



Original article

Are we really as good as we think we are?

AW Evans¹, B Aghabeigi¹, R Leeson¹, C O'Sullivan², J Eliahoo²

¹*Department of Oral and Maxillofacial Surgery, Eastman Dental Institute for Oral Health Care Sciences, University College London, London, UK*

²*Vezev Strong Wing, The University College London Hospitals NHS Trust, London, UK*

Differences are examined in assessment and self-assessment scores, in oral and maxillofacial surgery trainees and MSc postgraduates, following the surgical removal of lower third molar teeth. This study found evidence of a surprising and worrying over-rating of their own surgical skills by many trainees and postgraduates.

Key words: Surgical skills – Postgraduates – Oral and maxillofacial surgery – Assessment and Self-assessment

It is only by having a valid and accurate perception of our own performance that we can work on and improve our surgical skills. Self-assessment is an important tool in improving clinical practice, but there must be concurrence between self-claimed and external measures of performance.¹

This paper aims to examine the difference of assessment and self-assessment scores in oral and maxillofacial surgery trainees and MSc postgraduates following the surgical removal of lower third molar teeth.

Subjects and Methods

A total of 17 trainees and MSc postgraduates were assessed when surgically removing lower third molar teeth under general anaesthesia. The teeth were selected on the basis that their removal would necessitate raising of a flap and removal of bone. Assessors were members of staff of the department. One assessor was scrubbed, assisting and, where necessary, training the operator; the second observed

the procedure closely. Where necessary, the assessor/trainer instructed and/or took over the procedure in the normal way.

Operators were shown the assessment forms prior to the surgery. They were told that the assessment would not count in any way towards their continuous assessment.

Methods of assessment were:

1. An objective assessment of whether 20 components of the procedure were correctly or incorrectly performed.² In cases where the trainer corrected the operative technique or took over, the relevant parts of the procedure were judged incorrectly performed (maximum mark, 20).
2. An operative global rating scale (1–5).³ The scale is anchored by descriptors and measures different aspects of performance, *i.e.* respect for tissue, time and motion, instrument handling, knowledge of instruments, flow of operation, use of assistants, knowledge of procedure, overall performance (maximum mark, 40).

Correspondence to: AW Evans, Department of Oral and Maxillofacial Surgery, Eastman Dental Institute for Oral Health Care Sciences, University College London, 256 Grays Inn Road, London WC1X 8LD, UK

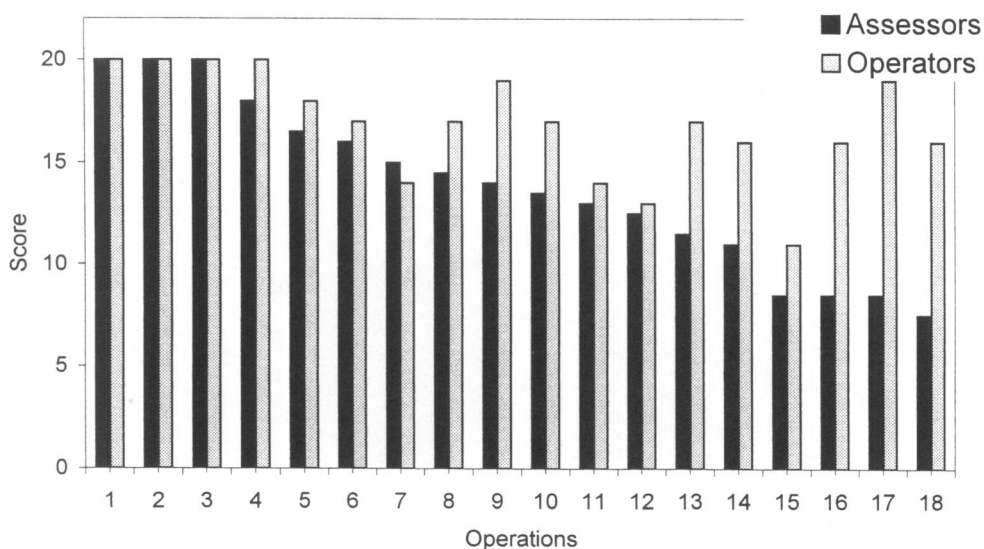


Figure 1 A comparison of objective checklist scores by assessors and operators.

Both types of assessment were marked by the two assessors during or immediately postoperatively. The operator was asked to assess his performance using the same assessment form immediately postoperatively. The results were correlated using standard statistical techniques.

Results

A total of 22 lower third molar teeth were removed by 17 different operators. There were 8 different assessors using both the objective checklist and global rating scales. In 18 cases, operators assessed their performance using both scales.

There was no evidence of a difference between the marks of the two assessors. Using a two-way analysis of variance $P = 0.70$ and $P = 0.68$ for the objective and global rating scales, respectively. The level of agreement between assessors was 86.36% ($\kappa = 0.79$, $P < 0.001$) in the objective checklist scale and 90.91% ($\kappa = 0.83$, $P < 0.001$) in the global rating scale.

Figure 1 shows self-assessment scores compared to the mean scores of each of the two assessors using the objective checklist scale. Figure 2 shows self-assessment scores compared to the mean scores of each of the two assessors using the global rating scale.

Two-way analysis of variance shows that there was evidence of a difference between assessors' and self-assessment marks using both types of scores (objective checklist score, $P < 0.001$; global rating score, $P < 0.001$).

Discussion

Although there was evidence of good agreement between assessors, there was poor agreement between assessors and operators when using both the objective checklist and global rating scales. Operators almost invariably scored themselves higher than the assessors. Some of these differences were substantial and some operators who were scored very low by assessors scored themselves extremely high. In the objective scale scores were up to 10.5 marks higher (maximum 20) than those of the assessors. They were up to 12.5 marks higher (maximum 40) in the global scale.

These results suggest that some operators have poor judgement and over-rate their surgical ability even when assessed for a specific procedure and given rigid criteria against which to mark.

Little work appears to have been done on self-assessment of specific clinical procedures, especially when marking the self-assessment **after** the procedure concerned was performed. There have, however, been reports⁴⁵ of relatively poor agreement between external measures of medical students' clinical performance and students' self-assessment of their performance. Additionally, lower performing medical students tended to rate their clinical performances higher than did their peers at initial self-assessment.⁵

In the present study, objective checklist scores although having very rigid criteria tended to be overscored more than the global rating scale where operators were perhaps reluctant to give themselves marks at the extremes of the scale. Certainly over-scoring of checklist criteria suggests that either operators did not know what was expected of them or in some cases exhibited a considerable degree of

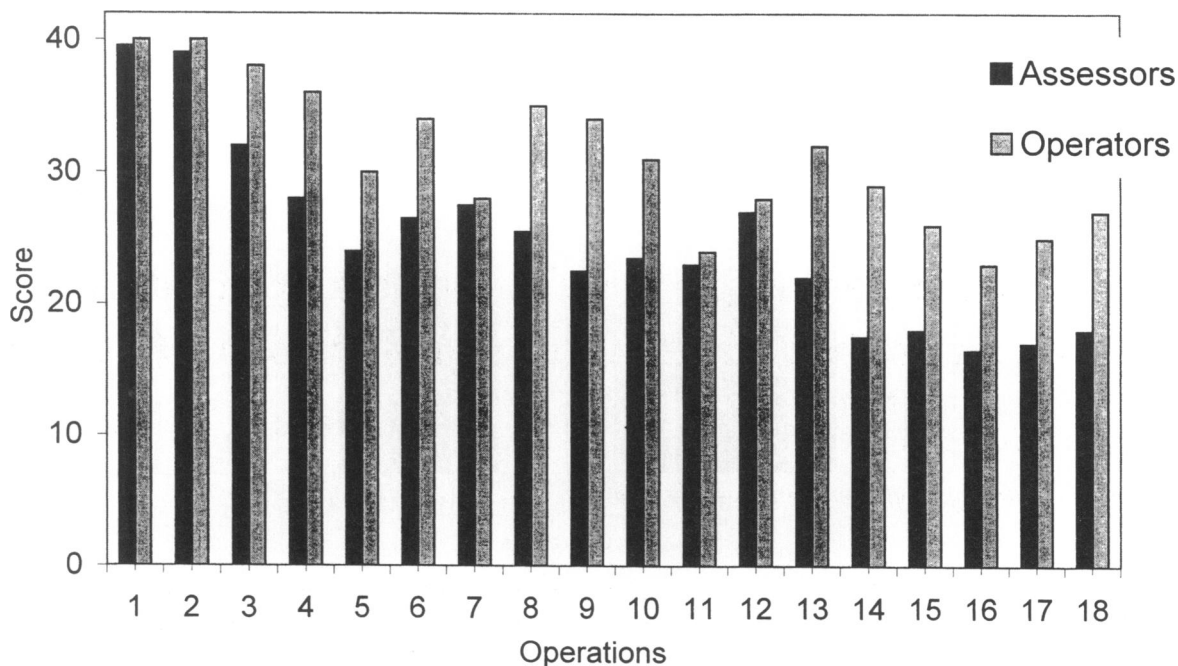


Figure 2 A comparison of global rating scores by assessors and operators.

self-deception. Alternatively, they may have scored potential or ideal performance or even tried to compensate for poor performance as a defence mechanism.^{5,6}

Conclusions

The results of this study found evidence of a surprising and worrying over-rating of their own surgical skills by many trainees and postgraduates in oral and maxillofacial surgery. There can be little doubt that there is a need to evaluate further the accuracy of self-assessment of operative skills. In conjunction with this, we must train surgeons to evaluate critically their performance and self-assessment can form an excellent basis for constructive feedback between trainer and trainee.

It may be found that some individuals will never develop the judgement to assess accurately their performance. It would be invaluable to have a way to identify these individuals so that they could be redirected at an early stage in their careers.

Acknowledgements

This work was undertaken by AW Evans, B Aghabeigi, R Leeson, C O'Sullivan and J Eliahoo with University College London Hospitals who received a proportion of its funding from the NHS Executive; the views expressed in this publication are those of the authors and are not necessarily those of the Trust or the NHS Executive.

References

1. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med* 1991; **66**: 762-9.
2. Evans AW. Assessing competence in surgical dentistry. *Br Dent J* 2001; **190**: 343-6.
3. Dath D, Reznick RK. Objectifying skills assessment: bench model solutions. In: Anon. *Surgical Competence. Challenges of Assessment in Training and Practice*. London: The Royal College of Surgeons, 1999; 52-66.
4. Antonelli MA. Accuracy of second-year medical students' self-assessment of clinical skills. *Acad Med* 1997; **72**: 563-5.
5. Woolliscroft JO, TenHaken J, Smith J, Calhoun JG. Medical students' clinical self-assessments: comparisons with external measures of performance and the students' self-assessments of overall performance and effort. *Acad Med* 1993; **68**: 285-94.
6. Arnold L, Willoughby TL, Calkins EV. Self evaluation in undergraduate medical education: the longitudinal perspective. *Med Educ* 1985; **60**: 21-8.