

---

## Intercollegiate specialty board oral examinations in surgery: purpose, content and marking criteria

M.H. Davis<sup>1</sup>, G.G. Ponnampereuma<sup>2</sup>, S. McAleer<sup>3</sup>, D. Galloway<sup>4</sup>

### Abstract

The traditional viva has been much criticised in the literature for lack of reliability and validity, but its use has continued in some postgraduate exams in the UK and commonwealth countries.

This study was designed to identify what examiners in nine sub-specialties of one discipline (surgery) thought they were testing in a viva and to introduce standardisation into the existing five-point marking scale used to grade candidates' performance in the viva.

Three hundred and one surgical examiners took part in one of eight similar workshops that used a plenary method to identify what the examiners thought they were assessing in the viva and a small group approach to identify descriptors for each point on the five-point marking scale. The examiners identified what they were assessing, and their answers were then categorised into three meta-competencies: professional ability/patient care; knowledge and judgement; and quality of response. The examiners used their experience of examining to identify descriptors for each point on the existing five-point rating scale. The descriptors were grouped under one of the three meta-competencies to create marking criteria for the new scenario-based oral exam that will be part of a larger clinical exam.

The meta-competencies are similar to those identified by other postgraduate oral exams. This study is reported as the methodology for the study and the meta-competencies being assessed may have wider application.

---

### Introduction

The oral examination (or viva voce), has been a traditional part of the examiner toolkit for many years. Examiners, sometimes working in pairs, ask the candidate random questions. Both the selection of questions and the scoring of the candidate answers are at the discretion of the examiners involved in the viva. The viva is usually unstructured and

does not have pre-validated questions and answers. The topics assessed, the level of difficulty of the questions asked, and the amount of help or prompting the candidate receives may vary widely (Davis & Karunathilake, 2005). The exam may thus be unfair to individual candidates.

There is evidence from studies of viva type oral exams that in addition to knowledge the format tends to assess candidate traits, such as personality (Bull, 1959; Holloway et al., 1967; Holloway et al., 1968); verbal style and dress (Rowland-Morin et al., 1991; Burchard et al., 1995); ethnicity (Roberts et al., 2000); and social background and gender (Esmail & May, 2000). This raises the question of what is being tested in the viva voce. Other concerns include cost effectiveness (McGuire, 1966), and examinee stress and acceptability (Jolly & Grant, 1997; Schiff, 2001). Norman (2000) reported that the

---

<sup>1</sup>Professor of Medical Education and Director, Centre for Medical Education, University of Dundee, Scotland, UK

<sup>2</sup>Lecturer in Medical Education, University of Colombo, Sri Lanka

<sup>3</sup>Senior Lecturer in Medical Education, Centre for Medical Education, University of Dundee, Scotland, UK

<sup>4</sup>Chairman, Joint Committee on Intercollegiate Examinations, Intercollegiate Specialty Boards, Central Administration, 2 Hill Place, Edinburgh EH8 9DS, UK

American exam boards discontinued the viva about 30 years ago because of these problems. The UK and other commonwealth

Colleges of Surgery and is the body responsible for the postgraduate surgical sub-specialty examinations in the United Kingdom and Ireland: Intercollegiate Specialty Board (ISB) exams. To identify why the ISB continued to use vivas, we asked the ISB examiners what they valued about the viva and what they assessed in the exam. Using the answers that we received for the above two questions we developed meta-competencies for the new ISB oral exams to replace the viva. A small group approach was used to develop marking descriptors for each point on a five-point scale for each meta-competence. The ISB refined the descriptors to create marking criteria. The new ISB oral exams take place within a structured, standardised clinical exam, part of which is scenario-based (oral exams) and part of which is based on real, standardised or simulated patients. The development of the structured, standardised clinical exam will be reported at a later date.

## Methods

Between August 2004 and February 2005 eight similar workshops were held for examiners in the ISB exams. In all, 301 ISB examiners attended a workshop. Each examiner was from one of the nine sub-specialties of the JCIE (i.e. general surgery, plastic surgery, cardio-thoracic surgery, urology, paediatric surgery, oral and maxillofacial surgery, otolaryngology, neurosurgery, and trauma and orthopaedic surgery). The number of examiners per workshop varied from 16 to 57 (mean = 38). Examiners worked in plenary session and in small groups during the workshops. Some small groups comprised examiners from a single sub-specialty while others were made up of examiners from more than one sub-specialty.

At plenary brainstorming sessions, the examiners were asked: what do you assess at the oral exam? The answers were written on a flip chart during the sessions. Later the answers were refined by removing duplications and categorised to identify the meta-competencies that the examiners assessed at the oral exam.

countries, however, continued with the viva for some exams.

The Joint Committee on Intercollegiate Examinations (JCIE) represents the Royal Small groups of approximately eight examiners were asked to develop a scenario and questions for a viva lasting approximately five minutes. Two groups then joined to form a group of 16. This group of 16 role played two vivas, one developed by each group of eight. Two examiners from the group that developed the scenario questioned a simulated candidate, who was a volunteer from the second group. Fifteen examiners scored each viva. The simulated candidate for one of the two vivas was asked to role play a borderline candidate. All 15 examiners were asked to reveal their scores. In no case was the scoring unanimous. In all cases where the examiner simulated a borderline candidate some of the examiners awarded a passing score and some a failing score. This exercise was used to convince the examiners of the need for marking descriptors for the scoring system to improve inter-rater reliability and was the only purpose of this exercise. The examiners, working in groups of 16, were then asked to make use of their experience to provide descriptors for each point on the 4 to 8 rating scale, currently used in the ISB exams.

After all the workshops were held, all descriptors generated by different examiner groups were collated to form a single set of descriptors for each rating point. The duplications were removed. The rating descriptors were then categorised under the meta-competencies that the examiners identified to formulate scoring criteria. These criteria were further refined by the JCIE for ease of use and were agreed by the JCIE for use in the ISB exams.

## Results

### *The meta-competencies*

Table 1 shows a summary of what the examiners said they were assessing in the oral exam, categorised into meta-competencies.

### *The marking descriptors*

All the descriptors that the examiners identified for each rating point are

available from the first named author on request. Duplicate descriptors were removed. The remaining descriptors for each point on the rating scale were

categorised under one of the three meta-competencies. Table 2 shows the descriptors for each rating point, linked to the three meta-competencies.

**Table 1 : What the oral examiners said they were assessing at the oral exam, categorised into meta-competencies**

Meta-competencies	Examiner answers to the question: what do you assess at the oral exam
Professional capability/ patient care	Professionalism  Demeanour / behaviour / personal attributes / attitudes / personality / honesty / integrity  Logical step-wise thought process, problem solving, i.e. higher order thinking related to surgery  Safety and competence for independent practice  Ethical approach to a clinical problem  Managing a patient with a surgical problem, e.g. history taking, examination, differential diagnosis, investigation and management  Reaction to "stress"  Approach to unknown situations / 'grey areas'  Integration of competencies  Determining the candidate experience
Knowledge and judgement	Justification of clinical decisions with evidence from literature  Clinical application of basic sciences  Wide range of topics  Exploration of candidate opinion
Quality of response	Logical step-wise thought process, problem solving, i.e. higher order thinking related to surgery  Communication skills  Reaction to "stress"  Approach to unknown situations / 'grey areas'

**Table 2 : The descriptors for each rating point categorised by the three meta-competencies**

Rating scale level	Descriptors		
	Professional capability / patient care	Knowledge and judgement	Quality of response
4	dangerous to patient care; unprofessional; dysfunctional; aggressive	Failed in all or most of the competences; factually incorrect information; fundamental lack of knowledge and higher order thinking skills	prompting does not work; no insight; disorganised, confused, inconsistent thought process
5	unacceptable professional behaviour; poor communication skills; not suitable for independent practice; may do more harm than good to the patient (i.e. unsafe)	cannot elaborate on an answer; poor higher order thinking skills; significant errors; poor deductive skills	slow responses; needs prompting for basic questions; does not volunteer correct answers; hesitant/unsure
6	can diagnose and manage common surgical conditions; can run a week-end shift; comfortable having in ward; just competent; flexible	good higher order thinking capacity; core information all correct, but no depth; inconsistent knowledge	confident; spontaneous; refuses to change track, but otherwise a 7; one answer outright fail; little prompting
7	can diagnose and manage common surgical conditions; can run a week-end shift; comfortable having in ward; just competent; flexible	good supporting evidence for answers; can cope with difficult topics/problems; sound interpretation and judgement; sound knowledge, but does not give more than needed; working out answers using the trainer's practice rather than their own practice	volunteers information; keeps telling what the examiners want to know; confident
8	happy to look after me and my family	clarity of thought; familiarity with the subject; "flies"; highly organised; extensive flawless knowledge with justification from the literature; has insight and good judgement	the examiner enjoyed the viva; viva inspired the examiner; examiner 'racing to keep up'

*Refinement for ease of use*

The JCIE and the authors refined the marking descriptors and the wording to

develop the final marking criteria, and the 'questions – answers - prompting' sub-categories under 'quality of response' in table 3, for ease of use.

**Table 3 : The rating rubrics for the JCIE oral examination**

Rating scale	Overall professional capability / patient care	Knowledge and judgment	Quality of response
4	The candidate demonstrated incompetence in the diagnosis and clinical management of patients to a level which caused serious concerns to the examiner	<ul style="list-style-type: none"> <li>• Did not get beyond default questions</li> <li>• Failed in most/all competencies</li> <li>• Poor basic knowledge/ judgment/understanding to a level of concern</li> <li>• Serious lack of knowledge</li> </ul>	<ul style="list-style-type: none"> <li>• Q<sup>1</sup>: Does not get beyond default questions</li> <li>• A<sup>2</sup>: Disorganised/ confused/ inconsistent answers, lacking insight</li> <li>• P<sup>3</sup>: Un-persuadable – prompts do not work</li> </ul>
5	The candidate failed to demonstrate competence in the diagnosis and clinical management of patients	<ul style="list-style-type: none"> <li>• Demonstrated a lack of understanding</li> <li>• Difficulty in prioritising</li> <li>• Gaps in knowledge</li> <li>• Poor deductive skills</li> <li>• Poor higher order thinking</li> <li>• Significant errors</li> <li>• Struggled to apply knowledge /judgment/management</li> <li>• Variable performance</li> </ul>	<ul style="list-style-type: none"> <li>• Q: Frequent use of default questions</li> <li>• A: Confused/ disorganised answers; hesitant and indecisive</li> <li>• P: Required frequent prompting</li> </ul>
6	The candidate demonstrated competence in the diagnosis and clinical management of patients	<ul style="list-style-type: none"> <li>• Good knowledge and judgment of common problems</li> <li>• Important points mentioned</li> <li>• Instils confidence</li> <li>• No major errors</li> </ul>	<ul style="list-style-type: none"> <li>• Q: Copes with competence questions</li> <li>• A: Methodical approach to answers; has insight</li> <li>• P: Requires minimal prompting</li> </ul>
7	The candidate demonstrated confidence and competence in the diagnosis and clinical management of patients	<ul style="list-style-type: none"> <li>• Ability to prioritise</li> <li>• Coped with difficult topics/ problems</li> <li>• Good decision making/ provided supporting evidence</li> <li>• Reached a good level of higher order thinking</li> <li>• Strong interpretation/ judgment but didn't quote the literature</li> </ul>	<ul style="list-style-type: none"> <li>• Q: Goes beyond the competence questions</li> <li>• A: Logical answers and provided good supporting reasons for answers</li> <li>• P: Fluent responses without prompting, but some prompting on literature</li> </ul>
8	The candidate demonstrated confidence and competence in the diagnosis and clinical management of patients to a level which would inspire confidence in the patient	<ul style="list-style-type: none"> <li>• At ease with higher order thinking</li> <li>• Flawless knowledge plus insight and judgment</li> <li>• Good understanding/ knowledge/management/ prioritisation of complex issues</li> <li>• Had an understanding of the breadth and depth of the topic, and quoted from literature</li> <li>• High flyer</li> <li>• Strong interpretation/judgment</li> </ul>	<ul style="list-style-type: none"> <li>• Q: Stretches examiners – answers questions at advanced level</li> <li>• A: Confident, clear, logical and focused answers</li> <li>• P: No prompting necessary</li> </ul>

1 – level of questions; 2– quality of answers; 3 – prompting and its effect

## Discussion

Much of the published work on postgraduate oral exam scoring rubrics is related to general practice. Wakeford et al. (1995) report that the general practice oral examiners identified "the candidates' approach to practice, their decision making skills and their justification for their decisions" as the areas assessed by their oral examination. Of the three meta-competencies in table 1 (i.e. overall professional capability/patient care; knowledge and judgement; and quality of response), the first two encapsulate what Wakeford et al. (1995) have identified. As reported by Ryding & Murphy (1999), Reinhart's (1995) review on oral examinations identifies 'assessing higher order thinking' as one of the key areas. Ryding & Murphy (1999) further state that Libert et al. (1993) have also used the oral examination at Harvard Dental School to assess the ability to think independently; to synthesise interdisciplinary information rapidly; and to exercise sound clinical judgement. All the above literature findings agree with what the surgical examiners identified in the present study.

As for the third meta-competency in table 1 (i.e. quality of response), Wakeford et al. (1995) lay down specific guidelines for the examiners regarding how the questions should be selected and phrased, the answers to be expected from the candidates and guidelines on prompting. 'Questions, answers and prompting' in table 3 are the three areas that the surgical examiners identified in relation to 'quality of response'.

The general practice examiners use a nine-point rating scale, whereas the surgical examiners use a five-point rating scale. We discussed the potential to change to a nine-point scale, but the examiners said they would find the change difficult as they were already calibrated on the five-point scale. The descriptors in the five-point, surgical exam scale, are more elaborate than the single phrase descriptors described by Yaphe & Street (2003). More elaborate descriptors will especially help examiner decisions around the pass-fail borderline; i.e. rating points 5 and 6 in table 3. Scales with too many rating points may also become unwieldy and may produce poor inter-rater reliability (Gray, 1996).

There are strong similarities between the meta-competencies assessed and the

scoring rubrics used in the oral examinations in surgical sub-specialties and in general practice. These meta-competencies may have wider application in postgraduate exams in other disciplines. If such application in other disciplines is confirmed, it may be possible to develop a generic model for medical postgraduate oral exams that will, in part, resuscitate the reputation of the oral exam.

### Practice points

- Viva voces are notoriously unreliable forms of assessment.
- Identification of what is being assessed in an oral exam may improve reliability.
- There is similarity between what general practitioners and surgeons say they assess in oral exams.
- The viva role play methodology demonstrated to the examiners the need for marking criteria.
- Identification of descriptors for rating scales may improve inter-rater reliability.

## References

- Bull, G.M. (1959) Examinations, *Journal of Medical Education*, 34, pp. 1154-1158.
- Burchard, K.W., Rowland-Morin, P.A., Coe, N.P.W. & Grab, J.L. (1995) A surgery oral examination: inter-rater agreement and the influence of rater characteristics, *Academic Medicine*, 70, 1044-1046.
- Davis, M.H. & Karunathilake, I. (2005) The place of the oral examination in today's assessment systems, *Medical Teacher*, 27, 294-297.
- Esmail, A. & May, C. (2000) Oral exams – get them right or don't bother, *British Medical Journal*, 320, 375.
- Gray, J.D. (1996) Global rating scales in residency education, *Academic Medicine*, 71, S55-S63.
- Holloway, P.J., Collins, C.K. & Start, K.B. (1968) Reliability of viva voce examinations, *British Dental Journal*, 12, 211-214.
- Holloway, P.J., Hardwick, J.L., Morris, J. & Start, K.B. (1967) The validity of essay and viva voce examining techniques, *British Dental Journal*, 123, 227-232.
- Jolly, B. & Grant, J. (1997) *The Good Assessment Guide – A Practical Guide to Assessment and Appraisal for Higher Education*, London: Joint Centre for Education in Medicine, 73.
- Libert, E.M., Fox, C.H., Henry, J.L. & Valachovic, R.W. (1993). Oral comprehensive examinations. Educational exhibit (Booth 502). Paper presented at the American

Association of Dental Schools (AADS) Annual Meeting, Chicago.

McGuire, C.H. (1966) The oral examination as a measure of professional competence, *Journal of Medical Education*, 41, 267-274.

Norman, G. (2000). Examining the examination: Canadian versus US radiology certification exam, *Canadian Associations of Radiologists Journal*, 51, 208-209.

Reinhart, M.A. (1995) Advantages to using the oral examination. In: Mancall, M.D. & Bashook, P.G. (Eds) *Assessing clinical reasoning: the oral examination and alternative methods*, American Board of Medical Specialties, Evanston IL. 31-39.

Roberts, C., Sarangi, S., Southgate, L., Wakeford, R. & Wass, V. (2000) Oral examinations equal opportunities, ethnicity, and fairness in the MRCGP, *British Medical Journal*, 320, 370-374.

Rowland-Morin, P.A., Burchard, K.W., Grab, J.L. & Coe, N.P. (1991) Influence of effective communication by surgery students on their oral examination scores. *Academic Medicine*, 66, 169-171.

Ryding, H.A. & Murphy, H.J. (1999) Employing oral examinations (viva voce) in assessing dental students' clinical reasoning skills, *Journal of Dental Education*, 63, 682-687.

Schiff, R. (2001). A short case prolonged, *British Medical Journal*, 323, 551.

Wakeford, R., Southgate, L. & Wass, V. (1995) Improving oral examinations: selecting, training, and monitoring examiners for the MRCGP. Royal College of General Practitioners, *British Medical Journal*, 311, 931-935.

Yaphe, J. & Street, S. (2003) How do examiners decide?: a qualitative study of the process of decision making in the oral examination component of the MRCGP examination, *Medical Education*, 37, 764-771.