



Does Changing Examiner Stations During UK Postgraduate Surgery Objective Structured Clinical Examinations Influence Examination Reliability and Candidates' Scores?

Peter A. Brennan, MD, FRCS,* David T. Croke, PhD,[†] Malcolm Reed, MD, FRCS,[‡] Lee Smith,* Evan Munro, FRCS,* John Foulkes, PhD,* and Richard Arnett, PhD[†]

*Intercollegiate Committee for Basic Surgical Examinations, The Royal College of Surgeons of England, London, United Kingdom; [†]Department of Quality Enhancement, The Royal College of Surgeons in Ireland, Dublin, Ireland; and [‡]Dean, Brighton Medical School, Brighton, United Kingdom

OBJECTIVE: Objective structured clinical examinations (OSCE) are widely used for summative assessment in surgery. Despite standardizing these as much as possible, variation, including examiner scoring, can occur which may affect reliability. In study of a high-stakes UK postgraduate surgical OSCE, we investigated whether examiners changing stations once during a long examining day affected marking, reliability, and overall candidates' scores compared with examiners who examined the same scenario all day.

DESIGN, SETTING, AND PARTICIPANTS: An observational study of 18,262 examiner-candidate interactions from the UK Membership of the Royal College of Surgeons examination was carried at 3 Surgical Colleges across the United Kingdom. Scores between examiners were compared using analysis of variance. Examination reliability was assessed with Cronbach's alpha, and the comparative distribution of total candidates' scores for each day was evaluated using *t*-tests of unit-weighted *z* scores.

RESULTS: A significant difference was found in absolute scores differences awarded in the morning and afternoon sessions between examiners who changed stations at lunch-time and those who did not ($p < 0.001$). No significant differences were found for the main effects of either broad content area ($p = 0.290$) or station content area ($p = 0.450$). The reliability of each day was not affected by examiner switching ($p = 0.280$). Overall, no difference was

found in *z*-score distribution of total candidate scores and categories of examiner switching.

CONCLUSIONS: This large study has found that although the range of marks awarded varied when examiners change OSCE stations, examination reliability and the likely candidate outcome were not affected. These results may have implications for examination design and examiner experience in surgical OSCEs and beyond. (J Surg Ed 73:616-623. © 2016 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: objective structured clinical examination, surgery, reliability, examiner, scenario

COMPETENCIES: Medical Knowledge, Practice-Based Learning and Improvement, Interpersonal and Communication Skills, Professionalism, Systems-Based Practice

INTRODUCTION

Undergraduate and postgraduate surgical examinations often utilize objective structured clinical examinations (OSCE) in an attempt to minimize variability and possible examiner bias as well as providing a consistent series of items and tasks for each candidate. The organization and execution of a successful OSCE needs considerable planning and knowledge of examination systems, and collaboration and effective interaction between the organizing Institute and examiners, with many of them undertaking the role in their own time.¹

A study of undergraduate medical OSCE students found no evidence that the duration of examining in a communication OSCE station influenced examiners and the marks

Correspondence: Inquiries to Peter A. Brennan, Maxillofacial Unit, Queen Alexandra Hospital, Portsmouth, PO6 3LY, UK; fax: (44) 2392-286089; e-mail: Peter.brennan@porthosp.nhs.uk

they awarded.² However, McLaughlin et al.³ reported that the point of entry to an OSCE circuit was significantly associated with scoring and could be a factor that may compromise the reliability of the marks awarded and the internal validity of an OSCE examination. Removing the first 2 stations from a candidates' final scoring in an attempt to eliminate examiner "warm up" did not influence this so-called differential rating over time, an effect that might be due to examiner fatigue as the OSCE continues.³

It is well known that examiner stringency or leniency, colloquially known as "hawk" and "dove" behavior, might also influence the mark awarded in any particular OSCE station.⁴⁻⁷ This can be minimized by pairing of examiners in those examinations or stations that are dual manned.⁶ However, with a comprehensive examination containing many different stations it is unlikely that this behavior influences the overall outcome for candidates, with these effects essentially canceling each other out.

In this time of increasing scrutiny of examinations, and with recent issues of potential bias around high-volume postgraduate medical examinations, an understanding of the potential role of these factors is needed.⁸⁻¹⁰ Candidates presenting themselves for examinations want reassurance of examination validity and that the processes are fair and unbiased.¹¹

The OSCE (part B) of the Intercollegiate Membership of the Royal College of Surgeons (MRCS) examination, a practical entry-level examination needed to enter higher surgical training in the United Kingdom and Ireland, consists of 18 stations (10 skills and 8 knowledge), each lasting for 9 minutes. This is a high-volume postgraduate examination delivered by the 4 Royal Colleges of Surgeons in the United Kingdom and Ireland with more than 1500 candidates taking it each year.¹² There are 2 broad content areas of knowledge and skills. Knowledge stations include 3 anatomy, 3 physiology or critical care scenarios, and 2 pathology scenarios. The clinical skills section of the OSCE includes 4 clinical examination stations, 4 communication stations, and 2 procedural stations.¹² For the purposes of quality assurance, standard setting statistics and the

establishment of the pass mark, these stations are interpreted as 11 distinct groups (Table 1).

A single examiner is present in most stations but 3 communication skills stations have 2 examiners each, with a final agreed mark being awarded. In all stations, there is an instruction sheet that candidates read to "set the scene" and once in the station the examiners have a pre-prepared list of questions to ask, with marks allocated throughout using a scoring checklist. There is a 1-minute break between candidates to assign the marks and prepare for the next candidate. A cohort of examiners typically works for up to 3 days at a time.

Each circuit lasts 180 minutes with a 20-minute break and a typical day involves 2 sessions of the examination. As the examination takes place for more than a variable number of days at up to 4 sites in the United Kingdom and Ireland, a standard blueprint is used for each examination and drawn from a bank of scenarios all of which are initially piloted. Stringent quality assurance protocols ensure that only questions with acceptable performance statistics are utilized and psychometric analysis is undertaken at individual question, examiner, and venue and session for each examination that is held.

It is recognized that human factors including repetition, tiredness, and boredom may influence examiner behavior during OSCE circuits, with MRCS examiners often having to stay in the same OSCE station all day.¹³

Several of the communication stations (such as history taking) require minimal examiner interaction, thereby increasing the likelihood of fatigue and disengagement. In response to feedback from examiners, it was agreed that examination departments could allow MRCS examiners to change stations at lunchtime, enabling a different scenario to be examined in the afternoon session. This option was not adopted uniformly by all examination departments, with some continuing to require examiners, wherever possible, to examine at the same station all day.

In the present study of more than 18,000 candidate-examiner interactions, we investigated whether marks awarded were comparable between examiners changing OSCE stations at lunchtime after a complete OSCE circuit (18 candidates examined), and those remaining in the same scenario all day. We assessed the effect of this examiner switching on station reliability and the overall scores that candidates obtained per examining day.

TABLE 1. The 11 Main Content Areas of the MRCS OSCE Examination (in 2 Main Groups). Some Areas Are Assessed in More Than 1 Station as Indicated

Broad Content Area	Content Area	Number of Stations
Knowledge	Anatomy	3
	Surgical pathology	2
	Data interpretation	2
	Critical care	1
Skills	Giving and receiving information	2
	History taking	2
	Physical examination	4
	Procedural skills	2
		18
Total		

MATERIALS AND METHODS

In the absence of a specific ethical committee responsible for postgraduate surgical examinations, the Intercollegiate Committee for Basic Surgical Examinations and Quality Assurance Committee approved the study. Data were collected from 3 examination periods of the MRCS part B OSCE during 2014 and 2015. Each scenario was given a score (of up to 20) by a single examiner in 15 stations, and in the 3 communication skills stations in which both a

Download English Version:

<https://daneshyari.com/en/article/4297519>

Download Persian Version:

<https://daneshyari.com/article/4297519>

[Daneshyari.com](https://daneshyari.com)