

COMMENTARY

The place of the oral examination in today's assessment systems

MARGERY H. DAVIS¹ & INDIKA KARUNATHILAKE²

¹Centre for Medical Education, University of Dundee, UK; ²Faculty of Medicine, University of Colombo, Sri Lanka

The oral examination or viva is a traditional form of assessment in which one or more examiners fire questions at the candidate. It typically takes the form of an interview or discussion between the examiners and candidate and happens in an examination hall or other such setting away from patients. It should be distinguished from other types of oral examination such as the long and short case, which take place in the presence of the patient or are focused around a patient seen by the candidate and the oral that is used for defence of written work such as a thesis. The oral examination is said to assess knowledge, to probe depth of knowledge and to test other qualities such as mental agility.

The use of oral examinations in high-stakes assessment systems has been criticized for many years because of low reliability (Colton & Peterson, 1967; Foster *et al.*, 1969; Kelly *et al.*, 1971). The low reliability relates, in part, to the examiner's active participation in the examination, which can introduce bias. In the traditional oral, each candidate may receive a different assessment with regard to content areas addressed, the difficulty of the questions asked, the level of prompting or help provided and the learning outcomes assessed; for example, knowledge of the basic sciences, patient investigation and management. These differences present difficulties not only in a norm-referenced system of assessment, where the intention is to rank the candidates, but also in a criterion-referenced system, where the intention is to assess whether or not the candidate has achieved a pre-determined standard.

The reasons for low reliability also have an adverse impact on validity (Schuwirth & Van der Vleuten, 1996) because of the potential for variation in content matter addressed and in the emphasis given to different content areas.

Oral examinations are usually employed in an attempt to assess the candidate's knowledge of a subject. What may be measured, however, are aspects of a candidate's personality (Bull, 1959). Holloway *et al.* (1967), Holloway *et al.* (1968) and Thomas *et al.* (1992) showed that viva marks correlated with personality scores. Rowland-Morin *et al.* (1991) and Burchard *et al.* (1995) showed that verbal style and dress of the candidates influence oral examination scores. Roberts *et al.* (2000) carried out a discourse analysis (a detailed study of language in use) of the oral component in the membership examination of the Royal College of General Practitioners (MRCGP) and pointed out that candidates

from ethnic minorities and those trained abroad may experience particular hidden difficulties with oral examinations leading to discrimination. Furthermore, the discrimination may not be limited to ethnicity. Esmail & May (2000) suggested that candidates from working-class backgrounds and, in some instances, female candidates may also be discriminated against.

The problems with oral examinations extend beyond poor reliability and validity. McGuire (1966) questioned the cost effectiveness of oral examinations, when the cost, in terms of professional time and energy, is weighed against its reliability and validity as a measure of professional competence. Any well-planned examination, however, is costly in terms of examiners' time and effort. The challenge is finding assessment instruments where the effort spent is educationally 'profitable'.

Orals can be highly threatening for candidates with resultant poor performance (Pokorny & Frazier, 1966; Cox, 1982; Thomas *et al.*, 1992; Jolly & Grant, 1997). It can be argued, however, that all examinations are stressful. The question is whether the viva is more stress provoking than other assessments. There is no evidence that orals are more stressful than other exams and, indeed, there is anecdotal evidence to the contrary. Schiff (2001), in a personal narrative, reported that the short case was more stressful than other parts of the MRCP clinical examination.

Norman (2000) pointed out that "most of the US boards abandoned the oral exam altogether about 30 years ago, based on evidence that, despite intuition to the contrary, it was adding little value to the evaluation process". Oral examinations, however, continue to be used in specialty board examinations in the UK, Canada and most parts of the former British Empire (Norman, 2000). Why is this?

First, the oral examination is a traditional form of assessment that has been used in undergraduate and postgraduate medical education for many years and breaking with tradition can be difficult. As Jayawickramarajah (1985) suggested, there are "difficulties in persuading examining boards and training medical examiners to employ appropriate alternative methods".

Correspondence: Prof. M.H. Davis, Centre for Medical Education, University of Dundee, Tay Park House, 484, Perth Road, Dundee DD2 1LR, UK.
Email: m.h.davis@dundee.ac.uk

Second, proponents of the oral examination suggest that it has several advantages over other forms of tests, including;

- (1) direct personal contact (Cox, 1982);
- (2) assessing problem-solving and reasoning (Sandars, 1998; Wass *et al.*, 2003);
- (3) recognition of safe and competent clinicians (Zelenock *et al.*, 1985);
- (4) assessing professionalism and ethics (Wass *et al.*, 2003);
- (5) opportunity to probe depth of knowledge (Cox, 1982; Gibbs *et al.*, 1988; Jolly & Grant, 1997);
- (6) flexibility in moving from one area to another (Deale, 1975; Schwartz & Sein, 1987; Gibbs *et al.*, 1998; Wakeford *et al.*, 1995);
- (7) feedback on curriculum (Colton & Peterson, 1967);
- (8) the ability to tailor the questions asked to the needs of each individual candidate (Gibbs *et al.*, 1993).

There are a few studies to support the claimed advantages of oral examinations. One such study conducted by McFarlane *et al.* (1989) used factor analysis to show that the oral examination measured the students' capacity to formulate ideas and their communication skills. In another study Zelenock *et al.* (1985) used multiple regression analysis to demonstrate that oral examinations can measure areas relevant to the clinical competence of medical students, such as problem-solving. Other claimed advantages remain unproven. For example, Colton & Peterson (1967) analysed oral examination scores of students at Harvard Medical School. They were unable to prove that student abilities such as problem-solving and decision-making can be reliably and validly assessed by an oral interrogation. Unproven does not, however, mean that the advantages are not real, but that the work to demonstrate the advantages has not been carried out.

Sandars (1998) and Wass *et al.* (2003) suggested that the ability to assess the candidate's decision-making skills is an advantage in oral examinations. But three major studies conducted separately by Evans *et al.* (1966), McGuire (1966) and Jayawickramarajah (1985), to analyse the cognitive level of questions asked during an oral examination, showed that oral examinations tend to test at a low taxonomic level; for example, recall of factual knowledge rather than problem-solving. In an important early study that analysed oral exam questions put to candidates who sat for specialty board examinations in USA, McGuire (1966) found that "These oral examinations measure a candidate's ability to recall isolated fragments of knowledge rapidly and under stress, that candidates rarely cite evidence for their answers...". Schuwirth & van der Vleuten (1996) also pointed out that oral examinations largely test factual knowledge, which can be better tested in written examinations.

Jolly & Grant (1997), Cox (1982) and Gibbs *et al.* (1993) suggested that during an oral examination it is possible to explore particular questions in more depth and to explore understanding further by raising more questions. The restricted nature of oral examinations with regard to time, however, and the evidence regarding the low taxonomic level of questioning suggest that while there is potential for oral examinations to explore understanding, that potential is underused.

Flexibility in moving from one area to another during the examination is cited as an advantage by Deale (1975), Schwartz & Sein (1987), Gibbs *et al.* (1988) and Wakeford *et al.* (1995). While this remains one of the advantages of this form of assessment and an attraction for examiners, the lack of standardization entailed is a major contributory factor to its low reliability rating.

Norman (2000) argues that problem-solving is content specific and the perception that the content of any one speciality can be adequately examined with only a few cases is inconsistent with the evidence. Difficulty in sampling from a wide area is cited as a disadvantage of orals (Brown *et al.*, 1996). In spite of being flexible, oral exams are restricted compared with assessment tools such as multiple-choice questions (MCQ) and the objective structured clinical examination (OSCE) that can sample widely. According to Jayawickramarajah (1985), even under the best circumstances the sample of competences that can be assessed by an oral examination within a given period of time is smaller than with a written test, MCQ or OSCE.

Colton & Peterson (1967) pointed out that many examiners consider that oral examinations are a useful feedback mechanism for the examiners, and by personally examining a sample of students the examiner can elicit valuable information on the strengths and weaknesses of the medical curriculum. This is not, however, a unique advantage of oral examinations and feedback regarding the curriculum can be obtained by using any good assessment tool.

So what, then, are our conclusions regarding the place of the oral examination in today's assessment systems? One of the first principles of assessment is that the assessment system should be capable of measuring the individual candidate's achievement of the course/curriculum outcomes. Southgate & Grant (2003) emphasized this principle in their recent document 'Principles and standards for an assessment system for postgraduate medical training'. Course outcomes in the healthcare professions are not restricted to knowledge and relate to the tasks that the professional carries out, the approach to the tasks and the individual's professionalism (Harden *et al.*, 1999). Furthermore these outcomes require to be assessed at different levels. Miller (1990) has taken forward our thinking about assessment by identifying the four levels at which assessment needs to take place; *knows*, *knows how*, *shows how* and *does*. To assess multiple outcomes at four different levels requires an examination toolkit or an assessment system and not merely one assessment instrument. If we select MCQs to assess *knowledge*, extended matching items or one of the many varieties of short-answer questions to assess application of knowledge or "*knows how*", the OSCE to assess "*shows how*" and direct observation of practice to assess "*does*", is there a need to use oral examinations? Jayawickramarajah (1985) pointed out that a decision for continuation of oral examinations should be considered in the light of the more valid and reliable assessment methods available.

Considering all the evidence, traditional oral examinations are probably not proven to be appropriate for high-stakes assessment. This does not mean, however, that they will not be employed in high-stakes examinations. The strength of the conviction of experienced oral examiners of the worth of orals cannot be ignored. Cox (1982) claimed in relation to the oral

examination that “direct personal contact allows assessment of appearance, manner, personality, alertness, confidence, honesty, self-awareness and other aspects of values and attitudes”. It is important that these aspects are assessed. Marshal & Ludbrook (1972) pointed out in relation to oral examinations that “we are not measuring anything, but merely judging that the student is or is not fit to join the club”. If this judgement includes fitness to practise medicine, perhaps it may be welcomed by patients and other stakeholders. Those using oral examinations may wish to consider what it is they wish to assess with the oral and whether there could be other assessment instruments that could do the job better. If the decision is taken to retain the orals in the examination system, there are several suggestions regarding best practice that should be followed.

(1) Structure the oral on clinical scenarios

Fabb & Marshal (1983) differentiated between structured and unstructured oral examinations. Structured oral examinations (SOE) based on a clinical case with well-defined goals can often give great insight into a candidate's knowledge, interpretive ability, problem solving and attitudes. Kearney *et al.* (2002) suggested SOEs can best evaluate the elements of problem-solving. Anastakis *et al.* (1991) measured criterion validity of an SOE for surgical residents by correlating SOE scores with MCQ and OSCE scores. Based on the results, which showed significant correlation, they suggested that the SOE is useful in the assessment of clinical knowledge and problem-solving. The same study reported inter-rater reliability results that were higher than previously reported results of traditional oral examinations. Swing & Bashook (2000) suggested in their ‘Toolbox of Assessment Methods’ that SOE score reliabilities range from 0.65 to 0.88. Wass *et al.* (2003) estimated that the reliabilities of the SOE conducted by the Royal College of General Practitioners were appropriate to high-stakes examinations.

(2) Use a number of orals

Stillman *et al.* (1983) and Daelmans *et al.* (2001) showed that reliability when using a number of orals is better than the reliability of a single oral examination.

(3) Use a number of examiners

Norman (2000) suggested that the oral examination must sample more broadly across cases and examiners. He suggests that one logistically defensible strategy is to use multiple vivas, with a single examiner per viva, and to base the final evaluation on independent assessments by multiple examiners. Swanson (1987) showed a reliability of 0.45 with the same examiner for each of four cases (role-playing orals), rising to 0.69 with a new examiner examining each of four cases and 0.82 with a new examiner for each of eight cases. Wass *et al.* (2003) estimated pass/fail reliabilities of 0.7 with one 20-min oral rising to 0.9 with four different examiners each oralling the candidate for 20 mins.

(4) Ask all candidates the same questions

If all the examiners ask every candidate the same questions on the same clinical material there are a number of benefits, including;

- advance planning to improve sampling of the syllabus;

- elimination of overlap between the orals and other components of the assessment.

Amiel *et al.* (1997) improved reliability in the SOE used with Israeli medical students to 0.64 using five standard questions for each oral.

(5) Use descriptors, rubrics or criteria for answers

Criteria for answers can provide clear guidelines on what is and is not an acceptable answer to the examiner's questions. Anastakis *et al.* (1991) showed that specified answers and a specific marking scheme in an SOE for surgical residents in Canada produced an overall reliability of 0.75.

(6) Train the examiners

Examiner training for oral examinations is crucial, particularly if all examiners are required to test clinical judgement and higher-order thinking across a range of tasks (Des Marchais & Jean, 1993; Wakeford *et al.*, 1995).

The oral examination is a ‘rite of passage’ and forms a common bond between members of individual Royal Colleges. Giving up such a tradition is not easy and it is likely that the oral examination will continue to have a place in assessment systems for some time to come. Substantial work, however, is needed to develop the traditional oral examination into a ‘best practice oral’ format appropriate for high-stakes examinations.

Notes on contributors

MARGERY DAVIS is the Director of the Centre for Medical Education, University of Dundee, UK.

INDIKA KARUNATHILAKE is a Lecturer in Medical Education, Faculty of Medicine, University of Colombo, Sri Lanka.

References

- AMIEL, G.E., TANN, M., KRAUSZ, M.M. & BITTERMAN, A. (1997) Increasing examiner involvement in an Objective Structured Clinical Examination by integrating a structured oral examination, *American Journal of Surgery*, 173, pp. 546–549.
- ANASTAKIS, D.J., COHEN, R. & REZNICK, R.K. (1991) The structured oral examination as a method for assessing surgical residents, *American Journal of Surgery*, 162, pp. 67–70.
- BROWN, M.H., REGEHR, G. & REZNICK, R.K. (1996) The effect of early performance on examiners' marking patterns during an oral examination, *Academic Medicine*, 71, pp. 73S–75S.
- BULL, G.M. (1959) Examinations, *Journal of Medical Education*, 34, pp. 1154–1158.
- BURCHARD, K.W., ROWLAND-MORIN, P.A., COE, N.P.W. & GARB, J.L. (1995) A surgery oral examination: interrater agreement and the influence of rater characteristics, *Academic Medicine*, 70, pp. 1044–1046.
- COLTON, T. & PETERSON, O.L. (1967) An assay of medical students' abilities by oral examination, *Journal of Medical Education*, 42, pp. 1005–1014.
- COX, K.R. (1982) How to improve oral examinations, in: K.R. Cox & C.E. Evans (Eds), *The Medical Teacher* (Edinburgh, Churchill Livingstone).
- DAELMANS, H.E.M., SCHERPBIER, A.J.J.A., VAN DER VLEUTEN, C.P.M. & DONKER, A.B.J.M. (2001) Reliability of clinical oral examination re-examined, *Medical Teacher*, 23, pp. 422–424.
- DEALE, R.N. (1975) Assessment and testing in the secondary school, *Schools Council Examinations Bulletin*, 32 (London, Evans/Methuen Educational).

- DES MARCHAIS, J.E. & JEAN, P. (1993) Effects of examiner training on open-ended, higher taxonomic level questioning in oral certification examinations, *Teaching and Learning in Medicine*, 3, pp. 24–28.
- ESMAIL, A. & MAY, C. (2000) Oral exams—get them right or don't bother, *British Medical Journal*, 320, p. 375.
- EVANS, L.R., INGERSOLL, R.W. & SMITH, E.J. (1966) The reliability, validity and taxonomic structure of the oral examination, *Journal of Medical Education*, 41, pp. 651–657.
- FABB, W.E. & MARSHAL, J.R. (1983) *The Assessment of Clinical Competence in General Family Practice* (Lancaster, MTP Press).
- FOSTER J.T., ABRAHAMSON, S., LASS S., GIRARD, R. & GARRIS, R. (1969) Analysis of an oral examination used in specialty board certification, *Journal of Medical Education*, 44, pp. 951–954.
- GIBBS, H., HABESHAW, S. & HABESHAW, T. (1988) *Interesting Ways to Teach: 53 Interesting Ways to Assess your Students* (Bristol, Technical and Educational Services).
- HARDEN, R.M., CROSBY, J.R. & DAVIS, M.H. (1999) AMEE Medical Education Guide No. 14: Outcome-based education, Part 1: an introduction to outcome-based education, *Medical Teacher*, 21, pp. 7–14.
- HOLLOWAY, P.J., COLLINS, C.K. & START, K.B. (1968) Reliability of viva voce examinations, *British Dental Journal*, 125, pp. 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, pp. 227–232.
- JAYAWICKRAMARAJAH, P.T. (1985) Oral examinations in medical education, *Medical Education*, 19, pp. 290–293.
- JOLLY, B. & GRANT, J. (1997) *The Good Assessment Guide—A Practical Guide to Assessment and Appraisal for Higher Specialist Training* (London, Joint Centre for Education in Medicine).
- KEARNEY, R.A., PUCHALSKI, S.A., YANG, H.Y.H. & SKAKUN, E.N. (2002) The inter-rater and intra-rater reliability of a new Canadian oral examination format in anesthesia is fair to good, *Canadian Journal of Anaesthesia*, 49, pp. 232–236.
- KELLY, P.R., MATTHEWS, J.H. & SCHUMACHER, C.F. (1971) Analysis of the oral examinations of the American Board of Anesthesiology, *Journal of Medical Education*, 46, pp. 982–988.
- MARSHAL, V.R. & LUDBROOK, J. (1972) The relative importance of patient and examiner variability in a test of clinical skills, *British Journal of Medical Education*, 6, pp. 212–217.
- McFARLANE, A.C., GOLDNEY, R.D. & KALUCY, R.S. (1989) A factor analytic study of clinical competence in undergraduate psychiatry, *Medical Education*, 23, pp. 422–428.
- MCGUIRE, C.H. (1966) The oral examination as a measure of professional competence, *Journal of Medical Education*, 41, pp. 267–274.
- MILLER, G.E. (1990) The assessment of clinical skills/competence/performance, *Academic Medicine*, 65, pp. S63–S67.
- NORMAN, G. (2000), Examining the examination: Canadian versus US radiology certification exam, *Canadian Association of Radiologist Journal*, 51, pp. 208–209.
- POKORNY, A.D. & FRAZIER, S.H. (1966) An evaluation of oral examinations, *Journal of Medical Education*, 41, pp. 28–40.
- 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, pp. 370–375.
- ROWLAND-MORIN, P.A., BURCHARD, K.W., GARB, J.L. & COE, N.P. (1991) Influence of effective communication by surgery students on their oral examination scores, *Academic Medicine*, 66, pp. 169–171.
- SANDARS, J. (1998) *MRCGP: approaching the New Modular Examination Approach to the Oral Examination Component* (Cheshire, Pasttest).
- SCHIFF, R. (2001) A short case prolonged, *British Medical Journal*, 323, p. 551.
- SCHUWIRTH, L.W.T. & VAN DER VLEUTEN, C.P.M. (1996) Quality control: assessment and examinations. Available at: http://www.oeghd.or.at/zeitschrift/1996h1-2/06_art.html (accessed).
- SCHWARTZ, P.L. & SEIN, K.T. (1987) A different twist to the 'staff:student ratio': administering medical oral examinations to students in groups, *Medical Education*, 21, pp. 265–268.
- SOUTHGATE, L. & GRANT, J. (2003) Principles and standards for an assessment system for postgraduate medical training, the Postgraduate Medical Education and Training Board. Available at: <http://www.pmetb.org.uk> (accessed 27 May 2005).
- STILLMAN, R.M., LANE, K.M., BEETH, S. & JAFFE, B.M. (1983) Evaluation of the student: improving validity of the oral examination, *Surgery*, pp. 439–442.
- SWANSON, D.B. (1987) A measurement framework for performance based tests, in: I.R. HART & R.M. HARDEN (Eds), *Further Developments In Assessing Clinical Competence* (Montreal, Can-Heal).
- SWING, S. & BASHOOK, F.G. (2000) *Toolbox of Assessment Methods* (Evanston, IL, Accreditation Council for Graduate Medical Education & American Board of Medical Specialties).
- THOMAS, C.S., MELLISOP, G., CALLENDER, K., CRAWSHAW, J., ELLIS, P.M., HALL, A., MACDONALD, J., SILFVERSKIOLD, P. & ROMANS-CLARKSON, S. (1992) The oral examination: a study of academic and non-academic factors, *Medical Education*, 27, pp. 433–439.
- WAKEFORD, R., SOUTHGATE, L. & WASS, V. (1995) Improving oral examinations: selecting, training and monitoring examiners for the MRCGP, *British Medical Journal*, 311, pp. 931–935.
- WASS, V., WAKEFORD, R., NEIGHBOUR, R. & VAN DER VLEUTEN, C. (2003) Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component, *Medical Education*, 37, pp. 126–131.
- ZELENOK, G.B., CALHOUN, J.G., HOCKMAN, E.M., YOUNG, L.C., ERLANDSON, E.E., DAVIS, W.K. & TURCOTTE, J.G. (1985) Oral examinations: actual and perceived contributions to surgery clerkship performance, *Surgery*, 97, pp. 737–744.