The Objective Structured Clinical Examination and student collusion: marks do not tell the whole truth

R Parks, P M Warren, K M Boyd, H Cameron, A Cumming and G Lloyd-Jones

doi:10.1136/jme.2005.015446

Updated information and services can be found at:
http://jme.bmj.com/cgi/content/full/32/12/734

These include:

References
This article cites 18 articles, 4 of which can be accessed free at:
http://jme.bmj.com/cgi/content/full/32/12/734#BIBL

Rapid responses
You can respond to this article at:
http://jme.bmj.com/cgi/eletter-submit/32/12/734

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Correction
A correction has been published for this article. The contents of the correction have been appended to the original article in this reprint. The correction is also available online at:
http://jme.bmj.com/cgi/content/full/33/3/186-b

Notes

To order reprints of this article go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to Journal of Medical Ethics go to:
http://journals.bmj.com/subscriptions/
Higher education institutions are paying increased attention to academic dishonesty among their students. Although cheating under any circumstances is a serious regulatory matter, its importance on courses leading to professional status is arguably greater. The General Medical Council states that doctors should be honest and trustworthy; so a question mark hangs over the suitability of any student found guilty of cheating for entry into the profession. Because little is known about the career trajectories of students who transgress ethical codes, a link between student dishonesty and future professional lapses cannot be assumed, however plausible the association may seem. Limited evidence from self-reported data from engineering students in the US points to an association between student dishonesty and future professional lapses. The evidence supports rising scores over intervals of a year, the evidence supports rising scores over time. Among the many responses was an anonymous confessional, from "recent graduates" of a UK medical school, describing collusion on a major scale in relation to an Objective Structured Clinical Examination (OSCE). Students had gained sight of a marking sheet 2 days before the exam and disseminated the information, enabling recipients to revise selectively for the OSCE. A greater number of merits and distinctions were awarded that year. However, there is genuine uncertainty about whether advance notice of content advantages a candidate sitting a performance assessment. The evidence from the literature is far from clear. Several studies have examined scores when similar OSCEs are administered over periods varying from days to 3 months and have shown that overall score trends do not rise over time. In two of these studies, accompanying self-reported data confirmed that students had discussed station content between administrations of OSCEs, although the nature of the interactions was not investigated. In another study, in which students were officially encouraged to discuss exam content, there was no evidence of rising scores over time. Contrary reports exist, however, of rising trends in overall marks with successive OSCE administrations and, in a repeat of their initial study, Colliver et al. analysed within station scores and showed that some component items did exhibit rising trends even when overall scores did not. In cases where OSCE stations have been reused over longer intervals of a year, the evidence supports rising scores over time.

The assumptions underlying these studies require closer scrutiny. Some assume that advance notice of station content will advantage candidates, resulting in higher scores for them...
than their peers. From this arises the claim that a trend for a rising score in a cohort reflects a potential breach of exam security and, conversely, that finding consistent scores over time indicates that no major violation of exam security has occurred.\textsuperscript{13} 14 16–19 The alternative view, that advance information can disadvantage candidates, resulting in lower marks over time, is considered, but features less prominently.\textsuperscript{15} 20 22 If information about content has the potential to either advantage or disadvantage a candidate, then inferences relating to trends in overall marks of the peer group to collusion become unreliable. Even consistent mean marks over time cannot be taken to indicate that exam security has been maintained.\textsuperscript{22} Detecting breaches of test security, particularly at the individual level, becomes difficult, if not impossible, under these circumstances. In addition, a focus on marks has seemingly diverted attention from the ethical implications of collusion and its potential significance for the development of professionalism in undergraduate students.

Aims
Against this background, we describe a critical incident that occurred in a UK medical school, in which student collusion at a third year OSCE was seen on a staff–student electronic discussion board. We consider the incident to be of wider relevance and interest owing to the many issues that it raises for students, teachers and educational institutions, in which the OSCE format is used in the assessment programme.

We aim to:

- describe the incident, its management and consequences for the school;
- challenge the view that trends in overall means or station marks, derived from sequential OSCE sittings, provide a sound basis for the presence or absence of collusion; and
- discuss ethical issues arising out of collusion from student and staff perspectives.

Because this paper relates to events in real time, which do not conform easily to the predetermined character of a research study, the following structure has been chosen:

- a description of the incident in context;
- a description of the staff response to the incident; and
- an analysis of OSCE marks.

Critical Incident

OSCE implementation

In all, 255 third year students sat for a summative OSCE consisting of 15 stations of 10 min each. The exam was conducted at three different sites over 2 consecutive days, resulting in three sittings per day at each site. Table 1 describes the individual station tasks and any intended variations between days 1 and 2. Marking sheets were standardised and the standard for scores was set using the Angoff method.

Students taking the OSCE on day 1 were explicitly advised not to pass on information about the exam to students at later sittings. Steps were taken to avoid successive groups of candidates talking to one another, although it was impossible to prevent day 1 and 2 students communicating during the assessment period.

Electronic discussion board

The medical school runs a series of password-protected, electronic discussion boards, organised by year of study, which are accessible to staff and students in the school. Less than 2 h after the end of the day 2 OSCE, students initiated a discussion about collusion between day 1 and 2 students, which continued until the publication of results 12 days later. Nearly 40 students contributed, about three quarters of whom remained anonymous.

The main topic was the perceived lack of equity between those students who had, and had not, received insider information about OSCE content. Students without insider information argued that colluding students retained an advantage over the rest of their colleagues, which devalued the achievements of honest students and the assessment process. There was concern that honesty was not being rewarded. Several posts criticised their peer’s dishonesty and its potential effect on professionalism and future practice. Although students reported either receiving insider information themselves or knowing of someone who had, no one confessed to conveying information. Those receiving insider information disclaimed any advantage, but others reported studying relevant electronic learning resources and memorising names of drugs in the interim. Clearly, collusion had occurred between students sitting the OSCE on the same day, as well as on different days.

Staff response

Under the circumstances, staff thought it imperative to respond to events, as the electronic discussion board is publicly accessible to the entire school. Although students who use the facility may elect to remain anonymous, staff reserves the right to uncover the identity of any student who fails to maintain proper standards. Staff associated with the

<table>
<thead>
<tr>
<th>Table 1 Description of the Objective Structured Clinical Examination stations in order of question consistency between days 1 and 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Station blueprint</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Gastrointestinal system</td>
</tr>
<tr>
<td>Locomotor system</td>
</tr>
<tr>
<td>Cardiovascular system</td>
</tr>
<tr>
<td>Respiratory system</td>
</tr>
<tr>
<td>Locomotor system: diagnosis and management</td>
</tr>
<tr>
<td>Pharmacology</td>
</tr>
<tr>
<td>Ethics</td>
</tr>
<tr>
<td>Practical skills</td>
</tr>
<tr>
<td>Communication skills</td>
</tr>
<tr>
<td>Resuscitation</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Pathology</td>
</tr>
<tr>
<td>Anatomy</td>
</tr>
<tr>
<td>Microbiology</td>
</tr>
<tr>
<td>ECG</td>
</tr>
<tr>
<td>Radiology</td>
</tr>
</tbody>
</table>

ECG, electrocardiogram; Q and A, questions and answers.
OSCE vigorously debated appropriate responses to the incident. Reprimanding students who had admitted receiving advance information was considered, but was narrowly rejected on the following grounds. Firstly, no student had admitted passing on information and these students were arguably as guilty as those who received information. Also, students would probably not divulge the names of informants. Secondly, the discussion board thread probably represented only the tip of the iceberg and the collusion may have been more widespread. Further problems arose from the interpretation of electronic interaction: without clues of non-verbal behaviour, irony and sarcasm are difficult to differentiate from factual statements.

Staff recognised that targeting the students who contributed to the discussion would make scapegoats of a few, whereas others, perhaps many more, would evade censure. Instead, they chose to post a response to all third-year students on the electronic discussion board, dealing with some of the students’ misconceptions about the implementation of the OSCE, recognising the inequity of the situation and promising to work towards a form of implementation that reduced opportunities for collusion to uphold the honest students. It was emphasised that the policy of non-disclosure of exam content would continue to remain an offence under the current assessment regulations.

**RESULTS ANALYSIS**

In the light of the student discussion, we reviewed the OSCE marks. The mean and SD for each day, differences in means (day 2—day 1) and the 95% confidence interval for the difference were calculated for the overall mark and the marks for each station (table 2). The day 1 and 2 marks were compared using the two sample t test (SPSS V.13). Although multiple comparisons were carried out, a Bonferroni correction was considered to be inappropriate because of the lack of independence of the overall mark and the station marks. Caution is therefore required when interpreting the reported p values.

The small increase in the mean overall mark between days 1 and 2 was not significant at the 0.05 level. For individual stations, roughly half showed significant differences between the average marks on the two days. It was impossible to attribute these findings to collusion, however, because other influences contributed to the average station mark. These included variations in examiners, student ability and case specificity arising from the participation of different patients at different sittings.25 Secondly, analysis of the results from all students sitting the exam on a specific day reflects an average effect for the class. If only a few students colluded then the effect would be masked by the results from the large majority who did not. Thirdly, collusion can affect the mark positively or negatively and the overall effect is diluted when considering the average values. Finally, the mark for an individual station was the sum of multiple sections, each of which may have been affected in opposite ways by collusion. For instance, a student may only lose or gain 2 marks out of 40 for a specific item. Thus, analysis of the marks for individual stations was not reliable for detecting or quantifying collusion.

We therefore decided to look in detail at the marks from those stations where examiners had commented spontaneously, without prompting by staff inquiries, that they had suspicions of collusion for their particular station (Communication skills, Pathology).

**Communication skills**

Examiners of the Communication skills station commented verbally to one of the authors (GL-J) that they had noted that day 2 students were offering the day 1 diagnosis, despite changes to the station scenario. Attempts to analyse the component items of the Communication skills answer sheets were foiled because the marks awarded for supplying the correct diagnosis were aggregated with marks for other aspects of the student-simulated patient interaction. It was therefore not possible to either support or refute the examiners’ assertions of collusion.

**Pathology**

We were alerted to possible collusion in the Pathology station by the examiners’ formal feedback comments, which were posted on the third year website. The examiners stated that “It seems possible that collusion was a confounding factor in day 2 answers to the second part of the question. The histology picture changed from tuberculosis on day 1 to squamous carcinoma on day 2.”

At this station, the scenario, questions and marking schedule were unchanged on days 1 and 2; only the photomicrograph specimen was changed (day 1 tuberculosis to squamous carcinoma).

---

**Table 2**: Total and stations scores for days 1 and 2

<table>
<thead>
<tr>
<th>Station</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Mean difference (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>128</td>
<td>127</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>389.9 (36.8)</td>
<td>396.5 (38.3)</td>
<td>6.6 (−2.66 to 15.9)</td>
<td>0.162</td>
</tr>
<tr>
<td>Stations with questions consistent between days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gastrointestinal system</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>30.5 (4.4)</td>
<td>31.1 (4.7)</td>
<td>0.7 (−0.4 to 1.8)</td>
<td>0.236</td>
</tr>
<tr>
<td>Total score</td>
<td>28.8 (6.8)</td>
<td>28.0 (6.4)</td>
<td>−0.9 (−2.5 to 0.8)</td>
<td>0.311</td>
</tr>
<tr>
<td>Stations with questions partially changed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>26.4 (6.4)</td>
<td>24.5 (5.3)</td>
<td>−1.9 (−3.4 to −0.5)</td>
<td>0.010</td>
</tr>
<tr>
<td>Total score</td>
<td>19.7 (4.5)</td>
<td>24.0 (4.4)</td>
<td>4.3 (3.2 to 5.4)</td>
<td>0.000</td>
</tr>
<tr>
<td>Stations with questions changed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>29.1 (5.4)</td>
<td>26.7 (5.5)</td>
<td>−2.4 (−3.8 to −1.01)</td>
<td>0.001</td>
</tr>
<tr>
<td>Total score</td>
<td>30.0 (5.3)</td>
<td>32.0 (4.4)</td>
<td>2.0 (0.7 to 3.3)</td>
<td>0.002</td>
</tr>
<tr>
<td>Stations with questions consistent between days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gastrointestinal system</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>26.4 (7.5)</td>
<td>26.2 (6.0)</td>
<td>−0.2 (−1.1 to 0.7)</td>
<td>0.220</td>
</tr>
<tr>
<td>Total score</td>
<td>24.1 (5.1)</td>
<td>23.6 (5.3)</td>
<td>−0.5 (−1.7 to 0.8)</td>
<td>0.483</td>
</tr>
</tbody>
</table>

Data are given as means and (SD) unless otherwise specified.

ECG, electrocardiogram.

---

Downloaded from jme.bmj.com on 16 January 2008
OSCE and student collusion

The incident also supports the contention that component items, when searching for breaches of test security, support the later work of Colliver et al. Instead, they support the later work of Colliver et al. providing a sound assumption for the presence or absence of cheating. The OSCE is now divided into two halves, each one being administered on one day only. One half consists of clinically based stations, involving interaction with simulated and real patients; the other is delivered via computer and includes questions on patient management, data handling and analysis, and interpretation of investigations. Although this will not exclude the possibility of collusion, it will reduce the opportunities for it.

For institutions, the more difficult problem is upholding honesty and protecting students who do not cheat, for they can feel disregarded by the system. From this incident and other reports, it seems that students are concerned about collusion and cheating, but may lack opportunities to voice their opinions. Possibly, the anonymous protection afforded by the discussion board may have encouraged students to speak out in this instance. To discourage cheating, Glick1 proposes that schools carefully articulate their expectations of students and take positive, public steps to encourage a culture in which unethical behaviour simply becomes unacceptable. Otherwise, there is the danger that turning a blind eye, however unintentional, may encourage the growth of undesirable norms in which peer group loyalty overcomes professional ethics. It is in this spirit that we report the incident, to publicise the uncertain effects of collusion in relation to OSCEs and to open a debate about these issues in relation to current OSCE implementation practices.

CONCLUSION

Although research examining the effects of collusion on OSCE marks has been conducted, there is little information about the nature of collusion and its ethical implications. An incident of collusion at a UK school showed that it could disadvantage students, although its effect on marks appeared to be limited. Honest students raised ethical concerns about cheating colleagues and the possible effect on their future professional practice. In the absence of any evidence against individual students, the school issued a public statement and has since modified the implementation of the OSCE to reduce opportunities for cheating. The potential influence of these issues on the development of social and cultural norms in medical school culture has been considered. In view of the global popularity of the OSCE format for assessing professional competence, these findings may have wider relevance.

ACKNOWLEDGEMENTS

We thank Mr Robert Lee of the Medical Statistics Unit, University of Edinburgh, UK, for advice on the statistics applied to the data given in table 2.

Number of correct or incorrect answers for diagnosis given at the Pathology station on days 1 and 2. Day 1 “TB” correct, number giving the correct answer of tuberculosis; day 1 Incorrect, number giving any incorrect answer; day 2 “Ca” correct, number giving the correct answer of squamous carcinoma; day 2 “TB” incorrect, number giving tuberculosis as an incorrect answer; day 2 Other incorrect, number giving an incorrect answer other than tuberculosis; not available, script on original mark sheet indecipherable.

DISCUSSION

These findings challenge the view that trends in overall mean or station marks, derived from sequential OSCE sittings, provide a sound assumption for the presence or absence of collusion. Instead, they support the later work of Colliver et al. in showing the need to examine specific station component items, when searching for breaches of test security. The incident also supports the contention that collusion has the power to disadvantage rather than to benefit students who indulge in collusion. The two premises, that overall marks are insufficiently sensitive as indicators of collusion, and the uncertain effect of advance notice on marks, go some way in explaining the variations in previous studies.

Some may criticise our conclusions on the grounds that the Pathology question preferentially tested knowledge rather than skills, as knowledge is easier to reproduce at short notice. There is some truth in this, but it is only partial. Although the OSCE is promoted as a skill-based assessment, skill must be constructed on, and integrated with, knowledge if it is to be of value in educating the safe, competent professional. Representing skills and knowledge as entirely different entities is unhelpful when, in reality, the relationship is far more complex. Furthermore, the anecdotal evidence from the Communication skills examiners shows how a station, perceived primarily as a skills test, is not immune to collusion.

It is reassuring that the marks awarded for the specific items are small and, standing alone, would have little effect on the outcome. It is, however, possible for multiple small breaches to have a cumulative effect for an individual student, which would probably go undetected. From the perspectives of the staff and honest students, there is comfort in knowing that malpractice was apparently not rewarded in this case. Yet, the incident threw up difficult issues for staff in handling the public, yet anonymous, electronic report and in protecting honest students. This is probably not an isolated case. When this incident was reported at a recent international conference in medical education, many members of the audience from around the world related similar experiences.

We describe below amendments to the OSCE aimed at reducing opportunities for collusion but, in reality, it will be impossible to block the flow of information, particularly electronic texting and emails. Trusting students will always be necessary when OSCEs are repeated day after day. Although plagiarism can be detected by software, detection of collusion in performance examinations presents an altogether more difficult problem. The deliberate inclusion of questions designed to test the possibility of collusion is one option for consideration. Colliver et al. follow a similar strategy in suggesting an examination of working diagnosis scores in the context of a simulated-patient examination in the US.

Steps have since been taken to modify the OSCE format to reduce opportunities for collusion. The OSCE is now divided into two halves, each one being administered on one day only. One half consists of clinically based stations, involving interaction with simulated and real patients; the other is delivered via computer and includes questions on patient management, data handling and analysis, and interpretation of investigations. Although this will not exclude the possibility of collusion, it will reduce the opportunities for it.
Authors’ affiliations
R Parks, Chancellors Building, Little France, Edinburgh, UK
P M Warren, H Cameron, Medical Teaching Organisation, University of Edinburgh, Edinburgh
K M Boyd, Medical School Buildings, University of Edinburgh
A Cumming, College of Medicine and Veterinary Medicine, University of Edinburgh
G Lloyd-Jones, Cranfield Professional Development, Cranfield University, Bedfordshire, UK

Competing interests: None.

REFERENCES
allocation considered from a lawyer’s point of view. It described resource allocation as a problem in the reasonable and legally accountable administration of a set of public bodies, rather than as a problem in clinical medicine, economics or moral philosophy. As such it was a very important contribution to a vexed field, and did much to clarify debates, at least in the academic field.

This second edition is therefore to be warmly welcomed. The NHS has undergone an even more extensive change since the first edition was published, under the banner of modernisation. Of particular importance from the point of view of resource allocation are the arrival of the National Institute for Clinical Excellence, the reorganisation of healthcare commissioning around primary care trusts and, within medical science, the rise of evidence-based medicine and the arrival of new, highly expensive classes of drugs for treating chronic diseases.

The approach taken by the author in the second edition is to consider the nature of resource allocation itself, then to look in detail at how resource allocation can be carried out through clinical judgement, expert guidelines, government regulation and direct democratic decision making. He then goes on to look at how the NHS is managed, considering, rightly, that the most important resource allocation decisions are taken at the mid-range level of primary care and hospital budgeting. He describes the current structure of the NHS in some detail, identifying the scope of the different powers that lie with different tiers or centres of decision making and the role these have in resource allocation. The bulk of the book is then devoted in turn to the analysis of review by the courts of administrative decisions relating to the allocation of resources, medical negligence law (whether rationing can create a case for action in negligence by causing the care offered to fall below a proper standard of care), the relationship between local decision makers and government policy, and governance and accountability processes. An important final chapter looks at the regulation of care provided under the auspices of the NHS by non-NHS providers at home and overseas.

The analysis and arguments are presented with admirable clarity and should be fully accessible to non-lawyers. However the structure of the NHS changes in the near future, short of outright privatisation, this book will remain a valuable guide to the issues of policy and practice.

Compared with the first edition, this edition is shorter and more focused on the guiding principles of the law and public policy in this area. The first edition was longer by about a quarter, and included chapters on complaints and professional discipline and on access to medical records (considered as means to make NHS decisions accountable to patients) as well as a more detailed treatment of the then effective “internal market”. There was also some discussion of the relationship of resource allocation decisions with decisions about the value of life and end-of-life care. Most of this is missing from the second edition. This makes it a more focused book and one with a longer shelf life in view of likely changes in the NHS. However, scholars would be advised to keep their first editions on their shelves, as some of the now omitted material remains of both historical and intellectual interest. The new book has gained in focus and grasp of general principles, but has lost some of the “black letter” lawyer’s attention to detail.

Overall, this as an important and fascinating book which deserves a wide readership among students, practitioners, managers, policy makers and lawyers. No work on resource allocation can overlook it.

R E Ashcroft

CORRECTIONS
doi: 10.1136/jme.2005.013904corr1
An error has occurred on page 569 in the October issue, in the paper titled Decisions at the end of life: an empirical study on the involvement, legal understanding and ethical views of pre registration house officers (J Med Ethics 2006;32:567–70). Line 24 in the right hand column should read “of intravenous hydration to be morally appropriate, even…” The journal apologises for this error.

doi: 10.1136/jme.2005.14720corr1
In the paper titled, Distributive justice in screening for hearing in newborns (J Med Ethics 2006;32:588–91), an error has occurred in line 8 of the second paragraph. The correct sentence is, “These approaches are unlikely to diminish the resources available to other healthcare needs inequitably and they may in fact result in raising the overall standard of healthcare across the communities.”

doi: 10.1136/jme.2005.15446corr1
In the paper titled, The objective structured clinical examination and student collusion: marks do not tell the whole truth (J Med Ethics 2006;32:734–9), the initial for the author Kenneth E Boyd is incorrect. It should be Kenneth M Boyd. The journal apologises for this error.