

Objective structured clinical examination compared with other forms of student assessment

L J SMITH, D A PRICE, AND I B HOUSTON

Department of Child Health, University of Manchester

SUMMARY Two hundred and twenty nine final year medical students were assessed in paediatrics using an objective structured clinical examination (OSCE) and a traditional viva voce examination, and the results were compared with other assessments of the students made during and at the end of the undergraduate course. Results of the OSCE correlated positively with other forms of assessment and more strongly than the viva voce examination. There was little correlation between the OSCE and viva results. Eighty per cent of students felt the OSCE to be a fairer system than other examinations and all external examiners commented favourably on it. An OSCE is an acceptable alternative to traditional means of examination in paediatrics and may be superior in certain aspects.

Traditional systems of examination of medical students may be subject to certain important criticisms. There may be considerable variation in the expectations of different examiners who may also vary in their skill in assessing the students' abilities. In addition, students may feel an examination to be unfair because of their perception of these variations. The objective structured clinical examination (OSCE) was described in 1975 by Harden *et al*^{1 2} in Dundee University and was modified for a pilot study in the assessment of paediatric students in Manchester.³ Subsequently, in 1982 and 1983, an OSCE has formed part of the paediatric examination in the University of Manchester. Each student passes through the same series of preordained 'stations', where clinical skills are assessed by the same examiners according to a predefined marking system. Marks are allocated to reflect the candidates' competence in taking histories, performing preset clinical tasks, evaluating physical signs, and interpreting various types of clinical data.

During 1983 final year medical students in Manchester were assessed in paediatrics using an OSCE and viva voce examination held on the same day. Those performing poorly in the OSCE and viva were referred to the final examination and those doing very well were encouraged to sit an honours examination. To evaluate the relative contributions of the OSCE and viva to the assessment of students' abilities we have compared the results of the OSCE and the viva with an in-course assessment in paediatrics, a general multiple choice question

(MCQ) examination (which included some paediatric questions), comparable but traditional assessments in other subjects, and the overall finals results. To assess acceptability of a novel form of examination, examiners and examinees were asked for their views on the OSCE.

Methods

OSCE. The assessment by OSCE of 229 students was completed in two mornings (I and II) using a different examination for each day; (the OSCE took place in three locations on each day). Each student attended one location and visited 21 stations consisting of two history stations (one of eight minutes, and one of four minutes), three clinical examination stations (one eight minutes and two of four minutes), five data stations (four minutes each) with questions dependent on previous history or examination stations, seven independent data stations (four minutes each), and four rest stations (four minutes each). Histories were of common complaints such as asthma or febrile convulsions, and students were asked to perform clinical tasks such as the examination of the newborn, a developmental assessment, or examination of a 'system', for example the respiratory system. The dependent data stations tested their ability to interpret their clinical findings and the independent data stations tested their ability to examine urine, interpret growth charts, radiographs, prescriptions, etc. Thirty marks were allocated to each eight minute and one four

minute station, 20 marks to each of the other four minute clinical stations, and 10 marks to each data station making a total of 250 marks.

One examiner was allocated to each clinical station, having received prior instruction and been provided with a carefully designed check list and marking schedule for each student attending the station. Written answers at the data stations were discussed and marks allocated in advance to improve uniformity of marking.

The OSCE mark was subdivided into a mark for all clinical stations and dependent data questions (OSCE clinical) and all data questions independent of clinical stations (OSCE data).

Viva voce. Students were given two case histories to read in four minutes and an examiner questioned them on these histories and related topics for eight minutes. A mark out of 50 was allocated. Altogether eight different examiners were involved in the viva.

In-case assessment in paediatrics. A grade had been allocated earlier to each student during the residential paediatric course based on clinical aptitude and a written project. This grade was assigned a numerical value on the scale 0 to 10.

Comparable examinations in other subjects. In other subjects, viz surgery, medicine, obstetrics and gynaecology, and psychiatry, traditional clinical examinations were held which had the same purpose as the OSCE, that is the results determined whether the student returned to a 'final' examination in the subject, was exempted, or was encouraged to sit an honours examination. The grades (A to E) awarded in each subject were assigned a numerical value (4 to 0).

MCQ examination. Similarly the marks for the students from the MCQ, which covered all clinical subjects, were assigned another numerical value on a scale of 0 to 6.

Overall performance in examinations (excluding paediatrics). This was assigned a numerical value on a scale 0 to 42 by combining marks from the MCQ and the comparable clinical examinations, with additional marks for honours and distinctions in subjects other than paediatrics.

Student and examiner opinion. Students' views were assessed by use of an anonymous questionnaire. External examiners were asked individually for their views immediately after the examination. These observations were recorded, and supplemented by written comments later.

Statistical methods. The results were analysed using the Statistical Package for the Social Sciences of the North Western University in the Regional Computer Centre of the University of Manchester. A test for normality, the Kolgorov-Smirnov goodness of fit test, was carried out on each item. When data was normally distributed, Pearson correlation coefficients and linear regression analyses were performed. When data was non-parametric, Kendall correlation coefficients were used.

Results

Complete OSCE marks. The marks for each OSCE were normally distributed. The mean (SD) for OSCE I was 151 (17) and for OSCE II was 167 (16).

There were highly significant ($P < 0.001$) correlations between each OSCE and overall examination performance; $r = 0.46$ and $r = 0.51$, for OSCE I and II respectively (Table 1). There were weaker but still highly significant correlations ($P < 0.001$) between each OSCE and marks from the MCQ, marks from comparable clinical examinations, and marks from previous in-course assessments. The mark from OSCE I, however, did not correlate with the viva mark and there was only very weak correlation between OSCE II and viva marks.

Subdivided OSCE marks. Table 2 shows that when the OSCE mark was subdivided into OSCE clinical and OSCE data marks, there was no correlation between OSCE clinical marks and the viva, and OSCE clinical marks had weaker correlations ($r = 0.11$ and $r = 0.14$) with the MCQ than OSCE data marks with the MCQ ($r = 0.25$ and $r = 0.23$).

Viva voce results. Marks from the viva examination correlated poorly with marks from the OSCE, either as a total or as clinical or data subdivisions (Tables 1

Table 1 Correlation of objective structured clinical examination (OSCE) I and II with other forms of assessment

	OSCE I		OSCE II	
	r or tau value	P value	r or tau value	P value
Overall performance	0.46 (r)	<0.001	0.52 (r)	<0.001
Multiple choice questions	0.34 (r)	<0.001	0.22 (t)	<0.001
Comparable clinical examinations	0.28 (t)	<0.001	0.36 (t)	<0.001
In course assessment	0.24 (t)	<0.001	0.32 (t)	<0.001
Viva voce	0.08 (t)	NS	0.11 (t)	<0.05

r = Pearson correlation coefficient.
tau = Kendall correlation coefficient.

Table 2 Comparison of the clinical and data components of the objective structured clinical examination (OSCE) and the viva voce examination with other assessments

		Overall performance	Multiple choice questions	Comparable clinical examinations	In course assessment	Viva voce
OSCE clinical	I	0.28*	0.11‡	0.21*	0.20*	0.09
	II	0.32*	0.14‡	0.29*	0.25*	0.07
OSCE data	I	0.21*	0.25‡	0.27†	0.06	0.03
	II	0.32*	0.23‡	0.30*	0.26*	0.15‡
Viva voce	I	0.14‡	0.03	0.20*	-0.02	—
	II	0.13‡	0.17‡	0.13‡	0.05	—

Kendall coefficients of correlation (tau values) are given.

* $P < 0.001$, † $P < 0.01$, ‡ $P < 0.05$.

and 2). Marks from the viva showed poorer correlations with all other forms of examination assessment in comparison with the OSCE.

Student opinion of the OSCE. Four fifths of the students perceived the OSCE to be fairer than other forms of clinical examination; two thirds found it less stressful; but opinion was equally divided about whether they had been examined on aspects stressed as important in the paediatric course.

Examiner opinion. All external examiners were enthusiastic about the OSCE format. They were in favour of a higher proportion of clinical stations.

Discussion

The logistics of examining a large number of students competently, fairly, and with a minimum of stress are difficult. In the Department of Child Health in Manchester, the OSCE has proved a useful adjunct to other forms of assessment. Ideally a single OSCE would be used for all students, but this would have overstretched the resources of the department in terms of examiners, patients, and accommodation. The normal distribution of marks in each OSCE facilitates comparison between examinations of different difficulty on different days and a grading of students can be made in relation to the mean and standard deviation for that particular examination.

The OSCE results correlate well with an overall assessment of the student's ability, contrasting with the poor correlation of viva voce assessment with almost all other forms of assessment. There is no correlation between the clinical part of the OSCE, and the viva and this may indicate that the latter is a less appropriate and potentially misleading technique for the assessment of clinical competence. Not only is it likely that the viva tests quite different student attributes but variation between examiners was very obvious and difficult to eradicate. Better

correlation might possibly have been achieved by using a single examiner but this is not practicable and was not possible within the constraints of the OSCE format. As might be expected, the marks from the data questions of the OSCE seem to reflect a skill seen in the MCQ examination and the correlation between the clinical component of the OSCE and the MCQ is poorer.

Based on student responses to a questionnaire and the comments of examiners, it is clear that students and examiners responded positively to the OSCE format and it was perceived by both groups to be more fair and less stressful than traditional clinical examinations. Each student is exposed to the same questions, situations, and examiners as their colleagues, and an individual examiner is not biased by the student's performance in another part of the examination. Of necessity patients with *common* disorders (in order to achieve sufficient patient numbers) are brought to the examination and therefore the student is not confronted with the traditional examination patient who has 'good signs' but a rare problem. The OSCE has to be planned carefully beforehand and it is simple to incorporate suggestions for improvement. Furthermore, the student expects to be examined in his clinical skills and therefore wishes to learn these skills from his tutors.

Disadvantages include the fairly lengthy preparation of OSCE questions, but a bank of good and discriminating questions can be established. Certain examiners may find the repetitive nature of their task more tedious than the varied nature of traditional examinations; the OSCE places more demand on observational skills of examiners. More importantly synthesis of information and problem solving by the student may not be so well tested and it is vital that data stations are designed with this in mind.

The OSCE has been shown to compare well with other forms of assessment and is superior to a viva examination in assessing clinical skills. The con-

siderable advantages outweigh the disadvantages and this system can be recommended as a useful form of assessment of the paediatric student.

We thank the tutors in child health and the external and internal examiners for their help, Mrs K Cordwell for preparation of the manuscript and Linda Hunt for statistical advice.

References

¹ Harden RM, Stevenson M, Wilson Downie W, Wilson GM. Assessment of clinical competence using objective structural examinations. *Br Med J* 1975;*i*:447-51.

² Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979;*13*:41-54.

³ Watson AR, Houston IB, Closs GC. Evaluation of an objective structured clinical examination. *Arch Dis Child* 1982;*57*:390-2.

Correspondence to Dr D A Price, Department of Child Health, Royal Manchester Children's Hospital, Pendlebury, Manchester M27 1HA, from whom free documentation of the design of this OSCE and samples of questions can be obtained on receipt of a stamped, addressed envelope.

Received 20 September 1984