Are general surgeons able to accurately self-assess their level of technical skills?

C Rizan¹, J Ansell², TW Tilston¹, N Warren², J Torkington³

¹Cardiff University, UK

²Welsh Institute for Minimal Access Therapy, UK

³Cardiff and Vale University Health Board, UK

ABSTRACT

INTRODUCTION Self-assessment is a way of improving technical capabilities without the need for trainer feedback. It can identify areas for improvement and promote professional medical development. The aim of this review was to identify whether self-assessment is an accurate form of technical skills appraisal in general surgery.

METHODS The PubMed, MEDLINE®, Embase™ and Cochrane databases were searched for studies assessing the reliability of self-assessment of technical skills in general surgery. For each study, we recorded the skills assessed and the evaluation methods used. Common endpoints between studies were compared to provide recommendations based on the levels of evidence.

RESULTS Twelve studies met the inclusion criteria from 22,292 initial papers. There was no level 1 evidence published. All papers compared the correlation between self-appraisal versus an expert score but differed in the technical skills assessment and the evaluation tools used. The accuracy of self-assessment improved with increasing experience (level 2 recommendation), age (level 3 recommendation) and the use of video playback (level 3 recommendation). Accuracy was reduced by stressful learning environments (level 2 recommendation), lack of familiarity with assessment tools (level 3 recommendation) and in advanced surgical procedures (level 3 recommendation).

CONCLUSIONS Evidence exists to support the reliability of self-assessment of technical skills in general surgery. Several variables have been shown to affect the accuracy of self-assessment of technical skills. Future work should focus on evaluating the reliability of self-assessment during live operating procedures.

KEYWORDS

Self-assessment – Technical skills – General surgery

Accepted 2 February 2015

CORRESPONDENCE TO

James Ansell, E: ansellj@cf.ac.uk

Self-assessment is an individual's ability to determine their capabilities and limitations.¹ This is a core skill in surgical training according to the Accreditation Council for Graduate Medical Education.² Self-assessment allows professional development by prompting the formulation of learning goals.³ Inaccurate self-assessment risks overconfident or unconsciously incompetent surgeons, shown to jeopardise patient safety.⁴ Accurate self-assessment may improve the cost effectiveness of simulation-based training while reducing residency programme costs.⁵

Although gauging performance is key to lifelong learning,⁶ evidence indicates self-assessment accuracy is poor.^{7,8} A meta-analysis of higher education self-assessment studies found a weak correlation between self and expert assessment (0.39) with a tendency for underestimation.⁹ Subsequent review work of healthcare professionals found similarly low correlations.^{6,10} The evidence for the accuracy of self-assessment of surgical technical skills is often contradictory. Moorthy *et al* found that senior surgical trainees accurately self-assess technical procedures in a simulated

operating theatre¹¹ whereas Pandey *et al* showed poor correlation between self and expert appraisal for similar skills.¹²

The primary aim of this systematic review was to appraise the evidence for the accuracy of technical skill self-assessment in general surgery. Secondary aims were to identify whether self-assessment accuracy is affected by the participant, the skill performed, the method of assessment or the assessors themselves.

Methods

A systematic review of the published literature was carried out in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.¹³ Studies evaluating the reliability of technical skills self-assessment in general surgery were searched for using the PubMed (1966 – present), MEDLINE® (1946 – July Week 3 2013), Embase™ (1947 – present) and Cochrane databases.

The search employed two domains of exploded Medical Subject Headings. These domains were combined by

“AND” while terms in each domain were combined by “OR”. Additional keywords were identified using seminal articles and expert consensus. The first domain comprised “self-appraisal”, “self-assess”, “self-assessment”, “self-confidence”, “self-evaluate”, “self-evaluation”, “self-perception” and “self-performance”. The second domain encompassed “surgical”, “surgical competence”, “surgical performance”, “surgical skill”, “surgical technique”, “technical competence”, “technical performance” and “technical skill”. Where a keyword mapped to further subject headings, relevant items were exploded. Two authors performed the literature search independently. Each investigator screened study titles, discarding irrelevant articles and duplicates. The remaining articles were retrieved for full text evaluation and the inclusion criteria applied. Cross-referencing was conducted against studies meeting the inclusion criteria. A third author resolved any discrepancies independently.

Inclusion and exclusion criteria

Original articles evaluating technical skills self-assessment were included if they compared self-assessment against another mode of evaluation. Eligible studies needed to either include at least one group of general surgeon self-assessors or involve self-assessment of generic surgical skills. The final inclusion criterion was providing sufficient detail of the assessed skill. Reviews, congress abstracts and opinion-based reports were excluded, alongside studies principally validating skill simulators or non-technical skills.

Data extraction, outcome measures and analysis

The endpoints extracted for each included paper consisted of the self-assessment score, the observer assessment score and the comparator or correlation between the two. The participant’s level of experience, skill assessed, setting, assessment tool, observation type, assessment timing and blinding were also recorded. For the expert observers, the number used, their relevant experience and inter-rater reliability (IRR) were recorded. Common endpoints between studies were identified and compared.

Self-assessment in each study was deemed ‘accurate’ if $p \geq 0.05$ for the difference between self-assessment and observer assessment (ie no significant difference between assessment scores), if $p < 0.05$ for the correlation coefficient between the two (ie significant correlation between assessment scores) or if Spearman’s rho/r ≥ 0.50 for the correlation coefficient (0.30–0.50 = moderate correlation; >0.50 = strong correlation).¹⁴ Included studies were rated based upon their levels of evidence in order to provide recommendations accordingly (Tables 1 and 2).¹⁵ For each included study, methodological shortfalls were identified and a value of 1 was assigned where each was avoided or 0 where the study was deficient. A meta-analysis was not performed owing to the heterogeneity of the skills under assessment and the tools of assessment.

Table 1 Summary of evidence levels¹⁵

Level	Diagnosis
1a	Systematic reviews (meta-analysis) containing at least some trials of level 1b evidence, in which results of separate, independently conducted trials are consistent
1b	Randomised controlled trial of good quality and of adequate sample size (power calculation)
2a	Randomised controlled trial of reasonable quality and/or of inadequate sample size
2b	Non-randomised trials, comparative research with parallel cohort
2c	Non-randomised trials, comparative research (historical cohort, literature controls)
3	Non-randomised, non-comparative trials, descriptive research
4	Expert opinions, including the opinion of work group members

Results

The primary search identified 22,292 articles (Figure 1). On screening of titles, 22,185 records failed to meet the inclusion criteria and a further 49 were duplicates. The remaining 58 studies were retrieved for full text review and of these, 46 were excluded. Cross-referencing revealed no further studies. Consequently, a total of 12 papers were included (Table 3).^{7,8,11,12,16–25} According to our criteria, eight studies (66.7%) found that self-assessment of technical skills is accurate (Table 4).^{11,16–22} All studies were evidence level 2b.

Participants undertaking the self-assessment process

The self-assessors were categorised according to surgical level in five studies (41.7%)^{8,16,17,19,22} and experience level in four (33.3%).^{7,11,12,21} The accuracy of self-assessment improved with experience in two studies (16.7%)^{11,25} (level 2 recommendation) and with surgical level in one study (8.3%)²² (level 3 recommendation). However, Munz *et al* found senior trainees were less accurate, overrating performance.¹⁷ Increasing age and non-European nationality were found to increase self-assessment accuracy in one study (8.3%)²² (level 3 recommendation). One study

Table 2 Summary of recommendation levels¹⁵

Level	Criteria
1	One systematic review (level 1a) or at least two independently conducted research projects classified as level 1b
2	At least two independently conducted research projects classified as level 2a or 2b, within concordance
3	One independently conducted level 2b research project or at least two level 3 trials, within concordance
4	One trial at level 3 or multiple expert opinions, including the opinion of work group members

(8.3%) suggested participant ‘impression management’ could explain poor self-assessment accuracy.¹¹ This refers to an individual’s attempt to influence others’ perceptions about themselves positively.

Type of skill performed during self-assessment

In order to assess technical surgical skills, six studies (50.0%) used bench models,^{8,12,17,20,22,23} three (25.0%) used virtual reality simulators,^{11,19,21} two (16.7%) used live animal models^{7,16} and one (8.3%) used a live operating setting.¹⁸ A variety of technical skills were assessed. Four studies (33.3%) assessed multiple surgical skills.^{12,17,22,23} Two of these reported better self-assessment accuracy in less advanced techniques, one study finding this among senior trainees¹⁷ and another in novice trainees²³ (level 2 recommendation). One study (8.3%) suggested that the use of an animal laboratory setting could underlie inaccurate self-assessment.⁷ The use of a stressful environment (eg examination setting) was proposed to reduce self-assessment accuracy by a further two studies (16.7%).^{12,20}

Assessment tools

The papers used a wide range of assessment tools (Table 5). Two studies used two tools.^{16,23} The following assessment tools were used in studies reporting that self-assessment is accurate: the operative component rating scale,¹⁶ objective surgical skills assessment,²² hierarchical task analysis,¹⁸ standardised forms¹⁹ and a self-assessment score of performance²⁰ (level 3 recommendation). Poor accuracy of self-assessment or divergent results were seen using the global rating scale,^{7,12,16,17} objective structured assessment of technical skills,^{11,21,23} a visual analogue scale⁸ and a global score.¹⁸ One study indicated that unfamiliarity with self-assessment rating tools was a potential reason for poor self-assessment accuracy.¹²

Self-assessors used retrospective video playback analysis in three studies (25.0%)^{16,19,23} and its use was ambiguous in a further study (8.3%).²¹ Self-assessment was accurate in three (75.0%) of these four studies.^{16,19,21} Only Ward *et al* compared direct versus video self-assessment.¹⁶ They found that the accuracy of self-assessment improved significantly

Table 3 Overview of the studies included

Study	Participants	Skill assessed (setting)	Assessor grade	Assessment tool
Ward, 2003 ¹⁶	Year 3–5 trainees (<i>n</i> =27) Surgical fellows (<i>n</i> =1)	Laparoscopic Nissen fundoplication (live animal)	Expert surgeon (<i>n</i> =3)	GRS OCRS
Munz, 2004 ¹⁷	Year 1–3 trainees (<i>n</i> =15) Year 4–6 trainees (<i>n</i> =15)	Cyst excision, bowel anastomosis, SFJ dissection (bench model)	Senior surgeon (number not specified)	GRS
Moorthy, 2006 ¹¹	Junior trainees (<i>n</i> =11) Middle grade trainees (<i>n</i> =9) Senior trainees (<i>n</i> =7)	SFJ ligation (VR simulated theatre)	Trained expert (<i>n</i> =3)	OSATS
Sarker, 2006 ¹⁸	Experts (<i>n</i> =10) (level not specified)	Laparoscopic cholecystectomy (live operation)	Trained expert (<i>n</i> =2)	HTA
Brewster, 2008 ¹⁹	Year 2 trainees (<i>n</i> =6) Year 4 trainees (<i>n</i> =1)	Laparoscopic dissection technique (VR simulated theatre)	Surgeon and faculty (<i>n</i> =1, <i>n</i> =2)	Standardised forms
Tedesco, 2008 ²⁰	Year 4–5 trainees: low experience (<i>n</i> =8) Year 4–5 trainees: high experience (<i>n</i> =9)	Renal angioplasty/stent procedure (bench model)	Interventionalist (<i>n</i> =1)	GRS SA score
Arora, 2011 ²¹	Inexperienced trainees (<i>n</i> =13) Experienced trainees (<i>n</i> =12)	Laparoscopic cholecystectomy (VR simulated theatre)	Senior surgeon (<i>n</i> =2)	OSATS
de Blacam, 2012 ²²	Year 1 trainees (<i>n</i> =114) Year 2 trainees (<i>n</i> =102)	Cyst excision, wound closure, anastomosis, SFJ ligation, laparoscopic skills (bench model)	Expert faculty (number ambiguous)	OSSA
Sidhu, 2006 ⁷	Junior trainees (<i>n</i> =10) Senior trainees (<i>n</i> =12)	Laparoscopic sigmoid colectomy (live animal)	Trained expert (<i>n</i> =2)	GRS
Pandey, 2008 ¹²	Surgical trainees (<i>n</i> =42) (level not specified)	SFJ ligation, arterial anastomosis (bench model)	Not specified (<i>n</i> =2)	GRS
van Empel, 2013 ⁸	Year 1–3 trainees (<i>n</i> =57) Year 4–6 trainees (<i>n</i> =42)	Hand tie exercises (bench model)	Senior surgeon (number not specified)	OSATS VAS
Hu, 2013 ²³	Interns (<i>n</i> =7) (level not specified)	Suturing and hand tie exercises (bench model)	Senior attending (<i>n</i> =1)	OSATS Global score

GRS = global rating scale; OCRS = operative component rating scale; SFJ = saphenofemoral junction; VR = virtual reality; OSATS = objective structured assessment of technical skills; HTA = hierarchical task analysis; SA = self-assessment; OSSA = objective surgical skills assessment; VAS = visual analogue scale

following review of videotaped performance (level 3 recommendation).

Retrospective video playback analysis was used by the observers in six studies (50.0%).^{7,11,16–18,23} A further two studies (16.7%) used both video analysis and direct observation.^{19,21} In four studies (33.3%), the observers used direct observation only.^{8,12,20,22} The observers were not blinded to the participants' identity in two studies (16.7%)^{8,22} and blinding status was omitted in two (16.7%).^{12,19} Observers were blinded only to surgeon's experience in a final study (8.3%).²⁰ Of these five studies lacking (full) blinding, two papers (40.0%) reported that self-assessment is inaccurate.^{8,12}

Observer assessment

The number of observers was <2 in two studies (16.7%)^{20,23} and either unspecified or ambiguous in a further three (25.0%).^{8,17,22} However, one such paper reported an IRR, implying the use of more than one observer.¹⁷ Of the remaining four papers, two (50.0%) reported poor

self-assessment accuracy.^{8,23} The IRR of observer assessment scores demonstrated at least substantial agreement (0.61–0.80) in all skills assessed and between all observers in five studies (41.7%).^{7,12,16,19,21} Almost perfect agreement (0.81–1.00) was seen in a further two studies (16.7%).^{17,18}

The IRR was not stated in five studies (41.7%)^{8,11,20,22,23} although it would have been inappropriate in two of these where only one observer was used.^{20,23} The IRR was deemed unnecessary in a further paper owing to the use of a previously validated assessment tool.⁸ The reason not to report the IRR was unspecified in the remaining two papers.^{11,22} Self-assessment was inaccurate in two (40.0%) of the papers that did not state the IRR.^{8,23}

Methodological quality

Variable methodological quality was found in the reviewed studies (Table 6). The setting appeared inappropriate in two papers (16.7%) as the technical skill was performed as part of an examination¹² or interview.²⁰ The assessment tools used by the self-assessors differed from those of the observers in

Table 4 Accuracy of self-assessment in the reviewed studies

Accuracy	Study	Level of evidence	Correlation for self vs observer assessment	Conclusions
Accurate	Ward, 2003 ¹⁶	2b	Strong overall at 3 timepoints ($r \geq 0.50$)	Accuracy improved by video review
	Munz, 2004 ¹⁷	2b	Strong for junior trainees for 3 skills ($p^* > 0.05$) Weak for senior trainees for 3 skills ($p^* < 0.05$)	Seniors overestimated performance; impression management reported
	Moorthy, 2006 ¹¹	2b	Weak for junior trainees ($\rho = 0.24$) Strong for senior trainees ($\rho = 0.52$)	Accuracy improved with experience
	Sarker, 2006 ¹⁸	2b	Strong overall ($k=0.79$)	Self-assessment is accurate in the operating theatre
	Brewster, 2008 ¹⁹	2b	Strong during procedure ($r=0.92$) Weak using video playback ($r<0.3$)	Direct expert assessment is accurate; videotape expert assessment is inaccurate
	Tedesco, 2008 ²⁰	2b	Moderate overall ($r=0.4$)	Stressful environment reduces correlation
	Arora, 2011 ²¹	2b	Strong for experienced trainees ($p \leq 0.05$) Strong for inexperienced trainees ($p \leq 0.05$)	Self-assessment is accurate regardless of experience
	de Blacam, 2012 ²²	2b	Moderate overall ($r=0.34$)	Performance underestimated; seniority, age and non-European nationality improve accuracy
Inaccurate	Sidhu, 2006 ⁷	2b	Weak overall ($p > 0.05$)	Performance overestimated
	Pandey, 2008 ¹²	2b	Weak overall for skill 1 ($r=0.04$) Weak overall for skill 2 ($r=0.09$)	Performance overestimated; impression management reported; participants unfamiliar with assessment tools
	van Empel, 2013 ⁸	2b	Weak for novices at 2 timepoints ($r=0.28, 0.24$) Weak for seniors at 2 timepoints ($r=-0.01, -0.33$)	Greater self-confidence in more senior residents; experience in a procedure increases self-confidence
	Hu, 2013 ²³	2b	Weak overall for 5 skills ($p^* \leq 0.05$)	Novice trainees overestimate performance; greater disparity in more advanced techniques

p = probability of differences between self-assessment and observer assessment
 p^* = probability of correlation coefficient between self-assessment and observer assessment

two studies (16.7%).^{8,20} One study (8.3%) failed to justify the expertise of the expert observer.⁸ The number of observers was <2 ^{20,25} or unspecified^{8,22} in four studies (33.3%). Blinding of the external observers was either not done,^{8,22} unstated^{12,19} or only partial²⁰ in five studies (41.7%). As mentioned above, the observer IRR was unstated in five studies (41.7%).^{8,11,20,22,25} Of the four studies reporting that self-assessment is inaccurate, three (75.0%) involved at least one of the methodological limitations mentioned above.^{8,12,23} The overall methodological quality score was higher for papers where self-assessment was accurate than for those finding inaccurate self-assessment (4.38 vs 3.75).

Discussion

This is the first systematic review evaluating the accuracy of technical surgical skills self-assessment. The literature

contained studies with divergent methodological designs and presentation of results. This review demonstrates that, broadly speaking, surgeons can self-assess accurately, contradicting existing evidence from other fields.^{6,9,10}

The evidence suggests that a number of features may improve ability to self-assess. Self-appraisal is more accurate with increased experience,^{11,25} surgical training level and age.²² Future studies should standardise these. A non-European nationality was also found to increase self-assessment accuracy.⁸ The impact of other demographic features on self-assessment (such as sex) would be interesting to pursue. Evidence suggests self-appraisal is more accurate for less advanced competencies.^{17,25} Future studies using surgical skills of varying difficulty could corroborate this.

The skill setting appears to affect self-assessment accuracy. Two papers collected data in stressful environments:

Table 5 Self-assessment versus observer assessment in the reviewed studies

Study	Self-assessment			Observer assessment			
	Tool	Observation	Timing	Tool	Observation	Timing	Blind
Ward, 2003 ¹⁶		GRS, OCRS	Direct, video	Timepoints 1–3*	GRS + OCRS	Video	After
Munz, 2004 ¹⁷		GRS	Direct	After	GRS	Video	After
Moorthy, 2006 ¹¹		OSATS	Direct	After	OSATS	Video	After
Sarker, 2006 ¹⁸		HTA	Direct	After	HTA	Video	After
Brewster, 2008 ¹⁹		Standardised forms	Video	After	Standardised forms	Obs 1: Direct Obs 2/3: Video	During + after
Tedesco, 2008 ²⁰		Self-assessment score	Direct	After	GRS	Direct	During
Arora, 2011 ²¹		OSATS	Not specified	After	OSATS	Obs 1: Direct Obs 2: Video	During + after
de Blacam, 2012 ²²		OSSA	Direct	Before + after	OSSA	Direct	During
Sidhu, 2006 ⁷		GRS	Direct	After	GRS	Video	After
Pandey, 2008 ¹²		GRS	Direct	After	GRS	Direct	During
van Empel, 2013 ⁸		VAS	Direct	Timepoints 1–3**	OSATS	Direct	After
Hu, 2013 ²³		Global score, OSATS	Video	After	Global score, OSATS	Video	After

GRS = global rating scale; OCRS = operative component rating scale; OSATS = objective structured assessment of technical skills; HTA = hierarchical task analysis; Obs = observer; OSSA = objective surgical skills assessment; VAS = visual analogue scale
 *1 = after skill performance; 2 = after review of videotaped performance; 3 = after review of 4 videotaped benchmark performances
 **1 = before first day of course; 2 = after first day of course; 3 = after 6-week autonomous training period

one as an adjunct to an examination¹² and another in an interview.²⁰ Future studies should aim to simulate, not exceed, stress levels found in the operating theatre. A further study highlighted that evaluating self-assessment

using animal models may not directly relate to theatre self-assessment.⁷ Only one study evaluated self-assessment in the live operating setting, finding self-appraisal to be accurate.¹⁸ Further work is needed to substantiate this.

Table 6 Methodological quality of the reviewed studies

Study	Appropriate assessment setting	SA tool = OA tool	Standard definition of expert observer	Number of observers ≥ 1	Blinding	Inter-rater reliability ≥ 0.4	Total
Ward, 2003 ¹⁶	1	1	1	1	1	1	6
Munz, 2004 ¹⁷	1	1	1	1	1	1	6
Moorthy, 2006 ¹¹	1	1	1	1	1	0	4
Sarker, 2006 ¹⁸	1	1	1	1	1	1	6
Brewster, 2008 ¹⁹	1	1	1	1	0	1	4
Tedesco, 2008 ²⁰	0	0	1	0	0	0	1
Arora, 2011 ²¹	1	1	1	1	1	1	5
de Blacam, 2012 ²²	1	1	1	0	0	0	3
Sidhu, 2006 ⁷	1	1	1	1	1	1	6
Pandey, 2008 ¹²	0	1	0	1	0	1	3
van Empel, 2013 ⁸	1	0	1	0	0	0	2
Hu, 2013 ²³	1	1	1	0	1	0	4

SA = self-assessment; OA = observer assessment

The studies used an array of assessment tools. However, it is unlikely that one standard self-assessment tool can be aptly applied to all technical procedures. Divergent results regarding the accuracy of self-assessment were seen when using the global rating scale^{7,12,16,17} or the objective structured assessment of technical skills.^{11,21,25} While both tools' use by trained assessors has been validated, this review questions its efficacy when self-assessing. Future studies must validate and evaluate technical skill assessment tools designed specifically for self-assessment.

The accuracy of self-assessment was increased significantly following retrospective video playback analysis. The majority of papers using video playback analysis found accurate self-evaluation.^{16,19,21} This supports the recommendation that self-assessment may be improved by asking participants to review their video recorded performance.

Methodological limitations were found in three^{8,12,25} of the four studies reporting that self-assessment is inaccurate. As a result, 'poor accuracy' of self-assessment cannot be confidently ascribed to the self-assessor as the reliability of the external assessment score was deficient. Future studies should consider using multiple observers, calculate the IRR and use appropriate blinding. The assignment of an individual as an 'expert', capable of judging the participants, varied, and this should be validated and justified.

This review was limited by the heterogeneity of methodology, outcomes and statistical tests used in the literature, making a meta-analysis unsuitable. This was overcome through use of stringent inclusion and exclusion criteria, allowing common endpoints to be sought. This may have resulted in the omission of some papers. However, the selected literature offers comparable, consistent evidence for the ability of surgeons to self-assess that is more relevant for the surgeon.

Conclusions

Surgeons are capable of evaluating their own technical skills accurately. Self-assessment is most appropriately used for more senior trainees practising less advanced skills and can be improved by using retrospective video playback analysis. Future studies should scrutinise the assessment tool used in more detail and evaluate the reliability of self-assessment in the live operating setting. Greater standardisation is required for the expert assessors, including adequate observer numbers, IRR and blinding.

References

1. Bandura A. Self-efficacy mechanism in human agency. *Am Psychol* 1982; **37**: 122–147.
2. Accreditation Council for Graduate Medical Education. *ACGME Program Requirements for Graduate Medical Education in General Surgery*. Chicago: ACGME; 2012.
3. Spencer JA, Jordan RK. Learner centred approaches in medical education. *BMJ* 1999; **318**: 1,280–1,283.
4. Evans AW, Leeson RM, Petrie A. Reliability of peer and self-assessment scores compared with trainers' scores following third molar surgery. *Med Educ* 2007; **41**: 866–872.
5. MacDonald J, Williams RG, Rogers DA. Self-assessment in simulation-based surgical skills training. *Am J Surg* 2003; **185**: 319–322.
6. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med* 1991; **66**: 762–769.
7. Sidhu RS, Vikis E, Cheifetz R, Phang T. Self-assessment during a 2-day laparoscopic colectomy course: can surgeons judge how well they are learning new skills? *Am J Surg* 2006; **191**: 677–681.
8. van Empel PJ, Verdam MG, Huirne JA *et al*. Open knot-tying skills: resident skills assessed. *J Obstet Gynaecol Res* 2013; **39**: 1,030–1,036.
9. Falchikov N, Boud D. Student self-assessment in higher education: a meta-analysis. *Rev Educ Res* 1989; **59**: 395–430.
10. Davis DA, Mazmanian PE, Fordis M *et al*. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* 2006; **296**: 1,094–1,102.
11. Moorthy K, Munz Y, Adams S *et al*. Self-assessment of performance among surgical trainees during simulated procedures in a simulated operating theater. *Am J Surg* 2006; **192**: 114–118.
12. Pandey VA, Wolfe JH, Black SA *et al*. Self-assessment of technical skill in surgery: the need for expert feedback. *Ann R Coll Surg Engl* 2008; **90**: 286–290.
13. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement. *PLoS Med* 2009; **6**: e1000097.
14. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd edn. Hillsdale, NJ: Lawrence Erlbaum; 1988.
15. Carter FJ, Schijven MP, Aggarwal R *et al*. Consensus guidelines for validation of virtual reality surgical simulators. *Surg Endosc* 2005; **19**: 1,523–1,532.
16. Ward M, MacRae H, Schlachta C *et al*. Resident self-assessment of operative performance. *Am J Surg* 2003; **185**: 521–524.
17. Munz Y, Moorthy K, Bann S *et al*. Ceiling effect in technical skills of surgical residents. *Am J Surg* 2004; **188**: 294–300.
18. Sarker SK, Hutchinson R, Chang A *et al*. Self-appraisal hierarchical task analysis of laparoscopic surgery performed by expert surgeons. *Surg Endosc* 2006; **20**: 636–640.
19. Brewster LP, Risucci DA, Joehl RJ *et al*. Comparison of resident self-assessments with trained faculty and standardized patient assessments of clinical and technical skills in a structured educational module. *Am J Surg* 2008; **195**: 1–4.
20. Tedesco MM, Pak JJ, Harris EJ *et al*. Simulation-based endovascular skills assessment: the future of credentialing? *J Vasc Surg* 2008; **47**: 1,008–1,011.
21. Arora S, Miskovic D, Hull L *et al*. Self vs expert assessment of technical and non-technical skills in high fidelity simulation. *Am J Surg* 2011; **202**: 500–506.
22. de Blacam C, O'Keeffe DA, Nugent E *et al*. Are residents accurate in their assessments of their own surgical skills? *Am J Surg* 2012; **204**: 724–731.
23. Hu Y, Tiemann D, Michael Brunt L. Video self-assessment of basic suturing and knot tying skills by novice trainees. *J Surg Educ* 2013; **70**: 279–283.