DOI: 10.21276/sjams.2016.4.11.3

Scholars Journal of Applied Medical Sciences (SJAMS) **ISSN 2320-6691 (Online)**
Sch. J. App. Med. Sci., 2016; 4(11A):3868-3875 **ISSN 2347-954X (Print)**

(An International Publisher for Academic and Scientific Resources)
www.saspublisher.com

Original Research Article

Psychometric properties of the Objective structured clinical examination in the Paediatrics 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,
[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 aim is to describe the psychometric properties of the July 2015 Paediatrics OSCE scores of 19, 3rd year Clinical Medicine students in a resource-limited medical school in Tanzania, East Africa, with the goal of improving assessment. This descriptive and cross-sectional study used literature review and structured observation as data collection methods. Students' performances were assessed using checklists by 4 examiners in 27 OSCE stations. Stations 24-27 were manned and stations 1-23 were in written format. Statistical analysis was carried out using SPSS computer packages. The written stations were too many in proportion to the clinical stations for an OSCE. The mean scores in the stations were between 5-78% and pass mark was set between 0-97percent. The tasks were either too easy or too difficulty in 26% (7stations, n=27) of the stations, the variance was generally too high for a criterion referenced test and none of the means was centrally located. Station 19 was extremely difficult. Examiners' error was moderately high at 66%, stability was too high (Alpha was 0.6, ideal is up to 0.5) and internal consistency ranged from -0.4 to 0.6 (alpha). G-coefficient was low at 0.22. Item analysis was good in 30% (8) but poor in 19% (5) of the stations. The analysis recommended that 19% of the stations was to be discarded, 52% of the stations to be reviewed and 30% of the station (whose properties were good) were to be banked for future use.
Keywords: OSCE scores, SPSS computer packages

INTRODUCTION

The objective structured clinical examination (OSCE) is increasingly being used as a major method of clinical assessment in sub Saharan Africa, yet its psychometric characteristics have not been well documented because its implementation can be challenging especially in resource limited institutions [1]. A number of issues (such as station length which may be too short to achieve reliable results, how to define pass/fail criteria, to mention a few) have been raised about how well the OSCEs practiced in resource limited medical schools is able to measure knowledge and skills necessary for competent clinical practice. The quality assurance machinery in place in most medical schools currently, involves mainly human raters which can be subjective, biased and inconsistent [2]. 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 [3].

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 the Paediatric department of a resource limited medical school in East Africa (Dar es Salaam, Tanzania) was described. 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.

METHODS

This study took place in Kampala International University, Dar es Salaam campus (hereafter referred to as KIU-D) but the OSCE was implemented in Amana Regional Referral Hospital. Kampala International University is a multi-campi university with campuses in Uganda, Tanzania and Kenya. 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 4 examiners in 27 OSCE stations in the paediatric ward. There were 2 medical officers, one specialist and one consultant. The consultant was the external examiner. Stations 24-27 were manned with real patients and stations 1-23 were in written format which included photographs. The study population was described as the current, active third year clinical medicine students, who passed the paediatric 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 19 of the 30 students in the faculty register were eligible to sit for the OSCE and all these 19 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, the examiners and students were briefed for 45 minutes and their consent for the research was obtained. The programme coordinator served the checklists to the examiners, and the students were ushered to the hospital paediatrics general ward according to the exam schedule.

The OSCE began at 09.00hrs. The live OSCE set-up (station design and station contents) and proceedings were documented 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. 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 testis 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 Paediatrics department of the KIU-D at the Amana regional referral Hospital is shown below.

Sociodemographic Characteristics of the Examiners and Students

There were 19 students (3 female and 16 male) amongst those eligible to sit for the Paediatrics OSCE. Of the 4 examiners who participated, 2 were Medical officers, 1 was a Specialist and the external examiner was a consultant. All the examiners and students were Tanzanians except for the specialist, who was a Nigerian.

OSCE Observations

The 19 eligible students worked on 56 tasks altogether distributed in each of the 27 OSCE stations. Each station was 5 minutes long, had between 1-4 tasks and the whole OSCE lasted 3hours. The manned stations were designed around the selected real patients' beds and the unmanned or written stations were set up on tables in the centre of the ward. Stations 24-27 were manned and stations 1-23 were in written format. The tasks in the manned stations were on history taking, physical examination covering the general, respiratory, cardiovascular and gastrointestinal systems only, diagnosis and procedural. The external examiner and the Specialist supervised the manned stations, while the medical officers marked the written stations after the OSCE.

Paediatrics Stations Metrics

The Paediatric 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.

In Table 1 below, the pattern of distribution of the scores were described by the skewness, kurtosis, checking for outliers and z-scores. The distribution of scores in all the stations was skewed but significantly so in 7 of them. Stations 19 and 26 were both significantly skewed and kurtosed. Station 19 was positively skewed while station 26 was negatively skewed. There were few extremely large values in station 19, Figure 1. After converting the raw scores into z-scores, the extreme values were normalised 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, Table 2. The mean scores ranged from 0.5 to 8.2 and were not centrally located in any of the stations. The standardized pass mark was from 0-9.7 across the stations. The coefficient of variation was generally high but extremely so in 4 stations, Table 1.

ANOVA (comparing means): The variance between the stations was significantly high, (>30%). Moreover, the variance between the station scores was significantly higher than the variance within the stations. The mean of station 19 was significantly lower than the rest (see Table 3).

Generalizability studies (variance components estimates): The students and the examiners contributed 7.6% and 65.8% respectively to the variance obtained in the paediatrics test. The interaction between students and examiners was 26.5%. G-coefficient=0.22 (see Table 4).

Inferential statistics

Station Analysis: Good Item Difficulty Index (IDI) is between 0.3 and 0.8. Station Discrimination Index (d) range from -1.0 to +1.0 (whilst d good range is between 0.3-0.5). Statistical Significance was used to determine whether the mean of those who passed was significantly different from the mean of those who failed the Paediatrics OSCE in each of the stations. Eight stations had good difficulty and discriminatory indices and significant differences in the pass/fail means. Stations 6 and 7 had negative 'd's (-0.2 and -0.4 respectively).

Reliability Checks:

The total alpha coefficient or stability for the Paediatrics OSCE was 0.6. There α correlations ranged from weak to moderate in strength (-0.4 to 0.6). Alpha coefficient (if-item-deleted) revealed redundancy in 13 of the 27 OSCE stations. There Pearson's correlations ranged from weak to moderate in strength (-0.2 to 0.7) with significant relationship with the total scores found in 10 stations.

Identifying Hawks and Doves in Paediatrics OSCE stations**:** Station 19 was strictly marked (hawk) based on the following evidences. The scores distribution in station 19 were significantly positively skewed and kurtosed due to the presence of few extremely large values (see Figure 1), while the majority of the scores were low. The station mean was significantly lower from the rest of the stations (see ANOVA in Table 3). The variability was too high (see Table 1). IDI was outside the good range and discriminating powers were very poor. Correlations and contribution are low (see Table 5).

Table 1: Descriptive statistics of the paediatric OSCE scores of 19 students in kIU-dar es salaam, July 2015.

Tests Stations	Mean	Standard setting	CV (%)	Skewness	Kurtosis	Outliers
1	4.3	4	25.7	-0.2	-0.3	Bottom
2	6.5	6	17.9	-0.3	0.5	Bottom, Top
3	7.8	8	20	-0.2	-1.2	---
4	4.6	5	32.4	-0.6	1.0	Bottom
5	4.9	5	28	-0.1	0.8	--
6	3.8	4	60.7	-0.3	-0.6	--
7	5.8	5.7	28	0.2	-1.4	--
8	1.7	0	174.4[ξ]	1.3*	-0.3	Top
9	5.3	4.9	46	-0.0	0.1	--
10	4.4	4.3	68.3	0.1	-0.2	Bottom, Top
11	6.1	6.3	50.7	-0.5	-0.3	--
12	7.3	8.3	44.2	-1.3*	0.9	---
13	7.4	9.7	56.6	-1.2*	-0.4	--
14	2.7	1.7	122.3[ξ]	1.1*	0.3	--
15	3.7	3.7	77.4	0.1	-1.1	--
16	3.4	4	48.7	-0.7	0.4	---
17	2.8	2	110.9[ξ]	0.8	-0.3	--
18	8.2	8	19.7	-0.9	1.2	Bottom
19	0.5	0	303.8[ξ]	3.0*	8.5[ϕ]	Top
20	6.4	7.1	50.5	-0.9	-0.1	Bottom
21	5.7	6.1	36	-0.6	0.2	--
22	5.3	5.4	36.9	-0.3	-0.6	--
23	6.3	6	14.9	-1.1*	0.2	Bottom
24	4.2	5	41.8	-0.7	-0.8	--
25	4.4	5	44.8	-0.6	-1.2	--
26	5.9	6	20.7	-1.6*	2.4[ϕ]	Bottom
27	5.9	6	10.9	0.1	-0.1	Top
Total	5.0	4.9	15.1	0.1	-1.1	--

*significant skewness. [ξ]ignificant kurtosis. [ϕ]xtremely high coefficient of variation.

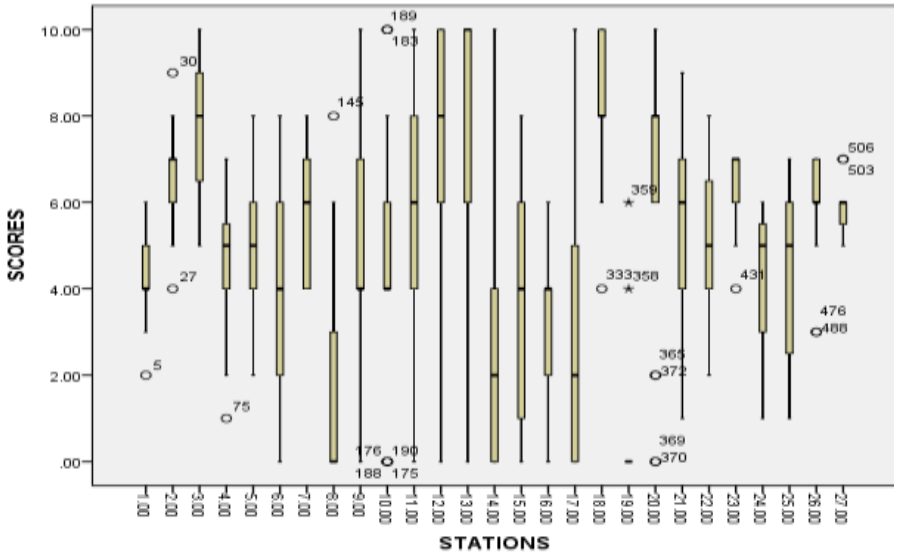

Fig 1: checking for outliers in paediatric stations of 19 students in kiu-dar es salaam, July 2015

Table 2: Summary of paediatrics z-scores of 19 students in kiu-dar es salaam, July 2015.

GRADES	GOOD	AVERAGE	POOR	TOTAL
GRADES	C	D	E	
Z-SCORE RANGES	>+1<+2 (13.6%)	+1≤0≥-1 (68.3%)	<-1>-2 (13.6%)	100%
Frequency	4(21%)	11(57.9%)	4(21%)	19(100%)
Raw Score Equivalent	5.8-6.2	4.4-5.7	3.7-4.1	

Table 3: Anova Table for Paediatrics Osce of 19 Students in Kiu-Dar Es Salaam, July 2015.

Variance	SS	Df	MS	F	Sig
Between Groups	1661.2(39.3%)	26	63.89	11.91	0.00
Within Groups	2569.1(60.7%)	479	5.36		
Total	4230.3(100%)	505			
	Turkey's HSD Post Hoc test: Station 19 mean was significantly lower than the rest of the stations.				

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 4: Generalizability Studies for Paediatrics OSCE of 19 Students in Kiu-Dar Es Salaam, July 2015.

	Components	**%**			
Students	0.36	7.6%			
Examiners	3.1	65.8%			
Students*Examiners	5/4 examiners	26.5%			
Total	4.71				
	G-coefficient= 0.36/1.61=0.22. Errors from examiners was 65.8%				

Table 5: Paediatrics OSCE stations analysis and correlations of 19 students in kiu-dar es salaam, July 2015.

Stations	IDI	d	S. Sig	α Corr.	α deleted	P.Corr(r)	r^2
1*	0.4	0.4	2.3	0.5	0.5	0.6$^{\wr}$	0.4
2	1.0	0.2	2.8$^{\wr}$	0.1	0.6	0.3	0.1
3$^{\phi}$	1.0	0.0	-0.3	-0.4	0.6	-0.2	0.0
4*	0.5	0.6	3.3$^{\wr}$	0.4	0.5	0.5$^{\wr}$	0.2
5*	0.7	0.6	5.3$^{\wr}$	0.4	0.5	0.6$^{\wr}$	0.4
6	0.3	-0.2	0.9	-0.0	0.6	0.1	0.0
7	0.6	-0.4	1.0	-0.2	0.6	0.0	0.0
8	0.3	0.6	-1.3	0.4	0.6	0.6$^{\wr}$	0.4
9	0.5	0.6	2.0	0.6	0.5	0.5$^{\wr}$	0.3
10	0.4	0.2	-0.2	0.0	0.6	0.2	0.1
11	0.6	0.4	0.2	0.1	0.6	0.4	0.1
12*	0.8	0.4	9.3$^{\wr}$	0.6	0.5	0.5$^{\wr}$	0.3
13*	0.7	0.6	13$^{\wr}$	0.5	0.5	0.6$^{\wr}$	0.3
14$^{\phi}$	0.2	0.2	1.0	-0.2	0.7	0.1	0.0
15	0.3	0.4	1.2	0.2	0.6	0.2	0.0
16$^{\phi}$	0.1	0.2	0.1	-0.2	0.6	0.1	0.0
17	0.3	0.2	1.8	0.1	0.6	0.3	0.1
18	1.0	0.2	3.6$^{\wr}$	0.5	0.6	0.5$^{\wr}$	0.2
19$^{\phi}$	0.1	0.2	1.5	0.2	0.6	0.3	0.1
20*	0.8	0.6	3.9$^{\wr}$	0.4	0.6	0.6$^{\wr}$	0.3
21*	0.7	0.6	6.5$^{\wr}$	0.5	0.6	0.7$^{\wr}$	0.5
22	0.7	0.2	1.4	0.2	0.6	0.2	0.1
23	1.0	0.0	2.5$^{\wr}$	0.1	0.6	0.1	0.0
24	0.5	0.4	1.3	0.3	0.6	0.4	0.2
25	0.6	0.6	0.8	0.1	0.6	0.2	0.1
26*	0.9	0.4	4.6$^{\wr}$	0.3	0.6	0.3	0.1
27$^{\phi}$	1.0	0.2	-1.0	-0.0	0.6	-0.2	0.1
Total	0.4			0.6			1.00

*stations with good difficulty index, discriminative index and statistical significance

$^{\phi}$ stations with poor difficulty index, discriminative index and statistical significance.

$^{\wr}$ significant p-value

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 2: 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 private 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 paediatrics department in KIU as well as in other medical schools of this region face huge human resource challenge. Amongst the three university teaching staff in the department, 2 were fulltime, 1 part-timer and at the time of this study, there was only one specialist. The department could not generate sufficient number of manned clinical stations because of the few number of examiners and the lack of standardized patients and blueprint. These deficiencies might negatively impact the validity of the OSCE. In this OSCE, standardized patients were not used because of the difficulty of recruiting well children to act as patients. Standardised patients (SPs) have been widely used to assess physicians' clinical competence. However, in paediatrics, the use of children in such a way has long been questioned with regard to ethics and the examination quality (validity, reliability, and feasibility) [4]. The manned stations covered barely 50% of the syllabus, but the other stations probably covered the rest. The 2 senior examiners supervised the manned stations while the medical officers (junior examiners) co-marked the written stations.

To achieve optimal patient outcomes and avoid medical errors in clinical practice, clinical assessments must be evaluated objectively and regularly. Hence, KIU-D is adopting the culture of regularly evaluating her assessments with the view of improving them and producing quality graduates. One of the major drawbacks of the OSCEs is the compartmentalisation of knowledge and discouraging candidates from looking broadly at patient's difficulties. Even Harden advocated additional testing using a long case or some form of work based assessment when OSCEs were used [5]. Hence, OSCE in combination with the long case assessment method has been adopted by KIU-D for clinical assessment. This study focused on describing the psychometric properties of the end of year OSCE administered on 19 third year undergraduate clinical students in the Paediatric department of KIU-D in July 2015.

The assessment depth in this promotional OSCE is likely to be superficial because the station length was uniform at 5minutes regardless of the nature of the task in the stations. Stations of 10 minutes or less within the OSCE inevitably mean that small component parts of paediatric skills will be tested and a holistic assessment of the whole person is unlikely to be possible [5]. Moreover, because of the excessive number of stations, examinee fatigue is a very likely facet influencing the scores. From the analysis, the mean was not centrally located and the coefficient of variation was too high for a criterion-referenced test [6, 7]. Moreover, station 19 was excessively difficult for the students. 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 not exactly an ideal assessment due to the presence of skewness, kurtosis and outliers [8], hence it is more appropriate to use the standardized pass mark (similar to the median in computation) to inform pass/fail decision and the normalized (z) scores to grade students' performances [9]. The ANOVA (in-between variance was more than 30%) and the G-coefficient (0.22) suggests that there were strong external factors influencing the students' scores such as errors from the examiners (65.8%) [3]. These external/extraneous factors may also include the environment where the OSCE was carried out, examinee's fatigue and the test itself. The examiners' errors documented in this study was very high compared to those recorded by several authors such as 12% in Britain and the USA, 10-17% in the Medical Council of Canada (MCC) examinations [10], 10.9% in the ECFMG Clinical Skills Assessment (CSA) [11] and 8.9% in Australia[12]. G-coefficient of 0.48 to 0.80 was documented and the bulk of correlation coefficients in the OSCEs, including high stakes examinations in developed medical schools is between 0.5-0.6 (moderate reliability) and do not reach the reliability coefficient threshold of 0.8 or over which is widely regarded as the marker of sufficiency[5, 13, 14]. The internal consistency was moderate-weak based on the alpha (-0.4 to +0.6) and the Pearson's correlation (-0.2 to 0.7) in this study compared to those of resource-rich medical schools in Saudi Arabia (0.76) [15]. In this study, stability was 0.6, which was comparable to the 0.64 obtained even in the most stringent ECFMG OSCEs. With the item analysis and reliability checks, only eight stations (30%, n=27 stations) had strong and useful tasks and therefore can be stored away in the OSCE bank for future use. Meanwhile 5 stations (19%, n=27 stations) need to be discarded and replaced with new tasks and the rest of the stations as indicated in Table 5 should be reviewed accordingly [16]. The examiner/s (especially the ones that marked station 19-hawkishly) need to undergo training [17, 10]. Contrary to earlier documentation that overall scores on the OSCE are often not very unreliable [18], in this study, we discovered that the overall scores per candidate were more stable than individual station scores; hence pass/fail decisions could be made conveniently based on the overall candidate scores.

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. 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. Odoi Adome R, Kitutu F. Creating an OSCE/OSPE in a resource-limited setting. Medical education. 2008 May 1; 42(5):525-6.
2. Crossley J, Davies H, Humphris G, Jolly B. Generalisability: a key to unlock professional assessment. Medical education. 2002 Oct 1; 36(10):972-8.
3. 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.
4. Tsai TC. Using children as standardised patients for assessing clinical competence in paediatrics. Archives of disease in childhood. 2004 Dec 1; 89(12):1117-20.
5. Marwaha S. Objective Structured Clinical Examinations (OSCEs), psychiatry and the Clinical assessment of Skills and Competencies (CASC) same evidence, different judgement. BMC psychiatry. 2011 May 16; 11(1):1.
6. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3118176/
7. Avijiit Hazra. Nithya, Gogtay. IJD Module on Biostatistics and Research Methodology for the Dermatologist. Indian Journal of Dermatology. 2016; 61(1): 10-20.
8. Kobayashi K, Sakuratani Y, Abe T, Yamazaki K, Nishikawa S, Yamada J, Hirose A, Kamata E, Hayashi M. Influence of coefficient of variation in determining significant difference of quantitative values obtained from 28-day repeated-dose toxicity studies in rats. The Journal of toxicological sciences. 2011; 36(1):63-71.
9. Kim HY. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restorative dentistry & endodontics. 2013 Feb 1; 38(1):52-4.
10. Taylor Courtney. How Do We Determine What Is an Outlier? About education. 2014.
11. Retrieved on 24 June 2016. Available from: (http://statistics.about.com/od/Descriptive-Statistics/a/How-Do-We-Determine-What-Is-An-Outlier.htm)
12. Bartman I, Roy M, Smee S. Catching the hawks and doves: a method for identifying extreme examiners on objective structured clinical examinations. Ottawa: Medical Council of Canada. 2011.

13. Arnold L. Assessing professional behavior: yesterday, today, and tomorrow. Academic medicine. 2002 Jun 1; 77(6):502-15.
14. Boulet JR, McKinley DW, Whelan GP, Hambleton RK. Quality assurance methods for performance-based assessments. Advances in Health Sciences Education. 2003 Mar 1; 8(1):27-47.
15. Roberts C, Rothnie I, Zoanetti N, Crossley J. Should candidate scores be adjusted for interviewer stringency or leniency in the multiple mini-interview? Medical education. 2010 Jul 1; 44(7):690-8.
16. Auewarakul C, Downing SM, Praditsuwan R, Jaturatamrong U. Item analysis to improve reliability for an internal medicine undergraduate OSCE. Advances in health sciences education. 2005 Jun 1; 10(2):105-13.
17. Tavakol M, Dennick R. Making sense of Cronbach's alpha. International journal of medical education. 2011; 2:53.
18. Al-Naami MY, El-Tinay OF, Khairy GA, Mofti SS, Anjum MN. Improvement of psychometric properties of the objective structured clinical examination when assessing problem solving skills of surgical clerkship. Saudi medical journal. 2011; 32(3):300-4.
19. Tavakol M, Dennick R. Post-examination analysis of objective tests. Medical Teacher. 2011 Jun 1; 33(6):447-58.
20. 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.
21. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. Medical education. 2011 Dec 1; 45(12):1181-9.