COMMENTARY

Objectivity and fair assessment in the clinical viva: is it possible?

Despite many recent changes in the assessment format and methodology of the postgraduate specialist qualifying examinations, the clinical viva and examination of patients in a clinical context remains an essential part of the overall assessment under the system in surgery in the United Kingdom as well as in Hong Kong. The rationale for this emphasis is related to the real need to gauge candidates in terms of their ability and skill to interact and communicate with patients in addition to the mere testing of knowledge.

Traditionally the results of assessments in the clinical vivas in the exit examination in plastic surgery in Hong Kong have been awarded in the form of a ranking scale of 4 to 8 points. Each of the numbers from 4 to 8 was assigned a label which were, respectively, poor, fail, pass, good, and excellent. Examiners would each give a candidate a mark, supposedly according to the candidate's performance. Thus, if an examiner's impression was that a candidate should pass, he would accord a score of 6. If he thought someone was good, a 7 usually. Rarely, if ever, was a candidate awarded a score of 8, because that was supposed to be given to the exceptional candidate, if not a genius, and it was better to award 7 in order to avoid having to explain too much at the examiners' meeting. Failing someone means giving a score of 5 and it was common to see a score of 5.5 as a strategic mark, particularly if you happened to be the only 'killer'. It would then be easier for you to refrain from being perceived the bad guy by making some qualifying remarks and upgrading the score to a 6. Giving reasons for the award of a particular score was encouraged but not mandatory.

Such were the vagaries of the past, owing to a primitive scoring system where the scale was directly linked to the outcome of the assessment, that is, pass or failure. Examiners were required to decide upon whether to 'stab' the knife or give a lift. This was an onerous burden, because of the potential implications to one's reputation. Most examiners do not resent being famous, but not for being nasty and harsh.

The author is the programme director and chief examiner in plastic surgery and the specialty is still in the process of working out if a conjoint examination could be held in Hong Kong together with the Edinburgh College. In order to be updated with the latest situation in the United Kingdom, on behalf of the College of Surgeons of Hong Kong, the chairman of the plastic surgery board and the author made a visit to observe the Intercollegiate Specialty Examination in plastic surgery held in Glasgow in March 2011.

We observed four important changes that were new to the Hong Kong system that appeared relevant:

(1) The use of pre-set viva questions using framed pictures and scenarios, which had been validated by consensus by a panel of examiners as to their fairness and appropriateness.

(2) Different stations were designated to examine specific areas of the syllabus. This was to ensure every candidate was scrutinised on the full spectrum of what they needed to know.

(3) The expected behaviour of examiners was set out to achieve an examination free from variability due to inappropriate conduct as was known to ensue in the past.

(4) The use of descriptors in the place of outcome parameters for the awarding of scores. Each score was defined and qualified by a set of descriptors which looked at the different aspects of a candidate's performance, including knowledge of the subject matter, judgement and decision-making, the quality of presentation, and handling of the patient. At the examination, these descriptors were tabulated clearly on a sheet of A4 paper for easy reference. Examiners looked for the set of descriptors which conformed most closely to a candidate's performance to allocate the corresponding score.

The above measures have all been discussed in the plastic surgery board, passed and then endorsed by the College of Surgeons of Hong Kong, and have been implemented for all our mock and exit examinations since October 2011. The experience from these examinations is summarised below.

We found that pre-set and validated questions helped to maintain a uniform and appropriate standard in the examination. Questions were structured and designed to warm up the candidate, to test for his/her standard as well as for outstanding abilities. There was no more room for unreasonably demanding questions or rare conditions that were never seen except in the examination hall. The familiar problem of examiners 'going astray' or 'wandering into extreme corners' can thus be prevented.

Provisions for ensuring that the full syllabus was covered obviously reduced the likelihood of candidates with major gaps to pass by luck and thus represented a better guarantee of fairness.

With regulation of examiners' behaviour, examiner idiosyncrasy and bias were largely eliminated. It was also anticipated that in due course, a system for the auditing examiners' techniques and behaviour will be instituted.

The most remarkable change, however, was for the awarding of scores. Scores were awarded objectively, since they were determined by checking performance against the descriptors for each mark, and not according to an examiner's subjective impression of whether a person should have a pass or fail. Strictly, the descriptors were so clear and discrete that it was extremely unlikely that examiners would award a wrong score by picking the wrong set of descriptors. In this sense, any examiner in a given situation would be expected to give the same score and thus the marking was supposedly assessor-independent. The new system also facilitated the award of scores away from the mean, which was a well-known failure in the past. Thus, it certainly helped to distinguish the better candidates. The award of half-hearted scores (in terms of 0.5) was no longer allowed and there was no up- or down-grading of scores. Scores were therefore objective, accurate, and more reproducible. Examiners were strictly required to jot down aspects of the candidates' performance, so that in the event of an appeal being filed, details were readily available to explain and substantiate why a particular score was given.

It is emphasised that these changes resulted from the implementation of what was in effect a 'package' of elements catering for structure, process, and human aspects. No assumption should therefore be made that similar improvements could be attributed to any of the four elements in isolation.

This model described here was taken from the specialty of plastic surgery, although potentially it has much wider application to all other specialties, in surgery, and in other medical disciplines. Our sister Colleges and the two local medical schools have not been slow to respond to this worldwide trend in objectivity, where similar efforts are in existence and are in various degrees of development. Our Hong Kong Colleges have all been seeking to achieve international recognition for our standards. We have very good trainees and have been working on our training schemes. An assessment system adopting the above safeguards and based on objective descriptors free from partiality, personal concerns, and inter-assessor variation seems to be the logical way to evolve.

David SY Wong, FHKAM (Surgery)
Email: sywong@surgery.cuhk.edu.hk
Department of Surgery
The Chinese University of Hong Kong
Prince of Wales Hospital
Shatin
Hong Kong

Answers to CME Programme
Hong Kong Medical Journal June 2012 issue

Hong Kong Med J 2012;18:228–37

I. Alzheimer's disease: early diagnosis and treatment?

A	1. False	2. False	3. False	4. True	5. False
B	1. True	2. False	3. False	4. False	5. False
C	1. True	2. False	3. True	4. True	5. True

Hong Kong Med J 2012;18:238–46

II. The reference framework for diabetes care in primary care settings

A	1. True	2. False	3. False	4. True	5. True
B	1. True	2. True	3. False	4. False	5. True
C	1. True	2. False	3. True	4. False	5. True