

LEARNING/TEACHING

Developing an OSCE Station to Assess the Ability of Medical Students to Share Information and Decisions with Patients: Issues Relating to Inter-rater Reliability and the Use of Simulated Patients

JILL E. THISTLETHWAITE, MMed BSc MB BS MRCPG DRCOG
ILTM

University of Leeds, UK

ABSTRACT **Context:** *Patient partnership is being promoted as an aid to compliance with treatment and to improve outcomes for patients. An integral part of this partnership is information-sharing between doctor and patient and negotiation of management decisions, together with an ability by the doctor to analyse critically treatment options. These skills are being taught to some extent in communication skills for medical students and assessments are being devised to test them.*

Objectives: *To develop an OSCE station to assess medical students' skills in the application of evidence and the negotiation of treatment options with a simulated patient. To evaluate the station by tests of reliability and comparison of the marking of observing examiners and simulated patients.*

Design: *An OSCE station was devised using a case scenario in which there were varying opinions as to the treatment options. The marking schedule was designed to assess the students' ability to assess evidence and discuss this with patients.*

Setting: *University of Leeds.*

Subjects: *Undergraduate medical students.*

Results: *A total of 194 students undertook the station. Inter-rater reliability amongst examiners was poor for individual student marks but there was better agreement as to whether a student passed or failed. There was poor agreement between the observing examiners and the simulated patients for individual marks. The examiners thought this was a useful attempt to assess students' ability to give information but agreed that further work was needed on the marking schedule.*

Address for correspondence: Dr Jill E. Thistlethwaite, Senior Lecturer in Community-based Teaching, Academic Unit of Primary Care and the Medical Education Unit, University of Leeds, 20 Hyde Terrace, Leeds LS2 9LN, UK. Tel: +44 (0) 113 233 4195. Fax: +44 (0) 113 233 4181. E-mail: j.e.thistlethwaite@leeds.ac.uk

Conclusions: *More work is needed to refine the station to increase its reliability. However, the station is useful in highlighting to students the importance of sharing information and decision-making on treatment options with patients.*

KEYWORDS *Shared decision-making, patient partnership, communication skills, assessment, simulated patients, reliability.*

Introduction

It has been claimed that paternalism is endemic in the health services (Coulter, 1999). To overcome this many medical curricula have begun to promote the concept of patient partnership, where patient and doctors work together to achieve common goals with their relationship being based on mutual respect for each other's skills and competencies. This has been shown to aid compliance with treatment and improve the outcome for the patient (Horder & Moore, 1990). Towle and Godolphin (1999) have defined a series of steps that should occur in consultations to enable doctors and patients to share in the decision-making process and to adopt the information for a shared decision-making model. These include eliciting the patient's views about treatment options, presenting technical information to the patient in an unbiased and clear way and the doctor sharing any treatment recommendation with the patient. Implicit in this process is the application of evidence, through critical analysis, to treatment options and discussion of this with the patient.

Communication skills are now an integral part of medical undergraduate curricula and one of the important skills learnt is the giving of information to patients. These skills are usually assessed by observing students interacting with patients or simulators, often by means of an objective structured clinical examination (OSCE) where students have been shown to perform poorly in information-sharing stations (Prislin *et al.*, 1998). Critical analysis is a skill that is fostered in medical education but students have been shown to do poorly in this area (Stern *et al.*, 1995). Previously in the Leeds course it has been assessed by written reports. Linking assessment of the two is an attractive proposition, as it requires application of information analysis to information-sharing. Bradley and Humphris (1999) have reported on the use of an OSCE station to assess the ability of medical students to apply evidence in practice, using a simulated patient (SP) who also assesses the students' quality of communication. This scenario used information regarding antibiotic treatment for sore throats and was relatively non-contentious with respect to treatment options.

This paper describes the development of an OSCE station to assess similar skills with added emphasis on the student's ability to share information and negotiate treatment with the SP in a community setting when there is conflicting evidence as to what treatment option is preferable. The evaluation of the station

compared the marks of the SPs to those of observing examiners and the inter-reliability of examiners themselves. Students were also asked their opinion of the station, as was a trained SP who in this situation acted as an examiner.

Methods

At the end of their third year the medical students at Leeds undergo a 20 station OSCE, lasting two hours, which they must pass to be able to proceed to their next year of training. The OSCE takes place on two sites. The OSCE has a mix of history, physical examination and data interpretation stations, and includes one communication skills station. The stations are each of six minutes' duration, and students move from one station to the next without a pause. The communication skills station has varied over the years from "giving information" to describing an endoscopy to taking a history from a patient with both physical and psychosocial problems.

The communication skills station was redesigned in 2000 and involved the management of a patient with non-ulcer dyspepsia in primary care. Overall the station was double the length of the other stations. For the first six minutes the students read an endoscopy report relating to Mr/Mrs Peters who had previously consulted a general practitioner because of severe indigestion associated with some weight loss. The endoscopy report read:

There is inflammation in the lower end of the oesophagus and evidence of reflux. There is gastritis in the antrum of the stomach; there are no abnormalities detected in the rest of the stomach and duodenum. Biopsies show moderate numbers of *Helicobacter pylori*.

The students were advised, in the station written instructions, that there is uncertainty as to the best treatment for this condition and were given two abstracts to read, one of which advocated the use of triple therapy (McCull *et al.*, 1998) and one that did not (Talley *et al.*, 1999). The students were told that they were acting as pre-registration house officers (PRHOs) in general practice and that if they wished to issue a prescription, they should inform the patient of this as they would need to get one of the GPs to sign it. They were advised that marks would be awarded for the explanation of the options to the patient, not whether they decided to prescribe or not.

After six minutes the students entered another room to consult with their patient played by an experienced simulated patient, who had prepared for the OSCE by studying detailed background notes about the role. The consultation lasted six minutes and was observed and marked by an examiner using the marking schedule shown in Table 1. The marking schedule was designed specifically for this station and was used in preference to a generic consultation skills measuring instrument, none of which are designed specifically to measure patient involvement, a key feature of this station (Elwyn *et al.*, 2001). The pass

Table 1. Assessment table

Details of task	Fully satisfied	Partially satisfied	Not satisfied
Student introduces self	2	1	0
Student clarifies reason for attendance	2	1	0
Explanation of endoscopy result	2	1	0
Checks that patient has understood	2	1	0
Outlines management plan	2	1	0
Outlines reason for plan	2	1	0
Checks that patient has understood	2	1	0
Discussed smoking & drinking	2	1	0
Asks patient if there are any other questions	2	1	0
Explains what will happen next	2	1	0
Total score: 20			

mark for this station was set at 10/20. For half the cohort of students, the SPs were also asked to mark their performance using the same marking schedule as the examiners. It had been intended that all the students would be assessed by the SPs but this was impossible at one site due to lack of time between students because of the configuration of the stations.

Of the examiners, two were medically qualified and the others were from various health and social care backgrounds (nursing, social work, communications skills lecturer) who teach consultation skills to undergraduates. The SPs were all experienced in role development, working with medical students and giving feedback and attend training sessions at least twice a year. They have worked with a variety of different check lists and this new one was explained before the examination. Following the OSCE examiners and SPs were asked for their comments on the station and after the final results had been published the students were asked for their comments by e-mail. The students had a week's holiday directly after their examination and did not return to university until their marks were posted; thus it was difficult to receive their opinions about the station until after they received their marks. This may have affected their comments.

Students were asked whether they thought the station was a reasonable way of assessing their communication skills; whether it was easy or hard; whether they felt they needed more time; and if they had any suggestions of other ways of assessing communication skills. SPs and examiners were asked for comments on the running of the station, the feasibility of the role and the marking schedule.

Statistical analysis was undertaken using a variety of different statistical tools and SPSS depending on the parameters under consideration.

Results

The OSCE took place at two different locations and at eight sittings. Overall 194 students undertook the examination. Seven simulated patients were used

(one took part in two sittings) and seven examiners (two took part in two sittings). For one sitting (25 students) two examiners marked to check for inter-rater reliability and for four sittings (98 students) the SP also marked the students using the same assessment sheet as the examiners (Table 2).

Statistical Techniques

Agreement was looked at in two ways. First, the marks given were compared by using Bland-Altman plots. These look at the mean mark given to a student by a pair of examiners compared to the difference in the marks. The “95% limits of agreement” are also calculated. These are an estimation of the likely difference in marks from the two examiners. Second, a pass mark of 10 was assumed and the examiners were compared as to whether or not they agreed a particular student had passed. The *kappa* (κ) statistic was calculated. κ varies between 0 (no agreement) and 1 (perfect agreement). κ of 0.6 or above is generally regarded as acceptable.

Examiner Agreement

E1 and E2 both examined the same 25 students. There was acceptable agreement at the pass-fail level (Table 3) with $\kappa=0.69$. In summary, there was poor agreement between the two examiners as to the mark an individual student should get, but somewhat better agreement as to whether or not a student passed.

Examiner/SP Agreement

The data presented here concern the five possible pairs of examiner/SP. In summary, there was poor agreement in the marks given by the examiners and the SPs and inadequate agreement over whether or not an individual student should pass.

Looking at agreement at the pass-fail level there were kappas ranging from 0.34 to 0.57 indicating inadequate to moderate agreement. (Table 4). (Kappa could not be calculated for E3/SP3 because they passed all the students, i.e. perfect agreement.)

Average Marks Given by Examiners

The mean mark given by each examiner is shown in Table 5. These figures were analysed using ANOVA. In summary most of the examiners gave broadly similar marks *on average* and were equally likely to pass a student *on average* with the exception of E3 who was a more generous marker and gave significantly higher marks ($p=0.001$).

Average Marks Given by SPs

The mean mark given by each SP is shown in Table 6. These figures were also analysed using ANOVA, which showed significant difference between SPs. A logistic regression was also carried out to see if the odds of SPs giving a pass

Table 2. Examiners and simulated patients

Student group	Examiner E	Simulated patient	Number of students
1	1,2	1	25
2	3	2	25
3	3	3	24
4	4	4	24
5	5	did not examine	24
6	6	did not examine	25
7	7	did not examine	24
8	7	did not examine	23

Table 3. Figure for kappa calculation, examiner 1 (E1) and examiner 2 (E2)

		E2		Total
		Pass	Fail	
E1	Pass	17	0	17
	Fail	3	5	8
	Total	20	5	25

Table 4. Summary pass–fail results comparing simulated patients and examiners

		Patients		Total
		Pass	Fail	
Examiners	Pass	90	12	102
	Fail	7	14	21
	Total	97	26	123

Table 5. Summary of marks given by each examiner

Examiner	Student number	Mean	Standard deviation
1	25	11.32	2.75
2	25	11.36	2.32
3	49	14.04	2.47
4	24	12.63	2.81
5	24	10.96	3.14
6	25	11.72	2.65
7	47	10.98	2.52
Total	219	12.01	2.86

mark were different. SP3 could not be included in this analysis as this SP gave all the students a pass mark. In summary, the patients gave very different

Table 6. Summary of marks given by each simulated patient (SP)

SP	Student number	Mean	Standard deviation
1	25	10.36	1.63
2	25	11.96	4.46
3	24	13.96	2.18
4	24	15.08	2.69
Total	98	12.81	3.43

average marks, but there were no significant differences between SP1, SP2 and SP4 in the likelihood of passing students. SP3 who passed all the students did not give the highest marks on average.

Comments from Examiners and SPs

Overall the examiners thought this was a useful attempt to assess the students' ability to give information in a situation without a clear-cut management path. However, they commented that the station was too complex and that the marking schedule did not adequately assess the students' abilities to share in the management decision. One examiner suggested that more communication skill tasks should be assessed, for example, "explains terms or uses simple language" and that marks should be given for the way in which students sought the patient's views on treatment. The SPs commented on the wild inconsistencies in the information concerning treatment given by different students, e.g. some wanted to treat, others said not necessary.

Comments from Students

All the students were e-mailed to ask them for their opinions about the station. Following their OSCE the students begin a five-week special study module that means that many are away from the university for some time; this may account for the poor response to the mailing, only 14 students replying, so these comments are not representative of the student body. However, they do give some indication of student opinion.

Overall the students thought that the station was too hard, that there was insufficient explanation of what was required of them and that it did not test their communication skills. Many gave suggestions as to better ways of assessing these skills.

Discussion

Medical students will rarely have the opportunity to discuss management issues with patients in part due to fears of litigation and in part due to lack of opportunity and supervision. Therefore, SPs have immense potential in helping students develop management skills including using the shared information

model. In order to highlight the importance of this model, students need to be assessed practising it, an assessment that could form part of an OSCE.

As the shared information model relies on the interaction between doctor and patient (Charles *et al.*, 1997), it would seem logical that patients would be involved in the assessment of the doctor or student. However, the results of this work have shown that SPs and examiners mark students differently using the same assessment scale. Given that the aim of such a station should be to standardize the examiner and the patient so that the student's performance can be seen as a measure of his or her clinical competence (Collins *et al.*, 1998), this discrepancy in results needs to be explained. It could be due to a number of factors: lack of training of examiners and/or SPs, a difference in outlook between the examiner observing from the outside and the SP observing from the viewpoint of the person intimately involved in the consultation, the examiner being able to mark during the consultation whereas the SP marks at the end. The last suggestion is possibly least likely as the accuracy of SPs in recording check list items has been found to be good and consistent over a one-day examination (Vu *et al.*, 1992).

For this station the assessment sheet was devised without prior discussion with the SPs. The tasks examined are fairly generic and marks are not awarded for knowledge. However, an SP may mark highly even if he or she was only given one option for treatment if that option was explained well. A medical examiner would know if the information being given was correct and whether both options were explained, and may mark down some candidates for incorrect information. This raises the question as to how much information the SP should be given prior to the examination. Should the SP have more knowledge than could reasonably be expected from a lay person? Or should the marking be purely from the perspective of a lay person and what went on in the consultation? From the comments of the medical examiners it was clear that the students did misinterpret the data and thus misinformed the patient, a situation that a non-medical examiner/SP may find difficult to detect. These considerations could be explored more fully when the station is run again and show that devising a valid and reliable station that tests the information for shared decision-making model is difficult. An advantage of having the SP mark solely is that the station is then less labour intensive. In order for the marking schedule to reflect both the medical and the patient's perspective, the ideal situation would be for both parties to have input into devising the marking schedule. For instance, in the development of simulated surgeries for the assessment of GP registrars the medical checklist is set by general practitioners but marked by the SPs, who also complete a patient satisfaction questionnaire which assesses the registrar's interpersonal skills and professional attributes (Allen *et al.*, 1998).

While there were differences in marks given by the examiners broadly speaking they agreed on which students should pass or fail. However, looking at the average marks awarded by examiners and SPs in this way does indicate that

some give consistently higher marks than others do, highlighting the need for training and feedback. Such training is time consuming and adds to the overall cost of running more complicated OSCE stations.

The medical students were critical of the station. It is true that this was the first time they had been asked to integrate critical analysis, sharing information with patients and deciding on management. However, each of these components had been taught previously in different parts of the course. In particular, in the second year communication skills course information giving was a key objective and students were encouraged to check understanding and to elicit patient's concerns when giving information. They were able to practise this approach with simulated patients, receiving feedback on their performance from both facilitators and SPs.

Obviously having only one OSCE station measuring the ability of students to share information and decisions with patients is not a reliable method of assessing their skills due to problems with case specificity; many more stations would be needed to do this (van der Vleuten & Swanson, 1990).

However, by having an OSCE station devoted to management in this way, the importance of sharing information and involving the patient in management decisions is highlighted. By using a clinical problem where there is no one "right" option students are required to present options and discuss their merits with the patient, allowing the patient to participate in the decision process.

Further work is needed to enhance the validity and reliability of such a station, improving the marking schedule in order to assess adequately the students' ability to involve the patient in the decision-making process. The SPs need to become more involved in devising the marking schedule and their role in assessing students needs to be clarified.

In Leeds we are developing simulated surgeries for medical students which assess both their clinical and consultation skills, including sharing information and decision-making.

Acknowledgements

The author would like to thank the examiners, SPs and students who commented on the station, Brett Scaife for help with the statistical analysis and Professor Trudie Roberts for her helpful comments during the preparation of the manuscript.

References

- ALLEN, J., EVANS, A., FOULKES, J. & FRENCH, A. (1998). Simulated surgery in the summative assessment of general practice training: results of a trial in the Trent and Yorkshire regions. *British Journal of General Practice*, 48, 1219–1223.

- BRADLEY, P. & HUMPHRIS, G. (1999). Assessing the ability of medical students to apply evidence in practice: the potential of the OSCE. *Medical Education*, 33, 815–817.
- CHARLES, C., GAFNI, A. & WHELAN, T. (1997). Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). *Social Science and Medicine*, 44, 681–692.
- COLLINS, J.P. & HARDEN, R.M. (1998). AMEE medical education guide no. 13: real patients, simulated patients and simulators in clinical examinations. *Medical Teacher*, 6, 508–521.
- COULTER, A. (1999). Paternalism or partnership? *British Medical Journal*, 319, 719–720.
- ELWYN, G., EDWARDS, A., MOWLE, S., WENSING, M., WILKINSON, C., KINNERSLEY, P. & GROL, R. (2001). Measuring the involvement of patients in shared decision-making: a systematic review of instruments. *Patient Education and Counselling*, 43, 5–22.
- HORDER, J. & MOORE, G.T. (1990). The consultation and health outcome. *British Journal of General Practice*, 40, 442–443.
- MCCOLL, K. *et al.* (1998). Symptomatic benefit from eradicating *Helicobacter pylori* infection in patients with nonulcer dyspepsia. *New England Journal of Medicine*, 339, 1869–1874.
- PRISLIN, M.D., FITZPATRICK, C.F., LIE, D., GIGLIO, M., RADECKI, S. & LEWIS, E. (1998). Use of an objective structured clinical examination in evaluating student performance. *Family Practice*, 30, 338–344.
- STERN, D.T., LINZER, M., O’SULLIVAN, P.S. & WELD, L. (1995). Evaluating medical students’ literature-appraisal skills. *Academic Medicine*, 70, 152–154.
- TALLEY, N.J., VAKIL, N., BALLARD II, E.D. & FENNERTY, M.B. (1999). Absence of benefit of eradicating *Helicobacter pylori* in patients with nonulcer dyspepsia. *New England Journal of Medicine*, 341, 1106–1111.
- TOWLE, A. & GODOLPHIN, W. (1999). Framework for teaching and learning informed shared decision making. *British Medical Journal*, 319, 766–771.
- VAN DER VLEUTEN, C. & SWANSON, D. (1990). Assessment of clinical skills with standardized patients: state of the art. *Teaching and Learning in Medicine*, 2, 58–76.
- VU, N.V., MARCY, M.M., COLLIVER, J.A., VERHULST, S.J., TRAVIS, T.A. & BURROWS, H.S. (1992). Standardized/simulated patients’ accuracy in recording clinical performance checklist items. *Medical Education*, 16, 99–104.