Ann Med Res 2020;27(10):2701-6

Annals of Medical Research

DOI: 10.5455/annalsmedres.2020.04.414

Original Article

Does the performance assessment of students during surgery clerkship in a medical faculty jeopardize the objectivity of unbiased structured clinical examination?

Emin Daldal[1], Hasan Dagmura[2], Huseyin Bakir[3], Fatih Dasiran[1], Ali Ihsan Saglam[1], Osman Demir[4], Murat Yidirim[1], Ismail Okan[1], Ertan Bulbuloglu[5]

[1]Department of General Surgery, Faculty of Medicine, Gaziosmanpasa University, Tokat, Turkey
[2]Department of General Surgery and Surgical Oncology, Kutahya Evliya Celebi Training And Research Hospital, Kutahya, Turkey
[3]Department of General Surgery and Surgical Oncology, Faculty of Medicine, Gazisomanpasa University, Tokat, Turkey
[4]Department of Biostatistics, Faculty of Medicine, Gaziosmanpasa University, Tokat, Turkey
[5]Department of General Surgery, Faculty of Medicine, Kahramanmars Sutcu Imam University, Kahramanmaras, Turkey

Copyright © 2020 by authors and Annals of Medical Research Publishing Inc.

Abstract

Aim: Objective structured clinical examination (OSCE) has become an important assessment method in medicine which is more reliable than traditional exams. Faculty members make an opinion about the students during their clerkship. We aimed to investigate whether their judgments affect the scoring in the OSCE application.

Material and Methods: The study was planned prospectively. 4th-year students participating in the OSCE was identified as the working group. At the end of the clerkship, the faculty members gave a performance assessment (PA) score including the professional attitude of the students. PA and OSCE scores were compared. Two other faculty members participated all OSCE stations as external evaluators.

Results: There was a difference between OSCE and PA scores. Fewer students were successful at OSCE than PA (p=0.002). While the mean PA scores of three of the five faculty members were statistically similar, other two were different. The scores given by the responsible faculty members and the external evaluators from the faculty of internal medicine were similar (p>0.05). The evaluators from the surgical faculty gave lower scores statistically different from both groups (p<0.05). There was a strong relationship between the scores given by the faculty members responsible for the OSCE application and the faculty members from the surgical faculty (r=0.936; p<0.001) and those from the internal medicine faculty (r=0.947; p<0.001).

Conclusion: PA scores of the faculty members did not affect the OSCE scores which were supported by the external observers. The OSCE assessment was not influenced by the PA, and was found to be reliable.

Keywords: Objective structured clinical examination; performance score; professionalism; reliability

INTRODUCTION

In addition to knowledge and technical skills, communication skills, multidisciplinary approach, adaptation to teamwork, and systematic work are required to provide healthcare service effectively. It is important that physicians receive training in line with this objective during their professional development. Effective evaluation of the training is also very important (1). Thus, inadequacies in the training can be compensated. An ineffective assessment does not show any deficiencies not providing sufficient feedback which may have a long-lasting effect on the student. For this reason, it is imperative to carry out the measurement and evaluation at the most appropriate professional standard (2). There is no standard in traditional clinical exams for the assessment of students: students are assessed by different examiners, subjected to a limited skill test, face different questions, and are scored in a non-standardized way (2). It is necessary to measure the student's knowledge, skills, clinical practice, and attitude in the effective evaluation of the targeted learning outcomes. For this purpose, the Objective Structured Clinical Examination (OSCE) was defined by Harden in 1975 as an alternative to clinical PA methods. The OSCE system aims to minimize the variables that affect the evaluation. In this way, all students go through the same assessment system and an effective assessment can hence be done.

Received: 01.05.2020 **Accepted:** 28.09.2020 **Available online:** 21.10.2020
Corresponding Author: Hasan Dagmura, Department of General Surgery and Surgical Oncology, Kutahya Evliya Celebi Training And Research Hospital, Kutahya, Turkey **E-mail:** hassen@hacettepe.edu.tr

This systemic assessment provides objectivity leading to a more accurate scoring as well as feedback to students and assessors (3). In the OSCE, topics and the learning objectives expected from the student are clear. With the established stations, all students are evaluated with the same questions in the same amount of time. Scoring is done with a standard and predetermined checklist (2,4,5).

The OSCE has become an important method of evaluation in undergraduate assessment and the medical field because it is more reliable than the traditional clinical exams and is successful in measuring various clinical competencies (6, 7). It aims to reduce the number of variables that may affect the student PA by providing standardization in the exam. In the OSCE, the student's scores should only be influenced by their performance (7). Although this is the goal, faculty members who are responsible for the clerkship may have judgments about the student's achievement. In recent years, professionalism has become increasingly important. Characteristics such as communication with the patient, respect for the patient, honesty, professional responsibility, and compliance with the team draw attention in the evaluation of professionalism (8, 9). During the clerkship, the faculty member observes the student including their professionalism in the clinic which may affect their thoughts about the student. We aimed to investigate whether the judgments of the faculty members about the student's professional behaviors during the clerkship affect the scoring in the OSCE application. To elucidate this interaction we invited external observers and we compared the grades of external observers to those given by the faculties responsible from the education Therefore, we aimed to determine the effects of the faculty member's views about the student on the OSCE application.

MATERIAL and METHODS

The OSCE application was initiated to evaluate the fourth-year students at the General Surgery Clinic, Faculty of Medicine, Tokat Gaziosmanpasa University in the 2017-2018 academic year. The major difference from traditionla examination methods was the objectivity. To achieve the prompt objectivity, the assessment must usn preformed structured evaluation forms and well trained examiners.. Prior to the implementation of the OSCE, we carried out preliminary preparations such as literature research, observation, and training. This prospective study was conducted with fourth-year students who had been receiving their undergraduate education in the 2018-2019 academic year. It included 23 students who participated in the OSCE application. Four clerkship groups are trained in our clinic during the year with an average of 25 students in each.

Clerkship Learning Objectives

The General Surgery clerkship includes eight weeks of training. In this process, the student's knowledge, skills, communication, and case management is aimed to develop to a sufficient degree. The clerkship students receive theoretical and practical courses in the field of surgery; they are expected to develop certain skills such as vascular access, nasogastric catheter application, Foley urinary catheter placement, and suturing as well as attitudes such as communication and professionalism.

Post-Clerkship Evaluation

The students must get at least 60 out of 100 points to be successful in the Faculty of Medicine, Gaziosmanpasa University. At the end of the clerkship in our clinic, student success is evaluated by the following: a multiple-choice exam, the OSCE, the evaluation of the portfolio (the list of interventional procedures required during the clerkship), and the PA score given by the faculty member to evaluate their attitudes during the clerkship. The multiple-choice exam accounts for 30 points, the portfolio evaluation for ten points, and the faculty member's PA for ten points, and the OSCE for 50 points. The PA, the portfolio evaluation including students' application and communication skills and voluntary participation, the theoretical multiple-choice exam, and the OSCE are conducted in the last week of the clerkship. All these assessment methods determine the success of the student.

Intra-Clerkship Performance Assessment

During the clerkship, the faculty members observe the students in different fields of application and courses such as the clinic, operating room, and the endoscopy unit. They also evaluate the student's performance based on their professional behaviors such as knowledge, attitude, communication, volunteering, professional responsibility, and honesty. The faculties members give the PA score over ten points. Students with a score of six or higher are considered successful. The student's success according to the PA score was evaluated by taking the average of the scores given by five faculty members.

Before the OSCE

We inform the students about the OSCE application and set out model practices. Five stations are established in the General Surgery Clinic for the OSCE. We inform the students of the way the OSCE works and show them the application area. The rules to be followed during the OSCE are explained.

The OSCE Stations

We establish five stations in the OSCE application. The students draw lots from the classified exam topics for the OSCE questions in the last week. The questions in the first station focus on basic topics such as fluid electrolyte, homeostasis, shock, surgical infections, and trauma. The second station topics are from the field of oncological surgery such as the esophagus, stomach, and breast cancer. The questions at the third station are about gastrointestinal diseases such as diverticulitis, acute appendicitis, and hemorrhoids. At the fourth station, the questions are based on the field of endocrine surgery such as thyroid, parathyroid, and adrenal. The last station addresses the skills that need to be acquired during a surgical clerkship, such as suturing, obtaining patient consent, and abdominal examination (Figure 1). We prepare ten questions prepared for each group. The students chose five questions out of 50 in the draw.

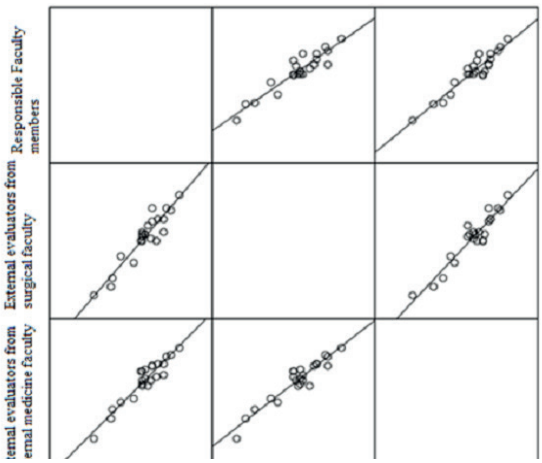


Figure 1. The schema of the OSCE stations

Evaluation before the OSCE Application

We invited five faculty members for each of the faculties of internal medicine and surgery as external observers to the OSCE. We had a meeting with them before the OSCE and informed them about it. One faculty member from general surgery, one from internal medicine, and one from surgery took place at each station. The general surgery faculty member at each station informed the other faculty members about the question and evaluation sheet. The checklist to be used for scoring was evaluated together.

The OSCE Application

At four stations, the students were asked to manage the patient through prepared scenarios. The questions drawn by lot were about parathyroid diseases, perianal abscess, soft tissue infection, and soft tissue sarcomas. The students were asked to perform an abdominal examination on a model at the application station. Scores were evaluated out of ten points: eight points for information and management, and two points for the smoothness of the presentation order and self-confidence.

The OSCE started with a ringtone. Students had five minutes at each station. When the surgery assistant rang the bell, the student left the station they were at and entered the next one. All the students completed the OSCE. Three faculty members evaluated each student at each station. The success of the exam was based only on the score of the general surgery faculty member.

After the OSCE

All the faculty members evaluated the exam at the end of the OSCE. To examine whether the general surgery faculty members' PA affects the outcome, the OSCE scores they gave were compared with those of the external observers for the same student. In addition, we investigated the relationship between the PA scores and the OSCE scores given by the responsible faculty members. When the scores obtained at the five stations were averaged, the students with a mean score of six and above were considered successful. In our study, we only measured the success in the OSCE and the PA scores. The test exam and the portfolio score were also taken into account in the success of the clerkship.

Ethics Committee Approval

We designed this study after obtaining the Permission No: 17713155-100 from the Deanship as well as the Permission No: 19-KAEK-147 of non-interventional clinical research from the Ethics Committee of Tokat Gaziosmanpasa University, Faculty of Medicine.

Statistical Analysis

The data are expressed as frequency and percent. McNemar's test was used to compare the categorical data between the groups. Pearson's correlation coefficient was used for the correlation between variables. A p-value <0.05 was considered significant. Analyses were performed using SPSS 19 (IBM SPSS Statistics 19, SPSS inc., an IBM Co., Somers, NY).

RESULTS

A total of 23 students participated in the OSCE, ten (43.5%) of whom were female students. The averages of the clerkship PA scores given by the responsible faculty members revealed that 15 (65.2%) out of 23 students were successful. As for the OSCE results, only five students (21.7%) were successful. Eight students unsuccessful in the PA also failed in the OSCE. Again, ten out of 15 students who were successful in the PA did not succeed in the OSCE. When we compared the PA scores with the OSCE scores, we noticed a difference and found that fewer students were successful in the OSCE results (p = 0.002) (Table 1).

Table 1. Students' score rates according to PA and OSCE

	Performance assessment score (n=23)	OSCE score (n=23)	p
<6	8(34.8%)	18(78.3%)	0.002
≥6	15(65.2%)	5(21.7%)	
Total	23(100%)	23(100%)	

The mean scores of three of the five faculty members were similar (p> 0.05). However, the average PA scores of the other two faculty members were statistically different from each other and the other three (p <0.05). The scores were also higher than those of the other three. Five faculty members responsible for the clerkship gave significantly higher scores in the PA than in the OSCE (p <0.001). When evaluated individually, we found that the PA scores of the four of them were higher than the OSCE scores and statistically significant (p <0.05). The remaining one faculty member's average PA score was also higher; however, the difference was not statistically significant when compared with the average OSCE score (p = 0.519) (Table 2).

We compared the scores given by the responsible faculty members and those by the external evaluators in the OSCE. An evaluation revealed that the scores of the responsible faculty members and those given by the external evaluators from the internal medicine faculty

were similar (p> 0.05). The external evaluators from the surgical faculty gave lower scores that were statistically different from both groups (p <0.05).

Table 2. Comparison of performance and OSCE scores given by the responsible faculty members

Variables	Mean ± Standard deviation	p
Total OSCE score	24.93±7.12	<0.001
Total PA score	32.61±6.57	
1. Faculty Member OSCE score	5.8±2.36	0.519
1. Faculty Member PA score	5.52±2.41	
2. Faculty Member OSCE score	3.7±1.72	<0.001
2. Faculty Member PA score	5.43±1.65	
3. Faculty Member OSCE score	5.09±1.41	0.018
3. Faculty Member PA score	5.65±1.07	
4. Faculty Member OSCE score	4.85±1.53	<0.001
4. Faculty Member PA score	8.78±1.31	
5. Faculty Member OSCE score	5.5±2.17	0.001
5. Faculty Member PA score	7.22±1.65	

There is a strong correlation between the scores given by the faculty members responsible for the clerkship and those of the external evaluators from the surgical faculty (r = 0.936; p <0.001) and the internal medicine faculty (r = 0.947; p <0.001). This relationship was statistically significant (Figure 2).

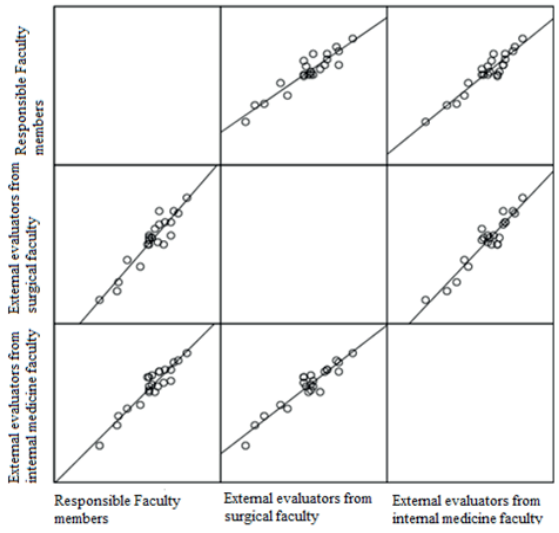


Figure 2. Distribution of OSCE scores given by the responsible faculty members and external evaluators

DISCUSSION

The OSCE has been increasingly used in the evaluation of education for nearly half a century and many articles have been published about it (10). The use of OSCE has become widespread in order to evaluate clinical skill in medical education. For example, in 1993–1994, only 38 out of 126 medical schools in United Stated reported that they took the OSCE exam, while almost all (121 out of 126) reported that these exams were used regularly until 2003-2004. OSCE was conducted to measure the clinical competence and clinically necessary skills of the second year students of Harvard Medical School and it was found to be useful in both subjects.(11) The objectivity depends on the same structured assessment of all students by the same examiners. A well-structured OSCE station must have a standard design to assess a clinical task performed by the students. We have been performing it in the Department of General Surgery for the last two years to evaluate fourth-year students. The shortcomings of traditional clinical exams to measure success have led the OSCE to gain so much value. Traditional exams have negative aspects such as inconsistency among the assessors, differences in questions, lack of clear instructions to candidates and assessors, and measuring success in a limited area (2). The two main features that make the OSCE important are its impartiality and exam structure. To ensure impartiality, checklists are used to standardize scoring and it is important that the assessors are trained. The goal is to use an objective assessment scale when the student fulfills the task by giving them clear instructions and to evaluate all students with the same exam questions. The OSCE allows students to be tested on multiple subjects in measuring clinical efficacy (2, 7). The excellent method to assess professional competence may not have been found yet. However, factors such as reliability, validity, impact on education, acceptability, and cost are important in a desirable evaluation model (12). We conducted the present study to evaluate the reliability of the OSCE that we had performed for the eighth time in our clinic. The faculty members who participated in this study as evaluators had participated in all the previous applications.

Although the PA in our study found 15 students (65.2%) as successful, only five (21.7%) were successful in the OSCE. This showed that the views of the faculty members about the students did not affect the OSCE result. In the PA, students' professional attitudes during their clerkship are important. Professionalism consists of respect for the patient, professional responsibility, compliance with teamwork, honesty, communication, and crisis management and is a part of the training (8,9,13,14). Faculty members' observations about students' professional attitudes are also important in the PA. In our study evaluating the reliability of the OSCE, we found that the OSCE scores did not correlate with the PA scores. In addition, the faculty members were not affected by their observation about the students. Some studies have reported that the evaluator's impartiality is affected by their impressions about the student (15-17). Some stated that the evaluator's first impression influences their assessment of the student, which is described as "the Halo Effect" in the literature. The Halo effect can reduce

the reliability of the assessment of competence and may cause inequality among students (18-20). A study found that knowing the student did not affect the assessment in the OSCE (21).

We evaluated the PAs of the faculty members about the students and found that two of them gave higher average scores in the PA. This difference may be due to their observations about the students in various fields. In addition, some evaluators may be more rigid while others may be more flexible. In the literature, this effect that makes a difference between the evaluators is defined as the "Hawk-Dove Effect" (22,23).

To control the effect of the PA on the OSCE, we invited faculty members from the faculties of internal medicine and surgery as external evaluators for the OSCE. The evaluation showed that the responsible faculty members and those from the internal medicine faculty gave similar scores. The external evaluators from the surgical faculty gave lower scores than both groups. There was also a strong positive correlation between the responsible faculty members and the external evaluators. In other words, if the responsible faculty member gave a high score, the external evaluators gave the same student a high score. Having more than one supervisor to evaluate students at the OSCE stations can increase reliability and reduce judgments about students (24). However, it should be kept in mind that this can increase the cost of the exams and require more workforces.

The major limitation of the study was the small number of the study group. The variable nature of OSCE results among the student groups could be another factor for limitation. The third reason could be the fact that OSCE was used only by surgery clerkship with possible adaptation problem by students resulting some degree of failure).

CONCLUSION

We found that clerkship PA scores given by the faculty members did not affect the OSCE scores as supported by external observers. The OSCE assessment was not affected by the PA, where we measured students' professional attitudes within the clerkship and was found to be reliable.

Acknowledgement: We would like to thank all of the general surgery and internal medicine faculty members who have been participated as external observers to this study. We also would like to express our gratitude to our 4th grade students who volunteered for the study.

Competing interests: The authors declare that they have no competing interest.

Financial Disclosure: There are no financial supports.

Ethical approval: We designed this study after obtaining the Permission No: 17713155-100 from the Deanship as well as the Permission No: 19-KAEK-147 of non-interventional clinical research from the Ethics Committee of Tokat Gaziosmanpasa University, Faculty of Medicine.

REFERENCES

1. Tabish SA. Assessment methods in medical education. Int J Health Sci, 2010;2.
2. Mohammad AH, Gominda P, Zubair A. The OSCE in the context of a holistic assessment, Hamza Mohammad Abdulghani, Essential Guide To Developing, Implementing, And Evaluating Objective Structured Clinical Examination, Singapore, World Scientific 2014:9-21.
3. Harden RM, Stevenson M, Downie WW, et al. Assessment of clinical competence using objective structured examination. Br Med J 1975;1:447-51.
4. McAleer S, Walker R. Objective structured clinical examination (OSCE). Occas Pap R Coll Gen Pract 1990:39.
5. Falcone JL, Claxton RN, Marshall GT. Communication skills training in surgical residency: a needs assessment and metacognition analysis of a difficult conversation objective structured clinical examination. J Surg Educ 2014;71:309-15.
6. Gopalakrishnan M. Objective Structured Clinical Examinations (OSCEs) as an Assessment Tool in Undergraduate Medical Education. J Natl Med Assoc 2018;110:553.
7. Khan KZ, Ramachandran S, Gaunt K, et al. The objective structured clinical examination (OSCE): AMEE guide no. 81. Part I: an historical and theoretical perspective. Med Teach 2013;35:1437-46.
8. Mazor KM, Zanetti ML, Alper EJ, et al. Assessing professionalism in the context of an objective structured clinical examination: an in-depth study of the rating process. Med Educ 2007;41:331-40.
9. Arnold L. Assessing professional behavior: yesterday, today, and tomorrow. Acad Med 2002;77:502-15.
10. Patrício M, Julião M, Fareleira F, et al. A comprehensive checklist for reporting the use of OSCEs. Med teach, 2009;31:112-24.
11. Simon SR, Bui A, Day S, Berti D, Volkan K. The relationship between second-year medical students' OSCE scores and USMLE Step 2 scores. J Eval Clin Pract 2007;13:901-5.
12. Van Der Vleuten CPM, The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ Theory Pract 1996; 1:41-67.
13. Ginsburg S, Regehr G, Hatala R, et al. Context, conflict, and resolution: a new conceptual framework for evaluating professionalism. Acad Med 2000;75:6-11.
14. Jha V, HL Bekker HL, Duffy SRG, et al. Perceptions of professionalism in medicine: a qualitative study. Med Educ 2006;40:1027-36.
15. Stroud L, Herold J, Tomlinsonet G, et al. Who you know or what you know? Effect of examiner familiarity with residents on OSCE scores. Acad Med 2011;86:8-11.
16. Iramaneerat C, Yudkowsky R. Rater errors in a clinical skills assessment of medical students. Eval Health Prof 2007;30:266-83.

17. Wood TJ. Exploring the role of first impressions in rater-based assessments. Adv Health Sci Educ Theory Pract 2014;19:409-27.
18. McKinstry BH, Cameron HS, Elton RA, et al. Leniency and halo effects in marking undergraduate short research projects. BMC Med Educ 2004;4:28.
19. Malouff JM, Stein SJ, Bothma LN, et al. Preventing halo bias in grading the work of university students. Cogent Psychol 2014;1: 988937.
20. Malouff JM, Emmerton AJ, Schutte NS. The risk of a halo bias as a reason to keep students anonymous during grading. Teach Psychol 2013;40:233-7.
21. Jefferies A, Simmons B, Regehr G. The effect of candidate familiarity on examiner OSCE scores. Med educ 2007;41:888-91.
22. McManus IC, Thompson M, Mollon J. Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP (UK) clinical examination (PACES) using multi-facet Rasch modelling. BMC Med Educ 2006;6:42.
23. Finn Y, Cantillon P, Flaherty G. Exploration of a possible relationship between examiner stringency and personality factors in clinical assessments: a pilot study. BMC Med Educ 2014;14:1052.
24. Gormley G. Summative OSCEs in undergraduate medical education. Ulster Med J 2011;80:127.