



Self-assessment of technical skill in surgery: the need for expert feedback

VA PANDEY, JHN WOLFE, SA BLACK, M CAIROLS, CD LIAPIS, D BERGQVIST

On behalf of the European Board of Vascular Surgery

ABSTRACT

INTRODUCTION Technical skill has been formally assessed in the Fellow of the European Board of Vascular Surgery Examinations (FEBVS) since 2002. The aim of this study was to examine the relationship between expert assessment and trainee self-assessment.

MATERIALS AND METHODS Forty-two examination candidates performed a saphenofemoral junction (SFJ) ligation and an anterior tibial anastomosis on a synthetic simulation. Each candidate was rated by two examiners using a validated rating scale for their generic surgical skill for both procedures. Candidates then anonymously rated their own performance using the same scale. Parametric tests were used in the statistical analysis; a P -value < 0.05 was considered significant.

RESULTS The maximum mark in each assessment was 40; 24 was considered a competent score. The interobserver correlation for examiners marks were high (SFJ ligation, $\alpha = 0.68$; distal anastomosis, $\alpha = 0.76$). Examiners' marks were averaged. The mean examiner score for the SFJ ligation station was 27.8 (SD = 4.1) with 36 candidates (85.8%) attaining a competent score. The mean self-assessment score for this station was 30.7 (SD = 4.66). The mean examiners' marks for the distal anastomosis station was 29.2 (SD = 4.2); 39 candidates (92.8%) attained a competent score. The mean self-assessment score was 32.1 (SD = 4.0). There was no correlation between examiner and self-assessment scores in either station (Pearson's correlation coefficient: SFJ, $r = 0.045$, $P = \text{NS}$); distal anastomosis, $r = 0.089$, $P = \text{NS}$). Bland and Altman plots assessed the agreement between examiner and self-assessment. These showed candidates marked themselves higher than examiners with a mean difference of 2.9 marks in each station.

CONCLUSIONS Candidates' self-assessment and expert independent assessment correlate poorly. Trainees overestimate their ability according to independent assessment; regular technical feedback during training is, therefore, essential.

KEYWORDS

Technical skill – Self-assessment – Expert feedback – Training

CORRESPONDENCE TO

JHN Wolfe, Consultant Vascular Surgeon, Regional Vascular Unit, 2nd Floor, Mary Stanford Wing, St Mary's Hospital, London W2 1NY, UK
E: john.wolfe@st-marys.nhs.uk

The ability to assess one's own performance critically in surgery is a valuable trait for surgeons throughout training and independent practice. Unfortunately, this remains an underdeveloped skill in surgical training and receives little attention from surgical educators.¹ For trainees, it allows identification of their strengths and, more importantly, weaknesses in their ability, to build upon previous performance and take the necessary remedial action. For surgeons in independent practice, the introduction of new surgical techniques necessitates focused self-assessment.

Evidence for self-assessment in surgery is poor,^{2,3} but studies in higher education have shown poor correlations between self and expert assessment. Falchikov and Boud⁴ conducted a meta-analysis of 44 self-assessment studies in higher education and found a mean correlation of 0.39

between self and expert assessment with participants generally under-rating their performance. The study by Harrington *et al.*¹ in 1997 demonstrated a poor correlation between residents and faculty at the end of an orthopaedic rotation. Risucci *et al.*⁵ compared self, peer and supervisor ratings with scores on the American Board of Surgery In-Training Examination (ABSITE). The results showed significant correlation between ratings by peers and supervisors ($r = 0.92$; $P < 0.001$). The average of peer and supervisor ratings showed a moderate correlation with ABSITE scores ($r = 0.58$; $P < 0.01$). Multivariate analysis suggested that supervisors were influenced mainly by the interpersonal skill of the resident and secondarily by their ability. Self-assessment was influenced mainly by the residents' perceptions of their own ability, followed by interpersonal skills and effort.

Self and peer ratings were lower than ratings by the supervisor, a finding consistent with the aforementioned self-assessment studies.

The purpose of this study was to examine the relationship between self and expert-examiner assessment of the technical skill of vascular surgeons at the end of their training. The Fellowship of the European Board of Vascular Surgery Examination (FEBVS, formerly known as the European Board of Surgery Qualification in Vascular Surgery, EBSQ-VASC) was introduced in 1996 by the Union Européenne des Médecins Spécialistes (UEMS) Board of Vascular Surgery.⁶⁻⁸ Pre-requisites to the examination included the possession of a Certificate of Completion of Specialist Training (CCST) issued by one of the member states of the European Union and personal experience of index procedures in vascular surgery evaluated using log-book accreditation.⁸ A technical skills component to this examination was validated in a 2-year pilot study^{9,10} and formally introduced in 2004 as an integral part of the fellowship examination.¹¹

Materials and Methods

Forty-two surgeons sitting the FEBVS examinations performed two surgical procedures:

1. The first task was saphenofemoral junction ligation on a synthetic model depicting a human saphenofemoral junction (Fig. 1). This model simulated a comparatively simple procedure performed by surgeons at all stages of their training.
2. The second, and more advanced, task was to perform a distal arterial anastomosis onto the anterior tibial artery of a synthetic leg model (Fig. 2). The second task was technically more challenging and is performed by surgeons at the latter stages of their training.

The candidates were given 25 min to perform the first procedure and 50 min to perform the second procedure.

Each examination candidate was marked for both procedures by two examiners using a modified global rating scale for surgical skill.^{12,15} This scale assessed eight components of generic surgical skill including suture, instrument and tissue handling as well as the flow of the operation and knowledge of the procedure on a 5-point Likert scale. The scale had descriptive comments at positions one, three and five representing a poor, competent and excellent performance, respectively. The minimum score from this scale was eight. A score of 24 represented an overall competent performance and a score of 40 was the maximum attainable.

Each candidate was also asked to complete the rating scale for both procedures confidentially at the end of the examination. The average examiner score was correlated with the self-assessment score for each candidate.



Figure 1 Bespoke model of the human saphenofemoral junction. Inset: the saphenofemoral junction dissected and the tributaries ligated.

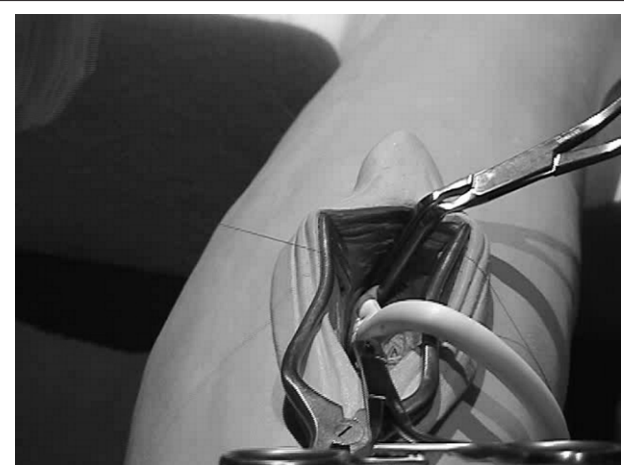


Figure 2 Synthetic leg model. The anterior tibial artery is exposed for a distal arterial anastomosis.

Examiners' scores were normally distributed; therefore, parametric tests were used in the statistical analysis. Interobserver correlation was assessed using Cronbach's alpha (α). The correlation between examiner and self-assessment scores was evaluated using Pearson's correlation coefficient; $P < 0.05$ was considered significant.

Results

All 42 examination candidates performed both procedures in the allotted time and all 84 self-assessment forms were completed correctly and returned at the end of the examination.

Interobserver correlation between examiners was high ($\alpha > 0.68$ for saphenofemoral junction ligation and 0.76 for the distal anastomosis).

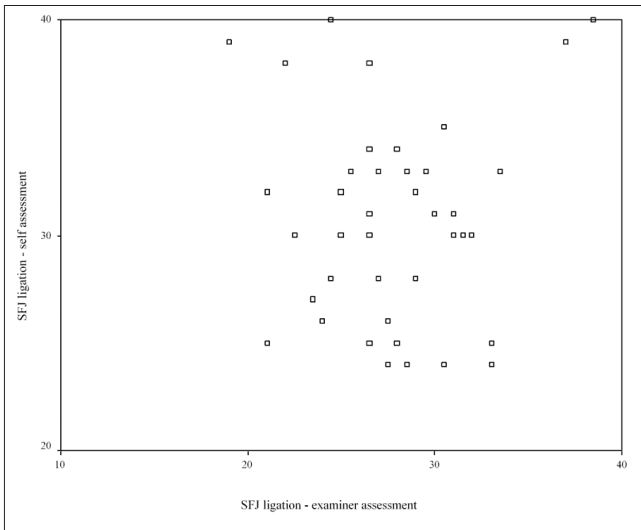


Figure 3 Saphenofemoral junction ligation. Correlation between examiner (horizontal axis) and self-assessment (vertical axis). $r = 0.045$; $P = NS$.

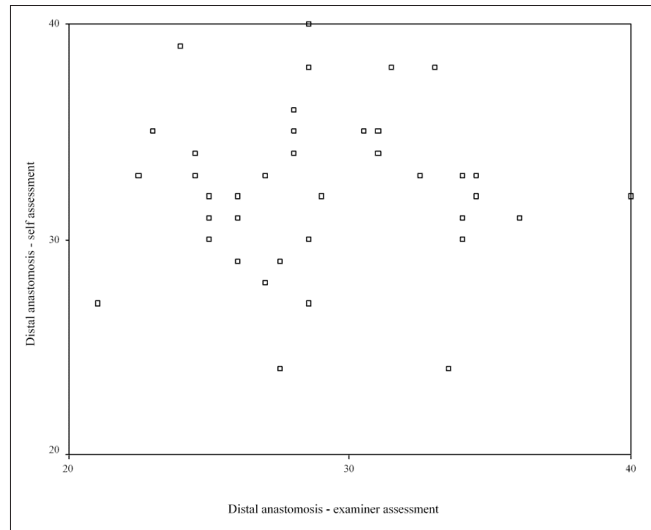


Figure 4 Distal anastomosis. Correlation between examiner (horizontal axis) and self-assessment (vertical axis). $r = 0.089$; $P = NS$.

The two examiners' scores were averaged and, in the first station, the mean candidate score was 27.8 (SD = 4.1). Thirty-six candidates (85.8%) attaining a competent score. The mean self-assessment score for this station was 30.7 (SD = 4.7). All candidates considered themselves competent performing this procedure.

The mean examiner score for the distal anastomosis station was 29.2 (SD = 4.2) Thirty-nine of 42 surgeons (92.8%) were considered competent in performing this procedure. The mean self-assessment score for this station was 32.1

(SD = 4.0). Furthermore, it was seen that the candidates who attained the lowest scores in the exercises marked themselves as highly as those candidates attaining the highest examiner scores for both assessments.

There was no correlation between self and examiner assessment in either station (saphenofemoral junction ligation - $r = 0.045$; $P = NS$; Fig. 3; distal anastomosis - $r = 0.089$; $P = NS$; Fig. 4).

Bland and Altman plots were used to assess the agreement between examiner and self-assessment. The mean

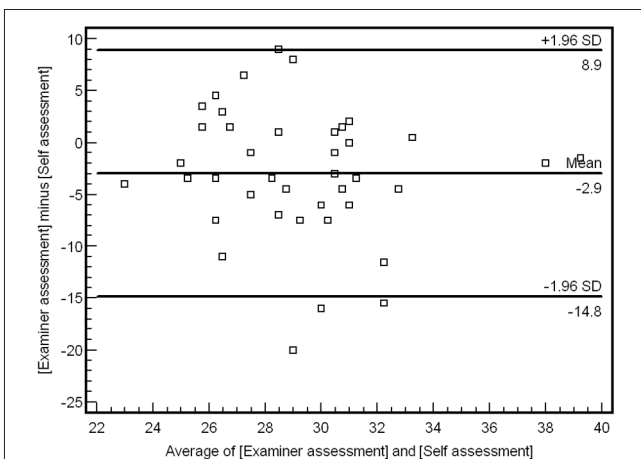


Figure 5 Bland and Altman plot for saphenofemoral junction ligation. The average of the examiner and self-assessment scores on the horizontal axis and the difference between examiner and self-assessment scores on the vertical axis. The central horizontal line represents the mean difference between the two scores (2.9 marks).

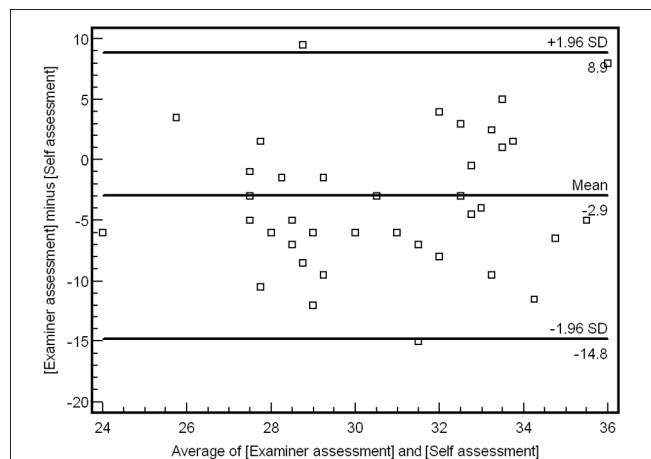


Figure 6 Bland and Altman plot for distal anastomosis. The average of the examiner and self-assessment scores on the horizontal axis and the difference between examiner and self-assessment scores on the vertical axis. The central horizontal line represents the mean difference between the two scores (again the difference was 2.9 marks).

difference between self and examiner scores was 2.9 for each procedure with the majority of examination candidates marking themselves higher than marked by the examiners (Figs 5 and 6).

Discussion

We have already shown that technical surgical skill is an ability independent of knowledge and critical evaluation.¹⁰ It does, however, remain possible that this skill can be adequately evaluated on critical self-assessment.

The poor correlation seen between self and expert assessment is consistent with previously published studies in higher education. The finding of candidates marking themselves higher than examiners was unexpected and not previously reported. This may have occurred for a number of reasons:

1. The previously published studies used medical students and more junior trainees. All examination candidates in this study were surgeons at the end of their training.
2. Senior trainees may overestimate their abilities when compared with expert examiner assessment. This may be related to the 'stressful' environment in which the assessment takes place and may not be true with workplace-based assessment.¹⁴
3. Despite the anonymous nature of the forms (numbers rather than names) and the confidential manner in which they were completed, the candidates may have been reluctant to mark themselves in a manner that would suggest that they were not competent to perform the procedure.
4. The examiners may have marked too harshly.

No previous educational study has examined the correlation between self and examiner assessment with such senior trainees. All surgeons had completed formal surgical training and were to embark on independent practice. Irrespective of their actual performance, it is unlikely that they would give themselves a mark that suggested that they were not competent to perform these procedures (*i.e.* recognising their fallibility but, nevertheless, believe they were competent). This was highlighted by a high proportion of examination candidates giving themselves a 'low pass' in self-assessment (3 out of 5 in all eight components of the generic global rating scale).

As self-assessment is rarely utilised in surgical training, it is likely that some of the examination candidates were unfamiliar with this technique and the rating tools used and subsequently overestimated their abilities.

All examiners were given training in the use of the rating scales: as the interobserver correlation was relatively high, it is unlikely that their marks were unfairly low.

Martin *et al.*,¹⁵ from the Centre for Research in Education in Toronto, advocated the use of videotaped benchmarks to improve self-assessment ability, suggesting that trainees compare their own performances with videotaped benchmarks. This may improve their ability to self-evaluate. This effect was seen more with the first year residents ($r = 0.22$ before benchmark review to $r = 0.45$ after review). Second-year residents were better able to evaluate their performance but improved less after benchmark review ($r = 0.53$ to $r = 0.65$).¹⁵ More recently Ward *et al.*,⁵ also from the Toronto group, identified significant improvements in the ability of residents' ability to assess their own performance after review of their own performance ($r = 0.5$ increasing to $r = 0.63$ after review of their own performance); the review of benchmark videotapes did not significantly improve self and examiner assessment correlations (increasing to $r = 0.63$).

The results of this study suggest that surgeons at the end of their training are inaccurate in the area of self-assessment and loathe accepting that their performance may be sub-optimal. It may, however, be related to the environment that the assessment takes place. Workplace-based assessment is now an essential part of the Intercollegiate Surgical Curriculum Project (ISCP) in the UK¹⁶ and it would be interesting to assess the influence of self-assessment in this relatively non-stressful environment compared to a 'high-stakes' surgical examination.

All surgical trainers should become accustomed to the current rating tools available and educational courses such as *Training the Trainer* (The Royal College of Surgeons of England) will undoubtedly provide a useful forum to implement such training in an era of reduced work hours and an ever-demanding patient population. Self-assessment with expert feedback throughout training appears to offer an efficient method of improving the technical performance of surgical trainees as an integral part of a structured surgical training programme.

References

1. Harrington JP, Murnaghan JJ, Regehr G. Applying a relative ranking model to the self-assessment of extended performances. *Adv Health Sci Educ Theory Pract* 1997; **2**: 17–25.
2. Moorthy K, Munz Y, Adams S, Pandey V, Darzi A. Self-assessment of performance among surgical trainees during simulated procedures in a simulated operating theater. *Am J Surg* 2006; **192**: 114–8.
3. Ward M, MacRae H, Schlachta C, Mamazza J, Poulin E, Reznick R *et al.* Resident self-assessment of operative performance. *Am J Surg* 2003; **185**: 521–4.
4. Falchikov N, Boud D. Student self-assessment in higher education: a meta-analysis. *Rev Educ Res* 1989; **59**: 395–430.
5. Risucci DA, Tortolani AJ, Ward RJ. Ratings of surgical residents by self, supervisors and peers. *Surg Gynecol Obstet* 1989; **169**: 519–26.
6. Buth J, Nachbur B. European Board of Surgery Qualifications in Vascular Surgery (EBSQ-VASC) assessments. Three years' experience. *Eur J Vasc*

- Endovasc Surg* 1999; **18**: 360–3.
7. 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.
 8. Liapis CD, Nachbur B. EBSQ-VASC examinations – which way to the future? European Board of Surgery Qualifications in Vascular Surgery. *Eur J Vasc Endovasc Surg* 2001; **21**: 473–4.
 9. 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.
 10. Pandey VA, Wolfe JH, Liapis CD, Bergqvist D. The examination assessment of technical competence in vascular surgery. *Br J Surg* 2006; **93**: 1132–8.
 11. Bergqvist D, Liapis C, Wolfe JN. The developing European Board Vascular Examination. *Eur J Vasc Endovasc Surg* 2004; **27**: 339–40.
 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. 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.
 14. Beard J, Bussey M. Workplace-based assessment. *Ann R Coll Surg Engl Suppl* 2007; **89**: 158–160.
 15. Martin D, Regehr G, Hodges B, McNaughton N. Using videotaped benchmarks to improve the self-assessment ability of family practice residents. *Acad Med* 1998; **73**: 1201–6.
 16. Intercollegiate Surgical Curriculum Project. <www.iscp.ac.uk/Assessment/WBA/Intro.aspx> [Accessed July 2007].