Technical skills continue to improve beyond surgical training

Vikas Pandey, MB, ChB,[a,b] **John H. N. Wolfe, MS,**[b] **Krishna Moorthy, MD, MBBS,**[b] **Yaron Munz, MD,**[b] **Mark J. Jackson, MD,**[a] and **Ara W. Darzi, MD,**[b] *London, United Kingdom*

Background: **There is growing focus on surgical technical competence and the means by which we are able to measure it. Ongoing studies have shown a plateau effect with increasing experience of the operator. The aim of this study was to assess the technical competence of five groups of surgeons with increasing experience and validate a new rating tool for use in surgical assessment.**

Methods: **Fifty surgeons performed a saphenofemoral junction ligation on a synthetic groin model. The procedure was videotaped, blinded, and reviewed independently by three assessors. Performance was assessed using a previously validated global rating scale of generic surgical skill. In addition, each procedure was rated with the procedure-specific Imperial College Evaluation of Procedure-Specific Skill (ICEPS) rating scale to establish the construct validity (ability to differentiate on the basis of skill) and inter-observer reliability.**

Results: **Both rating scales showed improved scores with ascending grades ($P < .001$) and demonstrated a high inter-observer reliability both for generic and procedure-specific skill ($\alpha = 0.97$ and $\alpha = 0.96$, respectively). Total operative scores demonstrated significant differences between surgeons in postgraduate years 1 and 2 and surgeons in years 3 and 4 and also between newly appointed and experienced consultants ($P < .041$). Procedure-specific performance showed a plateau effect at the registrar level. Generic skill continued to improve, and significant differences were seen between newly appointed and senior consultants ($P < .026$).**

Conclusion: **This study shows that surgical performance continues to improve significantly beyond consultancy, and the data suggest that generic and procedural performance continue to improve, with significant improvement in the former with increasing experience. The ICEPS rating scale demonstrates construct validity and a high inter-observer reliability supporting its use in formative and summative assessment. (J Vasc Surg 2006;43:539-45.)**

There is an increasing awareness for a need to assess technical competence. This has come about because of a variety of factors:

Reduced work hours. Across Europe, we are facing a reduction in the working hours of surgeons in training as a result of implementation of the European Working Time Directive.[1,2] In the United Kingdom as well as in most of Europe, there are no plans to increase the number of years to reach consultancy; therefore, current trainees will undoubtedly have less experience than previous generations at the end of their training period.[3] Those in surgical education are looking for ways to establish levels of competency during the training period to ensure that these less-experienced consultants will have sufficient skill to undertake independent practice.[4-13]

Regulation within the profession. Surgeons are assessed at all stages of their training from medical student to consultancy and beyond. Medical students are expected to demonstrate the core knowledge of surgery that is requisite of any medical practitioner. Junior surgical trainees are expected to demonstrate their knowledge of basic medical sciences as applied to surgery and also of clinical surgery in general. More advanced surgical trainees are expected to demonstrate this as well as detailed knowledge of operative surgery, surgical judgement, and decision-making. Once in independent practice, consultants are expected to undertake continuing professional development. Nevertheless, at no stage are structured validated criteria used to objectively assess the technical competence of surgeons.

Public demand. Increased openness within the profession has been exploited by the media both in the United Kingdom[14-16] and abroad. As a result, the general public has increased awareness and demands of the profession.

Over the last few years, a number of studies from institutions worldwide have looked at surgical competence across varying experience levels to validate numerous surgical rating tools. Current simulators and rating tools suggest that surgical technical competence shows a plateau at the senior trainee level, with no significant differences observed between groups beyond this level.[17-20] These rating tools have therefore been criticized for being insensitive.

The purpose of this study is to assess the surgical competence of all levels of experience, from senior house officers to experienced consultants, by using the newly developed procedure-specific Imperial College Evaluation of Procedure-Specific Skill (ICEPS) rating scale in conjunction with the Objective Structured Assessment of Technical Skills (OSATS) global rating scale.

METHODS

The task. Saphenofemoral junction ligation was selected, as it is an operation that is regularly performed by

From the Regional Vascular Unit[a] and the Department of Biosurgery and Surgical Technology,[b] Imperial College, St. Mary's Hospital.

Competition of interest: none.

Reprint requests: Mr John H. N. Wolfe, MS, Regional Vascular Unit, 2nd Floor Mary Stanford Wing, St. Mary's Hospital, Praed Street, London. W2 1NY United Kingdom (e-mail: john.wolfe@st-marys.nhs.uk).

0741-5214/$32.00

doi:10.1016/j.jvs.2005.09.047

Fig 1. Saphenofemoral junction groin model (Limbs & Things, Bristol, UK). Tissue consists of silicone with latex fluid-filled vessels.

most surgeons at all levels of experience in general and vascular surgery. The model was a newly developed synthetic model (Fig 1) depicting the human saphenofemoral junction and its tributaries. This model allows incision of the skin and dissection through the superficial fatty and deeper fascial layers. Once beyond the fascial layer, the surgeon has to identify the fluid-filled long saphenous vein with its four groin tributaries, ligate the tributaries, and then perform a saphenofemoral disconnection. Use of a synthetic model allows for standardization of the tasks and avoids the ethical constraints of this study. Previous studies in this institution have demonstrated concurrent validity of this model compared with operating theater performance.[21]

Participants. Fifty surgeons volunteered to take part in this study, 10 each from the following groups:

- Senior House Officers (SHO)—equivalent to surgeons in postgraduate years 1 and 2 of a surgical residency program. All the participants from this group had completed the Royal College of Surgeons Basic Surgical Skills course and had performed supervised saphenofemoral junction ligation.
- Junior Specialist Registrars (JSpRs)—equivalent to surgeons in years 3 to 4 of a surgical residency program. All the volunteers in this group were general surgical specialist registrars in their first to third year of training.
- Senior Specialist Registrars (SSpRs)—equivalent to a surgeon in a fellowship program. These surgeons were in subspecialty training in vascular surgery in their fourth to sixth year of training.
- Surgeons with recent Certificate of Completion of Specialist Training (Ncon)—equivalent to a newly appointed attending surgeon. All surgeons in this group had recently (<1 year) been awarded their Certificate of Completion of Specialist Training or had held a consultant post for <2 years.
- Senior Consultants (Scon)—surgeons who had been in independent practice for >5 years.

Consecutive trainees were recruited. Consultant surgeons practiced only vascular surgery or general surgery with vascular surgery as their speciality interest. Surgeons not in a training or consultant post were excluded.

Rating tools. We employed the rating scale used in OSATS and validated by the University of Toronto Centre for Research in Education.[4,22] This rating scale tests various aspects of generic surgical skill, including respect for tissue, instrument and suture handling, time and motion, flow and knowledge of procedure, and quality of end product. The scale is numbered 1 to 5, with 1 representing a poor performance, 3 (an average score) representing a competent performance, and 5 representing an excellent performance. Descriptive comments for each technical domain are given at each of these anchoring points. The minimum score for technical performance is 8, and 40 is the maximum attainable score. This rating scale is applicable to all open surgical procedures. Studies have been published from several centers using this rating scale.[8,20,22-24] It has demonstrated correlation between the performance on synthetic bench models[4] with performance on animal models and also correlation with operative theater performance.[25] The Toronto group has demonstrated a transfer of skill from synthetic models to operative performance[26] with this rating scale, and the Imperial College group has demonstrated a significant correlation of global scores with the results of electromagnetic motion analysis of surgeons' hand movements.[19]

The OSATS, however, only assesses the generic surgical skill of the surgeon, those aspects of surgical technique such as handling of instruments and sutures. Content validity of a technical skills assessment, or the extent to which an assessment tests the content domain (ie, does the surgeon identify and ligate the saphenofemoral junction?) it purports to, is generally assessed using checklists; however, Regehr et al[27] demonstrated a greater construct validity (the ability of this test to differentiate between surgeons of varying skill levels) and interobserver reliability of rating scales over checklists. We therefore developed the procedure-specific Imperial College Evaluation of Procedure Specific Skill (ICEPS) rating scale for assessing the content of a procedure on a standard 5-point scale. The domains tested are specific to saphenofemoral junction ligation (Fig 2). Combing the marks of OSATS and ICEPS (maximum 40 for each), we attained a total operative score for this procedure.

Video assessment. Video footage of saphenofemoral junction ligation performed on this synthetic bench model was blinded, with sound removed, and randomized. This video footage was reviewed and marked by three surgical research fellows (V. A. P., K. M., Y. M.) with experience in the use of rating scales for surgical assessment. Each assessor had rated >50 procedures with the OSATS global rating scale. All marking was independent, and videotape selection was random and identified only by codes.

Imperial College Evaluation of Procedure-specific Skill

Saphenofemoral Junction Ligation

Candidate no:
Assessor:
Date:

Please circle the candidate's performance on the following scale:

	1	2	3	4	5
Incision	**1** Does not use surface landmarks. Inappropriate placement of incision. Poor handling of scalpel	**2**	**3** Appropriate incision in terms of location and size. Looked at ease with scalpel	**4**	**5** Uses surface landmarks to make an appropriately located and sized incision. Handled scalpel expertly
Dissection	**1** Appeared unsure and excessively hesitant whilst dissecting. Caused trauma to tissues. Did not dissect into the correct anatomical plane.	**2**	**3** Controlled and safe dissection into correct anatomical plane. Caused minimum trauma of tissues. Used instrument satisfactorily while dissecting	**4**	**5** Superior and atraumatic dissection into the correct anatomical plane. Confident handling of instruments whilst dissecting.
Retraction	**1** Clumsy use of retractors. Did not allow visualization of the important structures making frequent changes to retractor setting.	**2**	**3** Good use of retraction allowing visualization of major structures. Had to change retractor position to visualize other structures	**4**	**5** Excellent use of retractptrs. Allowed good visualization of all necessary structures. Atraumatic
Tributaries	**1** Could not or did not try to identify any tributaries	**2**	**3** Identified all known tributaries. Did not seek other vessels	**4**	**5** Identified all known tributaries. Sought other possible tributaries.
Haemostasis	**1** Poor quality of knot tying. Knots frequently slipped or was excessively traumatic to vessels	**2**	**3** Competent knot tying. Minimal trauma to vessels. Minimal blood loss.	**4**	**5** Superior knot tying. Atraumatic. No knot slippage
SFJ Clearance	**1** Did not identify the Saphenofemoral junction or excessively traumatic dissection around that vessel	**2**	**3** Identified the Saphenofemoral junction. Safely dissected tissues away from vessel. Reasonsable clearance of vessel. Minimal trauma.	**4**	**5** Identified the Saphenofemoral junction. Exper dissection of tissues off the vessels. Atraumatic. Cleared well proximally and distally.
SFJ Ligation	**1** Did not ligate the SFJ of ligated CFV or caused excessive encroachment onto CFV after SFJ Ligation	**2**	**3** Good knot tying whilst ligating the SFJ. Minimal encroachment onto CFV following SFJ ligation.	**4**	**5** Excellent safe and secure ligation of the SFJ. Flush ligation with no encroachment onto CFV

Total score:

Fig 2. Imperial College Evaluation of Procedure-Specific Skill (ICEPS) rating scale for saphenofemoral junction ligation. Eight operative components on a 5-point scale. Descriptive comments at anchoring points aid marking.

Statistical analysis. The interobserver reliability for the three assessors was evaluated using Cronbach's α, the standard method of assessing the reliability of a rating scale. It is suggested that for research purposes, a reliability coefficient of $\alpha = 0.6$ to 0.8 is sufficient, and for a high-stakes assessment, such as a surgical examination, this coefficient should be >0.8.

Construct validity was assessed using the Kruskal-Wallis test. We identified significant differences between adjacent subgroups as an a priori hypothesis. This also allowed us to identify plateaus in surgical technical skill. We used the Mann-Whitney *U* test in the subgroup analysis. SPSS version 10.0 (SPSS, Inc, Chicago, Ill) software for Windows (Microsoft, Redmond, Wash) was used in the statistical analysis. $P < .05$ was considered a significant difference.

RESULTS

Eight was the minimum score (of a total of 40), a score of 24, deemed as competent in previous studies, was equivalent to an average score of 3 (competent) in each of the eight domains of the generic scale. Seven was the minimum ICEPS score (of a total of 35). A score of 21 was considered competent.

The OSATS scores ranged from 12.3 to 35.7 (maximum, 40), and 18 (36%) of the 50 surgeons (10 senior house officers [SHOs], 5 JSpRs, and 3 SSpRs) displayed poor generic surgical skill on this model, with an average mark of <24. ICEPS scores ranged from 13.7 to 34 (maximum, 40), and 15 (30%) of the 50 surgeons (10 SHOs, 2 JSpRs, 2 SSpR, and 1 Ncon) performed a substandard

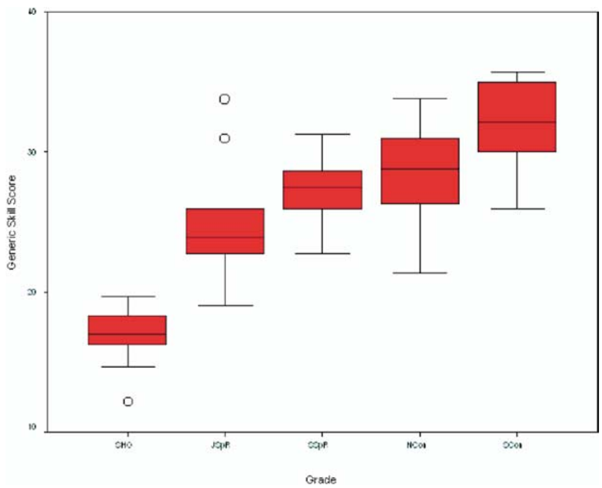

Fig 3. Box plot for generic surgical skills. As in the following box-plots, the horizontal line represents the median value, boxes represent the interquartile ranges and whiskers, the 95% confidence intervals. The circles represent statistical outliers and the asterisk represents extreme values. Significant differences seen between senior house officers (SHO) and junior specialist registrars (JSPR) ($P < 0.001$) and between new consultants (NCon) and senior consultants (SCon) ($P = 0.026$).

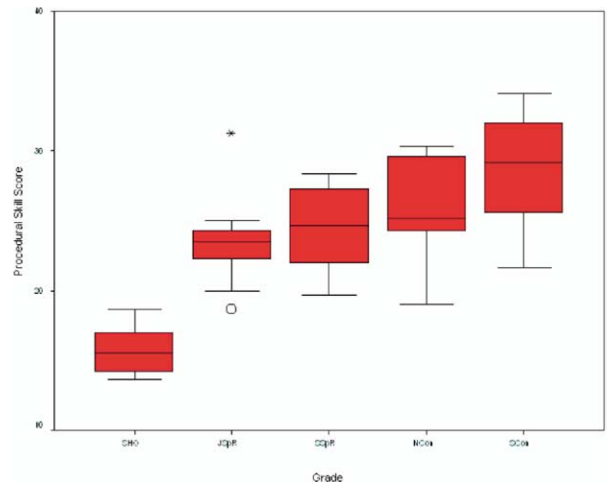

Fig 4. Box plot for procedural score. Horizontal line represents median value. Boxes represent interquartile ranges and whiskers, actual range. The circle represents statistical outliers and the asterisk represents extreme values. Although scores increase with grade, significant differences are only seen between the first two groups ($P < .001$). *SHO,* Senior house officer; *JSpR,* Junior specialist registrar; *SSpR,* senior specialist registrar; *Ncon,* new consultant; *Scon,* senior consultant.

procedure on the simulation, with scores of <21. The interobserver correlation for OSATS generic scores was high (Cronbach's $\alpha = 0.97$). A similar level of agreement was achieved with ICEPS scores for procedural skill ($\alpha = 0.96$).

Construct validity. The scores for generic surgical skill and for procedural performance increased significantly with ascending grade (Kruskal-Wallis; $P < .001$ for both).

Subgroup analysis. Senior house officers had significantly poorer generic surgical skill on the OSATS (scores 12.3 to 19.7) compared with junior specialist registrars (scores 19 to 33.7) (Mann-Whitney *U* test, $P < .001$) (Fig 3). The difference in generic surgical skill among junior and senior specialist registrars ($P = .121$) and senior specialist registrars (scores 22.7 to 31.3) and new consultants (scores, 21.3 to 33.7; $P = .345$) was not statistically significant. Senior consultants (scores, 26 to 35.7) had better generic surgical skill than their newly appointed colleagues ($P = .026$).

On the ICEPS, again the senior house officer group (scores 13.7 to 18.7) performed worse than their junior specialist registrar colleagues (scores, 18.7 to 31.3; $P < .001$) (Fig 4). Significant differences were not seen between junior and senior specialist registrars ($P = .425$) or between senior specialist registrars (scores, 19.7 to 28.3) and new consultants (scores, 19 to 30.3; $P = .306$). Significant differences in procedural performance were not seen in the two consultant groups (Scons scored 21.7 to 34; $P = .069$).

Total operative score. As OSATS tests a surgeon's generic skill and ICEPS the procedure-specific content of the procedure, we combined the two marks to attain a total operative score. We used this total score to establish when most technical improvement takes place and when this model shows the performance of the surgeon reaches a statistical plateau.

The median score, interquartile range, and actual range of scores within each subgroup are shown in Fig 5. There was a significant improvement in scores between the senior house officer and specialist registrar subgroups ($P < .001$). In clinical practice, this is reflected by the steep learning curve associated with commencing higher surgical training. Between junior and senior specialist registrars ($P = .162$) and also between senior specialist registrars and newly appointed consultants ($P = .426$), the differences between the subgroups were not statistically significant, although there was a general increase in the technical skill scores of the surgeons. Significant differences were seen between newly appointed and senior consultants ($P = .041$), explained by the significant improvement in generic skill.

DISCUSSION

Numerous methods have been described to assess surgical competence:

Direct observation. This is the method used by trainers, but it requires structured criteria to be valid and reproducible.[12,28] The performance of the trainee in this context may be affected by the trainer-trainee relationship; therefore, some regard this as a subjective method.

Of the structured criteria that have been established, among the most extensively validated is the global rating scale used in the OSATS, developed by the University of

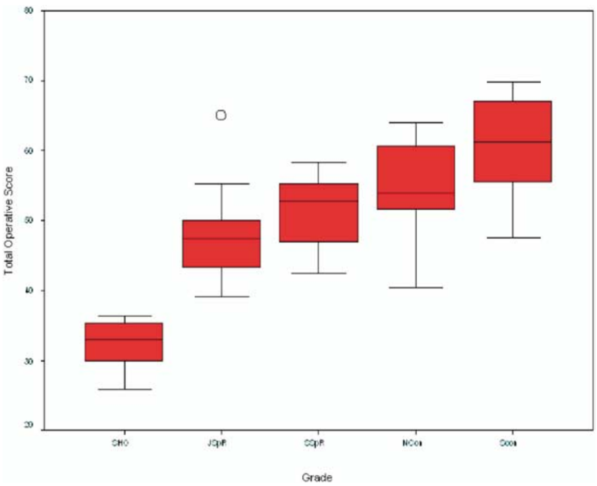

Fig 5. Box plot for total operative score. Horizontal line represents median value. Boxes represent interquartile ranges and whiskers, actual range. The circle represents statistical outliers. Significant differences are again seen between the first two groups ($P <$ 0.01). Performance then improves (although not significantly) until completion of training. There a significant improvement in score between the newly consultants *(Ncon)* and senior consultant *(Scon)* ($P =$ .04). *SHO,* Senior house officer; *JSpR,* Junior specialist registrar; *SSpR,* senior specialist registrar.

Toronto.[4,22] This scoring system has the advantage that it can be applied to videotape footage of operative procedure that can be blinded to ensure the objectivity of the assessor.

Motion analysis. The Imperial College Surgical Assessment Device developed in the United Kingdom consists of commercially available hardware (Isotrak II, Polhemus, Colchester, Vt) and patented software that quantifies hand movements made by the surgeon.[29-31] Data for the following parameters are given:

1. Motion trackers on the dorsum of both hands quantify the number of hand movements made, and software filters remove artefact movements such as physiologic tremor.
2. The distance travelled (path length) by the operator's hand in relation to a fixed point is displayed for each procedure.
3. Procedural time is measured.
4. The average speed for both hands is determined by dividing path length by time.

Motion analysis effectively assesses the economy of movement of surgeons' hands and has established that experienced surgeons have a greater economy of movement.[19,30,31] Economy of movement relates to elimination of purposeless movement rather than speed of the surgeon. ICSAD is arguably more objective than an examiner; however, the data are purely quantitative, which may be of great use when looking at the learning curves of surgeons or assessing longitudinal performance of a trainee during his or her training program.

Virtual reality. Virtual reality simulators[32-35] aim to realistically simulate surgical procedures, and the software can accurately assess parameters such as instrumental path length, simulated blood loss, simulated trauma to structures, and other variables. The increasing sophistication of these simulators allows for haptic force feedback to be delivered to the surgeons. This again allows for quantification of surgical skill in laparoscopic surgery, but the current technology does not allow for sophisticated simulation of open surgery.

Studies in the assessment of technical skill have previously demonstrated a plateau in the performance of surgeons during higher surgical training, and it is thought that this continues into consultancy. The reasons for this may include the following:

1. The rating tools may not be sufficiently sensitive at higher levels of experience.
2. The studies do not indicate the consultant's experience. In this study we aimed to seek differences between all levels of experience, from novice to expert, and actively recruited senior consultants into this study.
3. Performance on a particular task may actually plateau during higher surgical training, and the task is not sufficiently complex or technically demanding to demonstrate differences between the more senior groups of surgeons.

The plateau of performance has been seen using global scores of operative performance and even more so with checklists. We therefore abandoned using the latter in this study.

The two rating scales used in this study assess different aspects of surgical skill (generic and procedural content), allowing us to derive a total operative score for this procedure. Subgroup analysis of the total operative scores showed two points in a surgeon's career when most improvement in technical skill takes place. Between basic surgical training and higher surgical training, technical scores significantly increase. It appears that surgeons continue to improve after they have been awarded their certificate of completion of specialist training. Looking at subgroup differences in generic and procedural scores, it can be seen that this increase is not because they perform the procedure more thoroughly, but that their generic skill improves after consultancy.

Explaining variations between groups. Within some of groups, in particular the junior specialist registrars grade (years 1 to 3), the variation in marks is greater than within other grades. In the United Kingdom, this grade in particular is not a homogenous group with regards to technical skill. This reflects the variations in training and experience before attaining a National Training Number (NTN). Some regions regard experience favorably, whereas other regions may prefer to take on a younger, albeit less experienced surgeon for higher surgical training. It is obvious, therefore, that candidates in this grade will have a considerable range in technical skill, having spent from 2 to 5 years in basic surgical training. When trainees reach the stage of higher surgical training of senior registrars (year 4 to 6), the group has become more homogenous, although differences within this grade still exist.

Why assess content? It may be argued that it is unnecessary to develop a procedural rating scale if OSATS can elicit such differences. It is essential, however, for an assessment of technical skill to have content validity and be easy to use. Procedural rating scales have demonstrated improvements in specific components of abdominal aortic aneurysm repair following workshop-based training.[36] This is useful, not only for the workshop participant in providing formative assessment and feedback but also for the workshop organizer in identifying strengths and potential weaknesses of the course.

Applications in surgical training. Bench model simulators may provide an adjunct to operating theater training, but practicing on these models is likely to be ineffective without feedback. Simulator training has been used in other technically demanding professions such as aviation for many years.

Surgeons should read about the procedure and be conversant with the relevant surgical anatomy. They should then be able to demonstrate their competency in the various aspects of the procedure before performing the procedure on patients. This model of training and assessment on generic and procedural skill may be particularly effective with repetition on the model and the structured feedback provided with the rating scales.

CONCLUSION

The construct validity and interobserver reliability of the ICEPS and OSATS rating scales was high. In particular, the interobserver reliability for the OSATS global scores was higher than has been previously reported.[8,19] It is postulated that this may be due to the increased scrutiny that is required when marking with a procedural rating scale (ICEPS). In particular, use of a visual fast forward facility ("cue") was not used.

If an era when assessing the technical competence of surgeons is gaining momentum, we need to ensure the methods by which we do this are valid, robust, and helpful. This model of surgical assessment has been taken up in formative assessment of surgical skill in our institution and has paved the way for surgical examinations of technical competence.[24]

AUTHOR CONTRIBUTIONS

Conception and design: VAP, JHNW, KM, MJJ
Analysis and interpretation: VAP, JHNW, KM, YM, MJJ, AWD
Data collection: VAP, KM, YM
Writing the article: VAP, JHNW, MJJ
Critical revision of the article: JHNW, KM, YM, MJJ, AWD
Final approval of the article: VAP, JHNW, KM, YM, MJJ, AWD
Statistical analysis: VAP, JHNW, KM, MJJ
Overall responsibility: VAP

REFERENCES

1. The European Parliament and Council of the European Union. Council Directive 93/104/EC. Official Journal of the European Communities 1993:18-24.
2. EC Directive 75/362. Publications of the European Communities 1975;L167(30-6):1.
3. Wolfe JH. Training—vive la difference? Eur J Vasc Endovasc Surg 1999;18:1-3.
4. 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.
5. Buth J, Harris PL, Maurer PC, Nachbur B, van Urk H. Harmonization of vascular surgical training in Europe. A task for the European Board of Vascular Surgery (EBVS). Cardiovasc Surg 2000;8:98-103.
6. Darzi A, Smith S, Taffinder N. Assessing operative skill. Needs to become more objective. BMJ 1999;318:887-8.
7. Darzi A, Datta V, Mackay S. The challenge of objective assessment of surgical skill. Am J Surg 2001;181:484-6.
8. Mackay S, Datta V, Chang A, Shah J, Kneebone R, Darzi A. Multiple Objective Measures of Skill (MOMS): a new approach to the assessment of technical ability in surgical trainees. Ann Surg 2003;238:291-300.
9. MacRae HM, Cohen R, Regehr G, Reznick R, Burnstein M. A new assessment tool: the patient assessment and management examination. Surgery 1997;122:335-43; discussion 343-4.
10. MacRae H, Regehr G, Leadbetter W, Reznick RK. A comprehensive examination for senior surgical residents. Am J Surg 2000;179:190-3.
11. Pandey V, Jackson M, Bicknell C, Yapanis M, Darzi A, Wolfe J. Assessing competence in vascular surgery: a pilot study. Br J Surg 2003;90(S1):113.
12. Reznick RK. Teaching and testing technical skills. Am J Surg 1993;165:358-61.
13. Satava RM, Cuschieri A, Hamdorf J. Metrics for objective Assessment. Surg Endosc 2003;17:220-6.
14. Pringle M. The Shipman inquiry: implications for the public's trust in doctors. Br J Gen Pract 2000;50:355-6.
15. Smith R. All changed, changed utterly. British medicine will be transformed by the Bristol case. BMJ 1998;316:1917-8.
16. Smith R. Regulation of doctors and the Bristol inquiry. Both need to be credible to both the public and doctors. BMJ 1998;317:1539-40.
17. Munz Y, Moorthy K, Bann S, Shah J, Ivanova S, Darzi SA. Ceiling effect in technical skills of surgical residents. Am J Surg 2004;188:294-300.
18. Beard J, Donnelly J. Saphenofemoral disconnection: a good operation for assessing the basic surgical skills. BJS 2003;90(S1):33.
19. Datta V, Chang A, Mackay S, Darzi A. The relationship between motion analysis and surgical technical assessments. Am J Surg 2002;184:70-3.
20. Bann S, Kwok KF, Lo CY, Darzi A, Wong J. Objective assessment of technical skills of surgical trainees in Hong Kong. Br J Surg 2003;90:1294-9.
21. Datta V, Bann S, Beard J, Mandalia M, Darzi A. Comparison of bench test evaluations of surgical skill with live operating performance assessments. J Am Coll Surg 2004;199:603-6.
22. 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.
23. Ault G, Reznick R, MacRae H, Leadbetter W, DaRosa D, Joehl R, et al. Exporting a technical skills evaluation technology to other sites. Am J Surg 2001;182:254-6.
24. Pandey VA, Wolfe JH, 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.
25. Datta V, Bann S, Beard J, Mandalia M, Darzi A. Comparison of bench test evaluations of surgical skill with live operating performance assessments. Br J Surg 2002;89(S1):92.
26. Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M, et al. Assessment of technical skills transfer from the bench training model to the human model. Am J Surg 1999;177:167-70.

27. 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.
28. Watts J, Feldman W. Assessment of technical skills. In: Neufeld V, Norman G, editors. Assessing clinical competence. New York: Springer; 1985. pp 259-74.
29. Taffinder N, Smith S, Huber J, Russel R, Darzi A. Can a computer measure surgical precision? Reliability, validity and feasibility of the ICSAD. Surg Endosc 1993;13(S1):81.
30. Bann SD, Khan MS, Darzi AW. Measurement of surgical dexterity using motion analysis of simple bench tasks. World J Surg 2003;27:390-4.
31. 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.
32. Gallagher AG, Satava RM. Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Learning curves and reliability measures. Surg Endosc 2002;16:1746-52.
33. Moorthy K, Smith S, Brown T, Bann S, Darzi A. Evaluation of virtual reality bronchoscopy as a learning and assessment tool. Respiration 2003;70:195-934.
34. Reznick RK. Virtual reality surgical simulators: feasible but valid? J Am Coll Surg 1999;189:127-8.
35. Satava RM. Virtual reality surgical simulator. The first steps. Surg Endosc 1993;7:203-5.
36. Pandey VA, Black SA, Lazaris AM, Allenberg JR, Eckstein HH, Hagmuller GW, et al. Do workshops improve the technical skill of vascular surgical trainees? Eur J Vasc Endovasc Surg 2005;30:441-7.

Submitted Jun 18, 2005; accepted Sep 30, 2005.