Scholars Journal of Applied Medical Sciences (SJAMS) **ISSN 2320-6691 (Online)**
Sch. J. App. Med. Sci., 2016; 4(11A):3876-3882 **ISSN 2347-954X (Print)**

(An International Publisher for Academic and Scientific Resources)
www.saspublishers.com DOI: 10.36347/sjams.2016.v04i11.004

Original Research Article

Measuring the quality of the Objective structured clinical examination in the Obstetrics and Gynaecology department of a resource limited institution in East Africa

Ogah A.O[1], Jama M.P.[2] Brits H.[3], Ogah O.G.A[4]
[1]Department of Paediatrics, School of Health Sciences, Kampala International University, Dar es Salaam, Tanzania, P.O.Box 9790.
[2]PhD. Division Health Sciences Education, Faculty of Health Sciences, University of the Free State, Bloemfontein, South Africa.
[3]Professor of Medicine, Department of Internal Medicine, Faculty of Medicine, University of the Free State, Bloemfontein, South Africa
[4]Director of Quality Assurance, Kampala International University, Dar es Salaam, Tanzania. P.O.Box 9790.

***Corresponding author**
Ogah A.O
Email: nikeogah@gmail.com

Abstract: The objective is to improve assessments by measuring the quality properties of the Objective structured clinical examination scores of 10, 3rd year Clinical Medicine students in the Obstetrics and Gynaecology department of a resource-limited medical school in Tanzania, using psychometric methods, in July 2015. This descriptive and cross-sectional study used literature review and structured observation as data collection methods. Students' performances were assessed using checklists by 5 examiners in 4 OSCE stations. Stations 1-2 were manned and stations 3-4 were in written format. Permission to carry out the study and the ethical approval were obtained from Kampala International University and the Amana Hospital IRBs and management. Statistical analysis was carried out using Microsoft Excel and SPSS computer packages. The stations were too few and tasks in the stations barely covered 50% of the learning outcomes. The mean scores in the stations were between 34-91% and pass mark was set between 35-95%. The tasks were too easy in 3 of the stations, the variance was generally too high for a criterion referenced test and none of the means was centrally located. The scores were transformed to z-scores which eliminated all the extreme scores and prevented false awards of distinction and fail grades to students. Examiners' error was very high at 89% and the internal consistency was weak (Alpha was 0.008). Item analysis was poor in stations 1-3 and the examiner who marked station 3 was identified to be a dove. The analysis recommended that stations 1 and 3 be discarded, station 2 be reviewed and station 4(whose properties were good) be banked for future use. Moreover, the department is adviced to generate between 8-15 OSCE stations in subsequent OSCEs for better reliability coefficients. The properties of the OSCE experienced in resource limited settings may not be similar to those practiced in the established medical schools of developed countries.
Keywords: Psychometric Analysis, Resource Limited Medical Schools, Objective structured clinical examination, Obstetrics and Gynaecology department, post-examination evaluation

INTRODUCTION

Assessment is the heart of every training institution. Assessment drives learning [1]. Through assessments, the amount of learning acquired by the trainee can be measure and the programmes can be evaluated [2]. No assessment is ever the same with another one. OSCE being a form of clinical assessment which is resource intensive, its administration may be compromised in quality especially in resource limited Institutions [3]. The quality assurance machinery in place in most medical schools currently, involves mainly human raters which can be subjective, biased and inconsistent [4]. Psychometric analysis, offers a stable, objective and cheap means of measuring and improving consistently the quality of the OSCEs and all other forms of assessments [5].

A lot has been documented and published about the psychometric qualities of the OSCEs practiced in the Medical Schools of the developed countries, but even there, psychometric analysis has not yet been incorporated into the University policies for

regular use. Hence, every university has been producing medical graduates with different levels of competencies. Very little has been published about the real state of the OSCEs implemented in resource constrained Medical Schools in Sub-Saharan Africa. This might pose a risk to patient safety especially in countries that do not have a national qualifying examination to harmonise and regulate the quality of assessments and medical graduates certified and registered for practice. Hence, psychometric analysis should be fully integrated into the Quality Assurance examination policy of every Medical Schools, to harmonise and improve the quality of assessments, training, medical graduates and therefore patient care.

In this study, the psychometric qualities of the promotional OSCE implemented in a resource limited medical school in East Africa(Dar es Salaam, Tanzania) is presented, interpreted and discussed. The medical school selected in this study is private-owned, 4 years old as of the time of study and does not have its own Teaching Hospital yet (the Teaching Hospital is currently under construction). The clinical students rotate in nearby referral Hospitals (structured and function basically for patient care) which are affiliated to the university. Hence, the venue of this end of year OSCE was in one of the regional referral Hospitals in Dar es Salaam. With the challenges of limited resources experienced in the medical schools in Sub-Saharan Africa, should we expect the psychometric properties of the assessments implemented here to be similar to those published in medical institutions of high index countries

METHODS

This study took place in Kampala International University, Dar es Salaam campus (hereafter referred to as KIU-D). Kampala International University is a multi-campus university with campuses in Uganda, Tanzania and Kenya. Permission to carry out the study and the ethical approval were obtained from Kampala International University and the Amana Hospital IRBs and management. Participant anonymity was preserved. This descriptive and cross-sectional study used data collection methods such as literature review to obtain the psychometric tools and structured observation to observe the OSCE set-up, proceedings and student performances. Students' performances were scored using the Ministry of Health checklists by 5 examiners in 4 OSCE stations in the OBGY department. There were 3 medical officers, one specialist and one consultant. The consultant was the external examiner. Stations 3-4 were manned and stations 1-2 were in written format. The study population was described as the current, active third year clinical medicine students, who passed the OBGY theory paper and were therefore eligible to sit for the OSCE. The list of the eligible students and the invited examiners was obtained from the office of the Dean of the Faculty of Medicine. Only 10 of the 30 students in the faculty register were eligible to sit for the OSCE and all these 10 students were recruited for the study, hence no sampling was required. There was no standardized patient because according to the head of department, there was very limited time to recruit and train willing individuals to act as standardised patients. Moreover, the examiners were not familiar with global scoring. On the day of the OSCE, the examiners and students arrived at the Amana regional referral hospital at 07.00hours. The students were conveyed to and from the hospital which was 30minutes drive away from the university, by the school bus. After breakfast, we briefed the examiners and students for 45 minutes and obtained their consent for the research. All the examiners in OBGY were part-timers and therefore two of them missed the pre-OSCE briefing because they came late. The programme coordinator served the examiners, the checklists and the students were ushered to the OBGY hospital wards according to the exam schedule.

The OSCE began at 09.00hrs. The live OSCE set-up (station design and station contents) and proceedings were observed and briefly described during the OSCE using a checklist. The examiners observed and assessed the students' performance in the manned stations using a clinical checklist from the Ministry of Health, Tanzania. This checklist did not cover global grades; hence, the examiners did not record their global grades despite the pre-OSCE briefings. Hence, the analysis of the relationship between the global scores and the checklist scores was not done. The detailed checklist scores per examiner per candidate per station were obtained from the Head of Department after marking the written stations and compilation of scores from the examiners' checklists in the manned stations.

The psychometric analysis was carried out on the post-OSCE scores obtained from each station, examiner and candidate as well as overall students' performance. The psychometric methods used in this study were descriptive and inferential in nature under the classical test theory. The G-study under the item response theory was also used to determine the sources of errors in the OSCE. The descriptive statistics were used to describe the summarized data and included the frequency distribution of the scores, measures of central tendencies and measures of variation. The inferential statistics was used to inform decision that can be generalized on the population and included station analysis, reliability tests and identifying hawks and doves. The scores were subjected to statistical analysis in order to determine the reliability (indirectly the validity) of the scores. The dependent variable under study, to be measured, was the reliability of the OSCE scores, achieved by the students, as recorded by the examiners in the checklists. The variable (OSCE scores) is quantitative continuous in nature. The

independent variables were the facets and characteristics of the testers, tests and tastes operating during the OSCEs. The independent variables were categorical in nature.

Statistical and text analysis were carried out using Microsoft Excel and SPSS (version 17) computer packages. Information gathered from literature and documentations, together with the observations and findings from the psychometric analysis of the OSCE used to test the third year clinical medicine students were used to formulate post-examination remediation and recommendations for the improvement of the OSCE at the Faculty of Medicine, KIU-D and this study can be replicated in other institutions with constrained resources.

RESULTS

The results for the OSCE conducted by the OBGY department of the KIU-D and the Amana referral Hospital are shown below.

Sociodemographic Characteristics of the Examiners and Students

There were 3 female and 7 male students amongst those eligible to sit for the OBGY OSCE. Amongst the 5 examiners who participated, 3 were Medical officers, 1 was a Specialist and the external examiner was a consultant. All the examiners and students were Tanzanians.

OSCE Observations

The students worked on 10 tasks altogether distributed in each of the 4 OSCE stations (Table 1 below). Each station was 5minutes long and the manned stations (3 and 4) covered only history taking, general physical examination and genitourinary system.

OBGY Stations Metrics

The OBGY metrics consists of descriptive and inferential statistics under the classical test and item response theories. The tables below show the statistics of the stations and students (total column) scores. The values were rounded up to one decimal place.

Descriptive statistics

The descriptive statistics summarized and described the scores in the stations. This statistics include the scores distribution, measures of central tendencies and measures of variation.

Scores distribution:

In Table 1 below, the pattern of distribution of the scores was described by the skewness, kurtosis, checking for outliers and z-scores. The distributions of scores in stations 1-3 were significantly negatively skewed and positively kurtosed with few low extreme values, suggesting too easy tasks, see Figure 1. After converting the raw scores into z-scores, the extreme values disappeared and the best performance was in Grade C (Good: >+1<+2), while the weakest grade was E (Poor: +<-1>-2). The z letter grades corresponded with different scores in each station.

Measures of central tendencies:

The mean was not centrally located in any of the stations. Moreover, the fixed university pass mark of 50% was based on an ideal examination set-up, which was not the case here, as the distribution of the scores was significantly skewed. Hence, the standardized pass mark is a more appropriate score for making pass/fail decision in this OSCE, see Table 1 below.

Measures of variation:

The variance in all the stations was significantly high (coefficient of variation above 5%) especially in station 4. The variance was too high for a criterion referenced assessment such as this OSCE, where mastery of the subject is desired in every of the students. The standardized pass mark is more appropriate for making pass/fail decision rather than the fixed university pass mark of 50% in stations 1 and 3 because none of the 95% confidence intervals contain the university pass mark of 5(50%). Moreover, the 95% confidence interval of station 3 was totally different from the others and the standard error of the mean of station 4 was excessively high. see Table 1, below.

ANOVA (comparing means): The variance between the stations was very high (>30%) and significantly higher than the variance within the stations. The mean of station 3 was consistently and significantly higher than the rest. (see Table 2).

Generalizability studies:

Variance components estimates: The students and the examiners contributed 0.1% and 89.2% respectively to the variance obtained in the OBGY OSCE. The interaction between students and examiners was high at 10.7%. G-coefficient was low at 0.0 (see Table 3 below). The level of examiners' errors and student-examiner interactions experienced in the OBGY stations was significantly high enough to influence the OSCE scores.

Inferential statistics

The inferential statistics utilized in this study include the station analysis, reliability estimates and the procedures for identifying hawks and doves.

Station Analysis:

The Item Difficulty Index (IDI) has a good range between 0.3 and 0.8, which is seen in stations 2 and 4. The Station Discrimination Index (d) has a range of -1.00 to +1.00 (good range is between 0.3-0.5). Only station 2 had d within the good range with 0.3. d in

station 1 was negative. The situation in station 1 was such that the academically strong students were failing and the weak students were passing the tasks. The tasks in station 1 should be discarded. In the Statistical Significance analysis, the mean of those who passed was significantly higher than the mean of those who failed the OBGY OSCE only in station 4 (see Table 4). Pass/Fail means were not significantly different in the other stations. The pass/fail means were the same in station 3.

Reliability Checks:

The total alpha coefficient for the OBGY stations was 0.008, which was very poor. Stations 1 and 2 have negative correlations with the total alpha. Stations 3 and 4 had positive, but very poor correlations with the total alpha. Station 1 had the poorest correlation with total alpha with -0.1. When each station was deleted in turn, alpha in stations 1 and 2 improved to above the total alpha, while the alpha in stations 3 and 4 decreased. (see Table 5). Pearson's correlation (r) and contribution to the total score variance was significantly high in station 4 with 0.9. r is also a measure of the internal consistency in the stations. Correlations and contributions were low in the other stations (see Table 5). There was no relationship between the scores from the written stations and those from the manned stations(r=0.00), however, the mean of the written stations (4.0) was lower than that of the manned stations (6.6).

Identifying Hawks and Doves in OBGY OSCE stations: Station 3 examiner was a Dove based on the following evidences: The scores distribution was significantly negatively skewed. Has excess kurtosis. Have several outliers. Has significantly higher meant than the rest of the stations (see Table 1). Variability was high (see Table 1). IDI was too high. Discriminating power was very poor (see Table 4). Correlations and contribution were low (see Table 5).

Table 1: descriptive statistics of the Obgy Osce scores of 10 students in kIU-d, july 2015

Test	Station 1	Station 2	Station 3	Station 4	Total
Number of Tasks	3	3	2	2	10
Skewness	-1	-1.8	-1.8	0.1	0.5
Kurtosis	2.3	1.4	3.4	-1	-0.8
Outliers	Bottom	Bottom	Bottom	—	—
Mean	3.4	4.8	9.1	4.4	5.4
Median	3.5	5	9.5	5	5.3
Mode	3	5	10	6	4.3
S. Setting	3.5	5	9.5	4.5	5.3
Range	4	1	4	10	2.8
S. Deviation	1.1	0.4	1.3	3.4	1.0
SEM	0.3	0.1	0.4	1.1	0.3
95% CI(Mean)	2.6-4.2	4.5-5.1	8.2-10.0	2.0-6.8	4.8-6.1
C. Variation (%)	31.5	8.8	14.2	76.6	17.5

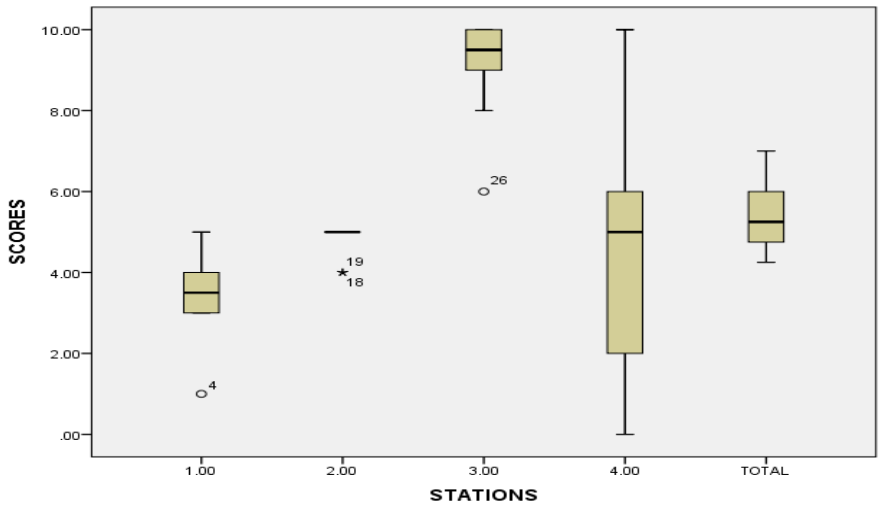

Fig 1: checking for outliers in OBGY OSCE scores of 10 students in KIU-D, JULY 2015

Table 2: Anova table of the OBGY Osce scores of 10 students in kIU-d, July 2015

Variance	SS*	Df**	MS***	F****	Sig*****
Between Groups	190.5(59.6%)	3	63.5	17.7	0.0
Within Groups	129.3(40.4%)	36	3.6		
Total	319.8(100%)	39			
	Turkey's HSD Post Hoc test= Station 3 mean is significantly higher than the rest.				

SS*: Sum of squares; Df**: Degrees of freedom; MS***: Mean of Squares; F****: Ratio of between mean of squares and within mean of squares; SS*****: Significance.

Table 3: Generalizability studies of the OBGY OSCE scores of 10 students in KIU-D, July 2015

	Generalizability studies: Variance components estimates				
	Components	**%**			
Students	0.0	0.1%			
Examiners	6	89.2%			
Students*Examiners	3.6/5 examiners	10.7%			
Total	6.7				
	G-coefficient= 0.0/6.715=0.0. Errors from examiners was 89.2%				

Table 4: Station analysis of the OBGY OSCE scores of 10 students in KIU-D, July 2015

	Stations					
Test	1		2	3	4	Total
IDI	0.1		0.8	1	0.5	0.7
D	-0.3		0.3	0	1	
S. Sig (P/F)	T=0.5		T=1.3	T=0.3,	T=5.5*	

*p is 0.00

Table 5: reliability estimates of the OBGY OSCE scores of 10 students in KIU-D, July 2015.

	Stations				
Test	1	2	3	4	Total
α Correlation	-0.1	-0.0	0.0	0.1	0.0
α Deleted	0.1	0.0	-0.0	-0.7	
Pearson Corr(r)	r=0.1	r=0.1	r=0.4	r=0.9*	
r^2	0.0	0.0	0.1	0.9	

*p is 0.00. Guidelines for the Interpretation of correlation coefficients: 0.75-1.00: strong; 0.50-0.74: moderate to high; 0.25-0.49: low to moderate; 0.00-0.24: weak.

Practice Points

- A perfect examination is practically impossible.
- No examination is exactly the same.
- Regular psychometric analysis of the OSCEs and other assessments is a must especially in resource limited medical schools.
- Examiners need psychometric tools which is stable and objective to validly evaluate assessments.
- Psychometric analysis of assessments helps to harmonise the competencyies of our graduates.
- Pass/fail decisions should be based on standardised passmarks and not on fixed university marks.
- The university grades should be based on z-scores rather than raw marks.
- There is ned for extensive training of teachers in resource limited medical schools in the OSCEs, global scoring and psychometric analysis
- More studies need to be carried in resource limited medical schools in Sub-Saharan Africa to further investigate the properties of their OSCEs.

Fig 1: Practice Points

DISCUSSION

East Africa is a challenged region in several areas which includes education, health, resource personnel, political and economic wise. KIU is one of the very few universities with a medical school in East Africa. OSCE is a very new ideology in the region of East Africa. Moreover, our examiners were not familiar with and the MoH checklist did not capture global scoring and therefore could not supply it despite the pre-OSCE briefing. The OBGY department in KIU as well as in other medical schools of this region face huge human resource challenge. All the internal examiners that participated in this OSCE were part-time staff of the university. The department could not generate sufficient number of stations because of lack of a stable fulltime staff to manage the affairs of the department. The 4 OSCE stations covered barely 50% of the syllabus. This poor coverage could have been as a result of lack of blueprinting before setting up the OSCE. The 2 senior examiners supervised the manned stations while the medical officers (junior examiners) co-marked the written stations. From the analysis, the OSCE was too easy, the mean was not centrally located and the CV was too high for a criterion-referenced test. The university currently uses a fixed pass mark of 50% (based on an ideal examination setting) to make pass/fail decisions and raw scores to grade students' peformances. As demonstrated in the study, this OSCE is far from being ideal due to the presence of skewness, kurtosis and outliers, hence it is more appropriate to use the standardized pass mark to inform pass/fail decision and the normalized (z) scores to grade students' performances. The non-overlapping confidence interval of station 3 suggested that the tasks in station 3 were probably of a different construct from the others. The very high SEM of station 4 suggests that the performances of the students in that station were not representative of the population. The ANOVA and the G-study suggests that there were strong external factors influencing the students' scores. These external factors include the examiners, the environment where the OSCE was carried out and the test itself. The internal consistency based on the alpha (-0.1 to +0.1) and the Pearson's correlation (0.1 to 0.4) were weak in stations 1-3, but significantly strong in station 4 (r=0.9). In this study, the overall alpha was 0.008, which is very low in comparison to the 0.64 obtained in the most stringent ECFMG OSCEs. With the item analysis and reliability checks, station 4 was the only strong and useful station in the OSCE. Stations 1 and 3 need to be discarded, station 2 reviewed and station 4 stored in the OSCE bank for future use. The examiner/s (especially the ones that marked station 3) need to undergo training [6]. Of concern is that there was no correlation between the scores from the written and manned stations and the mean of the written stations was lower than that of the manned stations suggesting that the knowledge base of the students was poor compared to their clinical skills achievement? This can be explained by the absence of a fulltime staff in the department and therefore delivery of knowledge during class lectures perhaps may have been significantly deficient. The limitations to the study include the absence of a Teaching hospital for the university, the examiners' checklist from the Ministry of Health did not cover global scoring and therefore could not be analysed alongside with the checklist scores. Moreover, we did not have direct access to the completed examiners' checklists; hence detailed analysis of the checklists could not be carried out.

CONCLUSION

This study evaluated an end of year OSCE in a resource constrained medical school in Tanzania. Published work on the OSCE practices in medical schools in the sub-Saharan region is very scarce. Examiners need robust psychometric tools to express valid, objective and consistent opinion of assessments that will lead to long term improvement of our examinations. The properties of the OSCE experienced in this study are not similar to those practiced in the established medical schools in the developed countries. Further studies need to be done on other examinations and in other medical schools in the same region to shed more light on the quality of the assessments used to evaluate future medical doctors in this region.

ACKNOWLEDGEMENTS

I wish to thank the ICT department of Kampala International University, Dar es Salaam for ensuring that my office has access to the internet to speed up this work.

REFERENCES

1. Arnold L. Assessing professional behavior: yesterday, today, and tomorrow. Academic medicine. 2002 Jun 1; 77(6):502-15.
2. Aranda, S. and Yates P. 2009. An overview of Assessment. Canberra: The National Cancer Nursing Education Project (Ed Ca N), Cancer Nursing. National Education Framework. Australia. Pg 2.
3. Mccrorie P, Boursicot KA. Variations in medical school graduating examinations in the United Kingdom: Are clinical competence standards comparable?*. Medical teacher. 2009 Jan 1; 31(3):223-9.
4. Crossley J, Davies H, Humphris G, Jolly B. Generalisability: a key to unlock professional assessment. Medical education. 2002 Oct 1; 36(10):972-8.
5. Pell G, Fuller R, Homer M, Roberts T. How to measure the quality of the OSCE: A review of metrics–AMEE guide no. 49. Medical teacher. 2010 Oct 1; 32(10):802-11.
6. Tavakol M, Dennick R. Post-examination interpretation of objective test data: Monitoring and

improving the quality of high-stakes examinations: AMEE Guide No. 66. Medical teacher. 2012 Mar 1; 34(3):e161-75.