

Assessment Methods of an Undergraduate Psychiatry Course at a Saudi University

*Mostafa Amr¹ and Tarek Amin²

أساليب تقييم مساق الطب النفسي للطلاب في جامعة سعودية

مصطفى عمرو، طارق أمين

المخلص: الهدف: يوجد عدد قليل من الدراسات حول أساليب التقييم في مجال الطب النفسي في الدول العربية، وهدف هذه الدراسة هو فحص نتائج أشكال التقييم المختلفة لمقرر الطب النفسي لطلاب السنة الخامسة بكلية الطب في جامعة الملك فيصل، المملكة العربية السعودية. الطريقة: تم فحص نتائج أداء 110 من الطلاب في الاختبار السريري الموضوعي والاختبار التقليدي السريري الشفهي وحافطة الأنشطة المنظمة وأسئلة الاختيار المتعدد التحريري. النتائج: تراوحت درجات الطلاب بين 32 - 50، 7 - 15، 5 - 10 و 22 - 45 في كل من الاختبار التقليدي السريري الشفهي والاختبار السريري الموضوعي وحافطة الأنشطة المنظمة وأسئلة الاختيار المتعدد علي التوالي. ووجد أن هناك ارتباطاً معتداً بين الاختبار السريري الموضوعي وباقي أساليب التقييم باستثناء أسئلة الاختيار المتعدد في تحليل الانحدار، حيث شكل الاختبار السريري الموضوعي 65.1% من التباين في الدرجات السريرية الكلية و31.5% من الدرجات النهائية ($P = 0.001$)، في حين مثل الاختبار التقليدي السريري الشفهي 74.5% من التباين في الدرجات السريرية. الخلاصة: تدل الدراسة علي وجود تكامل واتساق في أساليب تقييم الطلبة المستخدمة في مساق الطب النفسي، ولا سيما العنصر السريري. ومن شأن تلك المعلومات أن تكون مفيدة للتطورات المستقبلية في مجال التدريس الجامعي.

مفتاح الكلمات: طلاب الطب، تقييم، الطب النفسي، المملكة العربية السعودية.

ABSTRACT: Objectives: In Arab countries there are few studies on assessment methods in the field of psychiatry. The objective of this study was to assess the outcome of different forms of psychiatric course assessment among fifth year medical students at King Faisal University, Saudi Arabia. **Methods:** We examined the performance of 110 fifth-year medical students through objective structured clinical examinations (OSCE), traditional oral clinical examinations (TOCE), portfolios, multiple choice questions (MCQ), and a written examination. **Results:** The score ranges in TOCE, OSCE, portfolio, and MCQ were 32–50, 7–15, 5–10 and 22–45, respectively. In regression analysis, there was a significant correlation between OSCE and all forms of psychiatry examinations, except for the MCQ marks. OSCE accounted for 65.1% of the variance in total clinical marks and 31.5% of the final marks ($P = 0.001$), while TOCE alone accounted for 74.5% of the variance in the clinical scores. **Conclusions:** This study demonstrates a consistency among the students' assessment methods used in the psychiatry course, particularly the clinical component, in an integrated manner. This information would be useful for future developments in undergraduate teaching.

Keywords: Undergraduate medical students; Assessment; Psychiatry; Undergraduate; Saudi Arabia.

ADVANCES IN KNOWLEDGE

- In Arab countries there are few studies on assessment methods in the field of psychiatry, so this study contributes new information.
- The results of this study revealed that there was no significant difference in the students' scores that correlated with gender differences.
- The objective structured clinical examination (OSCE) as a method of student assessment accounted for a sizable part of the variance in total clinical and final marks, which indicated that the implementation of OSCE is a useful adjunct to other forms of clinical assessment.

APPLICATIONS TO PATIENT CARE

- This study showed that different forms of psychiatric assessment can improve student knowledge and the quality of their practice, and fulfill unmet information needs at the point of patient care.
- Students should be encouraged to become familiar with updated methods of assessment in psychiatry courses.

IN SAUDI ARABIA, THE COLLEGE OF MEDICINE at King Faisal University offers a six-year medical curriculum to selected Saudi students who have successfully completed one year of requisite general university studies following secondary school education. The first four years of the curriculum are devoted to pre-clinical (medical sciences and family medicine) learning. Students are exposed to behavioural sciences in the third year. In the fourth year, students are introduced to a problem-based learning (PBL) integrated curriculum. They practice communication, history taking, and the physical examination of different body systems, as well as relevant procedural skills. Training is conducted in a clinical skills laboratory using different types of simulators. They learn more about the interplay between the physical and psychological components of illnesses. The curriculum in years 5 and 6 is structured around a series of clerkship rotations in the departments of Internal Medicine, Surgery, Psychiatry, Obstetrics and Gynecology, and Pediatrics. Students graduate after successful completion of 12 semesters (229 hours per semester). The Department of Psychiatry attachment is a 6-week course based in a dedicated psychiatric hospital. Teaching-learning methods employed include lectures, small-group tutorials, and group discussions guided by department faculty. The major objectives of the Department for the attachment are that students 1) acquire a basic knowledge of the developmental aspects of psychiatric disorders; 2) identify and make use of all relevant sources of information when assessing each patient; 3) demonstrate competence in mental state examinations and physical assessments; 4) develop skills in appropriate communication with patients and colleagues, and 5) make a clear oral presentation of a case.

During the fifth year, students undertake six clinical rotations averaging 180 hours, arranged in two semesters of 3 rotations each. The group size for each rotation varies from 8 to 12 students. Since the Department was established in 2006, the rotating students have been evaluated through portfolios consisting of peer reviews, group work, case studies, ethics discussions, and critical reviews, and at the end of the course by a traditional oral clinical examination (TOCE) and an objective structured clinical examination (OSCE). At the end of the semester, a multiple choice question (MCQ)

examination is held [Figure 1]. Furthermore, in 2009, the Psychiatry Department conducted a survey that assessed the students' attitudes towards psychiatry that was published as an international education report.¹ The survey showed favourable changes in the students' attitudes following clerkship. However, less positive responses were seen in students' attitudes towards the quality of the medical school clerkship.

To improve the student learning/assessment experience we introduced the OSCE for the summative assessment of students, in conjunction with a traditional oral examination and portfolio. The potential marks for the written paper, MCQ, portfolio, OSCE and clinical examination are 50, 10, 15 and 25, respectively, for a total of 100 points. Although the use of OSCEs in psychiatry has been described as less widespread than in other medical fields, recent years have witnessed an increased interest in its use in psychiatry.^{2,3} The objective of this study was to assess the outcome of different forms of psychiatric course assessment among fifth-year medical students at King Faisal University, Saudi Arabia.

Methods

This was a cross-sectional survey carried out during the 2010–11 academic year, in two consecutive semesters, in which cohorts of male and female students (54 and 56, respectively) were invited to participate in the study. All students agreed to participate in the study, which was approved by the college authorities.

The MCQ paper at each examination contained 50 items worth one mark each. The initial item bank of 500 questions was designed to cover the following content areas: causes/risks, signs/symptoms, course, treatments, and mental health services. Two items were included to represent each content area. One item was answered through simple recall, and the other was designed to be answered interpretatively and commonly involved a brief, one to four sentence case presentation. Each MCQ item consisted of a stem no longer than five sentences in length (though typically only 1–2 sentences), along with four response options. Test items were developed following standard, well-described MCQ writing procedures, and were designed to avoid ambiguity, vagueness, and value-

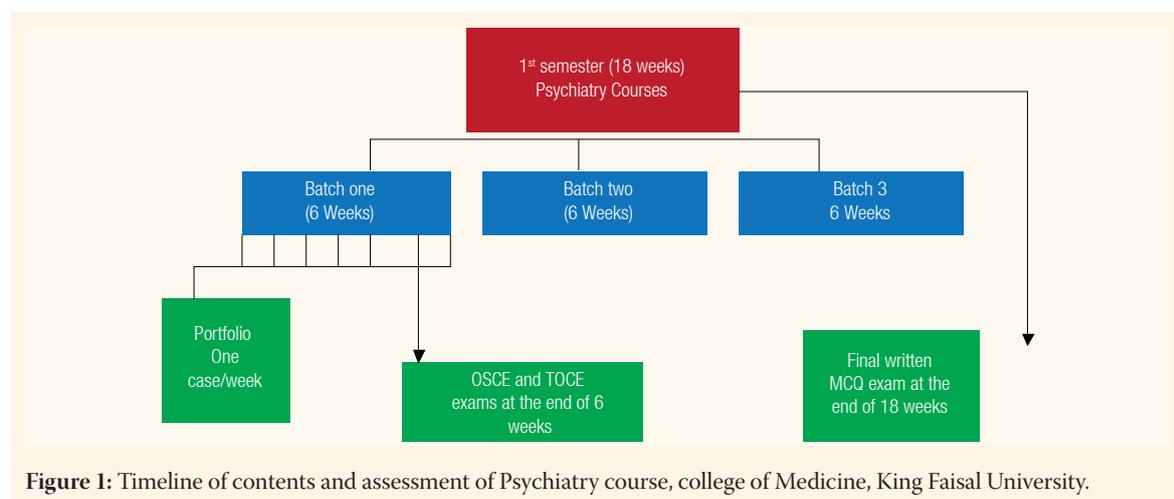


Figure 1: Timeline of contents and assessment of Psychiatry course, college of Medicine, King Faisal University.

laden language.⁴ Reliability (Cronbach's alpha) and concurrent validity (Pearson r) coefficients were obtained by correlating the scores of MCQ papers with the overall outcome of the examination. They were in the ranges of 0.83–.91 and 0.80–0.93 ($P < 0.05$), respectively. Indices of item facility and discrimination were in the ranges of 50–91, and 0.37–.45, respectively.

In TOCE, to explore the student's understanding of topics deemed relevant to curriculum, students interviewed and examined a real patient for over 45 minutes, and then summarised their findings to two examiners who questioned them by an unstructured oral examination on the patient's problem. The student's interaction with the patient was not observed. Reliability (alpha) and concurrent validity coefficients (Pearson r) were obtained by correlating the scores in the TOCE with the overall outcome of the examination. They were in the ranges of 0.58–.71 and 0.73–.81, respectively ($P < 0.05$).

The OSCE was based on the curricular constructs that included six thematic topics: mood disorders, anxiety disorders, child psychiatry, psychosis, personality disorders, and substance abuse. A blueprint was developed for each OSCE to capture the clinical competencies in the covered topics. A map for the stations was devised to guide the examinees and organisers with clear written instructions to the examiners, patients, and examinees. The OSCE was composed of nine stations which included two manned stations. A manned station (MS) referred to a station that had a real patient and an examiner. Students were allowed 15 minutes to perform tasks at each station. The first station included a psychiatric interview, where

students were to develop a rapport and conduct the interview within the assigned time frame for a male patient with schizophrenia. At the second station, the students assessed the mental status, with particular attention to the mood and affect, of a female patient with bipolar I mood disorder. In each station, two independent examiners rated the examinees independently according to checklists. The raters were selected from the lecturers who were not involved in the design and/or implementation of the station. Checklists contained the desired competencies to be examined (average 28 items). The scores were classified as 'done', 'not done', or 'done incorrectly', with questions on topics such as delusions, hallucinations, and performance. Each item was assigned a weight by the station's authors. At the end of each checklist, there were 4 questions with a 3-point Likert scale addressing the interview technique and included factors like empathy, degree of coherence, and verbal and nonverbal expression. Following the MS, students moved to an unmanned station (UMS) (4 minutes each) which included four dependent data stations (4 minutes each) with questions based on the previously taken history or examination stations, and three independent data stations. In these independent stations, students read a poster giving information regarding a history/examination and/or investigations, and he/she was required to answer questions related to diagnosis, further investigations, or management. Students moved between stations on time keepers' commands. Examiners supervised each station throughout the session and the whole group of students was assessed by a nearly identical process. At the end, the marking and answer sheets were

Table 1: Students' scores along the different types of assessment in the psychiatry course, College of Medicine, King Faisal University.

Assessment	Total (N = 110)	Gender		P value*
		Females (n = 56)	Males (n = 54)	
Written exam (50 marks)				
Mean ± SD	35.8 ± 4.2	35.7 ± 3.3	35.9 ± 5.1	0.679
Median	36.0	36.0	36.0	
Minimum-maximum	22.0-45.0	22.0-45	28.0-42.0	
Total clinical exam (50 marks)				
Mean ± SD	42.0 ± 3.5	41.6 ± 3.2	42.4 ± 3.7	0.143
Median	42.0	41.5	43.5	
Minimum-maximum	32.0-50.0	35.0-48.0	32.0-50.0	
OSCE exam (15 marks)				
Mean ± SD	11.7 ± 1.5	11.6 ± 1.6	11.9 ± 1.4	0.548
Median	12.0	12.0	12.0	
Minimum-Maximum	7.0-15.0	7.0-14.0	9.0-15.0	
TOCE exam (25 marks)				
Mean ± SD	21.7 ± 1.9	21.4 ± 1.9	22.1 ± 1.9	0.059
Median	22.0	22.0	22.5	
Minimum-Maximum	17.0-25.0	18.0-24.0	17.0-25.0	
Portfolio (10 marks)				
Mean ± SD	8.6 ± 1.2	8.6 ± 1.0	8.5 ± 1.4	0.872
Median	9.0	9.0	9.0	
Minimum-Maximum	5.0-10.0	6.0-10.0	5.0-10.0	

Legend: SD = Standard deviation; OSCE = objective structured clinical examination; TOCE = traditional oral clinical examination

* Mann Whitney test of significance.

collected from the examiners and students, respectively. The student answers for the UMS were corrected following a pre-designed checklist.

A portfolio was instituted to evaluate competency in designated topics specific to the curriculum. Students were expected to present one case per week at the ward rounds. Cases were discussed at the weekly group tutorial sessions according to the curriculum's schedule so, for example, students were presented with representative cases for mood disorders in week 1 and anxiety disorders in week 2. Two psychiatrists were trained to score each student's portfolio. For the 6 case areas, the scoring rubric was composed of a 6-point ordinal scale, where 1 = not competent, 3 = competent, and 6 = highly competent. Each student's performance was measured by averaging the two raters' scores for each case. Reliability (alpha) and concurrent validity (Pearson r) coefficients (obtained by correlating the scores of the MCQ papers with the overall outcome of the examinations) were in the ranges of 0.63–0.71 and 0.66–0.73 ($P < 0.05$), respectively. Weighted Kappa ranged from 0.84 to 0.95 for inter-

rater reliability.

Data analysis was carried out using the Statistical Software for the Social Sciences (SPSS) package (Version 15, Chicago, Illinois, USA). Median, mean, and standard deviations were calculated for examination marks. Statistical comparison was carried out using the Mann-Whitney test. Zero order and partial correlations were performed between test marks, and regression models were fitted to evaluate the predictive value of OSCE as an independent variable, either alone, or with other examinations, and total clinical score or total final marks as the dependent variables. To assess the reliability and credibility of the OSCE, statistical analyses of Cronbach alpha, Kappa, and Pearson's correlation coefficient were used.

Results

Table 1 displays the students' scores along the different assessment methods used to evaluate the outcome. The score range in the TOCE, OSCE, portfolio, and MCQ were 32–50, 7–15, 5–10 and

Table 2: Correlation matrix between medical student scores in different examinations and total clinical scores (N = 110)

Examinations	Pearson Correlation			
	Total clinical	OSCE	TOCE	Portfolio
Written: r coefficient	0.162	0.157	0.154	0.020
P value	0.091	0.102	0.108	0.835
Total clinical: r coefficient	--	0.807	0.863	0.558
P value	--	0.000	0.000	0.000
OSCE: r coefficient	--	--	0.493	0.196
P value	--	--	0.000	0.040
TOCE: r coefficient	--	--	--	0.379
P value	--	--	--	0.000

Legend: OSCE = objective structured clinical examination; TOCE = traditional oral clinical examination

22–45, respectively. There was no significant difference in scores earned by different genders. A significant positive correlation was seen between OSCE and all forms of psychiatry examinations except for the written/MCQ marks [Table 2]. Strong positive correlations were found between components of the total clinical examination (especially TOCE and OSCE), while moderate correlations were found between TOCE and OSCE and low correlations with the portfolio ($r = 0.86, 0.49$ and 0.20 , respectively). Figure 2 depicts the relationship between the students’ scores on the TOCE and written/MCQ examinations. There was no significant correlation between the two methods of assessment in students’ evaluations. On the contrary, Figure 3 shows a moderate and significant

correlation between TOCE and OSCE ($r = .493, P = 0.001$).

The Kappa concordance coefficient and the correlation between the scores of examinees were computed. They ranged from 0.75 for station 1 to 0.64 for station 2. The Cronbach’s alpha coefficients for station 1 and 2 were 0.82 and 0.78, respectively. In the generated linear regression model, OSCE accounted for 65.1% of the variance in total clinical marks and 31.5% of the final marks ($P = 0.001$). One unit of change was associated with a 1.63 point change in the total clinical score and a 2.05 point change in final marks. In multiple regression analysis, the TOCE alone accounted for 74.5% of the variance in the clinical scores. Conditioned on its presence, the OSCE explained an extra variance of

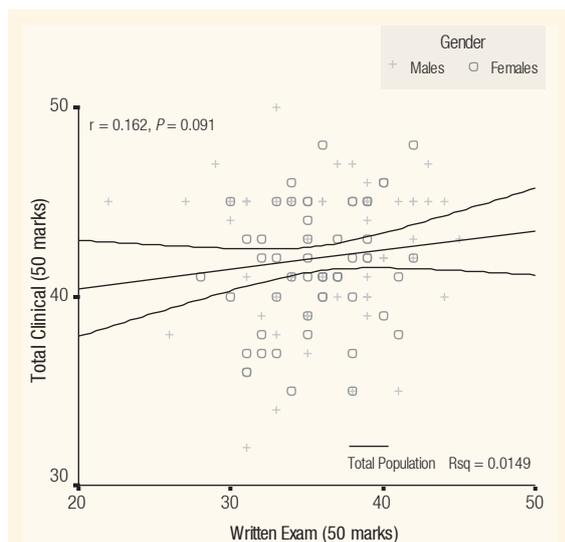


Figure 2: Correlation between written and total clinical assessment in Psychiatry course, Colleg of Medicine, King Faisal University.

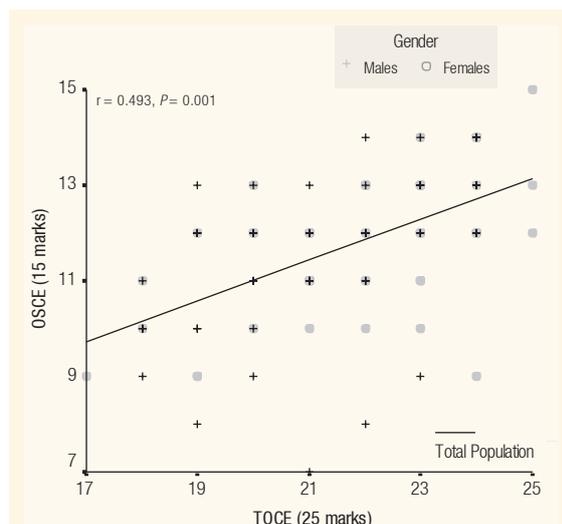


Figure 3: Correlation between TOCE and OSCE assessments of included students at Psychiatry course, King Faisal University.

19.2%. In regression analysis, the OSCE accounted for 65.1% of the variance in total clinical marks and 31.5% of the final marks ($P = 0.001$), while the TOCE alone accounted for 74.5% of the variance in the clinical scores.

Discussion

Findings from this study showed that the results of the MCQs are the most important predictors of final scores, as they accounted for 69.7% of student variability. These results are most likely due to the commonly observed relationship of a good quality MCQ test with other performance measures. It has been observed that general ability is the foundation of most performance measures and a well-constructed MCQ is the best estimator of this general ability.^{5,6} Also the results might reflect an unbiased evaluation of the medical students.⁷

The acquisition of clinical skills is paramount to the development of a safe and competent practitioner.⁸ OSCE as a performance-based assessment is a well-established assessment tool for many reasons: it is a competency-based, valid, practical, and effective means of assessing clinical skills that are fundamental to the practice of medicine, and to other health care related professions.⁹ While OSCE is in use in many medical disciplines in Saudi Arabia, particularly in general surgery, orthopaedics and internal medicine, psychiatric educators have been slow to adopt this method of evaluation.^{7,10-12} To the best of the authors' knowledge, this is the first report that addresses OSCE in undergraduate psychiatric assessment in Saudi Arabia. As expected, the implementation of OSCE in our Department has proved to be a useful adjunct to other forms of clinical assessment. The student' scores on the OSCE correlated well with the results in clinical examinations and explained a great part of the variance in total clinical marks. Similar findings have been reported in different specialties from different countries;¹³⁻¹⁵ however, these studies did not show a correlation between the results of the OSCE and the MCQs. This may be attributed to the fact that MCQs assess the students' cognitive abilities, covering the area of 'knows' and 'knows how' of Miller's pyramid of assessment, and possibly spanning the levels of Bloom's taxonomy of educational objectives, from the level of comprehension to the level of

evaluation.¹⁶ Additionally, the OSCE, like other forms of clinical examinations, tests a different domain of clinical skills (covering the area of 'shows how' of the Miller's pyramid of assessment) which is a prerequisite for physician performance in real life, such as history taking and physical examinations.³ Nonetheless, our results should be interpreted with caution as, according to previous studies in the literature, only two of the stations in our OSCE examination are considered classic OSCE stations.³

The results of the study show that the most significant predictor of overall clinical scores is the TOCE. It alone explained 74.5% of the variance in clinical scores. Conditioned on the presence of the TOCE, the OSCE explained an extra variance of 19.2%. The examiners awarded high marks to favour a more pleasurable student-teacher encounter which unfortunately created a 'halo effect' in the evaluation of the students. The OSCE significantly correlates with the TOCE, but still has an important role in predicting total clinical marks. It explained 65.1% of the variance in total clinical marks. A better designed OSCE and external examiners in the TOCE would help to increase the accuracy and reliability of clinical assessment.

Conclusion

This study demonstrates that different clinical methods used to assess medical students during their Psychiatry course were consistent and integrated. This information would be useful for future developments in undergraduate teaching of this subject.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Feroze Kaliyadan, Assistant Professor of Dermatology at King Faisal University, for his assistance in revising the manuscript.

CONFLICT OF INTEREST

No funding was received for this study, and no conflict of interest exists.

References

1. El-Gilany AH, Amr M, Iqbal R. Students' attitudes toward psychiatry at Al-Hassa medical college, Saudi

- Arabia. *Acad Psych* 2010; 34:71–4.
2. Vaidya NA. Psychiatry clerkship: Objective Structured Clinical Examination is here to stay. *Acad Psych* 2008; 32:177–9.
 3. Hodges B, Hanson M, McNaughton N, Regehr G. University of Toronto Psychiatric Skills Assessment Project. Creating, monitoring, and improving a psychiatry OSCE: a guide for faculty. *Acad Psych* 2002; 26:134–61.
 4. Haladyna TM, Downing SM. A taxonomy of multiple choice item-writing rules. *Appl Meas Educ* 1989; 37–50.
 5. Maatsch JL. Assessment of clinical competence on the Emergency Medicine Specialty Certification Examination: the validity of examiner ratings of simulated clinical encounters. *Ann Emerg Med* 1981; 10:504–7.
 6. Fabrega H Jr, Ulrich R, Keshavan M. Gender differences in how medical students learn to rate psychopathology. *J Nerv Ment Dis* 1994; 182:471–5.
 7. Shaheen M, Bader A, Al-Khudairy N. Experience of the orthopedic division at King Saud University in organizing and conducting the Objective Structured Clinical Examination (OSCE). *J KAU: Med Sci* 1991; 1:57–64.
 8. Zyromski NJ, Staren ED, Werrick HW. Surgery resident's perception of the objective structured clinical examination (OSCE). *Curr Surg* 2003; 60:533–7.
 9. Bakhsh TM, Sibiany AM, Al-Mashat FM, Meccawy AA, Al-Thubaity FK. Comparison of students' performance in the traditional oral clinical examination and the objective structured clinical examination. *Saudi Med J* 2009; 30:555–7.
 10. Guraya S, Alzobydi A, Salman S. Objective structured clinical examination: Examiners' bias and recommendations to improve its reliability. *J Med Med Sci* 2010; 1:269–72.
 11. Almoallim H. OSCE in Internal Medicine for Undergraduate Students Newly Encountered with Clinical Training. From: <http://services.aamc.org/30/mededportal/servlet/s/segment/mededportal/?subid=429>. Accessed: May 2011.
 12. Higazi M, Downing SM. Objective structured clinical examination as an assessment method in residency training: practical considerations. *Ann Saudi Med* 2008; 28:192–9.
 13. Townsend AH, McLlvenny S, Miller CJ, Dunn EV. The use of an objective structured clinical examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. *Med Educ* 2001; 35:841–6.
 14. Yu TC, Wheeler BR, Hill AG. Clinical supervisor evaluations during general surgery clerkships. *Med Teach* 2011; 33:479–84.
 15. Singhal N, Lockyer J, Fidler H, Keenan W, Little G, Bucher S. Helping babies breathe: Global neonatal resuscitation program development and formative educational evaluation. *Resuscitation* 2012; 83:90–6.
 16. Maatsch JL, Huang RR, Downing S, Barker D, Munger B. The predictive validity of test formats and a psychometric theory of clinical competence. *Proc Annu Conf Res Med Educ* 1984; 23:76–8.