Gaining an advantage by sitting an OSCE after your peers: a retrospective study

Short title: Gaining an advantage in an OSCE

Asim Ghouri[1], Charles Boachie[2], Suzanne McDowall[1], Jim Parle[3], Carol A. Ditchfield[1], Alex McConnachie[2], Matthew R. Walters[1*], Nazim Ghouri[1*]

1. School of Medicine, Dentistry and Nursing, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK
2. Robertson Centre for Biostatistics, Institute of Health and Wellbeing, University of Glasgow, Glasgow G12 8QA, UK
3. Institute of Clinical Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, B15 2TT, UK

* Joint senior authors

Address for correspondence:

Dr Nazim Ghouri
School of Medicine, Dentistry and Nursing
College of Medical, Veterinary and Life Sciences
University of Glasgow
Glasgow G12 8QQ
United Kingdom

Telephone: + 44 (0) 141 3303076
Fax: + 44 (0) 141 3302522
Email: nazim.ghouri@glasgow.ac.uk

Abstract

Purpose: To investigate if final year medical students undertaking an OSCE station at a later stage during an exam diet were advantaged over their peers who undertook the same station at an earlier stage, and whether any such effect varies by the student's relative academic standing.

Methods: OSCE data from six consecutive final year cohorts totalling 1505 students was analysed. Mixed effects logistic regression was used to model factors associated with the probability of passing each individual station (random effects for students and circuits; and fixed effects to assess the association with day of exam, time of day, gender, and year).

Results: Weaker students were more likely to pass if they took their OSCE later in the exam period. The odds of passing a station increased daily by 20%. Overall, the mean number of stations passed by each student increased over the 5 days.

Conclusions: Students undertaking the same OSCE stations later in an exam period statistically had better chances of passing compared to their peers, and the weaker students appear to be particularly advantaged. These findings have major implications on OSCE design, to ensure students are not advantaged by examination timing, and weaker students are not 'passing in error'.

Keywords: OSCE, undergraduate medicine, exam design, cheating, collusion

Introduction

Assessment in medical education has numerous purposes and indeed effects. Perhaps the most important purpose of summative assessment is to ensure that those candidates who pass are truly competent, or to put it another way and from a patient perspective, no-one wants to be treated by an incompetent doctor who passed their exams but should not have done. Data on the prevalence of cheating is limited but it is clearly an issue of concern to all stakeholders: Fargen et al (2016) in a literature review report substantial levels of cheating among medical students with the largest study reporting a prevalence of 39%. The Objective Structured Clinical Examination (OSCE), introduced around 40 years ago (Harden and Gleeson 1979) to improve the reliability and objectivity of the assessment of clinical skills, is an established mode of examination for medical students and trainee doctors (GMC 2014; Association of American Medical Colleges 2017). The OSCE consists of multiple stations simulating "real time" scenarios, which are intended to reflect clinical practice. Hence the OSCE examination allows assessment of a student's clinical skills: the 'shows how' of Miller's pyramid (Gormley 2011).

For organisational reasons the current configuration of OSCE delivery often involves the re-use of stations on consecutive days. This however gives candidates a potential opportunity to collude over the contents of the examination (i.e. cheat), potentially conferring an advantage to students undertaking the examination on later sittings (Parks et al 2006). Ultimately the suspicion of collusion can raise doubts over the validity of the exam grades, and draw into question the integrity of students suspected of participation in this behaviour, with consequent risk to the trust of the public in the medical profession (Smith 2000). In addition such collusion may advantage the weaker students more and result in some passing who should not.

Previous studies of OSCE results have not shown a significant improvement in student scores for stations repeated over time (Colliver et al 1991; Rutala et al 1991; Skakun et al 1992; Swartz et al 1993; Niehaus et al 1996; Brown et al 1999; Parks et al 2006; Kim 2010). These studies have been performed using third year and fourth year students in USA, UK and South Korea. The study with the largest number of students to date was performed by Parks et al (2006), who analysed the OSCE marks of 255 third year undergraduate medical students over a 2-day period. Student collusion was confirmed via an online discussion board set up by the medical school. However, no significant difference was observed in the total mark for the OSCE on day 1 compared with day 2. A clear indication of the effects of collusion could only be obtained from a single subsection of a pathology station, where 82 students on day 2 incorrectly gave the diagnosis which had applied on day 1 despite the slide having been changed.

The study with the most OSCE station repetitions was undertaken by Rutala et al (1991) at the University of Arizona. Seventy six fourth year medical students took an OSCE in which 14 stations were repeated over six sittings and 16 stations over four sittings. They did not find a significant change of scores over this period. A small number of long-term studies (Cohen et al 1993 and Jolly et al 1993) have shown evidence of progressive improvement in OSCE scores when the same stations are repeated over consecutive years. Gotzmann et al (2017) demonstrated an improvement in overall total scores in an artificially breached test security setting where two cohorts of students were given information about the stations beforehand.

The aim of this study was to investigate whether undergraduate final year medical students undertaking an OSCE station later on in an exam diet experience an advantage over their peers who undertake the same station at an earlier stage, and in particular to assess the potential impact of any advantage on the final result of the exam for individual students.

Methods

Student population

Exam performance of final year medical students (Year 5) undertaking their OSCE at the University of Glasgow, Scotland, UK, was analysed. For the purpose of this study, 6 consecutive final year cohorts were included (2009-10 to 2014-15 inclusive). Data from 1505 students was included (n=238, 262, 226, 261, 259, 259 in consecutive years). All students are given instructions prior to sitting the OSCE advising that they must not discuss the content of the OSCE with candidates sitting the OSCE at other sites or circuits until the final completion of the examination due to (1) issue of equity and respect to colleagues; (2) issue of failure to personally meet standards of Good Medical Practice; and (3) breach of University examination rules and GMC standards on probity. They are also advised that if they are found to be discussing the content of the OSCE by any means while the examination is running, they will be subject to disciplinary action in the form of referral to Senate and formal Fitness to Practice procedures. Additionally, all students are advised via the Professionalism tab on the University virtual learning environment (Moodle) - 'Social media: what does this mean for you' and this is linked to the following GMC information: http://www.gmc-uk.org/information_for_you/11851.asp.

OSCE format

For each year there were between 32 and 50 OSCE stations that contributed to the clinical component of the final year undergraduate medical curriculum. The examination was divided into four parts: A-D. Each part assessed different fields: A – Obstetrics and Gynaecology and

Psychiatry, B – Medicine and Surgery, C – Paediatrics, D – Other specialties. These were performed at various sites. In week 1 parts A and B run simultaneously, and in week 2 parts C and D run simultaneously. For each part, several identical circuits were performed throughout the day over 4-5 days to accommodate all students. Students were allocated a day and a circuit for each part at random

Each station in this analysis could involve assessment of communication skills, clinical examination, emergency care, performance of a practical procedure, identification of anatomical specimens, completion of a drug prescription chart or interpretation of clinical data in the form of an observation chart, ECG, or radiological images. Stations assessing history taking or communication skills used simulated patients (SPs). All SPs are trained actors who undergo training for the role. The majority of the SPs also take part in the teaching of communication skills so are highly experienced in the role of SP.

Each station had one examiner only during a circuit. All examiners were senior clinicians and all were trained in OSCE assessment. Whilst an examiner could mark the same station for more than one circuit on the same day or on different days, no individual examined the same station for the full 4-5 days. In addition, examiners may have assessed a different station during another circuit on the same or different days.

For each station, the student was marked out of 20 against an objective list of items. In addition, the examiner made a second, global judgement of the student's performance and categorised it as a "Pass", "Fail" or "Borderline". The pass mark for each station was then calculated by taking the numerical scores for all candidates who were rated as borderline and calculating the mean of these scores. Students passed or failed by their scores alone; the global judgement is not given any weighting in this regard.

The format of the stations was kept as consistent as possible throughout the days. Whilst clinical findings may have varied slightly due to different patients being used on different days, the station remained the same.

A minimum number of stations were required to be passed in order to pass the OSCE overall (66% from 2009-10 to 2013-14 and 75% for 2014-15).

Data protection and ethical approval

All data that was collated and subsequently analysed were part of standard data collected for each student undergoing the examination process. Student-identifiable information was removed from the data set before analysis. Ethical approval was not required as analysis of data is an external annual requirement of the Medical School to ensure satisfactory standards are being maintained with student-identifiable data remaining confidential to those faculty members who are entitled to access this information.

Statistical analysis and data presentation

Statistical analyses were carried out using SAS v9.3 for Windows. Student demographics (age and gender) and the mean pass and fail counts are summarised for each year group. The percentage of stations passed on each day are presented graphically for each year group. Mixed effects logistic regression was used to model factors associated with the probability of passing each individual station. Models included random effects for students and circuits. Fixed effects were included to assess the association with day of exam (1-5), time of day (am/pm), gender, and year. Associations are reported as odds ratios (ORs) with 95% confidence intervals (CIs) and p-values. A p-value less than 0.05 was used as an indication of a significant association.

We used our model to estimate the predicted probability of each student passing the OSCE overall. Then, for each student, we replaced the combination of days and times that each station was taken, with the corresponding combinations that each other student was given in the same year group. In this way, we were able to assess the potential impact of variation in pass rates by day and time on the overall probability of passing the OSCE, for each student, within the constraints of the way that stations were allocated to students in the same year.

In addition, within each year group, we divided students into groups according to their overall grade achieved, and looked at the days on which they took their exams, to assess whether the highest or lowest pass rates tended to be achieved by students who sat their exams earlier or later in the week. We calculated the mean day of all stations taken by each student as a measure of when the student took their exams, and tested for a correlation between the mean exam day and final score.

Results

Demographic variables and general OSCE performance

Table 1 summarises the students' demographic characteristics; age and gender distributions were similar across the six year groups. Table 1 also shows that the number of stations making up the OSCE reduced from a high of 50 stations in 2010-11, to 32 in 2014-15; however, the mean percentage of stations passed was consistent across the study period.

The reliability of each OSCE exam was measured using Cronbach's alpha (Cronbach 1951). Values are all >0.8 which demonstrates a high level of internal consistency for each individual exam.

[Table 1 near here]

Significance of influencing variables on likelihood of passing OSCE stations

Table 2 shows the estimated associations between academic year, gender, day, and time, and the probability of passing a station, derived from the mixed effects logistic regression analysis. There were no significant differences in the probability of passing stations across the six year groups. Male students were less likely to pass each station than female students ($p<0.001$). There was a trend across days, with the odds of passing a station increasing by 20% for each additional day. This association can also be seen in Figure 1. Overall, the mean number of stations being passed by each student rose over the 5 days. However, undertaking the same station in the afternoon as opposed to in the morning did not confer any statistically significant benefit.

[Table 2 near here]

[Figure 1 near here]

Predicted probability of passing the OSCE

Figure 2 shows the predicted probability of each student passing the minimum number of stations needed to pass the OSCE exam overall. This was derived from the mixed effects logistic regression model, based on the actual combination of days and times that each student was allocated, and under all possible alternative combinations, taken from the other students in the same year group. Only the 100 students with the lowest predicted probability of passing are shown for each year group; in each year, the remainder had predicted probabilities close to 100% regardless of when the stations were taken. Note that if a student in 2014-15 had a predicted probability of passing each station of 90%, then the probability of passing at least three quarters of the stations (i.e. at least 24 stations) would be 99.67%. The actual predicted probability of passing each station might vary slightly due to stations being taken on different days, but in general, for most students, these predicted probabilities are high enough that the predicted probability of passing overall is close to 100%; altering the days and times on which each station is taken has little impact on this overall probability.

[Figure 2 near here]

The mixed effects logistic regression model includes a random effect for students; in other words, there is an assumed distribution of students' abilities built in to the model. Only for those students at the lower end of this distribution does the overall predicted probability of passing dip noticeably below 100%; for these students, varying the days and times on which each station is taken has a greater effect upon overall probability of passing the exam. According to the model, this effect can be large: for some students, the probability of passing could vary between 10% and 90%, depending on the days and times on which they take their stations. Note, however, that these are predictions being made at the extreme fringes of the model, and may not be an accurate representation of the impact of when stations were taken for these students.

These predicted effects appear more marked since 2012; prior to this date, there were more stations in the OSCE (46, 50, and 45 in 2009-12) than there were in subsequent years (35, 35, 32).

Distribution of days and times by student performance

Table 3 shows the mean day on which students sat their stations each year, divided according to the final grade achieved. Between 2009-10 and 2012-13, there is no evidence that the days on which students sat their exams was associated with higher or lower scores. However, in 2013-4 and particularly 2014-15, these data suggest that the highest scoring students sat more of their stations later in the process.

[Table 3 near here]

Discussion

Our study suggests that final year medical students were significantly more likely to pass a station in the OSCE if undertaken on a later day compared to students undertaking the same OSCE station on earlier days of the exam period, and that this effect was particularly marked for weaker students. Our data also shows that this effect was more evident among the weaker students. This suggests that student collusion is a possible contributory factor in increasing the chances of passing an OSCE station. Although most probable, collusion is not the only potential explanation. Students may become more attuned to the examination process over time, and show improved performance even in the absence of collusion. Examiners may demonstrate trends in their marking over time, e.g. becoming more lenient on later days, though this is unlikely since the vast majority of all our examiners do not do more than one half-day session and there is no reason to expect this to occur. Regardless of the mechanism, our data suggests that the day on which an exam is taken may influence exam performance.

To our knowledge ours is the first study to report an effect of examination timing on OSCE performance. The annual dataset analysed was similar to the analysis undertaken by Parks et al (2006). Their analysis did not demonstrate student advantage on passing a station; however this could be a reflection of lack of days in the exam diet for a trend to emerge. As indicated in Figure 1, significant advantages start to appear when comparing students doing exams at least three days later, suggesting that any effect of collusion accrues gradually. The large dataset used in our study has made possible a robust assessment of factors associated with the likelihood of passing a station and therefore allowed for estimation of the associations with day and am/pm with greater precision.

The main strength of this study is that it is to our knowledge the largest undergraduate study yet which has examined the effect of OSCE timing on academic outcome. This study analysed results over six years, therefore giving a much larger student sample size overall. Furthermore, due to the number of days the OSCE is performed over, we attempted to examine for a trend in day to day performance to ascertain whether potential student collusion starts to have a significant effect on OSCE performance.

In response to the outcomes from our analysis, to ensure that no content becomes known to candidates sitting in later circuits and to ensure a fair examination for all candidates across all sites and circuits, all candidates are now quarantined during the OSCE. This involves the students do not have full electronic isolation - supervised by invigilators. Further, all new MBChB students also now sign a student agreement, which notes the following with regards to social media **–** '*I am aware that my responsibility to communicate professionally extends to digital media. I will not post images or text online which may cause concern or distress to any individual, in keeping with the personal attributes expected of a medical student or doctor. I will not access social media through NHS resources while on placement'*. Finally, as part of general measures, undergraduate students from across the school have engaged in a digital skills enhancement project. Students were invited to participate in the Digital Identity survey. The results identified key topics on which students want more guidance, including digital identity management, professionalism in an online environment, digital well-being, productivity skills (including management of digital distractions), and opportunities for communication and collaboration online. Forty student partners worked in collaboration with staff, focusing on potential challenges, solutions, and opportunities for curriculum developments in these areas. Face-to-face and online teaching resources were created based on their insights. This year these resources were used to develop workshops and teaching sessions to begin embedding digital skills into the undergraduate medical curriculum.

We acknowledge some important limitations of our study. Firstly, as this is an observational study, we are limited in the extent to which we can attribute causality to the associations we observe. Another major limitation of this study was that little information is known about the examiners for the stations for the period of analysis and therefore this could not be factored into interpreting the results. On the other hand we have no reason to suppose that examiners would be more lenient as the days go on, although examiner stringency and leniency can influence student pass rates and can account for up to 12% variance in OSCE scores (McManus et al 2006). In addition, as examiners are sourced from local hospitals where students undertook clinical attachments, it is likely that a portion of students may have encountered an examiner they were familiar with at some point during their OSCE. Studies of the effect of examiner familiarity as a source of bias have reported inconsistent findings (Jeffries et al 2007 and Stroud et al 2011). Although the configuration of OSCE examinations is broadly similar in undergraduate medical schools, differences in content and structure may limit the generalizability of our findings. However, we have no reason to believe that any such factors mitigate for or indeed against scores on different days. A final limitation is the generic structure of the marking structure used for history and examination stations. For the period of analysis, for example, if a student had a cardiovascular examination station, they could acquire enough marks to pass the station even though they may get the interpretation of the murmur or final diagnosis wrong. However this limitation is only applicable to certain history and examination stations and not to any procedural or investigation-based stations, for which marking schemes are more specific. And, again, there is no reason to suppose such effects would be of relevance to the timing of taking particular stations. Further studies in other medical schools would help to confirm or refute this.

In conclusion, our evidence suggests that students undertaking the same OSCE stations later on in an exam period have a better chance of passing compared to students earlier in an exam period, and this is particularly evident for the less able students. Whether this reflects collusion between students, trends in examiner assessments, or some other source(s) of bias, cannot be determined. We believe ours is the first study to look particularly at weaker students; it seems likely that any effects we are describing would be less apparent among the higher-performing students. And, from the point of view of the patient, this possible effect, which *could* result in doctors qualifying who would not otherwise do so, is of key significance.

These findings will be of interest to those charged with the organisation of OSCE examinations. Lapses in security of OSCES may have major local consequences which are the subject of national attention when they occur (Zamost S et al 2012; Glasgow medical students …2017). Further studies are justified to replicate our findings, and to investigate whether more stringent approaches to reducing opportunity for collusion, such as quarantining all students during an exam period, will reduce the apparent benefit conferred by sitting a later OSCE examination.

Practice points

- OSCEs are an established mode of examination for medical students
- The potential collusion in such examinations is a recognised limitation
- Data on the effect of collusion is divided
- We demonstrate over a six-year period of an OSCE consisting of multiple repeated carousels that weaker students were more likely to pass a station if they took their OSCEs later in the diet
- These findings have major implications for OSCE design; to ensure students are not advantaged by timing and weaker students are not "passing in error"

Author contributor statements

Asim Ghouri is an honorary clinical lecture at the University of Glasgow and contributed to drafting all versions of the manuscript.

Charles Boachie is a statistician at the Robertson Centre for Biostatistics at the University of Glasgow and analysed and interpreted data. Charles also contributed to drafting of the original manuscript.

Suzanne McDowall is Examinations and NHS Liaison Officer at University of Glasgow, Undergraduate Medical School. Suzanne helped extract the raw exam data used for analysis.

Jim Parle is Professor of Primary Care at the Institute of Clinical Sciences, University of Birmingham. Jim contributed to conceiving and designing the study and drafting the final manuscript.

Carol A. Ditchfield is a senior lecturer and is involved with the undergraduate medical curriculum at the University of Glasgow. Carol contributed to conceiving and designing the study. Carol also contributed to drafting of the original manuscript.

Alex McConnachie is a statistician at the Robertson Centre for Biostatistics at the University of Glasgow and oversaw data analysis and interpretation. Alex also contributed to drafting of the original manuscript.

Matthew R Walters is Head of School (Medicine, Dentistry and Nursing) at the University of Glasgow. Matthew conceived and designed the study and contributed to the drafting all versions of the manuscript.

Nazim Ghouri is Honorary Clinical Senior Lecturer at the University of Glasgow. Nazim conceived and designed the study, contributed to, and oversaw drafting of all versions of the manuscript.

Declarations of interest

All authors report no declarations of interest.

Ethical Approval

Ethical approval was not required and standard anonymised data collected by the university was the data source for analysis.

Funding

No external funding was sought for this study.

Reference List

Association of American Medical Colleges. Number of Medical Schools Requiring Final SP/OSCE Examination. Association of American Medical Colleges; [accessed August 31 2017]. **https://www.aamc.org/initiatives/cir/406426/9.html.**

Brown G, Manogue M, Martin M. 1999. The validity and reliability of an OSCE in dentistry. Eur J Dent Educ. 3(3):117-25.

Cohen R, Rothman AI, Ross J, Poldre P. 1993. Impact of repeated use of objective structured clinical examination stations. Acad Med. 68 Suppl:S73-S75.

Colliver JA, Barrows HS, Vu NV, Verhulst SJ, Mast TA, Travis TA. 1991. Test security in examinations that use standardized-patient cases at one medical school. Acad Med. 66(5):279-282.

Cronbach L. 1951. Coefficient alpha and the internal structure of tests. Psychomerika.16(3):297-334.

Fargen KM, Drolet BC, Philibert I. 2016. Unprofessional Behaviors Among Tomorrow's Physicians: Review of the Literature With a Focus on Risk Factors, Temporal Trends, and Future Directions. Acad Med. 91(6):858-864.

Glasgow medical students to resit exam after 'collusion'. 2017 Mar 27. BBC. [accessed May 19 2017]; **http://www.bbc.co.uk/news/uk-scotland-glasgow-west-39405833.**

GMC. 2014. How are students assessed at medical schools across the UK? [London (UK)]: General Medical Council; [accessed August 31 2017]. **http://www.gmc-uk.org/Assesment_audit_report_FINAL_pdf.pdf_59752384.pdf**.

Gormley G. 2011. Summative OSCEs in undergraduate medical education. Ulster Medical Journal. 80(3):127-132.

Gotzmann A, De Champlain CA, Homayra F, et al. 2017. Cheating in OSCEs: The Impact of Simulated Security Breaches on OSCE Performance. Teach Learn Med. 29(1):52-58.

Harden RM, Gleeson FA. 1979. Assessment of clinical competence using an objective structured clinical examination (OSCE). Med Educ.13(1):41-54.

Jefferies A, Simmons B, Regehr G. 2007. The effect of candidate familiarity on examiner OSCE scores. Med Educ. 41(9):888-891.

Jolly B, Newble D, Chinner T. Learning effect of reusing stations in an objective structured clinical examination. Teaching and Learning in Medicine: An International Journal 1993;5:66-71.

Kim JH. 2010. Does sharing information before a clinical skills examination impact student performance? Med Teach. 32(9):747-53.

McManus IC, Thompson M, Mollon J. 2006. Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling. BMC Med Educ. 6:42.

Niehaus AH, DaRosa DA, Markwell SJ, Folse R. 1996. Is test security a concern when OSCE stations are repeated across clerkship rotations? Acad Med. 71(3):287-289.

Parks R, Warren PM, Boyd KM, Cameron H, Cumming A, Lloyd-Jones G. 2006. The Objective Structured Clinical Examination and student collusion: marks do not tell the whole truth. J Med Ethics. 32(12):734-738.

Rutala PJ, Witzke DB, Leko EO, Fulginiti JV, Taylor PJ. 1991. Sharing of information by students in an objective structured clinical examination. Arch Intern Med. 151(3):541-544.

Skakun EN, Cook DA, Morrison JC. 1992. Test security on sequential OSCE and multiple choice examinations. In: Harden R, Hart I, Mulholland H, editors. Approaches to the assessment of clinical competence. Norwich: Page Brothers; p. 711-718.

Smith, R. 2000. Cheating at medical school. BMJ. 21:398.

Stroud L, Herold J, Tomlinson G, Cavalcanti RB. 2011. Who you know or what you know? Effect of examiner familiarity with residents on OSCE scores. Acad Med. 86 Suppl:S8-S11.

Swartz MH, Colliver JA, Cohen DS, Barrows HS. 1993. The effect of deliberate, excessive violations of test security on performance on a standardized-patient examination. Acad Med. 68 Suppl:S76-S78.

Zamost S, Griffin D and Ansari A. Updated 2012 Feb 6. Exclusive: Doctor cheating warnings expand to dermatology. CNN [accessed August 24 2017]; http://www.cnn.com/2012/02/03/health/doctor-cheating-dermatology/index.html.

Figure 1: Percentage of stations passed per exam day, by year group cohort

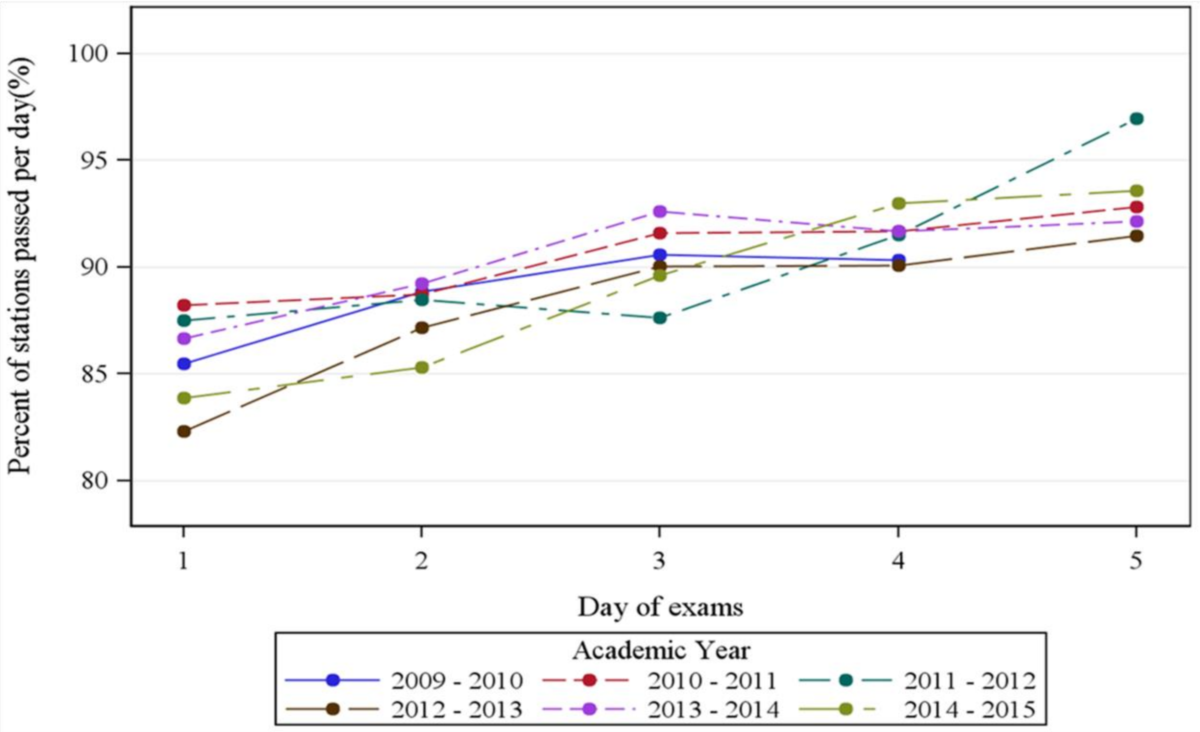

Figure 2: Overall predicted probability of passingfor the 100 lowest ranked students in each year group cohort. Error bars show range of predicted probabilities obtained by reassigning examination days from each other student in that year

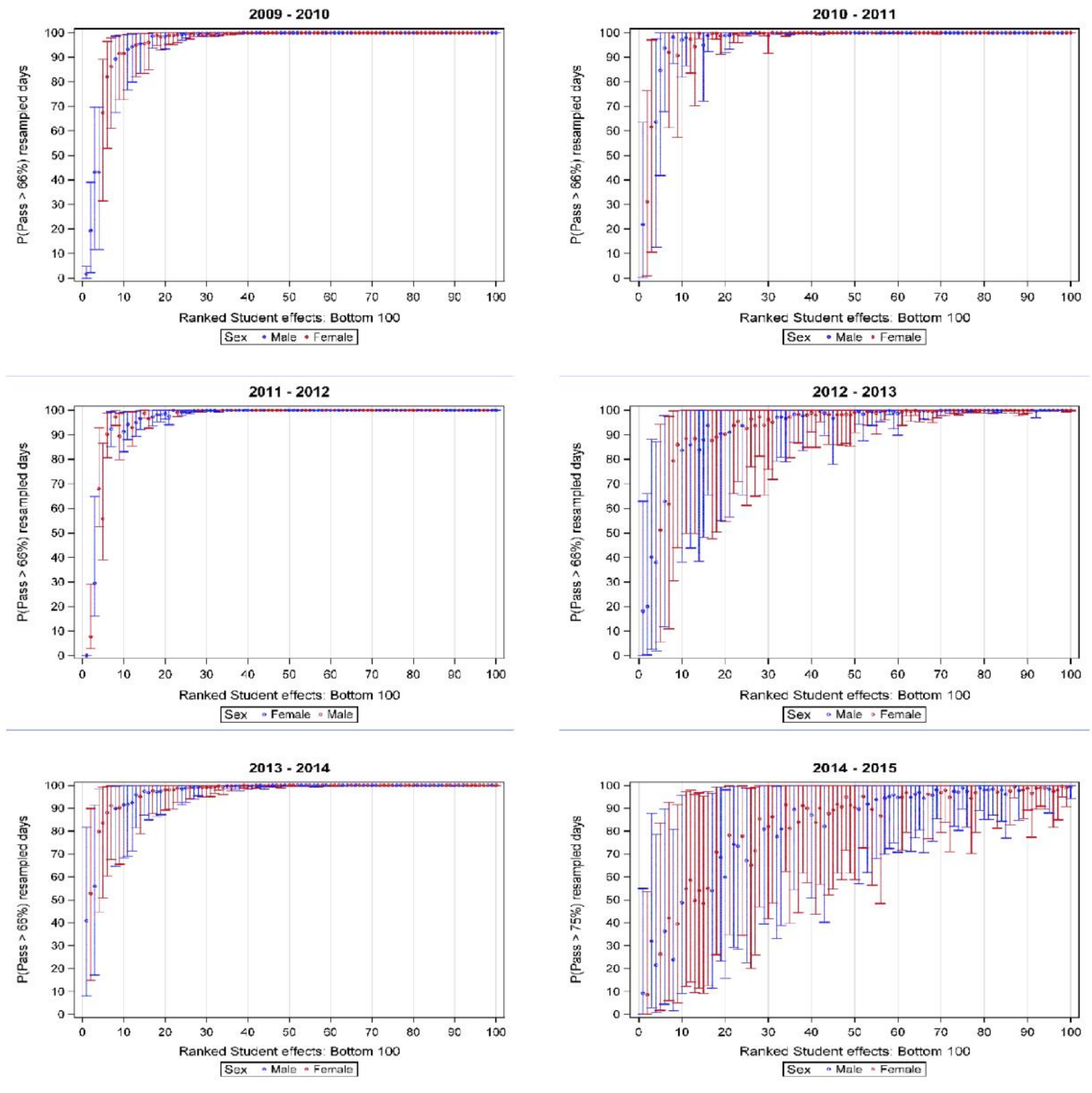

Table 1: Summary of student demographics, Cronbach's alpha and pass/fail counts, by year group cohort

	Year					
	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15
N Students	238	262	226	261	259	259
Age	24.7 (2.7)	24.6 (2.6)	25.0 (2.5)	24.9 (3.3)	24.9 (2.7)	25.0 (2.7)
N (%) Male	92(39%)	115(44%)	95 (42%)	115(44%)	106(45%)	103(44%)
N (%) Female	146(61%)	147(56%)	131(58%)	146(56%)	132(55%)	131(56%)
N Stations	46	50	45	35	35	32
Cronbach's alpha	0.88	0.87	0.93	0.84	0.83	0.81
Pass count- Mean (SD)	41.2 (4.2)	45.3 (4.2)	40.0 (4.5)	31.0 (3.4)	31.8 (2.9)	28.5 (3.0)
% Stations passed Mean (SD)	89.6 (9.2)	90.5 (8.3)	88.9 (10.1)	88.6 (9.6)	90.6 (8.4)	89.1 (9.4)

Table 2: Associations of key factors with probability of passing OSCE stations

		OR (95%CI), p-value
Academic Year	2009-10	1.15 (0.80, 1.65),p=0.443
	2010-11	1.33 (0.94, 1.89),p=0.109
	2011-12	1.18 (0.82, 1.69),p=0.377
	2012-13	1.00 (reference)
	2013-14	1.27 (0.88, 1.84),p=0.208
	2014-15	1.07 (0.74, 1.56),p=0.718
Gender	Male	0.72 (0.65, 0.79),p<0.001
	Female	1.00 (reference)
Day	per day	1.20 (1.17, 1.23),p<0.001
Time of Day	pm	0.99 (0.88, 1.10),p=0.795
	am	1.00 (reference)

Model includes random effects for student, circuit, and station, plus fixed effects for all row variables

OR: Odds Ratio

Table 3: Mean day on which exam stations were taken, according to students' final scores, with p-values from tests of (Pearson) correlation.

		Grade Achieved					Correlation, p-value
		A	B	C	D	E/F	
2009-10	N Mean Day	16 2.63	122 2.50	67 2.47	21 2.38	12 2.70	-0.008, p=0.897
2010-11	N Mean Day	148 2.97	78 2.94	21 2.85	11 2.85	4 3.09	0.052, p=0.402
2011-12	N Mean Day	96 2.53	88 2.55	26 2.47	12 2.59	4 2.73	-0.022, p=0.745
2012-13	N Mean Day	109 3.22	86 3.15	44 3.11	12 3.14	10 3.15	0.071, p=0.252
2013-14	N Mean Day	127 3.14	85 2.99	28 3.08	12 2.70	7 2.82	0.151, p=0.015
2014-15	N Mean Day	69 3.27	93 3.17	44 3.02	36 2.88	17 2.82	0.222, p=0.0003