

Bond University
Research Repository



A snapshot of current Objective Structured Clinical Examination (OSCE) practice at Australian medical schools

Heal, Clare; D'Souza, Karen; Banks, Jennifer; Malau-Aduli, Bunmi S.; Turner, Richard; Smith, Jane; Bray, Ebony; Shires, Lizzi; Wilson, Ian

Published in:
Medical Teacher

DOI:
[10.1080/0142159X.2018.1487547](https://doi.org/10.1080/0142159X.2018.1487547)

Licence:
Other

[Link to output in Bond University research repository.](#)

Recommended citation(APA):
Heal, C., D'Souza, K., Banks, J., Malau-Aduli, B. S., Turner, R., Smith, J., Bray, E., Shires, L., & Wilson, I. (2018). A snapshot of current Objective Structured Clinical Examination (OSCE) practice at Australian medical schools. *Medical Teacher*, 41(4), 441-447. Advance online publication. <https://doi.org/10.1080/0142159X.2018.1487547>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

For more information, or if you believe that this document breaches copyright, please contact the Bond University research repository coordinator.

A snapshot of current Objective Structured Clinical Examination (OSCE) practice at Australian medical schools

Abstract:

Objective: The objective of this collaborative study was to compare current practices of conducting high-stakes, exit-level Objective Structured Clinical Examinations (OSCEs) at all Australian medical schools. We aimed to document similarities and differences between schools, and compare existing practice against available gold standard, evidence-based practice. We also aimed to identify areas where gold standards do not currently exist, and could be developed in the future.

Methods: A 72-item semi-structured questionnaire was sent to all 19 Australian medical schools with graduating students.

Results: A total of 18/19 schools responded. Of these, 16/18 schools had summative exit OSCEs representing content from multiple medical specialties. The total number of OSCE stations varied from 8 to 16, with total OSCE testing time ranging from 70 to 160 minutes. All schools blueprinted their OSCE to their curriculum, and trained simulated patients and examiners. There was variation in the format of marking rubric used.

Conclusions: This study has provided insight into the current OSCE practices of the majority of medical schools in Australia. Whilst the comparative data reveals a wide variation in OSCE practices between schools, many recommended 'gold standard' OSCE practices are implemented. The collective awareness of our similarities and differences provides us a baseline platform, as well as an impetus for iterative quality improvement. Such discourse also serves to develop new gold standards in practice where none have previously existed.

Keywords: OSCE, clinical examination, medical education, Australian medical schools.

Introduction

Objective Structured Clinical Examinations (OSCEs) are used by all medical schools in Australia in order to assess clinical performance.

The criteria for good assessment, as outlined by the Ottawa 2010 Conference consensus statement (Norcini et al. 2011) should take into account the perspectives of all relevant stakeholders, including examinees, patients, teachers, educational institutions, healthcare systems and regulators, as part of a multi-source validation process (Messick 1989; Downing 2003).

These overarching principles have been distilled for the purpose of OSCEs by AMEE Guide 81 (Khan, Ramachandran, et al. 2013), which may be considered a gold standard, or best practice for those who administer this form of assessment.

An OSCE is thus defined as ‘an assessment tool based on the principles of objectivity and standardisation in which the candidate moves through a series of time-limited stations in a circuit for the purposes of assessment of professional performance in a simulated environment. At each station, candidates are assessed and marked against standardised scoring rubrics by trained assessors.’ The OSCE assesses performance at the ‘shows how’ level of Miller’s pyramid (Miller 1990) , and is most appropriately used alongside other assessment methods (Hamdy et al. 2010). Student responses to OSCE stations provide a reliable and valid means of assessing the clinical skills that they have acquired in their training, which will be vital for competence in their future careers.

The OSCE was originally described by Harden in 1975 (Harden RM et al. 1975), and was designed to improve the validity and reliability of the assessment of clinical performance. Until that time, clinical performance was assessed using long and short cases. The original OSCE, described by Harden, involved 18 test and two rest stations, each station being 4.5 minutes long with a 30 second break between each and a total examination time of 100 minutes (Harden RM et al. 1975). In contrast to common contemporary OSCE practice of integrating multiple clinical skills, the 1975 OSCE stations were designed to each test a single competency.

There is much in the medical education literature around OSCE theory (Swanwick 2014; Harden R et al. 2016). However, when OSCE faculty meet for workshops, conferences and collaboration meetings (nationally and internationally), questions remain about many of the finer details of how OSCEs are conducted *in practice* at each site – with faculty continuously striving to improve the quality of OSCEs conducted at their own medical school. To this end, the Australian Collaboration for Clinical Assessment in Medicine (ACCLAiM) was developed as an innovative venture between fifteen medical schools in Australia, using the OSCE to benchmark graduate outcomes and provide quality assurance in the assessment of medical students’ clinical performance. The collaboration grew from four participating medical schools in 2011 to 15 (out of the total 19 medical schools) in 2015. By August 2017, 56 collaboratively developed OSCE stations had been developed and embedded in summative exams in participating medical schools (Malau-Aduli et al. 2016). Whilst the Australian Medical Council (AMC) provides accreditation to each medical school within Australia, schools are permitted local differences in curricula, course duration and student entry requirements (graduate entry versus school leaver entry). National benchmarking provided by ACCLAiM promotes critical reflection on curriculum areas that potentially need greater emphasis or development. The principal outcomes of the project have been the development of a learning community, and an academic standards framework to compare and monitor students’ clinical competence across Australian medical schools. The project has also aided the sharing of assessment materials in a collaborative and flexible manner, and fostered the evaluation and quality assurance of clinical assessment practices.

Drawing upon the experience of ACCLAiM, this study aimed to compare current OSCE practices of Australian medical schools conducting high-stakes exit OSCEs by providing a ‘national snapshot’. Our definition of “exit” OSCE was a summative

examination covering multiple generalist and specialist clinical areas and administered to students in the final 2 years of the medical program as a final clinical examination. We sought to document similarities and differences in practice to available gold standards and use this as an impetus for quality improvement. We also aimed to identify areas where gold standards do not currently exist, and could be developed in the future in response to the results.

Methods

Sample

Fifteen geographically dispersed Australian medical schools participate in the ACCLAiM collaboration. All participating schools have similar horizontally and vertically integrated outcomes-based curricula, accredited by the AMC.

There are a further five medical schools in Australia in addition to the 15 ACCLAIM collaborators, one of which is a new school and yet to reach the stage of exit examinations. Therefore, only the 19 Australian medical schools with graduating students were invited to participate in this study.

Ethics

The project was approved by the James Cook University Human Research Ethics Committee (approval number H5595).

Questionnaire development

A focus group consisting of 14 ACCLAiM collaborators from 10 different medical schools was held in August 2016. This group nominated topics that they considered should be included in the questionnaire. The seven-page questionnaire was designed by one author (JB) with assistance from other authors (CH, KD, BMA), using current guidelines and literature to ensure content validity. The questionnaire development group comprised representatives of the three of the four original ACCLAiM schools. The questionnaire focused on exit level OSCEs - this was chosen because of comparable levels of intended graduate outcomes nationally.

The questionnaire was pilot tested in January 2017 by lead academics from three medical schools participating in ACCLAIM, and further refined prior to distribution, based on feedback. The questionnaire aimed to collect data on OSCE settings, structure, process and content. We considered AMEE guide no 81 (Khan, Gaunt, et al. 2013; Khan, Ramachandran, et al. 2013) to be the gold standard or best practice for the purposes of comparison.

Procedure

One key representative from each medical school was nominated to receive a link to the online questionnaire (Survey Monkey, San Mateo, CA, United States of America). Data was collected over one month (January to February 2017). The data collected was collated by JB and CH. Data clarification was performed by CH. The collated data was

reviewed in order for queries about data, to be clarified with the relevant school if required.

Results

Participation

Eighteen of the nineteen eligible Australian medical schools participated in this study.

Setting

Of the 18 schools, 11 schools reported having a 4-year course, 5 schools reported a 5-year course, and 2 schools reported a 6-year course. The Exit OSCE runs at various times within the course, from the end of year 5 of a 6-year course to the end of the final year of a 4- or 5-year course, and the end of year 3 of a 4-year course. Seven of the 18 schools were undergraduate entry, 10 graduate entry and one both graduate and undergraduate entry courses.

While 6 of 18 schools reported conducting their exit exams at one site, the other 12 schools described multiple sites, with the maximum being 8 different sites running the OSCE concurrently. The total number of students sitting the exam on any one occasion varied from 80 to 500.

Exam structure

The majority of schools have an official exit examination (defined as covering multiple generalist and specialist clinical areas, placed in the final 2 years of the program).

Two of the schools do not have an official multiple medical specialities exit examination, but have speciality specific OSCE examinations with 4-5 stations each in the final two years of their respective courses. For the remaining 16 schools with exit exams, the number of stations varied from 8 to 16 (median 12), with the length of stations varying between 5 minutes and 10 minutes (median 8mins). However, most (10/16, 62.5%) have 8-minute stations. Total OSCE testing time ranged from 70 to 160 minutes (median 96 mins) (Table 1). Most schools had 2-3 rest stations. One school programmed a 10-minute reading time at the beginning of the examination. All stations allowed reading time outside each individual station before entering the room, although this varied from 1 minute to 4 minutes (median 2 mins).

Exam content

In terms of blueprinting, all 18 schools mapped their OSCE against their own curriculum. In addition, seven mapped against the Australian Medical Council (AMC) Graduate Outcome Statements (Australian Medical Council 2012), three against the Australian Curriculum Framework for Junior Doctors (Confederation of Postgraduate Medical Councils 2012), three against the Medical Deans of Australia and New Zealand (MDANZ) Competency Projects final reports (Carmichael and Hourn 2011; Hays and Hourn 2012, 2014), and three against some combination of these. Content for the OSCE station was developed by faculty clinicians (11/18, 61%) and discipline heads (5/18, 27.7%), with course co-ordinators (2/18, 11%) and medical education unit (MEU) assessment staff also used (1/18, 5.5%). The majority of schools (13/18, 72%) included procedural skills in the OSCE content. (Table 2)

Critical errors (i.e. a single error resulting in the student failing the station overall) were not used by any of the participating schools.

Exam process - Students

There were varying controls put into place regarding personal digital and electronic devices. The most commonly excluded items were: mobile telephones 18/18 (100%), recording devices (17/18, 94.4%), smart watches (12/18, 66.6%) and fit-bands (11/18, 61.1%), although some institutions did not allow pens (5/18, 27.7%) or watches (3/18, 16.6%), out of concern for concealed recording devices. All schools utilised student sequestration to avoid students communicating exam content with each other. Students were allowed to write their own notes on paper outside stations during reading time in 9/18 (50%) schools, however none of this material was examinable. Most schools provided students with feedback, (15/18, 83.3%), mostly in the form of debriefing of station content (7/13, 53.8%) or written feedback on individual student performance in comparison to peers (10/13, 76.9%). The timing of the provision of this feedback varied from on the day of the OSCE to 4 weeks following the OSCE.

Exam process - Examiners

All schools had either one examiner, or a combination of one and two examiners per station. General practitioners and other specialists were used at all schools as examiners. Registrars (postgraduate year 4 or higher - 9/18 (50.0%) schools), junior doctors (9/18, 50.0%) and medical students (1/18, 5.5%) were used as examiners at some universities. Medical students were only used as examiners at a single school where the exit OSCE was done before the final year of the course, and only senior students were used in procedural stations only. Examiners were always aligned with their medical specialty at 4/18 (22.2%) schools, while 2/18 (11.1%) schools deliberately did not align examiners with their medical specialty.

Examiners were provided training through briefings and workshops, as well as provision of documents and online training prior to the examination. Training addressed issues pertinent to the specific station (15/18, 83.3%), OSCE theory and process (13/18, 72.2%), general examiner calibration (10/18, 55.5%) and calibration to the specific station (8/18, 44.4%). In the majority of cases, examiners were not paid (12/18, 66.6%). Those schools who did pay examiners provided items ranging from gifts such as a bottle of wine to an hourly university-level pay rate in excess of \$100 per hour, if their examiner duties were conducted in their own private time. Feedback was provided to the examiners by the majority of schools (10/18, 55.5%), however the format and timing of this varied from site to site. Routine verbal examiner feedback on the day of the OSCE was given at only one school, while at one school verbal feedback was only provided if there were discrepancies noted. However several schools (5/18, 27.7%) provided written feedback comparing the examiner's mean scoring of candidates in their station compared to the average score awarded by all examiners in the same station. The time range for providing this feedback varied from on the day of the OSCE up to 1 year (prior to the next exit OSCE). The minority of schools (2/18, 11.1%) moved examiners to another site to ensure that examiners were not overly familiar with the students they were examining. Most schools (14/18, 77.7%) required that examiners name students in advance of the OSCE with whom they had an expected conflict of interest.

Exam process - Simulated Patients

Simulated patients used in the OSCEs were most commonly reported to be professional actors (15/18, 83.3%) or members of the general public (11/18, 61.1%), while staff (7/18, 38.8%) and medical students (3/18, 16.6%) were also utilised. Simulated patients were also sourced from community volunteers (3/18, 16.6%), patients of clinical staff (2/18, 11.1%) and international elective students (1/18, 5.5%). While all schools trained their simulated patients, a variety of trainers were used including clinical educators (6/18, 33.3%), discipline heads (4/18, 22.2%), simulation co-ordinators (2/18, 11.1%), clinical skills nurses, station co-ordinators, assessment staff and even a performing arts academic. Training time ranged from 15-30 minutes to in excess of 3 hours and was generally undertaken prior to the exam (15/18, 83.3%) as well as on the day of the exam (7/18, 38.8%). The simulated patients were paid in a variety of ways ranging from reimbursement of travel costs and \$20 gift cards, to hourly rates between \$20 and \$50, and one school paying up to \$800 per day for professional actors. Feedback was provided to simulated patients by 10/18 (55.5%) of schools, usually on the day of the OSCE. This feedback was usually provided by the examiners, in the context of the role-play being inconsistent/of poor quality. One school reported the actor trainer observing all standardised patients and providing a rating and qualitative feedback on their role-play performance.

Simulated patients were asked to provide an independent rating of students' global performance at 4/18 (22.2%) of schools, particularly as pertains to patient-centred communication.

Where real patients were used in the OSCE, a variety of different physical signs were represented. (Table 3)

Standard setting

Schools most commonly used the borderline regression method (11/18, 61.1%), followed by borderline groups (5/18, 27.7%) and Angoff (3/18, 16.6%) methods. One school used a domain-based method, while Ebel and Hofstee were not used at any school for standard setting of OSCEs.

Marking sheets

Six medical schools (6/18, 33.3%) reported using only competency-based mark sheets, and 6/18 (33.3%) reported using only checklist mark sheets. The remaining 6/18 (33.3%) universities reported using a combination of both a checklist and competency-based mark sheet (2/18), or using an additional global rating scale (4/18). Examiners at 4/18 (22.2%) medical schools performed their marking using an iPad or tablet device, with the remainder completing paper mark sheets. The minority of medical schools reported that standardised patients' marks contribute to the students' station marks (4/18, 22.2%).

Exam analysis

Analysis of the OSCE results was reported as most commonly performed by members of academic staff (16/18, 88.8%), followed by psychometricians (6/18, 33.3%) and administrative staff (4/18, 22.2%). Sequential test format was used at 4/18 (22.2%) schools (Pell et al. 2013) Most schools used Microsoft Excel (Redmond, WA, United States of America - 7/18 (38.8%) schools) to perform their OSCE analysis, with SPSS Statistics (IBM, Armonk, NY, United States of America - 5/18 (27.7%) schools) being the second most commonly used software. Commercially-available, electronic exam

management software (EMS) was infrequently used by medical schools (3/18 (16.6%) schools).

Discussion

A total of 18 of the 19 Australian Medical Schools with graduating cohorts participated in this study, serving as an important example of national collaboration and collegiality. Due to the small sample size, the results are mainly descriptive. However, we feel that the findings are useful and allow Australian medical schools to compare their practices to other medical schools nationally, as well as with gold standard OSCE practice. International medical schools and other health professional courses will also find these ‘national snapshot’ results informative when compared to their own practices. For the purposes of our discussion, we have considered AMEE guide no 81 (Khan, Gaunt, et al. 2013; Khan, Ramachandran, et al. 2013), which condenses other sources of literature, to be the gold standard.

OSCE validity

Khan et al. (2013a) summarise the five sources of evidence for construct validity, which are of relevance to the OSCE. Our study explored three of these (test content, responses, and consequences). These are discussed in more detail below:

i) *The test content represents what the curriculum needs to assess, the tasks are realistic and the right domains are being assessed.*

Blueprinting is the process of determining the content and spread of skills included in an OSCE. The blueprint should take into account the context of the examination and should be mapped to the curriculum and other relevant materials (Khan, Gaunt, et al. 2013). All medical schools surveyed reported using a formal blueprinting process, with a broad range of guidelines used as references for blueprinting – from each school’s curriculum to national clinical curriculum guidelines.

ii) *The responses to the test items are accurately recorded, handled, stored and analyzed.*

It was interesting that dedicated psychometricians analyzed results at only a third of schools, although there is no recommendation that a psychometrician is required.

Standard setting is an integral component of results analysis. The recommended methods for OSCE standard setting are the borderline group, borderline regression or Cohen methods (Malau-Aduli et al. 2017) It was notable that three schools reported using the Angoff method of standard setting for the exit exam, which may not be appropriate in an OSCE setting (Khan, Gaunt, et al. 2013).

Although at the time of the questionnaire only three schools were using EMS, this area is rapidly evolving, and several schools have since reported adopting the use of an EMS.

iii) *The consequences of assessment are sound, i.e. the effect on learning is positive*

While most schools provided students with feedback, this was given immediately in less than half of cases. The educational impact of the OSCE is increased by immediate feedback, (Boursicot 2010) and our group recommends that lifelong learning would be increased by adopting this process.

OSCE reliability

Several factors may effect OSCE reliability, which include the number of stations and testing time, standardised rubrics and methods for standardising examiners and simulated patients.

i) The number of stations

The reliability and the validity of an OSCE are influenced by both the number of stations within each OSCE and the total length of the examination (Newble D 2004). Longer stations aid the validity of the content, while an increased number of stations (i.e. testing on a larger number of clinical cases) improves reliability (Roberts et al. 2006). Decisions around test length need to balance reliability coefficients (Cronbach alpha or G value) with feasibility and resource issues. The optimal number of stations for an examination therefore depends on the requirements and resources of an individual institution. However, as a general recommendation it is suggested that 14-16 x 5-10 min stations may provide adequate reliability (Epstein 2007). This figure is quite broad, with total testing time potentially ranging from 80 to 180 minutes. However, 2/16 (12.5%) schools reported a total testing time of less than 80 minutes, and 9/16 (56.0%) schools were less than the 100 minutes of total testing time that was first described by Harden (Harden RM et al. 1975). Only 3/16 (18.8%) schools ran 14 or more stations.

ii) Standardised scoring rubrics, to ensure that all examiners are marking against the same criteria for all candidates (Smee 2003).

A marking rubric is a scoring tool that lays out the specific expectations for a task or assignment, dividing it into component parts and providing a detailed description of what constitutes acceptable and unacceptable levels of performance (Hunter 1996)

While rating scales are categorized into analytic or holistic (Khan, Gaunt, et al. 2013), in reality, they exist on a continuum between two extremes. On one extreme, analytic scoring may comprise a 'checklist', which describes whether specific tasks have been performed as a binary variable (yes/no or performed/not performed). Alternatively, checklists can have a 5-7 point rating scale, which allows examiners to discriminate in the quality of the actions performed. Competency-based scales consist of descriptors for the level of performance of a student task, which may have three to five descriptors, usually incorporating descriptors for performances at fail, borderline fail, borderline pass, pass and exceptional level. On the other extreme, holistic scoring or 'global rating scales' allow the assessor to rate the process and content skills demonstrated by the student as a whole.

However there is discourse in the literature regarding the best type of rubric. Binary checklists were originally perceived to provide an objective assessment and to be more suitable for novice or non-physician examiners (Humphrey-Murto et al. 2005), although objectivity may not be related to increased inter-rater reliability (Wilkinson et al. 2003). It has also been suggested that the educational impact of the OSCE is

increased if it represents real life practice, rather than reducing clinical skills to checklists to be scored against (Miller 1990; Shumway et al. 2003; Boursicot 2010). Global scores have been shown to have greater reliability than checklist rubrics (Hodges and McIlroy 2003; Turner and Dankoski 2008). It has been proposed that holistic scales are more useful for assessing areas such as judgement, empathy, organisation of knowledge and technical skills (Morgan et al. 2001; Hodges and McIlroy 2003).

Considering this discourse, it is not surprising that there was variation in the marking sheets, with close to an equal division between schools using exclusively checklist-based marking sheets versus exclusively competency-based mark sheets. Some schools reported using a mixture or 'hybrid' of the checklist- versus competency-based mark sheets, and four universities also reported using additional global scoring. The group recommends that mark sheets in OSCE assessment would benefit from clearer definition of the rating scale used, and clearer recommendations regarding gold standard.

iii) Using trained examiners

Use of examiner training has been shown to reduce examiner variation in scoring (Newble et al. 1980; van der Vleuten et al. 1989; Epstein 2007). It is therefore appropriate that all medical schools reported providing examiner training. However, the number of hours dedicated to such training, and the content of the training varied at institutions, including the development of online training by a few institutions. This may be an area which would benefit from the development of standardized guidelines.

iv) Standardized patient performance

Poorly standardized simulated patients have been shown to reduce reliability (Nestel et al. 2008), and it was of note that although all simulated patients were given station-specific briefing for their specific roles, the time allocated for this varied from 15 minutes to 3 hours. Only 1/18 schools offered dedicated training in the improvisation skills required for medical simulation. Future work should focus on an optimal standard of training, both in length and content, for simulated patients. More work could also be done on giving simulated patients a voice in the assessment of future health providers. Evidence suggests that when adequately trained, simulated patients are capable of providing valid assessments of students' communication skills in OSCEs (Shirazi et al. 2014), representing the views of the community stakeholders in assessing their future doctors.

OSCE Feasibility

Khan et al (2013) stated that when compared to other assessment methods, the OSCE is seen as more resource intensive and time-consuming to set-up and run. It is therefore important to ensure that only stations that require an OSCE format are implemented to assess performance, while other content knowledge is tested in less resource-intensive examinations such as multiple choice question paper. The schools participating in this research blueprint their OSCE stations against national clinical skills frameworks, and often include the testing of procedural skills as OSCE stations.

This research highlights that variability in OSCE practice exists in the Australian context, where accreditation specifies graduate outcome statements and medical schools have various training pathways and settings.. Our study may serve as a useful tool for medical and health professional schools conducting OSCEs to properly consider feasibility in their own context.

Regarding exam security, so-called ‘smart devices’ are increasingly available to the student population. There was a notable variability in the extent to which schools excluded different devices. This may need to be reviewed to become increasingly restricted as more sophisticated devices become available; this is an area where a gold standard guideline could be developed. All universities reported some form of sequestering, during which time mobile phones are removed, to prevent leakage of information to fellow students. Interestingly, one of the universities removed the quarantine system previously, but reinstated it in response to demonstrated security lapses.

The results of our questionnaire showed that Australian medical schools agreed with each other, and aligned with AMEE gold standards, in the areas of blueprinting exam content and standard setting. However there was great variation in the number and length of OSCE stations used, and in the total OSCE testing time, which in some cases fell below the 80 minutes suggested in gold standard guidelines. There was also a great variation in the type of mark sheet used, and although there is discourse in the literature regarding the ideal marking rubric, generating a standardised mark sheet which could be used by all universities will be a subject of future collaboration and research for the ACCLAIM group. Although all schools trained simulated patients and examiners, the time and methods used varied; this is another area where standardised guidelines and collaborative resources could be developed. Whilst this research study is performed on medical schools in Australia, the findings may be generalisable to other medical and health professional postgraduate training institutions responsible for delivering OSCEs – both within Australia and internationally.

Conclusion

This study has highlighted areas of congruence among Australian Medical schools, in relation to OSCE assessment and these include blueprinting to the curriculum, standard setting, training of examiners and simulated patients. However, wide variations still exist in the timing and methods of training as well as the number and length of OSCE stations and total OSCE testing time.

The ACCLAIM collaboration has provided insight into the practices of nearly all Australian medical schools with respect to high-stakes OSCE assessments. Collective awareness of similarities and differences to gold standards provides an impetus for iterative quality improvements. Such discourse also serves to develop gold standards in practice where none have previously existed. Our recommendations are that standardised guidelines for the training of simulated patients and examiners are required, as well as a standardised marking sheet for the exit OSCE. Development of these gold standards is an area of future research for the ACCLAIM group.

Acknowledgements

The authors would like to thank administrative and academic staff from all participating schools

Declaration of interest: Non declared

Notes on Contributing Authors:

Dr Clare Heal MBChB, DRANZCOG, DipGUMed, MPHTM, PhD, is Professor of General Practice and Rural Medicine at James Cook University in Mackay, Queensland. In addition to her research activities, she is chair of Year five assessment and co-ordinator of general practice for JCU. She is a member of an expert exam item-writing panel for the Australian Medical Council and is co-chair of ACCLAIM.

Karen D'Souza MBBS (Hons), leads Clinical Curriculum and Assessment at the School of Medicine, Deakin University. She chairs the School's OSCE Committee, and is Co-Chair of the ACCLAiM Collaboration (Australian Collaboration for Clinical Assessment in Medicine). Her areas of research interest include OSCEs, simulation, selection, clinical and communication skills.

Dr Jennifer Banks PhD, MBS, BBS, is a senior research officer with College of Medicine and Dentistry, James Cook University.

Bunmi S. Malau-Aduli, BSc, MSc, GradCertMgt, GradCertULT, PhD, is an Associate Professor in Medical Education and the Academic Lead for Assessment and Evaluation at the College of Medicine and Dentistry, James Cook University, Queensland, Australia.

Richard Turner MBBS, BMedSC, FRACS, is Professor of Surgery in the College of Health and Medicine, University of Tasmania. He is a practising General Surgeon, whose research interests include anal neoplasia, pancreatitis, men's health, and clinical assessment. He has had a number of examiner roles with the Royal Australasian College of Surgeons and General Surgeons Australia.

Associate Professor Jane Smith MBBS, FRACGP, MHS, is an Associate Professor of General Practice and year 5 Head, Bond University Medical Program. Jane became a full time academic after years as a GP. As well as educating medical students in General Practice, she has been responsible for many Clinical Year Examinations, creating research interests in OSCE assessments, medical education and general practice.

Ebony Bray is a 6th year MBBS student with James Cook University, College of Medicine and Dentistry.

Lizzi Shires MBBS, FRACGP, is a Rural GP on the North West Coast of Tasmania and Director of the Rural Clinical School, University Tasmania. She is currently chair of Final Year with responsibility for exit examinations.

Professor Ian Wilson, Director of Learning and Teaching, School of Medicine,

University of Wollongong. MBBS; PhD; FRACGP; FAMEE. Following graduation from University of Adelaide, Professor Wilson became a GP. He has always had a significant interest in medical education and over the last 12 years has worked almost fulltime in education. Current research interests include student selection, professionalism and professional identity formation.

Conflicts of interest: none

A rubric is a scoring tool that lays out the specific expectations for a task or assignment, dividing it into component parts and providing a detailed description of what constitutes acceptable and unacceptable levels of performance (Hunter 1996)

Rubrics may be categorised into analytic or holistic (Khan, Gaunt, et al. 2013), in reality, they exist on a continuum between two extremes. Analytic scoring may comprise a 'checklist', which describes whether specific tasks have been performed as a binary variable (yes/no or performed/not performed). On the other extreme, holistic scoring or 'global rating scales' allow the assessor to rate the process and content skills demonstrated by the student as a whole.

References:

- Australian Medical Council. 2012. Standards for Assessment and Accreditation of Primary Medical Programs by the Australian Medical Council 2012. Australian Medical Council Limited.
- Boursicot KA. 2010. Structured assessments of clinical competence. *Br J Hosp Med (Lond)*. 71(6):342-344.
- Carmichael A, Hourn M. 2011. Developing a Framework of Competencies for Medical Graduate Outcomes. Medical Deans Australia and New Zealand Incorporated.
- Confederation of Postgraduate Medical Councils. 2012. Australian Curriculum Framework For Junior Doctors. [accessed 2017]. <http://curriculum.cpmec.org.au/index.cfm>.
- Downing SM. 2003. Validity: on meaningful interpretation of assessment data. *Med Educ*. 37(9):830-837.
- Epstein RM. 2007. Assessment in medical education. *N Engl J Med*. 356(4):387-396.
- Hamdy H, Telmesani AW, Wardy NA, Abdel-Khalek N, Carruthers G, Hassan F, Kassab S, Abu-Hijleh M, Al-Roomi K, O'Malley K et al. 2010. Undergraduate medical education in the Gulf Cooperation Council: a multi-countries study (Part 2). *Med Teach*. 32(4):290-295.
- Harden R, Lilley P, Patricio M. 2016. The definitive guide to the OSCE : the Objective Structured Clinical Examination as a performance assessment. Edinburgh ; New York: Elsevier.
- Harden RM, Stevenson M, Downie WW, Wilson GM. 1975. Assessment of clinical competence using objective structured examination. *Br Med J*. 1(5955):447-451.
- Hays R, Hourn M. 2012. Medical Graduate Competency Framework Stage 2 Medical Deans Australia and New Zealand Incorporated; Health Workforce Australia.
- Hays R, Hourn M. 2014. Developing a National Assessment Blueprint for Clinical Competencies for the Medical Graduate. Medical Deans Australia and New Zealand Incorporated; Health Workforce Australia.
- Hodges B, McIlroy JH. 2003. Analytic global OSCE ratings are sensitive to level of training. *Med Educ*. 37(11):1012-1016.

- Humphrey-Murto S, Smee S, Touchie C, Wood TJ, Blackmore DE. 2005. A comparison of physician examiners and trained assessors in a high-stakes OSCE setting. *Acad Med.* 80(10 Suppl):S59-62.
- Hunter DJ, R; Randhawa, B. 1996. The Use of Holistic Versus Analytic Scoring for Large-Scale Assessment Of Writing. *The Canadian Journal of Program Evaluation* 11(2):61-85.
- Khan KZ, Gaunt K, Ramachandran S, Pushkar P. 2013. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: organisation & administration. *Med Teach.* 35(9):e1447-1463.
- Khan KZ, Ramachandran S, Gaunt K, Pushkar P. 2013. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. *Med Teach.* 35(9):e1437-1446.
- Malau-Aduli BS, Teague PA, D'Souza K, Heal C, Turner R, Garne DL, van der Vleuten C. 2017. A collaborative comparison of objective structured clinical examination (OSCE) standard setting methods at Australian medical schools. *Med Teach.* 1-7.
- Malau-Aduli BS, Teague PA, Turner R, Holman B, D'Souza K, Garne D, Heal C, Heggarty P, Hudson JN, Wilson IG et al. 2016. Improving assessment practice through cross-institutional collaboration: An exercise on the use of OSCEs. *Med Teach.* 38(3):263-271.
- Messick S. 1989. Validity. In: Linn R, editor. *Educational Measurement*. New York: American Council on Education and Macmillan p. 13-104.
- Miller GE. 1990. The assessment of clinical skills/competence/performance. *Acad Med.* 65(9 Suppl):S63-67.
- Morgan PJ, Cleave-Hogg D, Guest CB. 2001. A comparison of global ratings and checklist scores from an undergraduate assessment using an anesthesia simulator. *Acad Med.* 76(10):1053-1055.
- Nestel D, Tierney T, Kubacki A. 2008. Creating authentic simulated patient roles: working with volunteers. *Med Educ.* 42(11):1122.
- Newble D. 2004. Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ.* 38(2):199-203.
- Newble DI, Hoare J, Sheldrake PF. 1980. The selection and training of examiners for clinical examinations. *Med Educ.* 14(5):345-349.
- Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrott V et al. 2011. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach.* 33(3):206-214.
- Pell G, Fuller R, Homer M, Roberts T. 2013. Advancing the objective structured clinical examination: sequential testing in theory and practice. *Med Educ.* 47(6):569-577.
- Roberts C, Newble D, Jolly B, Reed M, Hampton K. 2006. Assuring the quality of high-stakes undergraduate assessments of clinical competence. *Med Teach.* 28(6):535-543.
- Shirazi M, Labaf A, Monjazebi E, Jalili M, Mirzazadeh M, Ponzer S, I. M. 2014. Assessing medical students' communication skills by the use of standardized patients: emphasizing standardized patients' quality assurance. *Academic Psychiatry.* 38:354 - 360.
- Shumway JM, Harden RM, Association for Medical Education in E. 2003. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Med Teach.* 25(6):569-584.
- Smee S. 2003. Skill based assessment. *Bmj.* 326(7391):703-706.
- Swanwick T. 2014. *Understanding medical education : evidence, theory, and practice*. Second edition. ed. Chichester, West Sussex: John Wiley & Sons Inc.

Turner JL, Dankoski ME. 2008. Objective structured clinical exams: a critical review. *Fam Med.* 40(8):574-578.

van der Vleuten CP, van Luyk SJ, van Ballegooijen AM, Swanson DB. 1989. Training and experience of examiners. *Med Educ.* 23(3):290-296.

Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T. 2003. Objectivity in objective structured clinical examinations: checklists are no substitute for examiner commitment. *Acad Med.* 78(2):219-223.

Table 1: Number of stations and total OSCE testing time

# of schools N=16*	# of OSCE stations	Duration of station (mins)	Total OSCE testing time
1	16	10	160
1	12	6 x 10 6 x 15	150
1	16	8	128
1	12	9	108
1	12 for top 40% + 6 for remaining 60%	6 6	108
1	13	8	104
1	10	10	100
4	12	8	96
3	10	8	80
1	9	8	72
1	14	5	70

*These were the 16 schools with multidisciplinary exit OSCEs

Table 2: Procedural skills included in OSCE content

Procedural Skill	Number of schools N=17* (%)
Did not use procedural skill	5 (29.4)
Suturing	5 (29.4)
Pap smear	4 (23.5)
Urinary catheter	2 (11.7)
Injection	2 (11.7)
ABG	2 (11.7)
Life support (basic)	1 (5.8)
Life support (advanced)	1 (5.8)
Resuscitate a dummy	1 (5.8)
STI screen	1 (5.8)

*One school did not answer this question

Table 3: Real patients used in OSCEs

Real Patients	Number of schools N=16* (%)
Do not use real patients	8 (50)
Heart murmur, aortic stenosis	5 (31.2)
Chronic Obstructive Pulmonary Disease, pulmonary fibrosis	4 (25)
Diabetic Limbs, peripheral neuropathy	3 (18.7)
Rheumatic Hands	1 (6.2)
Parkinsons Disease	1 (6.2)
Splenomegaly	1 (6.2)

* Two schools did not answer this question