Frequently Asked Questions about the role of video in the MRCGP Clinical Skills Assessment

The Clinical Skills Assessment (CSA) is a performance based examination. It is one part of a tripos that contributes to the licensing of UK General Practitioners. It therefore has to have high performance characteristics for these significant decisions.

The examination was developed carefully and with a robust evidence base. It has continued to develop with constant evaluation and analysis, as is best practice in assessment and expected by our regulators.

Q: What is the structure of the examination?
A: The structure of the examination is that candidates are tested on 13 different cases representing different contexts of general practice. Judgements about the performance of the candidates are made by 13 different examiners. The examiners have been rigorously selected and trained for the purpose. In order to be feasible, to cope with demand and to be economic with costs, the examination is run simultaneously on three different circuits. The examining day begins with three examiners and three role players carefully calibrating their case for the day. This calibration takes over an hour. This is to ensure consistency of performance and marking on the three different circuits during the day. It allows the examiners to agree what are key components of the case and what would be required to achieve marks. The examiners are also contributing during the day to the borderline group method of standard setting for that day.

Q: Are videos used at all?
A: Yes: Part of the evaluation and analysis of CSA relies on the use of video for training and quality assurance of our examination. In the previous assessment centre in Croydon, RCGP had only had video capability in three of the 13 rooms on each circuit. With the move to Euston Square we now have video capability in all the rooms. This has given us enhanced opportunity for training materials and also rapid evaluation of any problems that might arise during the course of the examination.

Q: Why are videos not used for the examination?
A: The examination has never been designed to use video at any time to contribute to the process of marking of candidates which depends on the professional judgements of a skilled and calibrated examiner whose performance is under continuous scrutiny.

Q: How do you standardise examiners?
A: No single examiner is 100% perfect, 100% of the time. That is why we have 13 different judgements. We achieve an SEM of around 5% with Cronbach alphas of around 0.8. These are good reliability figures for a performance assessment of this length. In addition candidates are allowed four attempts at the examination. A generalizability analysis has been carried out to look for sources of variance (error) and the contributions of examiners to that variance is very small. Similarly we carry out regular comparative analyses of examiner performance. Examiner marks vary with the difficulty of the case as we would expect and there are no enduring characteristics of hawks (hard examiners) or doves (easy examiners).
Q: If a candidate has been a borderline fail why can we not remark at a later date with a different set of examiners?
A: We believe that the CSA exam is sufficiently robust to make decisions about candidates without the need to remark. We cannot use just one examiner to remark all 13 cases because of the problems of bias. Reliability is always a function of the number of separate examiners and the total testing time (Swanson, Norman et al. 1995). To achieve the same level of reliability as the CSA we would need 13 separate examiners. In effect that is a resit. It would also be very expensive. There will have been no calibration and the examiner would not be marking the case 26 times. Under the current situation examiners are blind to the candidate attempt number. As they enter the room the examiners do not know if the candidate is taking the exam for the first or the fourth attempt. They are also blind to the standard of the candidate. With the video resit model the examiners will know that they are marking a failing candidate with all the consequent bias that brings. In short we would be using a less reliable method.


January 2013