

Senior medical students as peer examiners in an OSCE

ANNETTE BURGESS, TYLER CLARK, RENATA CHAPMAN & CRAIG MELLIS

Sydney Medical School – Central, The University of Sydney, Australia

Abstract

Background: At Sydney Medical School, we have recently introduced a practice Objective Structured Clinical Examination (OSCE) where our junior medical students are assessed by the senior peers.

Aim: We sought to evaluate the efficacy of the programme.

Methods: The study took place in 2010 and 2011, with two cohorts of final-year students participating. A total of 40/98 (41%) of final-year students chose to participate as examiners. Following the completion of standardised marking sheets by the student examiners, the marking sheets were reviewed by a senior academic examiner, and the 'global' mark was adjusted in accordance with the marking criteria. Student examiners also completed an open and closed-ended questionnaire regarding their experience as an examiner. A total of 105/115 (91%) of year 2 medical students were examined in the OSCE over the two-year period.

Results: The senior academic changed a total of 94 'global' marks, reducing the majority (55%) from 'Satisfactory' to 'Borderline'; 12% were reduced from 'Satisfactory' to 'Not Satisfactory' and 33% from 'Borderline' to 'Not Satisfactory'. Student questionnaire results showed a high level of engagement with their examiner experience overall, and it was regarded as a useful learning experience.

Conclusion: Student examiners found peer assessment to be a very useful learning activity. However, our students need further training in how to globally assess a fellow student's overall performance objectively and to provide accurate feedback.

Introduction

Assessment and feedback by peers is becoming an increasingly recognised and valued method of enriching the student experience in medical schools around the world. It also has the potential to assist in preparing students for their professional lives as clinicians with peer review responsibilities (Cushing et al. 2011). It is known that assessment by peers can be both reliable and valid, as well as provide an effective learning experience for students (English et al. 2006). With growing demands on university and clinical staff within medical schools, the implementation of innovative and efficient assessment methods is a worthwhile endeavour (Jones et al. 2001). Exploring the efficacy of involving students as assessors of their peers may assist in addressing current resource challenges.

It is well recognised that peer assessment can be used to utilise the assessment of clinical competencies in fellow students (Finn et al. 2009). At the completion Stage 2 of the medical programme at Sydney Medical School, and as a barrier examination to enter Stage 3, students are required to undertake a summative Objective Structured Clinical Examination (OSCE), with clinical staff as examiners. In preparation for this exam, it is usual for Clinical Schools to run practice OSCEs in various formats, all with limited resources.

At Central Clinical School, we chose to administer a formal practice OSCE, with our final-year medical students as examiners of the year 2 students. The reason for doing so was twofold. First, we hypothesised that the experience as an

Practice points

- Peer assessment in mock-OSCEs provides a useful learning activity while limiting demands on clinical staff.
- While students feel confident in the accuracy of their ratings, they consistently rate their peers as performing better than do senior academics.
- Training is necessary to ensure accurate student examiner assessment.
- Question rotation and variability may help to keep student examiners engaged.

examiner would provide our senior students with knowledge and clinical skills revision and contribute to their training in professionalism where assessment and feedback are important attributes. Second, we simply do not have the resources to run a full-scale practice OSCE with clinical staff as examiners. We wanted to investigate whether it is of benefit to students to act as examiners of their peers in the practice OSCE, and also assess the level of agreement between student and academic examiner marking.

Methods

All final-year (year 4) medical students within Central Clinical School ($N=98$) were invited to act as examiners during a practice OSCE examination for year 2 students.

Correspondence: A. Burgess, Central Clinical School, The University of Sydney, Level 4, Building 63, Royal Prince Alfred Hospital, Missenden Road, Camperdown, NSW 2050, Australia. Tel: 61 2 9515 8172; fax: 61 2 9515 8173; email: annette.burgess@sydney.edu.au

The study took place over the course of two consecutive years, 2010 and 2011, with two cohorts of students participating. A total of 105 Stage 2 medical students undertook the practice OSCE examination, 54 (51%) from the 2010 cohort and 51 (49%) from 2011. In 2010, 54 out of 61 Stage 2 students participated. In 2011, 51 out of 54 students participated. While the practice OSCE is considered compulsory, some students were on leave for various reasons.

A total of 40 (41%) year 4 medical took part in the programme over the two-year period. In 2010, 19 out of 45 year 4 students took part (42%), and in 2011, 21 out of 53 year 4 students took part (40%). Examiners were provided with a 1 h training session detailing their assigned OSCE station, marking criteria, examination and feedback techniques. The OSCE consisted of five stations: two history-takings stations (chest pain and drug and alcohol); one physical examination station (lower limb examination); one procedural skills station (blood pressure) and one communication station (asthma puffer explanation). Each station ran for eight minutes (six minutes examination and two minutes feedback).

Examiner ratings

Examiners were instructed to assess year 2 students undertaking the practice OSCE using a standardised marking sheet. Each marking sheet contained 10–16 performance domains depending on the station. The marking criteria for these domains consisted of 'performed & correct', 'attempted, not correct' and 'Not attempted'. There was also one 'Global Judgement' marking criteria for each station: 'Satisfactory', 'Borderline' and 'Not Satisfactory'.

Following completion by the student examiner, marking sheets were reviewed by a senior academic. A second 'Global judgement' was made for each marking sheet by the senior academic, based on the number of criteria marked as 'Performed & Correct' by the student examiner. Academic global marks were calculated as follows: 'Satisfactory' when 80% or more of the criteria were marked by the student examiner as 'Performed & Correct'; 'Borderline' when 70% to less than 80% of the criteria were marked as 'Performed & Correct'; and 'Not Satisfactory' less than 70% or less of the criteria were marked as 'Performed & Correct'.

The global judgement scores of both student examiners and the senior academic were assigned a numeric value based on their mark (satisfactory = 1; borderline = 0.5; and not

satisfactory = 0) and analysed for differences in ratings using a sign test.

Student perception

Year 4 student examiners were also provided with a questionnaire regarding their experience as an examiner. The questionnaire consisted of 11 closed questions, five-point Likert scale questions such as 'Being an examiner developed my clinical skills' with responses ranging from 'Strongly Disagree' (1) to 'Strongly Agree' (5). The questionnaire also included two open-ended questions aimed at eliciting responses from students regarding the 'most useful' and 'least useful' aspects of the session.

Ethics approval was obtained from The University of Sydney Human Research Ethics Committee.

Results

Examiner ratings

Table 1 presents the distribution of marks made by both student examiners (Stu) and senior academics (SA). The senior academic changed a total of 94 marks, with the majority (55%) reduced from 'Satisfactory' to 'Borderline'; 12% were reduced from 'Satisfactory' to 'Not Satisfactory' and 33% from 'Borderline' to 'Not Satisfactory'. There were no instances in which a senior academic increased a mark from 'Not Satisfactory' to either 'Borderline' or 'Satisfactory' or from 'Borderline' to 'Satisfactory'. Station 3, which focused on communication, had the most number of mark changes, contributing almost a third (32%) of all the mark changes.

When ratings were assigned a numeric value and analysed using a sign test for paired data, there was strong evidence of a difference between those marks given by students when compared to those given by senior academics in all stations (Table 1), with student examiners consistently giving higher markings to the Stage 2 students. Results remained significant when both cohorts were analysed independently for all stations.

Student perception

Survey responses. All (100%) of the student examiners in both cohorts completed the post-OSCE questionnaire and median results for each of the 11 closed questions are reported

Table 1. Student examiner and senior academic ratings in practice OSCE stations and significance of difference.

	Station 1 (history)		Station 2 (physical examination)		Station 3 (communication)		Station 4 (history)		Station 5 (procedural)	
	Stu	SA	Stu	SA	Stu	SA	Stu	SA	Stu	SA
Satisfactory	69	67	90	71	93	72	81	72	93	81
Borderline	32	28	10	19	8	16	16	21	9	9
Not satisfactory	4	10	5	15	4	17	8	12	3	15
<i>p</i> -Value (sign test)	<i>p</i> = 0.008		<i>p</i> < 0.0001		<i>p</i> < 0.0001		<i>p</i> < 0.0001		<i>p</i> < 0.0001	

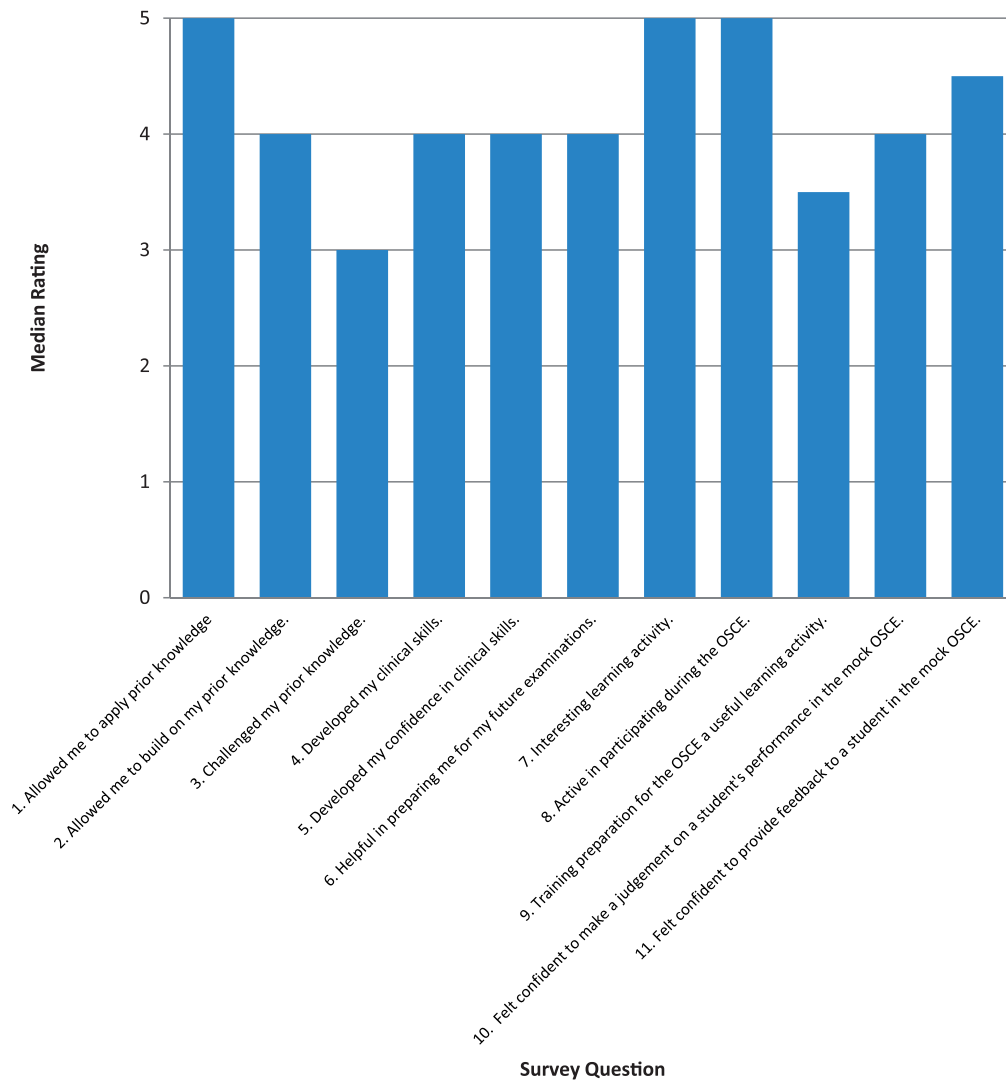


Figure 1. Closed question survey responses for year 4 student examiners in both 2010 and 2011.

in Figure 1. Students reported a high level of engagement with their examiner experience overall, especially with regard to 'using their previous knowledge', 'being an active participant in the process' and 'finding it interesting' (median = 5). Examiners were more ambivalent regarding the 'challenges the activity posed to their previous knowledge' (median = 3) and the 'usefulness of the training preparation as a learning activity' (median = 3.5). Examiners 'felt confident in making a judgement on the year 2 students' performance' (median = 4) and 'providing feedback' (median = 4.5).

Qualitative feedback. Year 4 students' qualitative feedback regarding the usefulness of acting as an examiner fell into three main themes. Almost half (40%) of both cohorts remarked how observing a variety of student styles/performances was both interesting and provided them with useful insight as to the examination process. Doing so enhanced 'Understanding [as to] what is expected from students as well as why' (Examiner).

Thirty-three percent of senior student examiners remarked on the benefits of providing feedback to junior peers. Students found this aspect of the examination enjoyable and valued the

opportunity to be able to share their knowledge and contribute to the education of their peers:

These sessions are good for fostering support of other students. Medicine can be confusing and isolating at times (Examiner).

Nice to get an opportunity to provide feedback and tips for students based on what I've learnt from my own exams and clinical (Examiner)

Learning how to give productive feedback to colleagues (Examiner).

Another 23% found that acting as examiners was useful in reviewing and reinforcing their own knowledge and clinical experience.

Seeing the same examination many times over – helped with remembering the examination (Examiner).

The drawbacks of acting as an examiner were primarily logistical and overwhelmingly fell into one of two categories. The first was that the repetition of being placed at one station throughout the exercise was repetitive and quickly became

dull. Examiners would have appreciated the opportunity to move about between stations.

[The worst feature was not] moving around stations – I spent all morning on Blood Pressure. It would have been interesting to do some other stations (Examiner).

Similar sentiments were expressed by 56% of those who responded to the question.

The second shortcoming was that student examiners (28%) felt that there was insufficient time to provide feedback to students, particularly those who were struggling to obtain a satisfactory result.

Discussion

Reliability and validity of peer marking

Our data suggest that while students are confident to make a judgement on a junior peer's performance by completing a standardised marking sheet and assessing each performance domain against set marking criteria, they appear less able to report an overall poor 'global' performance; hence, providing accurate and constructive feedback. This finding is particularly interesting considering the high median results on the questionnaire, indicating that examiners felt 'confident overall in providing a judgement regarding students' performance'. Thus despite being highly confident in their ability to assess performance, these novice examiners were inaccurate in providing a 'global' performance score. One obvious explanation is that student examiners were simply poor at providing an accurate overall assessment. However, it is also possible that while examiners were able to conceptualise students' performance with reasonable accuracy, they had misgivings about putting these scores down on paper, and being seen as responsible for a student's poor result. Indeed, research indicates that students can have concerns about passing judgement on colleagues' performance (English et al. 2006).

Regardless, it is of concern that in a practise OSCE students found it difficult to provide their peers with an honest global, overall score. It may mean that in a situation where a summative, overall mark was being provided, student examiners may mark even more leniently.

Provision of feedback to peers

While a recent literature review reported mixed opinions regarding the validity and reliability of peer feedback (Henning et al. 2008), it is thought that with the provision of adequate training by faculty, it can provide an effective learning experience for students (Kernan et al. 2005; Topping 2005; van den Berg et al. 2006). Survey responses from our student examiners showed that the faculty-led training in providing feedback and marking criteria for assessment was considered only moderately useful (median = 3.5). Indeed, it is clear from the student examiners' lack of ability to provide an accurate global mark for students that further training in this area is needed. Perhaps we underestimated their assessment

and feedback training needs, which may have contributed to their leniency in marking. It seems that it would also be of benefit to the students to alternate between at least two stations so that more can be gained from the experience in terms of revision of knowledge and engagement.

Preparation for future careers

Further training in assessment and feedback would not only better equip students to act as OSCE examiners, but also better prepare students for their future careers. Provision of feedback provides an important educational tool in developing professional competencies (Sluijsmans et al. 1999; Arnold et al. 2007). Peer assessment is not widely used formally in medical schools, even though peer review is a common requirement among junior and senior medical staff. As a result, physicians are often ill prepared for this aspect of their medical careers.

Development of professional attributes

Peer assessment can foster high levels of responsibility in students (Keaton & Richardson 1992). Qualitative data highlighted altruistic reasons for students taking part as examiners. The students' responses showed that students want to contribute to the education of their peers and help them develop the skills required for their high stake summative OSCE examination. It would be expected that the provision of further training and opportunities in this area would assist in fostering a supportive environment for medical students.

Revision of knowledge and clinical skills

Peer assessment is not only a judgement process, but also part of a process of developing skills, helping to inform self-assessment (English et al. 2006). Our students saw this process as a way to assess, review and develop their own knowledge and clinical skills. It also gave them an opportunity to observe different styles of clinical presentations.

Future research

Our study shows that this is a valuable exercise for our senior students acting as examiners. It also provides the junior students with an OSCE practise opportunity that would otherwise not be afforded to them due to scarcity of resources. It would, however, be worthwhile evaluating whether the provision of more intense training in feedback improves the accuracy of the global markings scores in the practise OSCE. Future research may also consider examining a single cohort through multiple iterations to determine the affects of practice on accuracy.

Conclusion

There are many reasons for the use of peer-based assessment. It not only acts as a meaningful learning activity for the student examiners, but also provides valuable experience in giving feedback. Student examiners found peer assessment a very

useful learning activity. It provided them with insight into exam technique and opportunity to review their own knowledge and clinical skills, and is a useful way to revise knowledge in an active learning environment. Our students, however, need further training in how to globally assess a fellow student's overall performance in an objective way, and provide meaningful and accurate feedback. When implemented correctly, peer examination can provide a very valuable resource.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Notes on contributors

ANNETTE BURGESS, BBusStud, MBT, MEd, is an Executive Officer at Central Clinical School, Sydney Medical School, The University of Sydney.

TYLER CLARK, BA (Psych), MA (AppSci-Hlth Psych), obtained his Masters in Health Psychology from The University of Sydney, where he currently works as an Associate Lecturer within the Office of Medical Education.

RENATA CHAPMAN, MD PhD, is a Senior Lecturer at Central Clinical School, Sydney Medical School, The University of Sydney, and also a general practitioner.

CRAIG MELLIS, MD, FRACP, is an Associate Dean, and Head of Central Clinical School, Sydney Medical School, The University of Sydney. He is a Respiratory Paediatrician, and a trained Epidemiologist and Biostatistician. Recent interests include medical education and evidence-based medicine.

References

- Arnold L, Shue CK, Kalishman S, Prislun M, Pohl C, Pohl H, Stern DT. 2007. Can there be a single system for peer assessment of professionalism among medical students? A multi-institutional study. *Acad Med* 82(6):578–586.
- Cushing A, Abbott S, Lothian D, Hall A, Westwood O. 2011. Peer feedback as an aid to learning – What do we want? Feedback. When do we want it? Now! *Med Teach* 33:105–112.
- English R, Brookes ST, Avery K, Blazeby JM, Ben-Shlomo Y. 2006. The effectiveness and reliability of peer marking in first year medical students. *Med Educ* 40(10):965–972.
- Finn G, Swandon M, Clipsham L, McLachlan J. 2009. Peer estimates of low professionalism correlate with low conscientiousness index scores. *Med Educ* 43(10):960–967.
- Henning JM, Weidner TG, Marty MC. 2008. Peer assisted learning in clinical education: Literature review. *Athl Train Educ J* 3:844–890.
- Jones R, Higgs R, de Angelis C, Prideaux D. 2001. Changing face of medical curricula. *Lancet* 357:699–703.
- Keaten JA, Richardson ME. 1992. A field investigation of peer assessment as part of the student group grading process. Paper presented at Western Speech Communication Association Convention. Conference Proceedings Feb 12–16 1992 Albuquerque NM.
- Kernan WN, Quagliarello V, Green ML. 2005. Student faculty rounds: A peer-mediated learning activity for internal medicine clerkships. *Med Teach* 27(2):140–144.
- Sluijsman D, Dochy F, Moerkerke G. 1999. Creating a learning environment by using self-, peer-, and co-assessment. *Learn Environ Res* 1:293–319.
- Topping KJ. 2005. Trends in peer learning. *Educ Psychol* 25(6):631–645.
- van den Berg I, Admiraal W, Pilot A. 2006. Peer assessment in university teaching: Evaluating seven course designs. *Assess Eval High Educ* 31(1):19–36.