

ABSTRACT NUMBER: OA-01

Can Peer Marking Replace Physicians As OSCE Examiners?

Marc P J McCafferty, Joseph T Garvin, Sarah McMahon, Karl J Sweeney, Carmel M Malone, Michael J Kerin

National University of Ireland, Galway, Ireland

Background: The role of the OSCE (Objective Structured Clinical Examination) format is now a well established, reliable and effective tool for assessing clinical skills and competence at undergraduate level. However, studies have criticised the limitation of the OSCE in assessing the integral approach to patients and complain that this format is both time and effort-consuming on the physician. The aim of this study was to assess the reliability and accuracy of peer examiners in evaluating student performance compared to the “gold standard” physician examiners.

Methods: This is a cross-sectional study of fourth year medical students undergoing a mid-term surgical OSCE. Sixty-eight students rotated through five stations designed to examine key components in information retrieval (two history stations); clinical skills (two examination stations); and information delivery (obtaining consent). Peers included second and fifth year medical students and were trained by the examiners to evaluate candidate’s skills based on a check-list of items designed for each station. Candidates were informed beforehand about the assessment formats. Statistical tests used were Pearson’s correlation coefficient and two-sample t-test. P<0.05 considered significant

Results: The overall mean score (+/-SD) for student performance as judged by physicians was 81.8% (+/-7.2) compared to 82.6% (+/-7.8) for peer examiners.

Table 1. Relationships between physician and peer scores

PHYSICIAN						
	Overall Mark	Information Retrieval		Clinical Skills		Information Delivery
		History 1	History 2	Exam 1	Exam 2	Consent
Peer	0.852 ^a	0.610	0.811	0.670	0.810	0.434
	P<0.001	P<0.001	P<0.001	P<0.001	P<0.001	P<0.001

^aPearson correlation coefficient

Conclusion

These results suggest that peer examiners may be an acceptable alternative to physicians in certain OSCEs which examine a student's ability to obtain information from patients or assess their examination skills. However, it is essential that clinical interactive stations requiring delivery of information to patients e.g., consent, are examined by skilled and experienced testers.

ABSTRACT NUMBER: OA-02

The Clinical Skills Confidence Index (CSCI-30): A Longitudinal Tool For Auditing Curricular Changes Affecting Clinical Skills Training

Richard Churchill

University of Nottingham Medical School, United Kingdom

Background: Medical students need to acquire a range of clinical skills by the time they graduate. Such skills may be taught at various stages in the curriculum and often across a range of teaching sites, and so course organisers need to be able to evaluate the effectiveness of teaching on both longitudinal and cross-sectional basis. In this presentation I will present a simple tool that has been developed to audit clinical skills teaching developments within the University of Nottingham Medical School curriculum in the UK.

Methods: An annual cross-sectional on-line survey of all medical students studying at Nottingham University Medical School. Students are asked to rate their confidence in performing a sample of 30 clinical skills at the end of each academic year. Confidence ratings for individual skills, and an aggregate score (the Clinical Skills Index: CSCI-30), both provide data on the effectiveness of the overall clinical skills teaching programme, and allow comparisons between student groups, teaching sites, curricular stage, and sequential years.

Results: The results to date from three consecutive years demonstrate that students’ confidence in specific skills is, as predicted, associated with the stage of the curriculum at which they are taught, such that the CSCI-30 increases significantly over time. There were no significant differences in scores between a new graduate entry cohort and students on a traditional course. However other curricular changes, such as the introduction of mandatory competency tests, have resulted in improved scores.

Conclusion: The Clinical Skills Confidence Index is a simple method for monitoring and evaluating curricular changes that may affect clinical skills acquisition. Its strengths and limitations will be discussed.

ABSTRACT NUMBER: OA-03

Evaluation Of The Impact Of Introducing Mandatory Assessment Of Core Clinical Skills (MACCS) Into An Undergraduate Medical Curriculum

Richard Churchill

University of Nottingham Medical School, United Kingdom

Introduction: Assessment and certification of competence in individual clinical skills is an increasingly important priority for medical educators who need to ensure that graduates are appropriately equipped for practice. Mandatory Assessment of Core Clinical Skills (MACCS) is one form of competency

assessment which is being introduced across the entire Nottingham medical curriculum. In addition to being able to assure competence for the purpose of certification, it was anticipated that the process will enhance student confidence in specific skills and also their overall performance in other types of clinical assessment.

Methods: Cohort study comparing Clinical Phase 1 medical students whose training had included MACCS (2008 cohort), with students from the year prior to the introduction of MACCS. The MACCS programme for this phase included 11 core skills. Outcome measures included results of end of phase OSCE examinations, results of a survey of self-reported confidence in clinical skills, and qualitative feedback from staff and students.

Results: There were no significant differences in the results of end of attachment clinical examinations between the two cohorts, possibly because of the design these examinations. However the aggregate confidence scores were significantly higher in the 2008 cohort, due specifically to an increase in self-reported confidence in some, but not all, of the skills that had been assessed. Feedback from students and staff was generally positive, but concerns were raised about the opportunity cost of implementing MACCS.

Conclusion: Mandatory skills assessments drives learning and improves confidence in specific skills, but may detract from development in other areas.

ABSTRACT NUMBER: OA-04

Validity Of Evaluation Form Suggested For Clinical Examination Of Nursing Students

Abdul-Kareem A Mahmood

Kufa University, Iraq

Background: There is a well prepared form suggested by WHO for practical evaluation of nursing students. The evaluation form includes three categories for scoring; nursing process, nursing action and attitude. This form was implemented in clinical evaluation of nursing students in Kufa university.

Objective: To verify the validity of the applied evaluation form in relation to OSCE and clinical micro- skills.

Design: A clinical trial in which a new evaluation form scores gained by the students were compared with their scores gained in OSCE.

Methods: A random sample of 80 students was selected to be evaluated at the intensive care units of Al-Sader teaching hospital in Najaf, Iraq during March and April 2008. Each student had been evaluated by two stages of examinations; one examiner applied the new evaluation form suggested by WHO and in the other stage the student was exposed to eight stages of OSCE including application of the clinical micro skills for evaluation. The final scores of the two tests had been compared to assess the validity of the new evaluation form. The new evaluation form included the categories of nursing process, nursing action and attitude.

Results: The new evaluation form was found to be 85% sensitive and 90 % specific. The total scores of the evaluation form was 180 with passing score of 90. Forty eight students

(60%) passing the evaluation test by the new evaluation form versus 51 (63.8%) passed in OSCE stages, with no statistically significant difference ($P < 0.05$).

Conclusion: The evaluation form of nursing process assessment is found to be good, sensitive and specific and shows no difference from the result of OSCE and cost-effective. So it can be applied of OSCE which is more difficult and needs more resources of examination staff and materials.

ABSTRACT NUMBER: OA-05

Consistency Of The Double-Marking In Assessments For Undergraduate Medical Students

Cho Min Naing, Daw Khin Win, Hematram Yadav

International Medical University, Kuala Lumpur, Malaysia

Background: The assessment procedures need to be reviewed so that students can be treated fairly and consistently in the markings by the assessors. The research questions were: What is the relationship of marks given by the assessors on the double-marking? If there should be any discrepancies, how can it be improved/resolved? A study was performed with two objectives:

- i. To investigate the relationship of double-marking between assessors; and
- ii. To provide suggestions for the improvement of the marking scheme.

Method: The data set was the double-markings of the selected four multiple essay questions (designated as Q5, Q7, Q8, Q10) used for the end of semester-5 assessment in June/July 2008 (M1/06 cohort) for undergraduate medical students at the International Medical University, Malaysia. Each question was given 10 marks. Pearson correlation coefficients were calculated to determine the relationships, while Paired t-tests were used to assess the differences between the marks of two assessors.

Results: A total of 175 students' marked data were assessed. The markings between the two assessors was significantly different for 2 sets of questions involved in this study ($t = -7.18, p < 0.001$ for Q5; $t = -2.06, p < 0.001$ for Q7), however it did not seem a large gap. Overall, there was a strong correlation between the marks of two assessors ($r = 0.971$ for Q5, 0.987 for Q7 & 0.994 for Q8, $p < 0.001$) except Q10, which had a weak relationship ($r = 0.482, P < 0.001$).

Conclusion: The findings showed that there was inter-assessor variation in marks awarded although their differences were small. It is recommended that more objective question types should be used, so as to enhance the degree of consistency between assessors. If discrepancy is the case, a consensus between assessors should be reached rather than averaging of the marks.

ABSTRACT NUMBER: OA-06

Using Simulated Patients In Orthopaedic OSCEs Is Associated With Higher Scores: A Retrospective Analysis Of Orthopaedic OSCEs In IMU Clinical School

Shong Hing Kock, Loh Keng Yin

International Medical University, Seremban Clinical School, Malaysia

Background: The orthopaedic Objective Structured Clinical Exam (OSCE) is one of the major components in the end of semester 7 and 9 examinations. One of the major concerns highlighted by the external examiners was the mean score of orthopaedic posting is among the highest of all postings where the majority of candidates scored above 60 to 90 out of total 100. Among the postulated reasons was that the orthopaedic department used many simulated patients with few clinical signs for the candidates to diagnose. The candidates were only tested on the steps of examination and not assessed on their diagnostic skills. The aim of this study was to determine the differences between the mean score of the candidates in orthopaedic OSCE stations in which the simulated patients (SP) were used versus the mean score of OSCEs in which the real patients were used.

Methods: A total of 5 cohorts of examinations score marks from August 2006 – August 2008 were analyzed. The mean score of all the candidates for each examination was determined. The type of patient used during the examination was determined from the orthopaedic OSCE questions. The overall difference in the mean score of simulated type of OSCE was compared to the mean score of the real patient OSCE. Student t test was used to determine the statistical difference of the mean score of these 2 groups.

Results: The mean score of both short and long OSCE for semester 7 & 9 orthopaedic OSCEs from August 2006 to August 2008 is shown in table 1 & 2. In the 5 examinations, simulated patients were used in 8 stations and real patients were used in 7 stations. The overall mean score for SP OSCE in semester 7 was 73.3 and for the RP OSCE was 57.9. The difference was statistically significant, $p = 0.025$ (Table 3). The overall mean score for SP OSCE in semester 9 was 78.3 and for the RP OSCE was 65.6. The difference was statistically significant, $p = 0.041$ (Table 3)

Shoulder, wrist and knee examination had been repeatedly tested in both semester 7 and 9 examination for both short and long OSCEs.

Conclusion: The high mean score in orthopaedic clinical OSCE for the semester 7 & 9 examinations was associated with the use of simulated patients. The use of simulated patients also out numbered real patients. As postulated by the external examiners, simulated patients do not provide adequate abnormal physical signs for the candidates to diagnose, therefore majority of the marks are given to the steps in performing the examination rather than ability to elicit findings. The results of this study demonstrated there is a need to re-look the types of patients selected for orthopedic

OSCEs in the future. More real patients with clinical findings should be used.

Table 1: Mean score for semester 7 Short & Long Orthopedic OSCEs

SEMESTER 7				
	Short OSCE		Long OSCE	
	Mean Score	Question: SP/RP	Mean Score	SP/RP
Aug 06	60.5	Malunion of ulna fracture (RP)	75.2	Wrist examination (SP)
Feb 07	73.8	Shoulder examination (SP)	76.5	Knee examination (SP)
Aug 07	61.6	Osteoarthritis knee (RP)	62.2	Avascular necrosis of hip (RP)
Feb 08	50.0	Supracondylar Fracture (RP)	55.2	Osteoarthritis (RP)
Aug 08	70.9	Knee examination (SP)	69.9	Wrist examination (SP)

SP: Simulated Patient RP: Real Patient

Table 2: Mean score for semester 9 Short Orthopedic OSCEs

SEMESTER 9		
Short OSCE		
	Mean Score	Question: SP/RP
Aug 06	66.4	Peripheral nerve injury (RP)
Feb 07	64.9	Wrist injury real (RP)
Aug 07	76.8	Knee examination (SP)
Feb 08	79.2	Shoulder examination (SP)
Aug 08	79.1	Shoulder examination (SP)

SP: simulated patient RP: Real patient

Table 3: Overall Mean score of SP OSCE station versus RP OSCE Station

	SP OSCE	RP OSCE	P VALUE
Semester 7 Mean score:	73.3	57.9	0.025
Semester 9 Mean score:	78.3	65.6	0.041

ABSTRACT NUMBER: OA-07

PBLs Or Large-Group Lectures? A Comparison Of Student Performance In Assessment

W Ranjit De Alwis, Nagarajah Lee

International Medical University, Kuala Lumpur, Malaysia

Background: Lectures and PBLs are the two main teaching/learning activities in the IMU medical curriculum. Medical as well as non medical teachers function as facilitators in PBL sessions. This study was done to compare the student performance in questions based on content covered in PBLs and in lectures, and to determine whether the facilitator's background affects the performance of students at examination.

Methods: The question paper of the nervous system end of course assessment was pre set so that the first part of the two problem case questions was based on the content areas covered only by PBLs, and the second part was based on the areas covered only by lectures. The document analysis method

was used in this exercise. Students' actual performance in the nervous system module PBL and Lecture based questions were used in the analysis.

Results: A total of 182 scripts were analyzed. For assessment on Parkinson Disease, the mean score for PBL based question was 8.36, significantly higher than the score for lecture based question (7.29), the p-value is 0.001. While for head injury, the mean score for PBL based question was 6.39 compared to 5.59 for lecture based question, the difference is significant at $\alpha = 0.05$. The difficulty index for the PBL based questions was higher (0.665) than the value for the lecture based questions (0.510). There was no significant difference in students' scores for both PBL based questions, regardless of whether the PBL sessions were facilitated by medical or non medical academic. The discriminative index for PBL and lecture based questions are 0.623 and 0.220 respectively.

Conclusion: The PBL mode of delivery seems to be more effective than the traditional lecture. The discriminative index provides evidence that the PBL based questions are better questions compared to lecture based questions. It has the ability to discriminate the low and high achievers, which is one of the criteria for test validity. The academic background of the facilitator did not have any influence on the student performance.

ABSTRACT NUMBER: OA-08

To Evaluate The Contents Of Modified Essay Questions (MEQ) For Testing Higher Order Cognitive Skills

Sheila Rani, Sivalingam Nalliah, Rumi Khajotia, Loh Keng Yin, Sambandam Elango

International Medical School, Seremban Clinical School, Malaysia

Background: The MEQs students take at the Final (Part 1) MBBS exams, is a discipline based theory examination attempting to test a broad base of subjects in all fields of Medicine. It has been found to be effective for testing higher cognitive skills. Concerns about whether the MEQ meets the objectives of the curriculum was tested by evaluating all MEQs set in 2007 to 2008.

Objective:

- i. To determine the level of cognitive skills attained (Bloom's taxonomy)
- ii. To determine the percentage of questions which suitably test higher levels of cognitive skills.

Method: Four cohorts of Final MBBS (Part 1) MEQs were analysed by four of the authors to identify the cognitive levels based on Revised Bloom's Taxonomy which identifies 6 different levels of thinking (1). Each component of the question was analysed and categorized according to levels of cognitive skills for the purpose of assessment. For ease of categorization, the highest level attained in each main stem question was accepted to reflect the level of cognitive skills attained.

Results: Of the twelve MEQ questions set during the duration of study (comprising 60 stems) analysed, 54 (90%) attained levels 4,5 or 6 of cognitive skills; 31 (51.66%) were at level 6, 15 (16.66%) were at level 5 and 8 (13.33%) attained level 4.

Conclusion: The MEQ questions set for the Final (Part 1) MBBS exams appear to test higher level of cognitive skills in keeping with the objectives of the medical curriculum at this stage of training.

ABSTRACT NUMBER: OA-09

Mini-CEX: Inter-Examiner Variability And A Comparison With 'Long-Case' Performance

Jayasinghe S, Tai K S, Kew S T, Cheong S K, Das Gupta E, Khajotia R, Koh K C, Chin S P, Morad Z

International Medical University, Seremban, Malaysia

Introduction: We introduced a modified Mini Clinical Evaluation Exercise (Mini-CEX) as a formative assessment during a 7-week posting in Internal Medicine in Semester 6. The first Mini-CEX (Mini-CEX-1, in week 3) and 'Long-Case' (LC, in week 7) were assessed by the same assessor, while the second Mini-CEX (Mini-CEX-2, in week 5) was rotated amongst the assessors. All assessors were IMU academic staff experienced in clinical assessments. They had access to a video on Mini-CEX and received written instructions on the procedures.

The objective was to investigate inter-examiner variability in Mini-CEX and correlate student performances in Mini-CEX with LC.

Methods: Complete data was available from 69 students who did both Mini-CEXs and the summative end-of-posting LC. The marks from Mini-CEX-1, Mini-CEX-2 and LC were correlated. The mean marks awarded by the assessors were compared using ANOVA.

Results: For Mini-CEX-1, the mean mark given by individual assessors had 2 outliers: A, who awarded a significantly lower ($p < 0.005$) mean mark of 39.6 (+/-9.6) than 6 others individually (range of means 56.1-73.9); and B who awarded a significantly higher ($p < 0.01$) mean mark of 73.9 (+/-11.1) than 4 others (range of means 39.6-63.2). For Mini-CEX-2, the mean mark given by B (83.1+/-14.1) was significantly higher ($p < 0.001$) than the rest (range of means 50.5-58.5). The mean LC marks did not differ among the assessors. There was a significant positive correlation between Mini-CEX-1 and LC (correlation coefficient 0.451; $p < 0.001$).

Discussion: The results show appreciable inter-examiner variability in the Mini-CEX but not in LC. The positive correlation between Mini-CEX-1 and LC maybe because of a common factor: each student is assessed by the same assessor on both occasions. It is necessary to have more teacher-training on conducting Mini-CEX in order to reduce inter-examiner variability.

ABSTRACT NUMBER: OA-10

Multiple Choice Versus Short Essay Questions In Assessment Of The Fifth Year Medical Students

Muhsin A Abdulmir Aldhalimi

University of Kufa- Najaf, Iraq

Background: The assessment of medical students is an important step in medical teaching. In spite of increasing global trend towards multiple choice questions, Many teachers still prefer short assay questions.

Objective: Assess the marks of the fifth year medical students in both types of questions.

Methods: Evaluate the trend of the marks of students in five subjects in which there are mixed types of questions and compare between them. The selected materials included internal medicine, dermatology, gynaecology, general surgery and paediatric.

Results: The overall success rate in Multiple choice questions was 32% compared with 74% for short assay questions.

Conclusion: The success rate is much lower in multiple choice questions. It is probably related to the type of the reading of the students rather than to the state of the level of the difficulty of the questions.

ABSTRACT NUMBER: OA-11

Frequent In-Course Assessment Is Feasible And An Effective Driver Of Learning

Toh Peng Yeow, Wai Sun Choo, Amir S Khir, Li Cher Loh

Penang Medical College, Penang, Malaysia

Background: A desirable assessment process should be reliable and valid in assessing clinical competence, and have sufficient education role and feasible in terms of resource consumption. Frequent assessment with feedback would also enhance learning.

Method: We implemented a clinical examination method structured after Gleason's OSLE (Objective Structured Long Examination Record), and ran it 2 weekly through a 6 weeks clinical rotation. Unlike the OSLE, direct observation of the consultation process was excluded. Assessors were encouraged to provide individualized feedback to the student at end of assessment interview. This assessment process was evaluated via questionnaire to participating students and assessors with reference to their perceived educational impact, feasibility and acceptability.

Results: 90.7% of students agreed that patient spectrum was appropriate. The questions asked were fair (88.2%) and tested adequately understanding and clinical reasoning (95.7%). Students perceived themselves to have improved with successive assessment in history taking and presentation skill (96.6%), clinical examination skills (89%) and clinical reasoning skills (90.7%). When compared with previous

module assessment which consist of a single long case that only took place at end of 12 weeks of clinical attachment, the new format was graded to have helped learning "tremendously" (64.7% students vs. 17.6% assessors), "Moderately" (32.8% vs. 82.3%) and marginally (16.8% vs. 0%). Overall, the positive perception is attributable more to the successive nature of the assessment and not the individualized feedback since only 63% of students were provided such. 58.8% of assessors agreed that there was time constraint for providing feedback during assessment. 96.6% of students and 94.1% assessors perceived the format created a stressful but positive learning environment. 52.9% assessors agreed that this format consumes significant resources but 88.2% rated it as manageable and 82.3% supported its continuation.

Conclusion: Frequent intermittent in-course clinical assessment with emphasis on individual feedback is feasible, acceptable and has significant positive educational impact.

ABSTRACT NUMBER: OA-12

Evaluation Of MCQ Papers Used In Basic Sciences Postgraduate Examination In Surgery

Thambidorai C Rajendrarao, Ismail Sagap

University Kebangsaan Malaysia, Kuala Lumpur, Malaysia

Background: Difficulty index (DFI) and discrimination index (DSI) have been used for assessing the MCQ items and provide information on the quality of the test items. However, the available regional literature on analysis of MCQ papers used for postgraduate (PG) surgery examinations is limited.

Material and Methods: MCQ items used for basic sciences PG examinations in surgery for three years from 2005 were analyzed. In each year, paper I had 150 true / false test items (30 questions of 5 items) each for anatomy and a physiology. Paper II had 150 items each for pathology and principles of surgery. The DFI & DSI of each of the 300 test items and the mean indices for each question (60 in every paper) were calculated. The correlation between DFI and DSI scores of the test items was analyzed.

Results: The mean DFI for the six papers ranged from 54 to 68 (SD 12-15) for each paper ($p < 0.005$). The DFI means for the component subjects in each paper ranged from 52 to 68 for anatomy, 51 to 63 for physiology, 61 to 72 for pathology and 58 to 64 for principles of surgery. The differences in the DFI between the individual subjects within each paper were statistically significant ($p < 0.005$).

The mean DSI for the six papers ranged from 0.1 to 0.25 (SD 0.09- 0.24) ($p < 0.005$). The DSI means for the subjects in each paper ranged from 0.15 to 0.44 for anatomy, 0.13 to 0.23 for physiology, 0.11 to 0.27 for pathology and 0.1 to 0.23 for principles of surgery and the differences were significant. ($p < 0.005$).

Of the total 1800 test items from six papers, 975 (54%) had DSI below 0.2, 1243 (69%) had DSI below 0.3 and in nearly

33% the DSI was either zero or in the negative. 381 test items of the total 1800 (21.1%) had DFI between 50-70 and the DSI in 120 of these was either 0 or in the negative. (*The results will be depicted in tables in the presentation*).

Conclusion:

1. The significant differences in DFI and DSI of MCQ items (from year to year and within the component subjects in each paper) despite the acceptable mean DFI level between 54 – 68 for each paper, show that in composite papers involving multiple subjects, analysis of individual test items and the subject components in each paper separately is more informative than the mean indices of the whole paper.
2. Contrary to the common view that moderately difficult items (DFI 50 to 70) are likely to have acceptable DSI (0.3), in this analysis nearly one third of the test items with DFI 50 – 70 had poor DSI suggesting that it is difficult to predict the DSI of a MCQ item from its DFI. Also, levels of DFI and DSI suggested in the literature as acceptable for UG examinations may not apply to PG or specialty medical examinations. This is possibly because PG examinations require more mastery and detailed knowledge of the subject than in UG medical examinations.
3. The need for periodic evaluation of MCQ items even in institutions using this assessment method for many years is reiterated.