East African Scholars Journal of Education, Humanities and Literature

OPEN ACCESS

Abbreviated Key Title: East African Scholars J Edu Humanit Lit
ISSN: 2617-443X (Print) & ISSN: 2617-7250 (Online)
Published By East African Scholars Publisher, Kenya

Volume-2 | Issue-6| June- 2020 | DOI: 10.36349/EASJEHL.2020.v03i06.010

Research Article

Comparison of a Task-Specific Checklist and End Exam Global Rating Scale for Scoring the Objective Structured Clinical Examination Used To Evaluate Sixth Year Medical Students in Surgery at Shendi University, Sudan

Mohanned Omer Abass[1]* and Mohamed Elimam Mohamed Ahmed[2]

[1]MD, Assistant Professor of Surgery, Shendi University, Sudan
[2]Professor of Urology, Faculty of Medicine, University of Gezira, Sudan

Article History
Received: 02.06.2020
Accepted:13.06.2020
Published: 23.06.2020

Journal homepage:
https://www.easpublisher.com/easjhcs

Quick Response Code

Abstract: ***Background:*** Objective structured clinical examination (OSCE) is one of the modern tools used to assess students' clinical competencies. The scoring systems for OSCEs vary widely, which may influence the reliability of this assessment tool. Both checklists and global rating scales (GRS) are widely used as scoring methods for student performance in the OSCE. ***Objective:*** To compare a task-specific checklist and end exam GRS in the scoring of OSCE for the evaluation of sixth year medical students in surgery at Shendi University in 2017 and 2018. ***Methodology:*** This cross-sectional study compared a task-specific checklist and end exam GRS scores in the OSCE for the evaluation of sixth year medical students in general surgery at the Faculty of Medicine of Shendi University over two consecutive years (2017 and 2018).The results from six stations were analyzed, three from each year. SPSS was used for data analysis. Cronbach's alpha was used to assess OSCE reliability and Spearman's correlation coefficient was used to identify correlations between the scoring methods. ***Results:*** The reliability of both scoring methods was 0.60 across all stations. The majority of students scored B or above (61%) when using the checklist while 46.5% scored B or above when the GRS was used. Spearman's rho correlation coefficients (ρ) between checklist scores and GRS scores for the six stations ranged between 0.63 and0.88, and these correlations were found to be statistically significant ($p = 0.001$). ***Conclusions:*** The OSCE is a feasible and reliable approach for assessment of the clinical competence of undergraduate medical students. There was a strong correlation between the two scoring systems (checklist and GRS). Based on the results of this study, both ways of scoring the OSCE are considered acceptable.

Keywords: Objective structured clinical examination (OSCE); Checklist; Global rating scale (GRS), Correlation.

INTRODUCTION

Methods for assessing students in medical education have changed dramatically over the past few decades. These methods have advanced from standard pen-and-paper tests of knowledge toward a more complex system of evaluation(Howley, L. D. 2004).

The goal of assessment in medical education remains the development of reliable measurements of student performance which, in addition to having predictive value for subsequent clinical competencies, also have a formative, educational role(Wass, V. *et al.*, 2001).

One of the modern ways to assess students' clinical competencies is by objective structured clinical examination (OSCE). It was first described by Harden and Gleeson in 1975 (Harden, R. M. *et al.*, 1975; & Harden, R. M., & Gleeson, F. A. 1979) and has become one of the most widely used assessment methods.

The OSCE is a formal examination with a specified set of tasks that every candidate is expected to perform in the presence of examiners. These tasks are uniform for all and highly structured, and the expectations of the candidates are explicit and unambiguous (Reronr, R. 1998).It involves the assessment of clinical skills and behaviors throughout a series of timed stations requiring the demonstration of practical skills, problem-solving strategies and behaviors (Harden, R. M., & Gleeson, F. A. 1979).

Scoring methods for OSCEs vary widely, influencing the reliability of this assessment tool. Checklists have been standardized in many established OSCE programs and have intuitive value as an assessment tool; however, they also have many limitations. Another method of scoring is a global rating scale (GRS), scored by experts, which has been claimed to show higher interstation reliability, better construct validity and concurrent validity than checklists (Turner, J. L., & Dankoski, M. E. 2008). Global ratings are station-independent scales that identify areas of competence such as organization, communication and rapport, together with similar constructs that may not be captured in a set of binary checklist items. When examining performance, global ratings of OSCE performance show impressive psychometric characteristics, whether used in conjunction with a checklist or on their own (Hodges, B., & McIlroy, J.H. 2003).

The correlation between OSCE scoring using a checklist and a GRS was previously elucidated in a study that demonstrated that the GRS is comparable to a checklist for evaluating the professional behavior of physical therapy students. The correlation between the checklist and GRS appears to become stronger when assessing more advanced students (Turner, K. *et al.*, 2014).

The purpose of this study was to determine the correlations between the standard checklist and GRS as scoring methods in the OSCE for surgery in final year medical students over two years(2017 and 2018).

METHODOLOGY

This descriptive cross-sectional study compared a specific checklist and end exam GRS in the OSCE for surgery for the assessment of sixth year medical students in the Faculty of Medicine, Shendi University, in two consecutive years (2017 and 2018).

The surgery OSCE in the Faculty of Medicine of Shendi University comprises seven active stations across three domains: three stations for history taking, three for performing clinical examination, and one procedural/communication skills station. There are also five rest stations, making a total of 12 stations per set.

Four sets containing the same 12 stations in the same order run simultaneously. The time allocated for each station is 8 minutes. The whole batch exam includes only two rounds, starting in the morning and ending in the afternoon.

The OSCE stations were reviewed by an external examiner who was selected by the faculty in collaboration with the Sudanese Medical Council to follow the intended national standards. Then, a group of expert external and internal examiners determined the minimal pass level for each station (using the modified Angoff's method) and the mean was calculated to decide the cut-off for a pass grade in the overall performance of the OSCE. A briefing session on the OSCE was held for examiners before the exam, then they were distributed to the stations.

The students had gone through an OSCE in their fifth year, simulating their final year OSCE (mock exam), and were trained to go through this type of exam.

A stratified random sample was taken for each year.
In the year 2017:

- station 1, focused clinical examination of a lump;
- station 2, focused history of a patient presenting with obstructive jaundice;
- station 3, procedural (conducting primary survey in a polytrauma patient).

In the year 2018:

- station 1, focused history of a patient presenting with hematemesis;
- station 2, focused clinical examination of hernia;
- station 3, breaking bad news.

In the active stations, the student is requested to perform a certain task for which the examiner gives a mark using two types of scoring systems, namely a task-specific checklist atthe beginning of scoring and GRS at the end. The checklist includes detailed steps specific to each station which the candidate needs to complete in order to earn marks.

RESULTS

Overall, 81 students in the year 2017 and 96 students in the year 2018 were included in the study. The OSCE scores of those students were taken from three stations for the two batches (2017 and 2018), totaling six stations, and analyzed. The two scoring systems (checklist scores and end exam GRS) were compared.

Thirty examiners with a mean of 15 years post-qualification experience (range 1–40 years) participated in the 2017 OSCE, while 31 examiners with a mean of 13 years post-qualification experience (range 1–41 years) participated in the 2018 OSCE.

Reliability

Cronbach's alpha was used to test the reliability of the two scoring systems (checklist and end exam GRS) for the six stations, which revealed that the reliability of both scoring methods across all stations was 0.60 (range 0.71 to 0.84), as shown in Table 1.

Table 1. The reliability of objective structured clinical examination using Cronbach's alpha.

	Station 1, 2017	Station 2, 2017	Station 3, 2017	Station 4, 2018	Station 5, 2018	Station 6, 2018	All stations
Reliability	0.92	0.77	0.80	0.80	0.84	0.71	0.60

Students' grades using checklist scores

When using the checklist, most students scored A, with a percentage of about 39%, followed by grade B, which was about 22%, and finally grade F, which was about 9%, as shown in Figure 1.

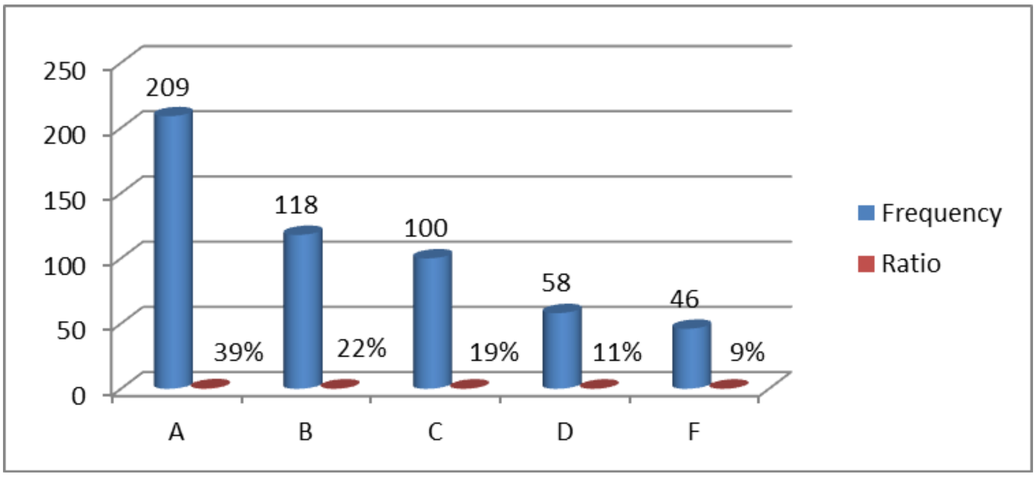

Figure 1. Students' grades assessed using the checklist method for all stations.

Students' grades using end exam global rating scale scores

When scored according to the end exam GRS, most students scored C, with a percentage of about 37.3%, followed by grade B, which was about 30.3%, and finally grade F, which was about 3.2%, as presented in Figure 2.

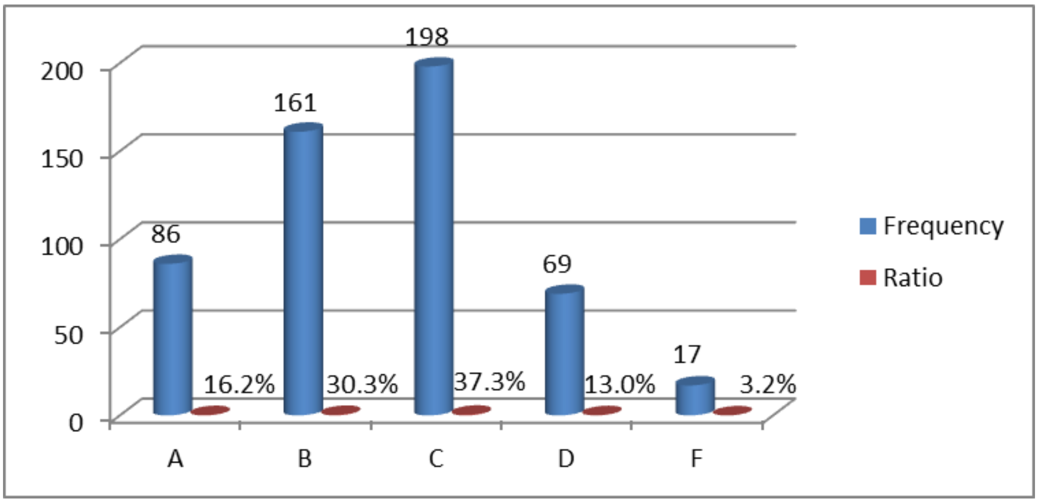

Figure 2. Students' grades using the end exam global rating scale method for all stations.

Correlation between the two scoring systems

Spearman's rho correlation coefficients (ρ) between checklist scores and end exam GRS scores for the six stations ranged between 0.63 to 0.88, and the correlations were found to be statistically significant (p = 0.000), as shown in Table 3.

Table 3. Correlation coefficients between the two evaluation methods

		2017		
		N	**Correlation**	**P-value**
Station 1	Checklist and GRS	81	0.88	0.000
Station 2	Checklist and GRS	81	0.64	0.000
Station 3	Checklist and GRS	81	0.72	0.000
		2018		
		N	**Correlation**	**P-value**
Station 1	Checklist and GRS	96	0.68	0.000
Station 2	Checklist and GRS	96	0.73	0.000
Station 3	Checklist and GRS	96	0.63	0.000

GRS: Global Rating Scale

DISCUSSION

The OSCE is widely used in health professions to assess students' clinical competencies. It has been an integral part of the assessment of clinical competencies of medical students in the Faculty of Medicine of the University of Shendi for the past 10 years. During this period, the OSCE construction and scoring system has undergone considerable changes to achieve the highest levels of validity and reliability, as well as to meet the learning outcomes and community needs.

Reliability of the objective structured clinical examination

Reliability refers to the reproducibility of a set of measurements, consistency or stability of measures over time and over test forms, including different samples of items(Kim, J. *et al.,* 2009).

In this study, the reliability assessment of OSCE for all stations gave an alpha value of 0.6, indicating that the OSCE as a whole had an acceptable reliability. A study by Joong *et al.,*. reported an alpha value of 0.68 across 16 stations for 185 candidates(Barman, A. 2005). Also, in a systematic review on the reliability of OSCE, the overall alpha value across stations was 0.66 (95% confidence interval [CI] 0.62–0.70) and the overall alpha within stations across items was 0.78 (95% CI 0.73–0.82). The better than average reliability in this study was attributed a greater number of stations and a higher number of examiners per station (Sim, J. H. *et al.,* 2015). In another study, the overall Cronbach's alpha was 0.66 (varying from 0.56 to 0.79) and 0.59 (varying from 0.51 to 0.66) in a summative OSCE over two consecutive years (Brannick, M. T. *et al.,* 2011).

The scoring method used in the OSCE has a great impact on student behavior. A study by McIlroy *et al.,*. found that students adapt their behaviors to the system of evaluation, thus, it is possible that their awareness of marking in the OSCE may affect the way they learn medicine. Many factors have been found to affect the reliability of the scoring method for the OSCE. For example, a study that assessed student perceptions of the evaluation method showed that an individual's ability to adapt to the system of global rating forms is relatively station specific, possibly depending on their expertise in the domain represented in each station (Chisnall, B. *et al.,* 2015).

Students' scores in the objective structured clinical examination

A large proportion of students scored B or above (61%) when using the checklist, while 46.5% scored B or above when the GRS was used. In a study that reported the experience of OSCE at the Faculty of Medicine of Alzaeim Alazhari University, Sudan, 67.7% of students were awarded grade B or above, which is higher than in the present study. The scoring system used in this previous study was a checklist score(McIlroy, J. H. *et al.,* 2002). Thus, the differences may be attributed to many factors, such as a different exam setting, normal variability between students and the different scoring system.

Another factor was identified in a study by Iris Schleicher *et al.,*., who reported weak inter-rater reliability, found that the scoring was correlated only weakly with the examiner's level of experience, and identified some gender effects. The findings of examiner effects, even in standardized situations, may influence the outcome even when students perform equally well. Examiners need to be made aware of these biases prior to examination (Idris, S. A. *et al.,* 2014).

Many studies on the influence of first impressions on subsequent ratings in OSCE have suggested that first impressions could play a role in explaining variability in judgments(Schleicher, I. *et al.,* 2017; & Wood, T. J. *et al.,* 2017).

Correlation between checklist and global rating scale scores

Checklists are used to evaluate the thoroughness of performance, such that greater competence of a skill is evidenced by completion of a greater number of checklist items(Wood, T. J. 2014). On the other hand, the GRS is a non binary evaluative tool, requiring assignment of a score along a predetermined scale of many points to assess the overall performance of the examinee (Cunnington, J. P. W *et al.,* 1996). Given these facts, the two scales should be correlated because they assess the candidate in a specific task in almost the same environment.

Disparities between checklist and GRS scores are illustrated in the study of Pell *et al.*,., which revealed the extent of misalignment between assessors' checklist decisions and their 'predictions' (i.e., the global grades) across a range of different academic cohorts and levels of assessment in a large-scale OSCE. This misalignment could be the result of a number of problems, for example, assessor training, support materials and 'rogue' assessors, and provides deeper insight into stations that might have been judged as 'acceptable' based on pre-existing metrics. The disparity was found to exist mainly in the borderline group(Reznick, R. K. *et al.*, 1998).

In our study, the Spearman's correlation coefficient was higher than 0.5 for all stations, meaning that there is strong correlation between checklist and GRS scores. This correlation was found to be statistically significant.

Given that performance on checklist items is typically consistent throughout OSCEs and that the correlation between the checklist and GRS becomes stronger as students gain experience, checklists may be a redundant tool in the examination of practical skills later in academic programs. Therefore, it has been suggested that as students advance through an educational program and gain more experience, the GRS may be a more appropriate tool to evaluate professional behavior.

Many factors may affect the correlation observed in the current study. Those stations that did not correlate well may have involved more complex skills or problem-solving strategies, potentially leading to completion of the clinical skill at the expense of professional behavior items. Alternatively, examiner bias or error may have accounted for the differences (Pell, G. *et al.*, 2015).

Acknowledgements

Many thanks go to Shams-eldin, the head of the statistics department of Shendi University, for his help and advice in statistical analysis. We also acknowledge and thank the dean of the Faculty of Medicine of Shendi University for providing the resources required to complete this work. Special thanks go to Widad, Fatima, Awatif and Alaa, who assisted with data collection.

Funding: None.

REFERENCES

1. Barman, A. (2005). Critiques on the objective structured clinical examination. *Annals-Academy of Medicine Singapore*, *34*(8), 478-482.
2. Brannick, M. T., Erol‐Korkmaz, H. T., & Prewett, M. (2011). A systematic review of the reliability of objective structured clinical examination scores. *Medical education*, *45*(12), 1181-1189.
3. Chisnall, B., Vince, T., Hall, S., & Tribe, R. (2015). Evaluation of outcomes of a formative objective structured clinical examination for second-year UK medical students. *International journal of medical education*, *6*, 76-83.
4. Cunnington, J. P. W., Neville, A. J., & Norman, G. R. (1996). The risks of thoroughness: reliability and validity of global ratings and checklists in an OSCE. *Advances in Health Sciences Education*, *1*(3), 227-233.
5. Harden, R. M., & Gleeson, F. A. (1979). Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical education*, *13*(1), 39-54.
6. Harden, R. M., Stevenson, M., Downie, W. W., & Wilson, G. M. (1975). Assessment of clinical competence using objective structured examination. *Br Med J*, *1*(5955), 447-451.
7. Hodges, B., & McIlroy, J.H. (2003). Analytic global OSCE ratings are sensitive to level of training. Med Educ, 37, 1012–1016.
8. Howley, L. D. (2004). Performance assessment in medical education: where we've been and where we're going. *Evaluation & the health professions*, *27*(3), 285-303.
9. Idris, S. A., Dimitry, M. E., Mohammed, B. H., Elkheir, I. S., Mansour, A. O., Mutwalli, I. M., ... & Mohamed, A. A. (2014). Experience of OSCE in faculty of medicine, Alzaeim Alazhari University, Sudan. *Asian Pacific Journal of Health Sciences*, *1*(4), 419-424.
10. Kim, J., Neilipovitz, D., Cardinal, P., & Chiu, M. (2009). A comparison of global rating scale and checklist scores in the validation of an evaluation tool to assess performance in the resuscitation of critically ill patients during simulated emergencies (abbreviated as "CRM simulator study IB"). *Simulation in Healthcare*, *4*(1), 6-16.
11. McIlroy, J. H., Hodges, B., McNaughton, N., & Regehr, G. (2002). The effect of candidates' perceptions of the evaluation method on reliability of checklist and global rating scores in an objective structured clinical examination. *Academic Medicine*, *77*(7), 725-728.
12. Pell, G., Homer, M., & Fuller, R. (2015). Investigating disparity between global grades and checklist scores in OSCEs. Medical Teacher 1-25.
13. Reronr, R. (1998). Comparing the psychometric properties of checklists and global rating scales for assessing performance on an GSCE-format examination. *Acad Med*, *73*, 993-997.
14. Reznick, R. K., Regehr, G., Yee, G., Rothman, A., Blackmore, D., & Dauphinee, D. (1998). High-stakes examinations: what do we know about measurement? Process-rating forms versus task-

specific checklists in an OSCE for medical licensure. *Academic Medicine*, *73*(10), S97-99.

15. Schleicher, I., Leitner, K., Juenger, J., Moeltner, A., Ruesseler, M., Bender, B., ... & Kreuder, J. G. (2017). Examiner effect on the objective structured clinical exam–a study at five medical schools. *BMC medical education*, *17*(1), 71.
16. Sim, J. H., Aziz, Y. F. A., Vijayanantha, A., Mansor, A., Vadivelu, J., & Hassan, H. (2015). A closer look at checklist scoring and global rating for four OSCE stations: Do the scores correlate well?. *Education in Medicine Journal*, *7*(2),39-44.
17. Turner, J. L., & Dankoski, M. E. (2008). Objective structured clinical exams: a critical review. *Fam Med*, *40*(8), 574-578.
18. Turner, K., Bell, M., Bays, L., Lau, C., Lai, C., Kendzerska, T., ... & Davies, R. (2014). Correlation between global rating scale and specific checklist scores for professional behaviour of physical therapy students in practical examinations. *Education Research International*, *2014*.
19. Turner, K., Bell, M., Bays, L., Lau, C., Lai, C., Kendzerska, T., ... & Davies, R. (2014). Correlation between global rating scale and specific checklist scores for professional behaviour of physical therapy students in practical examinations. *Education Research International*, *2014*.
20. Wass, V., Van der Vleuten, C., Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *The lancet*, *357*(9260), 945-949.
21. Wood, T. J. (2014). Exploring the role of first impressions in rater-based assessments. *Advances in Health Sciences Education*, *19*(3), 409-427.
22. Wood, T. J., Chan, J., Humphrey-Murto, S., Pugh, D., & Touchie, C. (2017). The influence of first impressions on subsequent ratings within an OSCE station. *Advances in Health Sciences Education*, *22*(4), 969-983.