# Student ability and learning experience in assessing peers alongside supervisors in the long case

*A. Burgess[1], C. Roberts[2], K. Black[1] & C. Mellis[1]*

## Abstract

***Background:*** Peer assessment can provide a valuable method of enriching students' learning experience, particularly when students act as the assessors, which provides a highly effective enquiry-based learning experience and is increasingly being utilised in medical schools internationally. In preparation for their summative examinations, senior medical students at the Sydney Medical School are required to assess their peers, alongside an academic co-examiner, in the formative long case examinations. This study sought to assess the level of agreement in marking and decision making between student peer and academic assessors, to evaluate the impact of peer assessment on examination performance and to investigate students' perception of their experience as assessors.

***Methods:*** Medical students examined their peers, alongside an academic co-examiner. We randomly allocated half of the student participants to examine a peer, alongside an academic assessor, *prior* to being examined themselves. The level of agreement in marking was determined by comparing the independent marking sheets of student and academic co-examiners. Data on whether the student was examined before or after assessing a peer (order of examination) were collected and compared to measure whether prior participation as a peer assessor improved student examination performance. Questionnaires and focus group discussions were used to evaluate student peer assessor perceptions.

***Results:*** Over a 3-year period (2010 to 2012), 197 students participated as co-examiners and were also examined by their peers, with students marking significantly more leniently than their academic co-examiners. Order of examination had no

1 Sydney Medical School—Central, University of Sydney
2 Sydney Medical School—Northern, University of Sydney

*Correspondence:*
Dr Annette Burgess
Sydney Medical School—Central
University of Sydney
Building 63, Level 4, Royal Prince Alfred Hospital
Missenden Road
Camperdown, NSW 2050
Australia
Tel: +61 2 9515 8172
Email: annette.burgess@sydney.edu.au



ISSN 1442-1100

significant bearing on student examination performance. Students identified several benefits of acting as an examiner, including better insight into examination technique, opportunities for self-reflection and knowledge acquisition, and development of some of the attributes of professionalism. However, students identified difficulty in providing critical feedback to peers.

***Conclusion:*** Engagement as an assessor alongside an academic supervisor provides a rich learning experience for students. Additional training in both peer assessment and feedback may increase students' professional and educational outcomes for future iterations of the activity.

***Keywords:*** peer assessment, peer assisted learning, medical students, long case examinations.

## Introduction

Much has been published internationally about the relationship between peer assessment and student learning. Important associations include enhanced self-reflection about one's own skills, development of a deeper understanding of subject knowledge, improvement in professionalism and potential resource benefits for universities (Cassidy, 2006). Peer assessment has been described as an educational arrangement in which students assess the quality of their fellow students' work and provide each other with feedback (Van den Berg, Admiraal, & Pilot, 2006, p. 19). While the accuracy of peer assessment is debatable (Topping, 2009), and whether it improves student performance is unknown (English, Brookes, Avery, Blazeby, & Ben-Shlomo, 2006), peer assessment can provide a valuable method of enriching students' learning experience (English et al., 2006; Topping, 2009) and is becoming increasingly used in medical schools throughout the world.

Although the educational and professional benefits to students of acting as assessors of their peers are acknowledged, it is important to determine the accuracy of peer marking (Topping, 2009). While some studies have found a trend of students marking more leniently than teachers (Burgess, Clark, Chapman, & Mellis, 2013; Heywood, 2000; Reiter, Rosenfeld, Nandagopal, & Eva, 2004), others found students grade accurately and consistently (compared to professional teachers) provided they had defined marking criteria and sufficient training (Bucknall et al., 2008; Marcoulides & Simkin, 1995). Although peer assessment is widely reported as being more accurate than self-assessment (Topping, 1998), several authors suggested that the process lacks objectivity and is subject to bias (Brindley & Schofield, 1998; Bushell, 2006).

By providing an authentic environment that mirrors a real-life examination situation, both the student learning experience in peer assessment and the future benefits of practical application are enhanced (Kneebone & Nestel, 2005). Even though it is not always reflected in students' marking accuracy, the act of assessing a peer provides insight into marking procedures and standards of expectation that may otherwise be difficult to gauge (Cassidy, 2006). Thus, peer assessment gives students insight into examination technique, providing a useful tool for students preparing for their own assessment (English et al., 2006).



ISSN 1442-1100

Peer assessment requires a student to engage with the knowledge and clinical skills of another student, providing an effective enquiry-based learning experience, thus promoting a deeper engagement with learning (Ramsden, 1992; Silbert & Lake, 2012). By assessing a peer, students are provided with an opportunity for decision making, which has been shown to improve the development of clinical skill competencies (Bradley & Postlethwaite, 2003). During the process of peer assessment, the observation and judgement of another student's performance provides a model for internal self-assessment of a student's own learning needs, encouraging the development of critical self-awareness (Ramsden, 1992). By encouraging student assessors to take responsibility for identifying their own knowledge gaps and utilising appropriate resources, autonomy in their learning is extended (Ten Cate & Durning, 2007).

Students are often apprehensive about both their own clinical ability and the responsibility associated with making a judgement on a peer's performance (Cassidy, 2006). Assessing and providing relevant, useful feedback to a fellow student, who may also be a friend, can be intellectually challenging and socially uncomfortable (Topping, Smith, Swanson, & Elliot, 2000). Despite these reported difficulties, learning to assess and provide feedback to one's peers are important skills for medical students to develop, and are likely to be required throughout their medical careers (Cushing, Abbott, Lothian, Hall, & Westwood, 2011; Topping, 2009). By empowering students to make independent judgements, professional self-awareness is developed (Ramsden, 1992).

There are potential advantages to encouraging peer assessment at the institutional level. With growing demands on university and clinical staff within medical schools, the implementation of innovative and efficient assessment methods has the potential to reduce institutional costs for clinical learning and teaching, and in this regard is considered a worthwhile endeavour (Jones, Higgs, De Angelis, & Prideaux, 2001; Topping, 1998). Involving students as assessors of their peers in order to develop professional competencies may provide a model to assist in addressing current resource challenges.

### *Context*

The clinical school in which this study occurred is a large tertiary teaching hospital and one of six schools to which students are allocated in the final two years (known as Stage 3) of a four-year graduate-entry problem-based medical program. As part of the assessment strategy, students are required to undertake a formative long case clinical examination in preparation for their summative long case clinical examination. The summative long case examination requires students to perform an unobserved history-taking and physical examination of a real patient for 1 hour. Then they spend 20 minutes preparing their findings, and finally, students' presentations are assessed for 20 minutes by a pair of clinician examiners. Defined criteria are used to assess students on six domains: history, examination, summary and problem list, differential diagnosis and investigation, management, and impact of illness on patient and family. Scores are allocated to each criteria in each domain (poor performance = 1; short of standard = 2; expected standard = 3; better than expected = 4; much better than expected = 5). In order to pass the long case, a student must score 3 ("expected standard") or more for



ISSN 1442-1100

two of the domains within history, examination and summary, and problem list; and two of the domains within differential diagnosis and investigation, management, and impact of illness on a patient. The formative examination is designed to mirror the procedure of the summative process, with students acting as examiners of their peers, alongside an academic co-examiner.

We sought to investigate any perceived and observable benefits to students when acting as examiners of their peers in these formative examinations. Our specific research questions were:

- What is the level of agreement in marking between student and academic assessors in the formative long case?
- Does acting as a peer assessor in formative long case examinations improve student performance on their own formative long case examination?
- How do students perceive their experience as assessors of their peers in the formative long case?

## Methods

We conducted the study over a 3-year period, from 2010 to 2012. Mixed methods were used for data collection and analysis to inform a richer understanding of our results (Creswell, 2002). While quantitative methods allowed us to systematically measure certain factors considered important in research literature, qualitative methods allowed us to tap into participants' perspectives and obtain useful and meaningful answers to our third research question (Johnson & Onwveghbuzie, 2004).

All Stage 3 students (i.e., Years 3 and 4) who had not previously sat any formative long case examination (n = 197) were allocated to co-examine their peers and to be examined themselves (2010, n = 97 students; 2011, n = 51 students; 2012, n = 49 students). Due to the changes within the curriculum in 2010, there were a greater number of students who had not previously sat the formative long case examination. We randomly allocated the assessment–examination order so that half of the students examined their peer, alongside an academic assessor, *prior* to being examined themselves.

Prior to the assessment, all students were provided with a 1-hour briefing and training session to take them through the six steps of the process. All student co-examiners were required to find a suitable patient on the ward, take a history from the patient, perform a physical examination and then present the patient to the academic co-examiner; formulate and lead the questioning of their peer; mark the examination independently of the academic co-examiner (using a standardised marking sheet); and lead feedback to their peer. Students were specifically briefed in the use of Pendleton's positive critique method to provide feedback to their peers (Pendleton, Schofield, Tate, & Havelock, 2003).

Ten senior academic staff with extensive undergraduate long case examination experience acted as co-examiners and were instructed, at a briefing prior to the assessment, to allow student co-examiners to lead questioning and feedback. These staff members included one female colorectal surgeon, two male vascular surgeons, one male



ISSN 1442-1100

cardiothoracic surgeon, one male immunologist, one female general practitioner, one female haematologist, one male oncologist, one male respiratory paediatrician and one female gynaecologist.

### *Data collection*

In order to address our three research questions, we used a validated survey and focus groups to collect data on the level of agreement between academic and student examiners, the impact of the intervention on student examinees' performance scores and perceptions of the impact of the intervention on student learning. All analyses were conducted using SPSS v19.0 and p values less than 0.05 were considered statistically significant. Cohen's kappa was calculated using Excel. Because our data were not normally distributed, we took a non-parametric approach.

#### **Level of agreement in marking.**

Student and academic examiners independently marked examinations using their own marking sheet. There were six domains for marking, three within *"Case Presentation"*, including history, examination, and summary and problem list, and three within *"Case Discussion"*, including differential diagnosis and investigation, management, and impact of illness on patient and family. Within each domain, the marking criteria ranged from 1 to 5 (5 being "much better than expected"). It was decided that two scores (by the academic and the student examiners) would be considered to have attained agreement if the two scores were both fails (1 and 2) or if the two scores were both passes *and* adjacent by one mark (3 and 4 or 4 and 5).

Both mean and median (with interquartile ranges) ratings were determined for peers and academic examiners. We used chance corrected agreement (Kappa statistics, K) to determine the level of agreement in marking between student and academic co-examiners, which established the level of inter-rater agreement between two independent raters using our standard marking sheet (Hasnain, Onish, & Elstein, 2004). Kappa statistics were generated to determine the level of agreement between the two marks (possible range 1 to 5) for each domain. We used the accepted standard to interpret Kappa measurements (Landis & Koch, 1997), where K = 0.01 to 0.20 is regarded as "slight" agreement, 0.21 to 0.40 "fair"; 0.41to 0.60 "moderate", 0.61 to 0.80 "substantial agreement" and above 0.80 "almost perfect agreement".

A comparison was made using the Wilcoxon signed-ranks z test (a non-parametric analysis for paired data) to examine whether the difference in ratings between peers and academic examiners were statistically significant. Cronbach's alpha was used to determine internal consistency of marking for student examiners and supervisors.

#### **Student performance—order of examination.**

In order to measure whether student performance in the formative long case examination was improved when a student had acted as an examiner prior to being examined themselves, data were explored to determine whether the order of examination (acting as a candidate first or examiner first) affected the examination scores. This is a non-graded,



ISSN 1442-1100

pass/fail examination. In order to pass, students must have passed two of the three domains within *"Case Presentation"* and two of the three domains within *"Case Discussion"*. Chi-square analyses were used to assess whether there was a statistically significant association between the order of examinations (being an examiner or a candidate first) and whether or not the student passed or failed their examination. Where expected cell sizes were too small, Fisher's Exact Test p value is reported.

### The student experience.

#### *Student peer examiner survey questionnaire.*

Surveys were distributed to all student examiners immediately following each long case examination. The survey questions were based on Brookfield's Critical Incident Questionnaire, which has been validated to provide significant feedback on student experiences in the learning environment (Brookfield, 1995). Students were asked to respond to 10 closed items (only nine items in 2010) using a Likert-scale of 1 to 5, with 1 being "strongly disagree", 2 being "disagree", 3 being "neutral", 4 being "agree" and 5 being "strongly agree". The tenth item, "I feel confident providing feedback to my peer", was added in the second year of the study, following initial student comments during the 2010 focus groups. Descriptive statistics were used to analyse these data (Creswell, 2002).

#### *Student peer qualitative data.*

At the completion of each set of long case examinations, students were invited to attend focus groups. The focus group questions were designed to explore aspects of the students' learning experience in greater depth. The questions specifically focused on the perceived benefits of the assessment program and areas of concern. Focus group data were transcribed verbatim with each participant being assigned an anonymous identifier (1 to 53). Thematic analysis was used to build an understanding of the students' experience as assessors. A portion of the data was read by the first author and analysed to identify initial themes. Following negotiation of meanings with the second and third authors, a coding framework was developed and applied to the full dataset (Creswell, 2002). QSR Int. Pty Ltd. NVivo (version 10, 2013) qualitative data analysis software was used for data analysis and management. Ethics approval was obtained from the University of Sydney Ethics Committee.

## Results

### *Level of agreement in marking*

Data were available from 182/197 student examinations (n = 82 in 2010, n = 51 in 2011 and n = 49 in 2012). Since there was uncertainty about student identification in 15/197 marking sheets, these 15 were removed from the dataset.



ISSN 1442-1100

Table 1
*Peer and Academic Co-examiner Examination Ratings by Domain (n = 182)*

| Marking domain | Peer examiner | | Academic examiner | | Statistics* |
|---|---|---|---|---|---|
| | Median (IQR) | Mean (SD) | Median (IQR) | Mean (SD) | |
| History | 4 (3–4) | 3.9 (0.7) | 4 (3–4) | 3.6 (0.8) | $z = -4.53$, $p < 0.001$ |
| Examination | 4 (3–4) | 3.7 (0.7) | 3 (3–4) | 3.4 (0.7) | $z = -5.04$, $p < 0.001$ |
| Summary & problem list | 4 (3–4) | 3.6 (0.8) | 3 (3–4) | 3.4 (0.8) | $z = -2.86$, $p = 0.004$ |
| Differential diagnosis & investigation | 3 (3–4) | 3.4 (0.7) | 3 (3–4) | 3.2 (0.7) | $z = -3.39$, $p = 0.001$ |
| Management | 3 (3–4) | 3.5 (0.7) | 3 (3–4) | 3.3 (0.7) | $z = -3.94$, $p < 0.001$ |
| Impact of illness on patient & family | 4 (3–4) | 3.7 (0.9) | 3 (3–4) | 3.3 (0.8) | $z = -5.80$, $p < 0.001$ |

3 = "expected standard"; 4 = "better than expected". IQR is the interquartile range.
*paired samples Wilcoxon Signed Ranks test

Table 2
*Level of Agreement Between Peer and Academic Co-examiners 2010 to 2012*

| Marking domain | 2010 to 2012 (n=182)* | | Kappa values |
|---|---|---|---|
| | n | % (95% CI) | |
| History | 157/180 | 87.2 (81.6–91.3) | 0.29 |
| Examination | 166/181 | 91.7 (86.8–94.9) | 0.26 |
| Summary & problem list | 153/181 | 84.5 (78.6–89.1) | 0.14 |
| Differential diagnosis & investigation | 155/181 | 85.6 (79.8–90.0) | 0.27 |
| Management | 158/180 | 87.8 (82.2–91.8) | 0.27 |
| Impact of illness on patient & family | 150/179 | 83.8 (77.7–88.5) | 0.26 |

*some n's varied due to missing data.

Peer and academic co-examination ratings by domain are shown in Table 1. Median ratings were either 3 ("expected standard") or 4 ("better than expected"). There were statistically significant differences between the marks provided by peer and academic co-examiners ($p < 0.01$ for all domains), with the students consistently providing higher marks.

The level of agreement between peer and academic co-examiners is shown in Table 2. The proportion of examinations where the two examiners agreed is presented along with the 95% confidence interval (95% CI). Level of agreement according to kappa was "slight" for the summary and problem list domain, but "fair" for all others.

Table 3 displays the agreement in overall pass or fail marking between academic and peer co-examiners (n = 172). The number was reduced from 182 to 172 because of missing data from individual criteria needed to accurately determine a pass or fail result on each



ISSN 1442-1100

Table 3
*Agreement in Overall Pass or Fail Marking Between Academic and Peer Co-examiners (n = 172*)*

| | | Academic co-examiner | | |
|---|---|---|---|---|
| | | Overall pass | Overall fail | Total |
| Peer examiner | Overall pass | 155 | 12 | 167 |
| | Overall fail | 2 | 3 | 5 |
| Total | | 157 | 15 | 172 |

*some n's varied due to missing data.

individual marking sheet. We found that over the 3-year period, a total of 15 students (out of 172) were failed by academic examiners, while only five students were failed by student examiners. Additionally, there were 12 cases where an academic co-examiner failed a student, but the peer examiner passed the student. There were two cases where an academic co-examiner passed a student, but the peer examiner failed the student.

Internal consistency (Cronbach's alpha) for student examiner marking was r = 0.82 and r = 0.89 for academics.

### *Student performance—order of examination*

Data on order of examination (candidate or examiner first) were available for n = 181 (n = 82 in 2010, n = 50 in 2011 and n = 49 in 2012). We were uncertain about the order of examination for one candidate, and this student's data was excluded. Students acted as examiners prior to being examined in 52% (95/181) of examinations.

The proportion of students who were "failed" by a peer examiner was similar for both candidates who were and were not an examiner first (3.2% and 3.5%, respectively; Fisher's Exact Test, p = 1.000). The proportion who were "failed" by the academic examiner was higher among students who were candidates prior to being examiners (12.8% versus 7.4%), but there was not a statistically significant difference (p = 0.288).

### *The student experience*

The response rate to the feedback questionnaire was 93% (n = 183/197) over the 3 years and ranged from year to year from 90% to 95%. The responses are displayed in Figure 1.

Eight focus groups, with a total of 53/197 (27%) participating students were held over the 3-year period (2010, n = 27; 2011, n = 10; 2012, n = 20).

Analysis of the qualitative data revealed four main themes: insight into examination techniques and examiner expectations, opportunity for reflection and knowledge acquisition, difficulty in marking and provision of feedback, and development of professionalism attributes.



ISSN 1442-1100

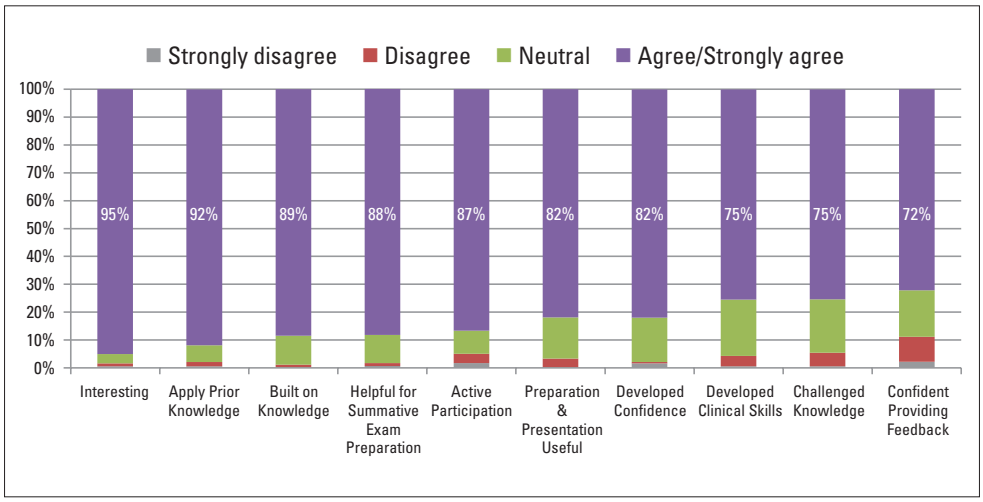


*Figure 1.* Stacked bar chart: Responses to questions regarding the experience of being a peer examiner, using a 5-point Likert scale ranging from 1 being "strongly disagree", 2 being "disagree", 3 being "neutral", 4 being "agree" and 5 being "strongly agree".

**Insight into examination techniques and examiner expectations.**

This theme refers to students' understanding of the examination format, procedures and marking criteria. Students remarked how discussion and advice from the academic co-examiner allowed them to think about their peer's presentation of the case from the examiner's perspective, providing a greater understanding and clarity of the performance domains and marking criteria. Students perceived that they benefited from the peer assessment process by gaining a unique insight into the examiner's expectations (Cassidy, 2006):

> *It added to my learning experience by sitting on the other side of the fence. I felt that helped give me a sort of context … as to levels of expected performance.* (S9)

Students also found it useful to be able to benchmark their own performance against that of their peer:

> *It is the first time you get to look at another case in an objective way and get to think about the case. Then you reflect on your own performance and think how to improve.* (S16)

**Opportunity for reflection and knowledge acquisition.**

When acting as co-examiner, students found that seeing the patient, preparing the questions and then summarising and presenting to the academic co-examiner was useful. Student assessors perceived that preparation as an examiner served as a means for revision of their medical knowledge and clinical skills (Burgess et al., 2013):



ISSN 1442-1100

*I read up about the case, and thought about some questions. It gave me a good opportunity to think. Knowing you had to ask questions made me learn a lot more and think a lot more.* (S1)

These findings support the theory that students learn by reflecting and identifying their own learning needs through enquiry-based learning (Ramsden, 1992).

**Difficulty in provision of feedback to peers.**

This theme refers to how students felt about providing feedback to their peers, which was a requirement during the assessment process. Students reported conflicting personal and professional difficulty in assessing and providing feedback to a colleague. This feeling of a level of responsibility for another student's mark resonates with students' experience as assessors in other settings (Burgess et al., 2013). Students repeatedly stated that they felt uncomfortable providing any negative feedback to their peers and that this part of the assessment process was the most difficult for them:

*You can either be objective and honest or you can just kind of get through it. And if you're honest, it can create some difficulties.* (S44)

This finding might be expected due to the social dimensions and pressures that exist within student groups (Topping, 2009).

**Development of professionalism attributes.**

This theme refers to students' recognition of skills related to professionalism that they perceived were developed through their assessment responsibilities. Students remarked on their privileged position as assessors, with the opportunity to observe a different style of presentation, self-reflect, make a judgement and demonstrate clinical leadership in articulating appropriate feedback (Cassidy, 2006).

Students found that they could be more reflective as an examiner than as a student being examined. Observing another student helped to inform their own self-reflection (English et al., 2006):

*When you're taking part in a discussion, if you're being examined, it's hard to sort of remember what questions were asked, but when you're being an examiner, you could remember it more from an objective point of view and … it gets easier to reflect on it when … you weren't nervous.* (S41).

As assessors, students became more engaged with the experience. These findings align with Topping's (1998) view that the additional time spent in assessing peers provides a greater opportunity for "thinking, comparing, contrasting and communicating" (p. 254).

Many students showed recognition that while giving feedback is difficult, particularly when students are likely to know each other, it is a professionalism attribute that can be developed through acting as an examiner in these formative examinations:

*I think it's difficult but it still needs to be done. You have to be able to give co-workers feedback from here on in.* (S49)



ISSN 1442-1100

Through assessing their peers, students develop transferable skills in professionalism that may be applied in future teaching and assessment encounters (Ten Cate & Durning, 2007).

## Discussion

This study provides evidence of the level of agreement between peer and academic examiner marking in the formative long case examinations, evidence of effects on student performance resulting from the order of examination and insight into the students' learning experience as examiners of their peers. Our students proved to be more lenient markers than the academic examiners, and we could not establish that prior participation as a peer assessor improved subsequent examination performance. However, the activity was highly valued by students, with students identifying greater insight into examination techniques and examiner expectations, opportunities for reflection and knowledge acquisition and the chance to benchmark their own performance. The ability to provide honest feedback to peers was identified as an area of concern to students, even though they recognised this as a professionalism competency.

### *Level of agreement in marking and decision making*

Others have found that students graded consistently compared to professional teachers in a context where they were provided with defined marking criteria and sufficient training (Bucknall et al., 2008; Marcoulides & Simkin, 1995). However, we were unable to demonstrate a good level of agreement in marking between peer and academic examiners, since students were consistently more lenient. It is possible that we had underestimated the amount of training prior to such a formal peer assessment activity that is required for effectiveness. Although the marking domains and criteria were provided to students, spending more time in articulating the criteria may have improved student accuracy in marking (Topping, 2005; Van den Berg et al., 2006).

Student examiners were also significantly less likely than academic examiners to fail a student, with 12/172 (7%) of instances where an academic co-examiner failed a student, but the student examiner passed the student. The low correlation in marking is likely to reflect both the reluctance of students to judge their peers (English et al., 2006) and perhaps the inherent difference of opinion that occurs in scoring a clinical long case examination, even between two academic staff (Wilkinson, Campbell, & Judd, 2008).

### *Order of examination*

We found no significant difference in student results when students acted as examiners of their peers before being examined themselves. This finding resonates with the findings of others who have shown that students' examination performance is not improved by participating as peer assessors (English et al., 2006; Sluijsman, Dochy, & Moerkerke, 1999; Topping, 1998). Although English et al. (2006) suggested that one possible



ISSN 1442-1100

explanation may be that students are more focused on the skill of "marking", rather than identifying their own areas of weakness, our students repeatedly commented on the benefits of self-assessment and reflection.

### *The student experience*

It is thought that peer assessment is demanding on the assessor's cognitive skills, leading to a deeper understanding of knowledge (Van Lenh, Chi, Baggett, & Murray, 1995). Our students were able to recognise educational benefits that occurred through active learning before, during and after the peer assessment event (Ballantyne, Hughes, & Mylonas, 2002; Topping, 2009; Vickerman, 2009). Students reported benefits from preparing as an examiner, the actual examination, and from reflection. Having students act as co-examiners of their peers, alongside an academic examiner, allowed the student learning experience in assessment to encompass a broad, active, learning process. In a context of a student assessor and faculty member assessing a student–patient encounter, there were two sources of interaction that promoted student engagement with learning as a co-examiner. First, the peer assessors were engaged as active questioning participants, and second, they modelled the inquiry processes of the disciplinary experts, constructing clinical knowledge for themselves. In addition, alongside academic examiners, students felt that they were able to gain a unique insight into the examiners' expectations and develop a greater understanding of the marking domains and criteria (Cassidy, 2006).

Peer assessment is not only a judgement process but also part of a process of developing skills that help to inform self-assessment (English et al., 2006). While acting as an examiner of their peers in the long cases, students are given a unique opportunity to observe another student's performance and a different style of presentation. Students were able to improve their clinical reasoning skills by virtue of observing other students' clinical reasoning in their presentation and discussion of the long case. Having students examine each other in the clinical long case provided an enquiry-based learning opportunity, where students were able to recognise knowledge areas in which they themselves were deficient and focus on their own learning, finding suitable strategies to seek information (Ramsden, 1992).

Students commented on the challenges of critiquing and providing feedback to fellow students, and recognised their tendency to mark leniently. It is possible that while student examiners were able to gauge students' marks with reasonable accuracy, they were reluctant to honestly critique their peers, with a fear of being seen partially responsible for a student's poor performance. Indeed, it appears well documented that students can have concerns about passing judgement on a colleague's performance (English et al., 2006). Students might not be willing to accept responsibility for assessment of their peers where students know each other well or if the assessment is formative (Falchikov & Goldfinch, 2000). Social pressures within student groups can influence and affect the accuracy of peer assessment, with perception of criticism as being socially uncomfortable (Topping, 2009). However, learning how to give feedback forms part of social and assertion skill development, facilitating greater employee abilities in these areas (Marcoulides & Simkin, 1995). During focus group discussions, students acknowledge that while providing feedback to their peers is



ISSN 1442-1100

difficult, it is an important professionalism attribute. Providing opportunities in higher education for students to assess a peer's work provides a powerful tool to develop important work place skills (Boud & Falchikov, 2006).

## Implications

Our findings suggest that students assessing peers alongside supervisors in a formative long case examination is a valuable learning activity, which is generalisable to other clinical schools. Although peer assessment alongside an academic is a resource intensive activity, this study has shown that it is beneficial to have the academic examiner present throughout the examination to maximise the learning outcomes for students.

Others have reported that adequate training by faculty in peer assessment, including organisation, outcomes, criteria and, particularly, skills in giving feedback, is required to optimise the student learning experience (Boud & Falchikov, 2006; Kernan, Quagliarello, & Green, 2005; Topping, 2009; Van den Berg, 2006). An important consideration is whether refining the training students receive would prevent them from consistently overestimating the standard that student examinees have achieved. Our 1-hour training session may have underestimated students' assessment and feedback training needs, contributing to their leniency in marking and difficulties in providing feedback.

## Conclusion

Students found peer examination alongside an academic co-examiner in the formative long case examinations a useful learning activity in preparation for their own summative long case examinations. It provided students with insight into examination techniques and an opportunity to develop their medical knowledge and clinical skills. It also provided students with a point of reference to measure and reflect on their own performance (Topping, 1998). The ability to provide honest feedback to peers was identified as an area of concern for students. Further training in both assessment and feedback skills, as well as guidance by academic supervisors, may enhance the educational and professional outcomes of the peer assessment activity for students in the future.

## References

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education*, *27*, 427–441.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, *31*, 399–413.

Bradley, P., & Postlethwaite K. (2003). Simulation in clinical learning. *Medical Education*, *37* (supplement 1), 15.

Brindley, C., & Schofield, S. (1998). Peer assessment undergraduate programmes. *Teaching in Higher Education*, *3*(1),169–176.



ISSN 1442-1100

Brookfield, S. (1995). *Becoming a critically reflective teacher*. San Francisco: Jossey-Bass.

Bucknall, V., Sobic, E. M., Wood, H. L., Howlett, S. C., Taylor, R., & Perkins, G. D. (2008). Peer assessment of resuscitation skills. *Resuscitation*, *77*(2), 211–215.

Burgess, A., Clark, T., Chapman, R., & Mellis, C. (2013). Senior medical students as examiners in the OSCE. *Medical Teacher*, *35*, 58–62.

Bushell, G. (2006). Moderation of peer assessment in group projects. *Assessment & Evaluation in Higher Education*, *31*(1), 91–108.

Cassidy, S. (2006). Developing employability skills: Peer assessment in higher education. *Education and Training*, *48*(7), 508–517.

Creswell, J. (2002). *Educational research: Planning, conducting and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

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! *Medical Teacher*, *33*, 105–112.

English, R., Brookes, S. T., Avery, K., Blazeby, J. M., & Ben-Shlomo, Y. (2006). The effectiveness and reliability of peer marking in first year medical students. *Medical Education*, *40*(10), 965–972.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*(3), 287–322.

Hasnain, M., Onish, H., & Elstein, A. S. (2004). Inter-rater agreement in judging errors in diagnostic reasoning. *Medical Education*, *38*(6), 609–616.

Heywood, J. (2000). *Assessment in higher education*. London: Jessica Kingsley Publishers.

Johnson, R., & Onwveghbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26.

Jones, R., Higgs, R., De Angelis, C., & Prideaux, D. (2001). Changing face of medical curricula. *Lancet*, *357*, 699 –703.

Kernan, W. N., Quagliarello, V., & Green, M. L. (2005). Student faculty rounds: A peer-mediated learning activity for internal medicine clerkships. *Medical Teacher*, *27*(2),140–144.

Kneebone, R., & Nestel, D. (2005). Learning clinical skills: The place of simulation and feedback. *The Clinical Teacher*, *2*(2), 86–90.

Landis, J. R, & Koch, G. G. (1997). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Marcoulides, G. A., & Simkin, M. G. (1995). The consistency of peer review in student writing projects. *Journal of Education for Business*, *70*(4), 220–223.

Pendleton, D., Schofield, T., Tate, P., & Havelock, P. (Eds.). (2003) *The new consultation: Developing doctor–patient communication.* Oxford, UK: Oxford University Press.



ISSN 1442-1100

Ramsden, P. (1992). *Learning to teach in higher education*. London: Routledge.

Reiter, H. I., Rosenfeld, J., Nandagopal, K., & Eva, K. W. (2004). Do clinical clerks provide candidates with adequate formative assessment during objective structured clinical examinations*? Advances in Health Sciences Education*, *9*, 189–199.

Silbert, B. I., & Lake, F. R. (2012). Peer-assisted learning in teaching clinical examination to junior medical students. *Medical Teacher*, *34*, 392–397.

Sluijsman, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self-, peer-, and co-assessment. *Learning Environment Research*, *1*(3), 293–319.

Ten Cate, O., & Durning, S. (2007). Peer teaching in medical education: Twelve reasons to move from theory to practice. *Medical Teacher*, *29*(6), 591–599.

Topping, K. J. (1998). Peer assessment between students in college and university. *Review of Educational Research*, *68*, 249–276.

Topping, K. J. (2005). Trends in peer learning. *Educational Psychology*, *25*(6), 631–645.

Topping, K. J. (2009). Peer assessment. *Theory Into Practice*, *48*(1), 20–27.

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education*, *25*(2),149–169.

Van den Berg, I., Admiraal, W., & Pilot, A. (2006). Peer assessment in university teaching: Evaluating seven course designs. *Assessment & Evaluation in Higher Education*, *31*(1), 19–36.

Van Lehn, K. A., Chi, M. T. H., Baggett, W., & Murray, R. C. (1995). *Progress report: Towards a theory of learning during tutoring.* Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.

Vickerman, P. (2009). Student perspectives on formative peer assessment: An attempt to deepen learning? *Assessment & Evaluation in Higher Education*, *34*: 221–230.

Wilkinson, T. J, Campbell, P. J., & Judd, S. J. (2008). Reliability of the long case. *Medical Education*, *42*(9), 887–893.



ISSN 1442-1100