item's content, and weight of each item. The first two rounds were *via* a video conference, and the last was conducted through offline discussion. The tool was reviewed for its completeness, relevance, and representativeness and normalized (17). The final form (maximum score = 100) included the following three domains: part 1: Prepare (five items, max score 8); part 2: Steps of the Procedure (13 items, max score 82); and part 3: Diagnosis (two items, max score 10). The final form is shown in Supplementary Table 1. We based specific detailed descriptions of angiography steps with emphasis on the entirety of the procedure, including checking patient identity, asepsis, handling catheter and wire, the capture of the standard picture, correct use of contrast, and reading angiogram.

2.2. Assessment procedure

A total of 38 vascular residents and neurosurgeons from the Department of Neurosurgery, Guangzhou First People's Hospital, and Tianjin Medical University General Hospital were recruited for this research. The trainees were divided into two groups: residents (residents with no endovascular experience) and surgeons (neurosurgeons with endovascular experience of 100–200 cases). The residents received a 2-day (3 h) endovascular skill training course, including didactic teaching and diagnostic cerebral angiography operation training on simulator Mentice Vascular Intervention Simulation Trainer (VIST; Gothenburg, Sweden, Guangzhou center) or Simbionix ANGIO Mentor (Simbionix, Cleveland, OH, Tianjin center).

Before training, a didactic course, including instruction on catheter handling, device selection, endovascular techniques, and manipulating simulators, was given to residents. The training time of all residents was 3 h. One-on-one training was provided

TABLE 1 Background information of participants.

	Residents			Surgeons		
	Guangzhou	Tianjin	Person correlation coefficient (p-value)	Guangzhou	Tianjin	Person correlation coefficient (p-value)
M	8	10		9	10	
F	1	0		0	0	
Age (mean)	26	26.2		30.1	35.3	
Clinical working years (mean)	2.8	0.6	0.329	3.6	9.5	0.444
Experience with vascular/year (mean)m	0	0	/	158	275	0.050

M, male; f, female.

TABLE 2 Comparison of mean checklist scores across each group between the two centers.

	Guangzhou				Tianjin			
	Pre-training	Post-training	Surgeon	Sig	Pre-training	Post-training	Surgeon	Sig
Prepare	5.6	7.9	6.3	0.051	4.2	7.6	7	0.003
Procedure	41.6	56.1	63.7	0.004	41	59	72.5	0.002
Diagnosis	1.6	8.4	7.3	0.013	1.1	6.2	9.2.	<0.001
Total Score	48.8	72.4	77.3	0.000	46.3	72.8	88.9	<0.001

throughout the training sessions by senior neuro-interventional doctors. Performance on cases with aortic arch type I (to avoid case difficulty bias) was used to assess early on the 1st day after didactic teaching, and this was repeated after the training by two highly experienced surgeons (mean experience of 20 years). Surgeons were also tested after learning to manipulate the simulator for 30 min.

The evaluation forms were used by on-site observation and video (screen recording) alike with two assessment tools: the checklist and the GRS of endovascular performance (Supplementary Tables 1, 2). Each examinee was evaluated by two raters (a total of four from two hospitals with a mean of 20 years of experience). The raters were trained by lecture and literally before the assessment. Screen recording recorded the overall core part of the procedure (Steps of the Procedure) to ensure an overview of the entire critical performance sequence; then, they were scored by all raters from two centers using the two tools. Video scoring was reviewed a month after direct observation (18, 19). The order of videos was randomized, so the raters were blinded to the identity of operators.

2.3. Statistical analysis

All data were imported into SPSS 21.0 for analysis. Evidence of validity was provided by comparing scores between different groups with different clinic experiences in two centers. Data of DI for the checklist were evaluated aspect of construct validity. Another piece of evidence was the relationship between the total scores obtained from the two assessment tools. The "inter-rater" and "intra-rater" reliability were tested with Spearman's rho tests. The internal consistency of the checklist was evaluated by assessing the value of Kendall's harmonious coefficient, including different assessment

forms, centers, and tools. The significance level in all hypothesis testing procedures was predetermined at $p = 0.05$.

3. Results

The basic background information such as age, sex, clinical experience, and angiography cases is listed in Table 1, and the two cohorts were identical in terms of background, except for angiography cases.

Average scores across each group at every time are shown in Table 2. The scores in the surgeons group in Tianjin were significantly higher than any other cohort ($p < 0.05$), which positively correlated with the number of angiography experiences (Table 1).

The scores were placed based on average scores per item in Table 3. The average discrimination index (DI) and difficulty coefficient (DC) of each item were also calculated. DI was more than 0.6 indicating good homogeneity. Based on data presented in Table 4, the total scores of the raters using the two tools, with both assessment forms, were gratifying reliability indices (Cronbach's α more than 0.6). The comparison with the score sheet showed that the score across the two evaluation tools had a significant positive correlation (Figure 1), indicating that the checklist tool is consistent with the GRS internationally approved.

According to data presented in Table 5, for each rater, there was a strong positive statistically significant intra-rater reliability between the two evaluation forms (for all raters, $rho > 0.9$, $p < 0.001$). The same result was also observed in raters between different centers ($p < 0.001$, $rho > 0.9$).

The vast majority of the scores of the examiners using the two tools for both periods of the assessment showed that the checklist tool was more satisfactory for intra- or extra-reliability indices than

TABLE 3 Checklist item performance averages of difficulty coefficient and differentiation index for all participants.

Items (total score)	Mean score	SD	Difficulty coefficient (DC)	Differentiation index (DI)
1 (1)	0.8	0.36	0.71	0.57
2 (1)	0.8	0.39	0.64	0.71
3 (2)	1.8	0.50	0.76	0.43
4 (3)	1.7	0.52	0.75	0.5
5 (3)	1.5	0.64	0.68	0.64
6 (6)	5.6	1.27	0.88	0.88
7 (5)	4.4	1.39	0.63	0.45
8 (6)	4.0	2.13	0.6	0.81
9 (5)	4.3	1.19	0.77	0.46
10 (6)	4.3	1.10	0.69	0.33
11 (6)	5.1	1.65	0.76	0.48
12 (6)	5.0	1.60	0.75	0.50
13 (6)	4.7	1.89	0.68	0.64
14 (6)	4.7	1.68	0.7	0.60
15 (6)	4.9	0.75	0.76	0.48
16 (12)	7.2	2.61	0.51	0.97
17 (8)	7.3	1.17	0.71	0.57
18 (5)	3.1	1.55	0.36	0.71
19 (5)	3.5	2.22	0.5	1
20 (5)	3.1	1.83	0.5	0.89

TABLE 4 Rater reliability (Cronbach's α coefficient) of on-site observation and video recording-based scores by GRS and checklist.

Raters	On-stie observation		Video recording	
	GRS	Checklist	GRS	Checklist
Rater 1	0.721	0.895	0.85	0.893
Rater 2	0.705	0.893	0.831	0.893
Rater 3	0.653	0.829	0.661	0.849
Rater 4	0.694	0.821	0.793	0.979

GRS (Table 6; Kendall's harmonious coefficient is 0.849, while GRS is 0.684).

4. Discussion

In this study, we designed and implemented a new objective checklist to assess the residents' technical ability to operate the standard procedure of cerebral angiography, for it is the basic skill of cerebrovascular intervention. This is the first assessment tool designed based on Chinese expert consensus for Chinese trainees. The assessment is based on VR simulators, which can be performed in all steps of intervention except for arterial puncture. This is because this procedure is complex and better with another assessment tool. For a simulation module to be considered valid, it should ideally be assessed for validity, and the contemporary meaning of validity is a unitary concept with multiple aspects that considers construct validity as the whole of validity (20).

Validity evidence refers to data collected to assign a meaningful interpretation of assessment scores (21). The checklist was developed stepwise by a Delphi method with nine experts, which could ensure quality content validity. Unfortunately, a rate of agreement (RoA) as a measure of consensus among the experts was not calculated in this study, which is an appropriate measure of consensus particularly when Likert scales are used (22).

The data from this study support the construct validity of the assessment scale because the scale was able to reflect the increased competence of residents, to differentiate with and without experienced trainees, or with different cases experience ($p < 0.05$). The study showed a significant improvement in all domains of residents' skills, but mainly in the steps of the procedure and then diagnosis. The detailed nature of the checklist makes it an interesting tool for training purposes or surgical plans with complicated cases, to find out the obstacle steps.

After training, there was no significant difference between experienced and non-experienced trainees in GZ, which means group residents take a shorter time to achieve the same scores compared with the surgeons group. A study by the Aggarwal team (23) showed that there is an expected learning curve in performing simulated endovascular tasks. The results also suggested that no correlation exists between an individual's operative experience as reported by case logs and their technical performance (24). Another explanation is that cerebral angiography is a basic skill for intervention, and trainees can grasp it in a short time with standard training. While there was a different result between the residents and surgeons in TJ, it may be due to the surgeons having more angiography experience. The inconsistent results indicated that the volume of operations should
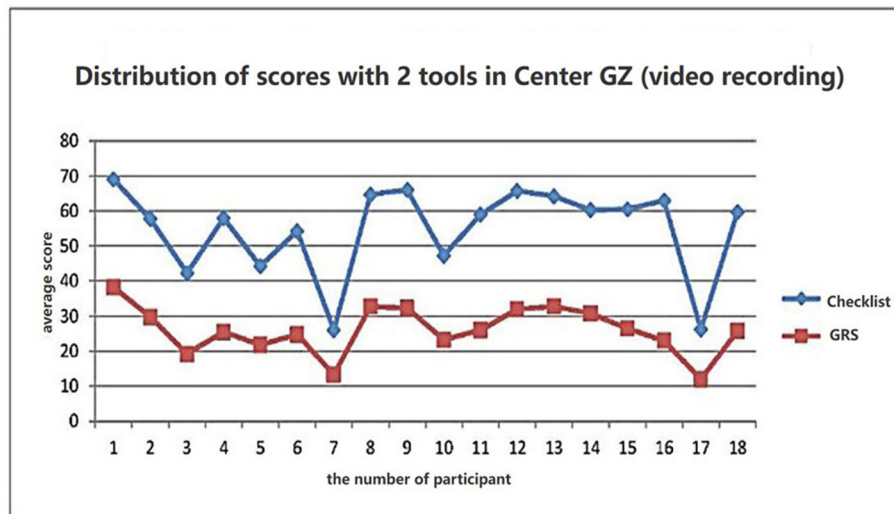
FIGURE 1
Mean score distribution of the resident group with GRS and checklist. Score differentiation between the two assessment tools, Pearson's correlation 0.9.

TABLE 5 Reliability for the mean score on different centers, intra-rater, and different tools for the same examinee inter-hospitals.

	Spearman's rho	p-Value
1. Centers		
Guangzhou (checklist)	0.975	<0.001
Tianjin (checklist)	0.930	<0.001
2. Intra-rater		
Rater 1	0.974	<0.001
Rater 2	0.933	0.053
Rater 3	0.961	<0.001
Rater 4	0.929	<0.001
3. Tools		
GRS	0.948	<0.001
Checklis	0.61	0.061

TABLE 6 Reliability results of the two tools used by the four raters.

	Kendall's harmonious coefficient	p-Value
GRS	0.684	0.001
Checklist	0.849	<0.001

not be a reliable and direct measure of technical skill (25). Of course, it will be more comprehensive if the learning curve was measured by patient outcomes (morbidity or mortality) or by procedure-specific metrics (blood loss and operative time) (26). To judge the score of the checklist can reflect the achievement of the examinee, we need to evaluate the performance of residents with the checklist in the operative room (OR) compared to simulations in the future.

As can be seen from the difference in scores before and after the training, the scores of the residents group in key items 3, 5, and 10 increased significantly after training, which affected the total score. The variation trends of 11 and 12 were consistent with the total score. The checklist not only helps identify which trainees have not yet achieved competence but also reveals specific areas of angiography in which a trainee needs further reinforcement and experience.

The scores of the experienced group were also different, especially because the score was positively correlated with the number of angiography cases, which was consistent with actual clinical experience. Meanwhile, it is combined with automatic metrics on the simulator, including "catheter scraping vessel wall" and "catheter moving without support of wire," to avoid subjective deviation or omission of examiners in key steps.

The data of correlation with GRS and Cronbach's α suggest that the checklist is not redundant, and the latter could be a useful tool in the assessment of surgical competence. Cronbach's α and Spearman's rho of GRS were both lower than that of the checklist. Similar results could be seen in other research (27, 28). Sarker et al. (29) used error-based checklists to evaluate the performance of the senior surgeons and showed high inter-rater reliability (k value 0.79–0.84; $p < 0.05$). Procedure-based assessments (6) also possessed high inter-rater reliability ($G > 0.8$, using three assessors for the same index procedure) but were very procedure-specific and long (checklist of up to 62 items), which limited its practicality for use in evaluating common endovascular procedures.

With the gradually increasing introduction of newly developed tools for the assessment of technical skills into practice, it is important to define the role of examiners as well (30). To minimize rater bias, all raters in this study were experts in neurovascular disease. The scores of the new objective checklist showed a strong positive intra-rater reliability than GRS ($rho > 0.9$). The geographic variation, non-uniform standards, and different medical education backgrounds of raters cannot be ignored after all. There was no specific and clear criterion in every item for GRS, raters exhibited more random variability than with the checklist (31). The positive results indicating the new assessment tool may be used to measure angiographic skills on the simulator with a rater bias-free. Its superiority needs to be

validated by a different level of raters, and limited experts cannot meet a large number of examinees after all.

Discussions have arisen in the past as to whether one should rate surgical performance in an on-site or video-based manner (32–34). On-site assessment enables all aspects of the procedure to be rated, and video-based assessment enables the rater to view the videos at their convenience, which has led to the increased popularity of blinded video assessment of surgical skills (33). In this study, screen recording showed the same results as on live, indicating that consistency can be guaranteed even if the same rater is reassessed over a period of time. The objective checklist could minimize the relative influence of subjection on-site or incompleteness under video recording. Another detail that should be noticed is that three residents did not get increased scores at live observation because of nervousness, and this situation did not happen during the video recording. Video recording should be improved to record the whole scene when the examinee operates beginning with Prepare.

There is no doubt as to the value of surgical skills assessment, not only for evaluating training effect and qualification but also for estimating case difficulty in the clinic. The purpose of our study is to provide the examiner with a standardized tool in licensure and certification SBA to assure the individual who passes the examination can also adapt to practice. Following the above-mentioned checklist design did bring almost negligible subjective rater bias, repeating the study with multiple centers involved may provide more significant results.

It is worth noting that the data mentioned earlier only verified to reliability and validity of the checklist. In the future, to apply for certification examination, we must need more samples of residents from different centers to determine the passing line forward. Furthermore, we have to evaluate the reliability of diverse raters, such as peers (those of the same career tract) or junior faculty, to ensure enough competent raters to meet the large examinations. In this regard, a novel and step-by-step study need to be performed.

5. Conclusion

In this study, we designed an objective assessment tool for the cerebrovascular angiography performance of residents/trainees. After that, we made a preliminary feasibility study, and the results indicated three issues: first, the score can be differentiated between trained and untrained trainees, while our proposed tool also positively correlates with the GRS. Second, the checklist assesses the delicate step of performance. Third, the new assessment tool was feasible and acceptable to both examiner and faculty with valuable high consistency.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Guangzhou First People's Hospital, Guangzhou, China.

Author contributions

JL concepted and supervised the study. XY, NZ, and GW carried out the study. XY analyzed, interpreted the data, and drafted the manuscript. WS critically revised the manuscript for important intellectual content. All authors contributed to the article and approved the submitted version.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2023.1122257/full#supplementary-material

References

1. Rudarakanchana N, Van Herzeele I, Desender L, Cheshire NJ. Virtual reality simulation for the optimization of endovascular procedures: current perspectives. *Vasc Health Risk Manag.* (2015) 11:195–202. doi: 10.2147/VHRM.S46194

2. Jun L, Daming W. An overview of the development of neuro-endovasvular Intervention in China. *Chin J Neuroimmunol Neurol.* (2019) 26:237–9. doi: 10.3969/j.issn.1006-2963.2019.04.001

3. National Health Commission of the People's Republic of China. *Chinese Technical Management Standard for Neuroendovascular Intervention Treatment.* (2019). Available online at: http://www.nhc.gov.cn/yzygj/s3585/201911/cd7563ffac1e4ceeaf465c396d990755.shtml (accessed September 26, 2019).

4. Menglin L, Gang W, Jianping L. Healthcare simulation in China: current status and perspectives. *Chin Med J.* (2019) 132:2503–4. doi: 10.1097/CM9.0000000000000475

5. Barsuk JH, Cohen ER, Feinglass J, McGaghie WC, Wayne DB. Residents' procedural experience does not ensure competence: a research synthesis. *J Grad Med Educ.* (2017) 9:201–8. doi: 10.4300/JGME-D-16-00426.1

6. Irfan W, Sheahan C, Mitchell EL, Sheahan MG 3rd. The pathway to a national vascular skills examination and the role of simulation-based training in an increasingly complex specialty. *Semin Vasc Surg.* (2019) 32:48–67. doi: 10.1053/j.semvascsurg.2018.12.006

7. Tombleson P, Fox RA, Dacre J. Defining the content for the objective structured clinical examination component of the professional and linguistic assessments board examination: development of a blueprint. *Med Educ.* (2000) 34:566–72. doi: 10.1046/j.1365-2923.2000.00553.x

8. Boyd MA, Gerrow JD, Duquette P. Rethinking the OSCE as a tool for national competency. *Evaluation.* (2004) 8:95–95. doi: 10.1111/j.1600-0579.2004.338ab.x

9. Chaer RA, DeRubertis BG, Lin SC, Bush HL, Karwowski JK, Birk D, et al. Simulation improves resident performance in catheter-based intervention: results of a randomized, controlled study. *Ann Surg.* (2006) 244:343–9. doi: 10.1097/01.sla.0000234932.88487.75

10. Maertens H, Aggarwal R, Desender L, Vermassen F, Van Herzeele I. Development of a PROficiency-Based StePwise Endovascular Curricular Training (PROSPECT) program. *J Surg Educ.* (2015) 73:51–60. doi: 10.1016/j.jsurg.2015.07.009

11. Hseino H, Nugent E, Lee MJ, Hill ADK, Neary P, Tierney S, et al. Skills transfer after proficiency-based simulation training in superficial femoral artery angioplasty. *Simul Healthc.* (2012) 7:274–81. doi: 10.1097/SIH.0b013e31825b6308

12. Gosling AF, Kendrick DE, Kim AH, Nagavalli A, Kimball ES, Liu NT, et al. Simulation of carotid artery stenting reduces training procedure and fluoroscopy times. *J Vasc Surg.* (2017) 66:298–306. doi: 10.1016/j.jvs.2016.11.066

13. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill *via* an innovative "bench station" examination. *Am J Surg.* (1997) 173:226–30. doi: 10.1016/S0002-9610(97)89597-9

14. Chinese Medical Association Neurology Branch, Chinese Medical Association Neurology Branch Neurovascular Intervention Cooperation Group. Chinese expert Consensus on the Operation Specification of Cerebral Angiography. *Chin J Neurol.* (2018) 51:7–13. doi: 10.1186/s40035-018-0116-x

15. Willaert WI, Cheshire NJ, Aggarwal R, Van Herzeele I, Stansby G, Macdonald S, et al. Improving results for carotid artery stenting by validation of the anatomic scoring system for carotid artery stenting with patient-specific simulated rehearsal. *J Vasc Surg.* (2012) 56:1763–70. doi: 10.1016/j.jvs.2012.03.257

16. McPherson S, Reese C, Wendler MC. Methodology update: delphi studies. *Nurs Res.* (2018) 67:404–410. doi: 10.1097/NNR.0000000000000297

17. McKinley RK, Hastings AM. Tools to assess clinical skills of medical trainees. *JAMA.* (2010) 303:332 doi: 10.1001/jama.2010.20

18. Schlager A, Ahlqvist K, Rasmussen-Barr E, Bjelland EK, Pingel R, Olsson C, et al. Inter- and intra-rater reliability for measurement of range of motion in joints included in three hypermobility assessment methods. *BMC Musculoskelet Disord.* (2018) 19:376. doi: 10.1186/s12891-018-2290-5

19. Siddiqui NY, Galloway ML, Geller EJ, Green IC, Hur HC, Langston K, et al. Validity and reliability of the robotic objective structured assessment of technical skills. *Obstet Gynecol.* (2014) 123:1193–9. doi: 10.1097/AOG.0000000000000288

20. Sarker SK, Kumar I, Delaney C. Assessing operative performance in advanced laparoscopic colorectal surgery. *World J Surg.* (2010) 34:1594–603. doi: 10.1007/s00268-010-0486-4

21. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ.* (2003) 37:830–7. doi: 10.1046/j.1365-2923.2003.01594.x

22. von der Gracht HA. Consensus measurement in Delphi studies review and implications for future quality assurance. *Technol Forecast Soc Change.* (2012) 79:1525–36. doi: 10.1016/j.techfore.2012.04.013

23. Van Herzeele I, Aggarwal R, Neequaye S, Darzi A Vermassen F, et al. Cognitive training improves clinically relevant outcomes during simulated endovascular procedures. *J Vasc Surg.* (2008) 48:1223–300. doi: 10.1016/j.jvs.2008.06.034

24. Moorthy K, Munz Y, Sarker SK. Objective assessment of technical skills in surgery. *BMJ.* (2003) 327:1032–7. doi: 10.1136/bmj.327.7422.1032

25. Warf BC, Donnelly MB, Schwartz RW, Sloan DA. Interpreting the judgment of surgical faculty regarding resident competence. *J Surg Res.* (1999) 86:29–35. doi: 10.1006/jsre.1999.5690

26. Cook JA, Ramsaya CR, Fayersb P. Statistical evaluation of learning curve effects in surgical trials. *Clinical Trials.* (2004) 1:421–7. doi: 10.1191/1740774504cn042oa

27. Seo S Thomas A,Uspal N. A Global rating scale and checklist instrument for pediatric laceration repair. *MedEdPORTAL.* (2019) 15:10806. doi: 10.15766/mep_2374-8265.10806

28. Stavrakas M, Menexes G, Triaridis S, Bamidis P, Constantinidis J, Karkos PD, et al. Objective structured assessment of technical skill in temporal bone dissection: validation of a novel tool. *J Laryngol Otol.* (2021) 2021:1–11. doi: 10.1017/S002215121001201

29. Sarker SK, Hutchinson R, Chang A, Vincent C, Darzi AW. Self-appraisal hierarchical task analysis of laparoscopic surgery performed by expert surgeons. *Surg Endosc Other Interv Tech.* (2006) 20:636–40. doi: 10.1007/s00464-005-0312-5

30. D'Souza N, Hashimoto DA, Gurusamy K, Aggarwal R. Comparative outcomes of resident vs attending performed surgery: a systematic review and meta-analysis. *J Surg Educ.* (2016) 73:391–9. doi: 10.1016/j.jsurg.2016.01.002

31. Iramaneerat C, Yudkowsky R. Rater errors in a clinical skills assessment of medical students. *Eval Health Prof.* (2007) 30:266–83. doi: 10.1177/0163278707304040

32. Nickel F, Hendrie JD, Stock C, Salama M, Preukschas AA, Senft JD, et al. Direct observation versus endoscopic video recording-based rating with the objective structured assessment of technical skills for training of laparoscopic cholecystectomy. *Eur Surg Res.* (2016) 57:1–9. doi: 10.1159/000444449

33. Anderson TN, Lau JN, Shi R, Sapp RW, Aalami LR, Lee EW, et al. The utility of peers and trained raters in technical skill-based assessments a generalizability theory study. *J Surg Educ.* (2022) 79:206–15. doi: 10.1016/j.jsurg.2021.07.002

34. Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AMN, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med.* (2013) 369:1434–42. doi: 10.1056/NEJMsa1300625