



Universiteit  
Leiden  
The Netherlands

## Examiners' perceptions in surgical education: the blind spot in the assessment of OSCEs

Want, A.C. van der; Bloemendaal, P.M.; Hage, J.A. van der

### Citation

Want, A. C. van der, Bloemendaal, P. M., & Hage, J. A. van der. (2021). Examiners' perceptions in surgical education: the blind spot in the assessment of OSCEs. *Journal Of Surgical Education*, 78(2), 590-596. doi:10.1016/j.jsurg.2020.07.024

Version: Publisher's Version

License: [Creative Commons CC BY-NC-ND 4.0 license](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3280120>

**Note:** To cite this publication please use the final published version (if applicable).



# Examiners' Perceptions in Surgical Education: The Blind Spot in the Assessment of OSCEs

Anna C. van der Want, MA, MPhil, PhD,\* Peter M. Bloemendaal, MD,<sup>†</sup> and Jos A. van der Hage, MD, PhD<sup>†</sup>

\*HAN University of Applied Sciences, Nijmegen, the Netherlands; and <sup>†</sup>Leiden University Medical Center, Leiden, The Netherlands

**PURPOSE OF THE ARTICLE:** Interpreting objective structured clinical examination (OSCE) scores remains a challenging task.<sup>1</sup> It remains unclear how examiners make a decision to score an OSCE the way they do. The examiners' thoughts and decision-making processes remain a hidden part of the OSCE assessment. One cannot assume that all examiners have a shared belief about what is important in an OSCE or what is the aim of an OSCE assessment. These differences are problematic, since they lead to students not getting the 'correct' score for their OSCE.

**MATERIALS AND METHODS:** Using a qualitative case study design,<sup>2</sup> this study aims to gain in-depth insight into examiners' perceptions of OSCEs and their role as an examiner. For this purpose, semi-structured interviews with OSCE observers (n = 6) were conducted.

**RESULTS AND CONCLUSIONS:** The results show that the perception of examiners on the aim of OSCEs and their motivation and task as examiner are crucial in understanding the differences among OSCE scores of examiners. Training and discussion among the examiners are needed to decrease the differences among the OSCE scores of examiners. (J Surg Ed 78:590–596. © 2020 The Author(s). Published by Elsevier Inc. on behalf of Association of Program Directors in Surgery. This is an open access article under the CC BY-NC-ND license. (<http://creativecommons.org/licenses/by-nc-nd/4.0/>))

**KEY WORDS:** OSCEs, OSCE examiners, Qualitative case study, Task perception, Surgical education

**COMPETENCIES:** Practice-Based Learning & Improvement, Professionalism

Funding: The authors of this paper have not received any funding or benefits to conduct this research.

Correspondence: Inquiries to A.C. van der Want, MA, MPhil, PhD, HAN University of Applied Sciences, Kapittelweg 35, Nijmegen 6525, EN, the Netherlands; e-mail: [Annavanderwant@gmail.com](mailto:Annavanderwant@gmail.com)

## INTRODUCTION

Like in many other fields of education (medical/dental/nursing),<sup>3,5</sup> objective structured clinical examinations (OSCEs) are a crucial, valuable part of undergraduate surgical education. OSCE is a form of assessment of practical skills in medical education. It is '...an approach to the assessment of clinical competence in which the components of competence are assessed in a planned or structured way (...) the student is assessed at a series of stations with one or two aspects of competence being tested at each station'.<sup>6</sup> However, interpreting OSCE scores remains a challenging task.<sup>1</sup> Even with a training for examiners and digital scoring system to score medical students' performance, differences exist among OSCE scores of examiners.<sup>7,9</sup> Previous research concerning OSCEs mostly focused on the OSCE itself,<sup>9,10</sup> the student perception of OSCEs<sup>5,11,12</sup> and the stress OSCEs might causes to students.<sup>13-15</sup>

The existing differences among OSCE scores and the accompanying inter-observer variability of examiners are problematic, since they lead to students not getting the 'correct' score for their OSCE. Moreover, the differences among examiners question the reliability and validity of OSCEs.

Previous research on the role of the examiner of OSCEs underlines the differences among examiners<sup>16</sup> and describes various factors that possibly explain the differences in scores among OSCE examiners (Table 1).

For instance, examiner *commitment* is important to decrease the differences among OSCE scores.<sup>17</sup> Also, next to the assessment expertise of the examiner and the degree to which examiners read the information on OSCEs,<sup>18</sup> other factors influencing the differences among examiners include whether an examiner knows the students whose OSCE he/she is assessing,<sup>19</sup> the number of students someone is assessing, and the position of the examiner (e.g., resident, faculty member, part-time/full-time working examiners).<sup>4</sup> Furthermore, Hope and

**TABLE 1.** Variables Determining the Differences Among Examiners

Variable	Reference
Commitment of examiners	Wilkinson et al. (2003)
Examiners reading information on OSCEs	Berendonk et al. (2013)
Examiners' familiarity with students	Stroud et al. (2011)
Number of students an examiner is assessing	Park et al. (2015)
Position of examiner (resident/faculty member, part-time/full-time)	Park et al. (2015)
Assessment time during the day	Hope and Cameron (2015)
Interpretation of scoring schemes	Chahine et al. (2016)

Cameron (2015) found that OSCE examiners are more lenient in the first OSCE they assess of the day than during their last OSCE of the day.<sup>20</sup> A study by Chahine et al.<sup>21</sup> showed that if there is no difference among OSCE scores of examiners, then there is no guarantee of similar student performance during the OSCE. Their study showed that identical ratings can be based on different interpretations of scoring schemes or reasoning.<sup>21</sup>

Despite the findings of previous studies, examiners' role perception remains unclear. Moreover, it remains unclear how examiners make a decision to score an OSCE the way they do. The examiners' thoughts and decision-making processes remain a hidden part of the OSCE assessment. One cannot assume that all examiners have a shared belief about what is important in an OSCE or what is the aim of an OSCE assessment.<sup>21,22</sup>

Therefore, this paper aims to focus on the decision-making process and thoughts of examiners of OSCEs. By studying how examiners perceive OSCEs and their role as examiners, as well as their beliefs of their tasks as an examiner, this paper aims to gain insight into the differences among OSCE scores of examiners. Understanding differences in the interpretation of the examiner's own role will help us improve the way we apply OSCEs in our medical education curriculum. The research questions of this study are as follows:

- 1) What are the differences in scores among OSCE examiners?
- 2) How do examiners of OSCEs in surgical education perceive their role as OSCEs examiners?
- 3) How can the differences among examiners of OSCEs be explained by their perception of OSCEs and their role as examiners?

## MATERIALS AND METHODS

Using a mixed-method study design,<sup>2</sup> this study aims to gain insight into the differences among examiners of OSCEs by studying examiners' perceptions of OSCEs and their role as examiner.

First, we studied the differences among the scores of examiners of OSCEs in an undergraduate surgical clerkship programme in a university hospital in the Netherlands. In order to do so, the scores of examiners on OSCEs were collected using an online OSCE assessment form between 2016 and 2018.

The OSCEs were used to assess the students' skills and knowledge regarding physical examination; each student randomly selected (by throwing dice) the topic of the OSCE (ranging from the examination of knee joints, veins, and thorax trauma to pulsation of the carotid artery and orthopaedic or cardiovascular physical examination). Each student was assessed during 1 OSCE, which lasted about 10 minutes. During the OSCE, the student performed the physical examination on a peer (medical student), while the OSCE examiner was present and scored the skills and knowledge of the student using an online scoring form. All the examiners received a similar instruction to examine OSCEs with the online scoring form. The instruction of the examiners consists of an individual oral instruction by the coordinator, combined with written instruction about the OSCEs and the digital scoring form. The online scoring form was developed by the university hospital and consisted of 2 parts. The first part of the online assessment form consisted of 5 general scoring items that were applicable for every OSCE, irrespective of the topic of the OSCE (regarding doctor-patient communication, instruction, empathy, and basic hygiene). The second part of the online assessment form consisted of specific questions regarding the topic of the OSCE (for the topic 'knee', e.g., the scoring form consisted of items regarding specific tests like varus/valgus, appley, McMurray).

The examiner was expected to indicate on a 3-point scale (0 = 'not at all shown by the student'; 1 = 'partly shown by student, 2 = 'clearly shown by the student') the extent to which the student had shown his/her skills and knowledge regarding the items on the assessment form. A mean score from all the scoring items was calculated, which resulted in the final score for the student. For our research, all mean scores per examiner between January 2016 and February 2018 were calculated for each OSCE 'topic' separately and for all 6 OSCE topics together. This enabled us to gain insight into the differences in scores among OSCE examiners (research question 1).

Based on the differences in OSCE scores among the examiners, semi-structured interviews were conducted

**TABLE 2.** Description of Participants (Examiners)

Examiner	Sex	Number of Years of Working Experience (in Health Care)	Number of Years of Teaching Experience	Resident/Staff Member
A	M	5-10	<5	Resident
B	M	10-20	10-20	Staff member
C	M	>20	>20	Staff member
D	M	>20	>20	Staff member
E	M	>20	>20	Staff member
F	M	5-10	<5	Resident

to gain insight into the examiners' perception of OSCEs (research question 2) and possible explanations about the differences among OSCE scores (research question 3). These interviews were conducted directly after an evaluator scored an OSCE of a medical student. The interviewer was present during the OSCE and observed the evaluator during the OSCE. The interview guideline consisted of open-ended questions on background information of the participant (i.e., the examiner of OSCEs) and open-ended main questions related to examining OSCEs followed by follow-up questions. On each of the following themes, the participants were asked to elaborate on their perception of their role as an OSCE examiner, which we divided into 3 parts: (1) the aim of OSCEs, (2) their task as an OSCE examiner and (3) their motivation for examining OSCEs. In addition, participants were asked about a possible explanation for differences among examiners. This interview guideline was piloted in February 2018, which resulted in minor adjustments for the formulation of the questions.

The semi-structured interviews were audio-taped and transcribed verbatim afterwards. The analysis consisted of several steps: First, all interviews were read thoroughly by the authors to explore the data and to look for possible additional relevant themes that came up from the data. One additional theme was found and added as a category for further analysis: all participants spoke about the medical content of the OSCEs. In a next step, a matrix was constructed consisting of (1) the participants (in the rows of the matrix, 1 participant per row), the categories (in the columns of the matrix, 1 category per column), and the aim of OSCEs; (2) their task as an OSCE examiner; (3) their motivation for examining OSCEs; (4) medical content of the OSCEs and (5) possible explanation for differences among examiners. Using this matrix, all interviews were coded, resulting in a matrix with summaries of the interviews for each category (step 3). In a fourth step, a condensed version of the matrix was constructed in an iterative process of data reduction. This resulted in a concise matrix that enabled comparison among the participants within each

of the categories. Using the condensed matrix, the results were described for each category.

### Participants: Selection Procedure

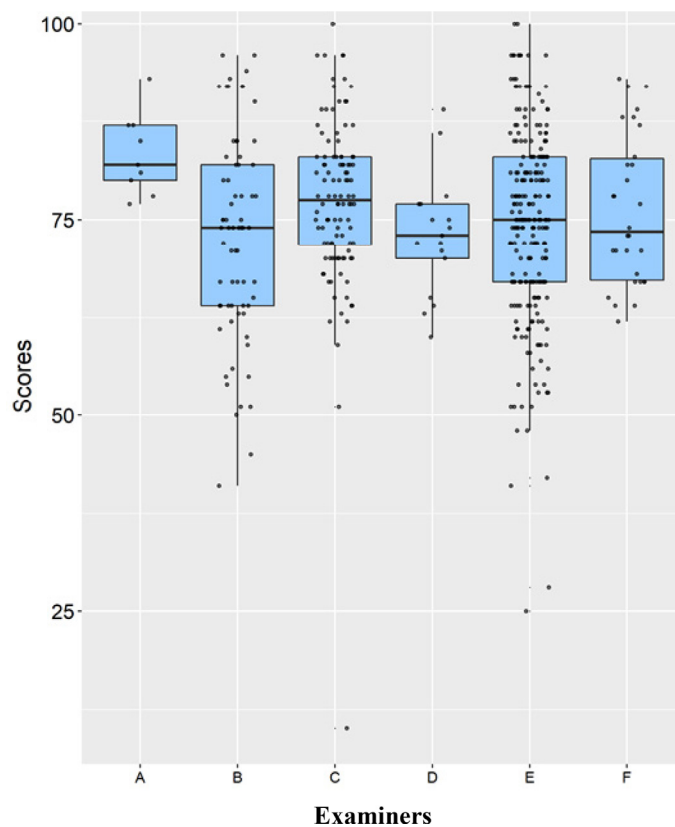
All examiners of the department of surgery ( $n=6$ ; Table 2) of a large university hospital in the Netherlands were invited personally during a departmental meeting in 2016 to participate in the first part of this study, where the online OSCE assessment forms were collected to study the differences among OSCE examiners. All 6 examiners agreed to participate with the data collection of their OSCE scores in order to study the differences among the examiners. In 2018, these examiners were invited by email to participate in an interview focusing on their role perception as an examiner of OSCEs and their possible explanations for the differences among examiners. All ( $n=6$ ) examiners participated in individual semi-structured interviews, which were conducted between April and July 2018. All examiners of OSCEs in surgical education in this study were residents or staff members in the department of surgery at a university medical hospital centre in the Netherlands. One participant in this study also has a task of coordinating the education within the department and organizes the OSCEs and the instruction of the examiners.

## RESULTS

This paper aimed to gain insight into the differences in scores among OSCE examiners (research question 1), the role perception of OSCE examiners (research question 2), and possible explanations for the differences among examiners of OSCEs (research question 3).

### Differences in Scores Among OSCE Examiners

In Figure 1, OSCE scores per examiner for 2016 to 2018 are depicted. On the vertical axis, the final OSCE scores are depicted, with 100 points being the highest score and 0 points being the lowest scores. Students need at least 55 points to pass the OSCE. As can be seen in



**FIGURE 1.** OSCE scores per examiner, 2016 to 2018.

Figure 1, the results show differences among the examiners. Whereas examiners B and E assessed some OSCEs below 55 points, the other examiners did not. Examiners B and E also have the widest range in scores (varying from 45-100 points) as compared to, for instance, examiner A, whose scores range from 90 to 100 points. Examiner A gave the highest scores for the OSCEs.

### Role Perception

Table 3 provides an overview of the examiners' perception of the aim of OSCEs, their task, the content and the motivation.

### Aim of OSCEs

Concerning the second research question, the examiners' perception of the aim of OSCEs is to prepare students for surgery clerkship and to motivate them to read the course materials. Examiners F and A mentioned the high difficulty level of the OSCEs. For students, OSCEs are 'serious business' (examiner F). Examiners D and B referred to OSCEs in the wider perspective of education, underlined the importance of education, since everything changes fast in the medical domain, and referred to the fact that not all students will become surgeons

and therefore should only know some basics about surgery in order to, for instance, refer a patient to a specialist properly later in their career.

**TABLE 3.** Overview of the Results

Variable	Result
Aim of OSCEs	To prepare students for surgery clerkship and to motivate them to read the course materials
Task of examiners	To provide suggestions and give advice to the students in order for them to learn. Differences were found regarding how examiners interpret the assessment form (rubric).
Content of OSCEs	Three of the six examiners mentioned that some content or tests when examining a patient are outdated, proven unreliable, or not useful. All examiners in this study mentioned that they enjoy their role as examiner and taking part in education. Although education is felt to be important by the participants, at the same time the participants perceive it as a 'compulsory' part of their job.

## Task of Examiners

Concerning the tasks of the examiners, all examiners find it important to provide suggestions and give advice to the students in order for them to learn. Differences can be found regarding how examiners interpret the rubric: Whereas examiner A saw it as an important task to identify students who have not prepared for the OSCE, examiner C stated that everybody passes the OSCEs he examines, since an OSCE is mostly a training session and not an assessment as such. Examiner F's main aim is to provide, as he stated, 'objective assessment', whereas examiner E stated it is most important that students learn something.

## Content of OSCEs

When it comes to the surgical content of the OSCEs, 3 of the 6 examiners mentioned that some content or tests when examining a patient are outdated, proven unreliable, or not useful, since they are never done in practice (e.g., when studying the knee, students have to take a certain test, whereas in daily practice, additional imaging would be made). One examiner referred to some tests as 'folklore'. In addition, 3 examiners prefer to leave out female breast examination, since they find it awkward for both students and examiners and not meaningful to practice on healthy women. In addition, students only take a randomly selected OSCE, so not all students take the same OSCE, and some OSCEs are easier for some students (veins) than others (e.g., thorax trauma).

## Motivation of Examiners

When looking at their motivation as examiners, all examiners in this study mentioned that they enjoy their role as examiner and taking part in education. Examiners D, C and E all have formal roles in the educational domain as a coordinator of education and OSCE designer (examiner E), supervisor of PhD students and research programme development (examiner C), or researcher in surgical education (examiner D). Examiners F and A are residents-in-training in the surgical department and participate as examiners voluntarily on top of their programme. However, for all examiners, teaching and examining OSCEs are a part of their job as a surgeon in a university hospital. Research and patient care are perceived as higher priorities. Although education is felt to be important by the participants, at the same time the participants perceive it as a 'compulsory' part of their job. One-on-one contact with students is especially appreciated (examiners D and C), and all examiners like to provide suggestions to students.

## Possible Explanation for Differences Among Examiners

Asked about possible explanations for the difference among the grades of examiners (research question 2), examiners F and E felt that they are possibly stricter than their colleagues. Apart from some short informal discussions on the way to perform the exams, the examiners do not discuss or evaluate their experiences with the grading procedure. Another possible explanation provided by the participants of this study is that the examiners interpret the digital assessment form in different ways. Examiners feel that it can be helpful and a possible solution to decrease the differences among examiners to discuss and evaluate their experiences with the grading procedure. Another possible explanation is the different areas of expertise among the examiners: some examiners have decades of experience, while others just a few years. Since a lot of OSCEs in this programme focus on orthopaedics, examiners feel that orthopaedic specialists might contribute more to the learning process of the students than they can. In addition, examiner E stated that possibly 'wanting to be the nice guy' can influence the (differences among) grades of facilitators, since facilitators feel that they are not nice if they grade a student as 'fail'. Possible solutions perceived by the examiners to decrease the differences among examiners are taking time to meet as examiners and discuss with all examiners (1) the aim of the OSCEs, (2) their role as examiners, (3) the content and process of the OSCEs and (4) the interpretation of the form. In addition, 2 examiners mentioned that it could be helpful to examine a few OSCEs together.

## CONCLUSION AND DISCUSSION

The main aim of this study was to gain insight into the differences in scores among examiners, the role perception of OSCE examiners, and the possible explanation for the differences among examiners of OSCEs. This study shows that the scores of OSCE examiners differed among the examiners. These findings are in line with previous research that underlined that whereas scoring forms can be helpful to decrease the differences among OSCE examiners,<sup>8,9</sup> the ultimate decision about how to use the scoring form seems to be made by the examiner.<sup>21</sup>

Concerning the role perception of OSCE examiners, the examiners differ widely in their perceived aim of the OSCEs, ranging from 'assessment of learning' (assessing students' level of clinical reasoning and testing whether students are ready for an internship) to 'assessment for learning', to providing students an opportunity for

learning during the OSCE.<sup>23</sup> On this topic, differences were found among the residents (assessment of learning) and the staff members (assessment for learning). This is possibly related to the years of experience as an OSCE examiner and the 'examiner's expertise'.<sup>4,18</sup> The examiner's expertise increases over the years as examiners expand their repertoire and develop their own interpretation of the aim of OSCEs and the scoring form. It might be that OSCE examiners with a few years of experience interpret their role as an examiner quite literally as 'to assess students' skills and knowledge', whether examiners with more experience have a broader perspective on their role as an examiner. In addition, the difference between examiners might also be explained by the differences in the background of the examiners of this study (being an OSCE developer, a staff member or a resident). The commitment of examiners might also play a role in this respect.<sup>17</sup>

This exploratory mixed-method study has several limitations. The data were limited to self-perceived data, and a small number of 6 participants from 1 hospital participated in this study. Future research could focus on a larger number of participants from multiple hospitals and study the possible effect of hospital context. This study and previous studies focusing on OSCE examiners have all been small-scale studies that focused on specific aspects of examining OSCEs or creating models based on qualitative research. Large-scale mixed-method studies in various contexts with multiple interventions are needed to see what really makes a difference in terms of the differences among examiners' OSCE scores. Previous research found that a training can reduce the variation between OSCE scores of examiners.<sup>24-26</sup> The effective elements of a training are still under researched.<sup>8</sup> Reid et al.<sup>8</sup> argue for an online compulsory course for examiners yet the content of the course remains unspecified. In addition to the previous research, it would be interesting to study what kind of intervention (e.g., a training or supervision meetings for examiners), which activities (e.g., practicing, discussing with other examiners, watching videos of OSCEs) and what content (e.g., focused on the aim of OSCEs, on the content of OSCEs or on assessment strategies) would influence the difference among examiners.

A crucial element in the explanation of the differences seems to be the examiners' perception: the perception of the aim of OSCEs, their role, and the scoring form. Therefore, for OSCE examiners, it is essential to be aware of their perceptions regarding OSCEs and to discuss them with other examiners in order to decrease the differences among OSCE scores. In our university hospital, the results of this study were shared and discussed with all the examiners in the department and plans are made to further adjust the training for examiners. In a

future study, the effect of the training for examiners will be reported. This study underlines the fact that using a qualitative approach helps recognize the different interpretations examiners have and different actions examiners perform in using OSCEs.

## ACKNOWLEDGMENTS

The authors would like to thank the participants of this study.

## REFERENCES

1. Hurley KF, Giffin NA, Stewart SA, Bullock GB. Probing the effect of OSCE checklist length on inter-observer reliability and observer accuracy. *Med Educ Online*. 2015;20:29242. <https://doi.org/10.3402/meo.v20.29242>.
2. Yin RK. *Case Study Research: Design and Methods*. Thousand Oaks, California: Sage; 2003.
3. Qureshi NS. Examiners' perceptions of the objective structured clinical examination in colposcopy. *J Obstet Gynaecol (Lahore)*. 2013;33:188-190. <https://doi.org/10.3109/01443615.2012.737050>.
4. Park SE, Kim A, Kristiansen J, Karimbux NY. The influence of examiner type on dental students' OSCE scores. *J Dent Educ*. 2015;79:89-94.
5. Johnston ANB, Weeks B, Shuker MA, et al. Nursing students' perceptions of the objective structured clinical examination: an integrative review. *Clin Simul Nurs*. 2017;13:127-142. <https://doi.org/10.1016/j.ecns.2016.11.002>.
6. Harden RM, Hart IR. An international virtual medical school (IVIMEDS): the future for medical education? *Med Teach*. 2002;24:261-267. <https://doi.org/10.1080/01421590220141008>.
7. Rothman AI, Blackmore D, Cohen R, Reznick R. The consistency and uncertainty in examiners' definitions of pass/fail performance on OSCE stations. *Eval Health Prof*. 1996;19:118-124.
8. Reid K, Smallwood D, Collins M, Sutherland R, Dodds A. Taking OSCE examiner training on the road: reaching the masses. *Med Educ Online*. 2016;21. <https://doi.org/10.3402/meo.v21.32389>.
9. Shokoohi S, Emami AH, Mohammadi A. Medical education online. *Med Educ Online*. 2014;19:4-6.
10. Setyonugroho W, Td K, Kennedy IFM, Td T, Kropmans IFJB. Patient education and counseling reliability and

- validity of OSCE checklists used to assess the communication skills of undergraduate medical students: a systematic review. *Patient Educ Couns*. 2015;98:1482-1491. <https://doi.org/10.1016/j.pec.2015.06.004>.
11. Nasir AA, Yusuf AS, Abdur-Rahman LO, et al. Medical students' perception of objective structured clinical examination: a feedback for process improvement. *J Surg Educ*. 2014;71:701-706. <https://doi.org/10.1016/j.j Surg Educ.2014.02.010>.
  12. Small LF, Pretorius L, Walters A, M Ackerman, Tshifugula P. 'Students' perceptions regarding the objective, structured, clinical evaluation as an assessment approach'. *Heal SA Gesondheid*. 2013;18:1-8.
  13. Muldoon K, Biesty L, Smith V. "I found the OSCE very stressful": student midwives' attitudes towards an objective structured clinical examination (OSCE). *Nurse Educ Today*. 2014;34:468-473. <https://doi.org/10.1016/j.nedt.2013.04.022>.
  14. Hadi MA, Ali M, Haseeb A, Mohamed MMA, Elrggal ME, Cheema E. Impact of test anxiety on pharmacy students' performance in objective structured clinical examination: a cross-sectional survey. *Int J Pharm Pract*. 2018;26:191-194. <https://doi.org/10.1111/ijpp.12389>.
  15. O'Carroll PJ, Fisher P. Metacognitions, worry and attentional control in predicting OSCE performance test anxiety. *Med Educ*. 2013;47:562-568. <https://doi.org/10.1111/medu.12125>.
  16. 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. <https://doi.org/10.1186/1472-6920-6-42>.
  17. Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T. Objectivity in objective structured clinical examinations. *Acad Med*. 2003;78:219-223. <https://doi.org/10.1097/00001888-200302000-00021>.
  18. Berendonk C, Stalmeijer RE, Schuwirth LWT. Expertise in performance assessment: assessors' perspectives. *Adv Heal Sci Educ*. 2013;18:559-571. <https://doi.org/10.1007/s10459-012-9392-x>.
  19. Stroud L, Herold J, Tomlinson G, Cavalcanti RB. Who you know or what you know? Effect of examiner familiarity with residents on OSCE scores. *Acad Med*. 2011;86(10 Suppl):8-11. <https://doi.org/10.1097/acm.0b013e31822a729d>.
  20. Hope D, Cameron H. Examiners are most lenient at the start of a two-day OSCE. *Med Teach*. 2015;37:81-85. <https://doi.org/10.3109/0142159X.2014.947934>.
  21. Chahine S, Holmes B, Kowalewski Z. In the minds of OSCE examiners: uncovering hidden assumptions. *Adv Heal Sci Educ*. 2016;21:609-625. <https://doi.org/10.1007/s10459-015-9655-4>.
  22. Joe JN, Harmes JC, Hickerson CA. Using verbal reports to explore rater perceptual processes in scoring: a mixed methods application to oral communication assessment. *Assess Educ Princ Policy Pract*. 2011;18:239-258. <https://doi.org/10.1080/0969594X.2011.577408>.
  23. Taras M. Using assessment for learning and learning from assessment. *Assess Eval High Educ*. 2002;27:501-510. <https://doi.org/10.1080/0260293022000020273>.
  24. Cook DA, Dupras DM, Beckman TJ, Thomas KG, Pankratz VS. Effect of rater training on reliability and accuracy of mini-CEX scores: a randomized, controlled trial. *J Gen Intern Med*. 2009;24:74-79. <https://doi.org/10.1007/s11606-008-0842-3>.
  25. Holmboe ES, Hawkins RE, Huot SJ. Effects of training in direct observation of medical resident's clinical competence: a randomized trial. *Ann Intern Med*. 2004;140. <https://doi.org/10.7326/0003-4819-140-11-200406010-00008>.
  26. Schwartzman E, Hsu DI, Law AV, Chung EP. Assessment of patient communication skills during OSCE: examining effectiveness of a training program in minimizing inter-grader variability. *Patient Educ Couns*. 2011;83:472-477. <https://doi.org/10.1016/j.pec.2011.04.001>.