

Developing Clinical Skill Videos as an Instrument to Assess the Objective Structured Clinical Examination (OSCE) Examiners' Effect



Oscar Gilang Purnajati, Rachmadya Nur Hidayah,
and Gandes Retno Rahayu

Abstract Video-based assessment is a reliable method for testing clinical skills performance. Several published studies have different results because of various bias factors. This study aimed to describe the development and use of videos to assess the effect of the Objective Structured Clinical Examination (OSCE) examiners' backgrounds. Cardio-Pulmonary Resuscitation (CPR) was chosen for this study because it has a guideline from the American Heart Association. The development steps included: the assessment guidelines were rewritten by two cardiologists; two standardized simulated CPR procedure videos were made with their supervision. The CPR video showed performance following the guidelines and the other showed CPR not according to guidelines. The cardiologist gave feedback after watching the two videos. Finally, 51 OSCE examiners in the Medical Faculty, Duta Wacana Christian University assessed the CPR performance in the videos using standardized assessment guidelines. Examiners were categorized according to their backgrounds and the average results of the assessment based on their background characteristics were analyzed by the Kruskal–Wallis test. The results show that the two videos were developed and the assessment on those two videos did not significantly differ between examiners' background categories ($p > 0.05$). The clinical practice experience and educational background category had a significant score difference ($p = 0.04$; $df = 3$ and $p = 0.03$; $df = 2$, respectively). There were no score differences between examiners, except in clinical practice experience and educational background categories. Video-based assessment can foster the objectivity of OSCE hence it can be applied in OSCE scoring assessor training. However, there are still sources of biases that academics need to be aware of and consider.

Part of this paper was presented at INA-MHPEC 2022 and received The Best Poster Presentation.

O. G. Purnajati (✉)

Faculty of Medicine, Universitas Kristen Duta Wacana, Yogyakarta, Indonesia
e-mail: oscargilang@staff.ukdw.ac.id

R. N. Hidayah · G. R. Rahayu

Department of Medical Education and Bioethics, Faculty of Medicine, Public Health,
and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

Keywords Video-based assessment · CPR · OSCE examiners · Background variability · Clinical skill performance

List of Abbreviation

OSCE Objective Structured Clinical Examination

1 Introduction

The Objective Structured Clinical Examination (OSCE) is a central component in assessing the clinical skills of medical students, and because the results provide information about the competencies of the students being assessed, the process must be ensured to be rigorous and accurate [1]. However, several factors interfere with the assessment in the OSCE, namely the inconsistency of the checklists and differences in the details of the assessment on each item, including the global rating scale [2], inequality in making checklists and their constructs [3], the level of difficulty of the material tested in the OSCE [4], and simulation patients which had a positive impact on student performance during OSCE compared to the use of student role-plays [5].

The OSCE blueprint plays an important role in the OSCE assessment, ensuring that exam candidates are comprehensively tested for competence [6]. However, a hidden pattern in examiners may influence them in conducting OSCE assessments [7]. Those hidden patterns are the perception of doctor-patient communication [8], various cultural factors of the examiner [9], the contrast effect of the previous student which becomes a benchmark for judging the next student [10], and the use of different assessment references in the OSCE [11]. Some of the factors for the inaccuracy of OSCE results can stem from the imprecision of the test, the variability of the examiner, and all of the other psychometric properties (simulated patients, assessment materials, scoring guides, etc.) [12].

Video-based assessment is considered a reliable method for testing clinical skills performance. Students can learn and prepare clinical skills with the help of video examinations, as a benchmark for clinical skills competency [13]. The use of video-based assessments of simulated examinations shows that these assessments can provide a valid and reliable method for testing the clinical performance of students [14]. The examiner's background, related to social and psychological processes, the examiner's clinical practice experience, the experience of assessing the OSCE, and the examiner's gender appropriateness, had a major role in the inaccuracy of the assessment even though the OSCE was administered under the most standard conditions [11, 12]. However, several studies that have been conducted on this aspect still found different results due to various bias factors. In this study, the considered OSCE examiners' backgrounds were gender, education, clinical practice experience and

duration, OSCE experience, and their OSCE training. This study aimed to describe the development of the videos and to analyze the developed video examination results from the OSCE examiners regarding their backgrounds. The findings of this study can add a reliable way to foster the objectivity of OSCE.

2 Methods

This study described how the process of making a video-based assessment is done (Fig. 1). First, Cardio-Pulmonary Resuscitation (CPR) skill assessment was chosen because it already has a specific guideline from the American Heart Association [15, 16]. To be usable by our OSCE examiners, it was rewritten by two cardiologists, adapted in Bahasa Indonesia, and they revised the assessment rubrics that already matched the OSCE requirement. The validity of the content in the rubric and assessment guide was achieved when the assessment instrument was reviewed by the cardiologists. Then, we developed standardized simulated CPR procedure videos with their supervision based on the guideline. Our students served as the actors in both videos; one video that portrayed CPR according to the guidelines and the other one that did not comply with the guidelines. The cardiologists gave feedback after watching the videos and revisions were completed where appropriate.

A total of 51 OSCE examiners from the Faculty of Medicine, Duta Wacana Christian University were enrolled in the study using total sampling, to assess the CPR performance in both videos using standardized assessment guidelines. These OSCE examiners were pre-clinical and clinical teaching lecturers from various scientific groups in the medical faculty. The Faculty of Medicine, Duta Wacana Christian University (UKDW) uses the OSCE as a regular clinical skills examination every semester for undergraduate medical education.

This study used a quantitative method, in the form of a cross-sectional study of the assessment of OSCE examiners on the Cardiopulmonary Resuscitation (CPR) competency video. In giving the assessment, the results of the OSCE examiner's assessment based on each background characteristic were analyzed by the Kruskal-Wallis test because the distribution of the data was not homogeneous. This study

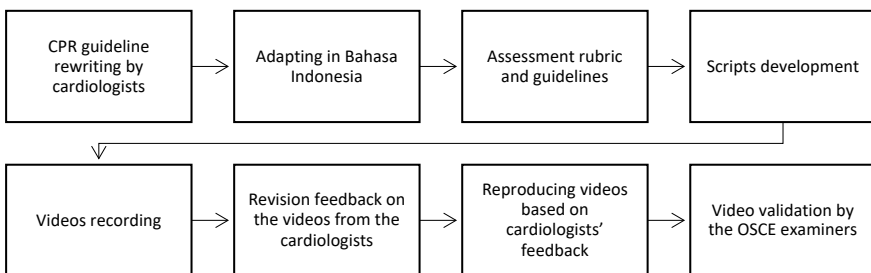


Fig. 1 The development of the clinical skill videos

was submitted to the Health Research Ethics Committee, Faculty of Medicine, Duta Wacana Christian University, while data collection was initiated after receiving approval (Reference No.1068/C.16/FK/2019).

3 Results

3.1 Script Development

Both video scripts were written and acted according to the American Heart Association's standardized rubrics and scoring guidelines. The scripts for these two videos were compiled by researchers, then reviewed and revised by two cardiologists. The two scripts were further developed into rubrics and assessment guides by the two cardiologists. Rubrics and assessment guides were prepared to evaluate student performance in CPR competencies. The validity of the content on the rubrics and assessment guides was achieved when the assessment instruments were reviewed by experts, who are cardiologists. This rubric and guideline for assessing CPR competencies coherently assessed three competencies, namely the primary survey, CPR procedures, and professional behavior that must be achieved on each value scale. The three competencies were defined in detail with specific explanations in the assessment guide. For CPR scripts that are not following the guidelines, standardized examinees performed <70% of clinical skills in the rubric, while for CPR scripts according to guidelines, standardized examinees performed >70% of clinical skills on the checklist.

3.2 Video Recording

The CPR video that showed performance following the guidelines and the other showed CPR not according to guidelines were recorded which contained the following: primary survey, CPR procedures, and professional behavior. All videos were recorded in the Skills Laboratory Faculty of Medicine, Duta Wacana Christian University with a digital Canon photographic camera. The sequence of video scripts was supervised by the researchers. The scripts were filmed by Medical Information Technology (IT) staff and were repeated several times to achieve the best situation that was written in the scripts. The cardiologists gave revision feedback on those two videos, then we reproduced the videos based on their feedback.

3.3 Video Validation

The validation of those two videos was conducted by the OSCE examiners as participants in this study. Participants in this study were 51 examiners described below in Table 1.

In giving the assessment, the median of two videos scoring results of the OSCE examiner’s assessment based on each background characteristic and the significance from the Kruskal–Wallis analysis can be seen in the following Table 2.

The CPR videos that showed performance following the guidelines provided results that were not significantly different in the average value of the assessment results between each characteristic of OSCE examiners. Significant differences occurred in the two groups of examiners’ characteristics, namely education and clinical experience when examiners assessed CPR competencies that were not following the guidelines. The median score for those groups was the same (33.33) with a *p*-value of 0.04; *df* 3 and *p*-value of 0.03; *df* 2, respectively.

Table 1 Study subjects’ characteristics

Background		Number of participants (<i>N</i> = 51)
Gender	Male	22 (43%)
	Female	29 (57%)
Education	Bachelor undergraduate	19 (37%)
	Master’s degree	16 (31%)
	Doctoral Degree	3 (6%)
	Specialist doctor	13 (25%)
Clinical practice experience	General practitioner	28 (55%)
	Specialist	14 (27%)
	No clinical practice	9 (18%)
Duration of clinical practice experience	<2 years	9 (18%)
	2–5 years	17 (33%)
	>5 years	25 (49%)
OSCE experience	<2 years	9 (18%)
	2–5 years	24 (47%)
	>5 years	18 (35%)
OSCE examiner training	<3 times	21 (41%)
	3–5 times	17 (33%)
	>5 times	13 (25%)

OSCE, Objective Structured Clinical Examination.

Table 2 Video assessment results

Video	Characteristics	Median	Max–Min score	<i>p</i> -value
According to guideline	Male	88.89	48.15–100.00	0.30
	Female	100.00	66.67–100.00	
	Bachelor undergraduate	100.00	66.67–100.00	0.17
	Master degree	88.89	48.15–100.00	
	Doctoral degree	100.00	100.00–100.00	
	Specialist doctor degree	88.89	59.26–100.00	0.50
	General practitioner	88.89	48.15–100.00	
	Specialist	94.44	59.26–100.00	
	No clinical practice	100.00	74.07–100.00	
	Clinical experience <3 years	88.89	66.67–100.00	0.90
	Clinical experience 3–5 years	100.00	48.15–100.00	
	Clinical experience >5 years	92.59	59.26–100.00	
	OSCE <2 years	100.00	66.67–100.00	0.83
	OSCE 2–5 years	94.44	59.26–100.00	
	OSCE >5 years	90.74	48.15–100.00	
	OSCE examiner training <3 times	100.00	66.67–100.00	0.83
	OSCE examiner training 3–5 times	100.00	48.15–100.00	
	OSCE examiner training >5 times	88.89	66.67–100.00	
	Not according to the guideline	Male	40.74	11.11–74.07
Female		33.33	0.00–59.26	
Bachelor undergraduate		33.33	0.00–59.26	0.04*
Master degree		37.04	0.00–55.56	
Doctoral degree		33.33	29.62–44.44	
Specialist doctor degree		44.44	33.33–74.07	0.12
General practitioner		33.33	0.00–59.26	
Specialist		42.59	29.63–74.07	
No clinical practice		33.33	11.11–55.56	
Clinical experience <3 years		33.33	11.11–40.74	0.03*
Clinical experience 3–5 years		33.33	0.00–55.56	

(continued)

Table 2 (continued)

Video	Characteristics	Median	Max–Min score	<i>p</i> -value
	Clinical experience >5 years	40.74	0.00–74.07	0.30
	OSCE <2 years	33.33	11.11–59.26	
	OSCE 2–5 years	33.33	0.00–51.85	
	OSCE >5 years	40.74	0.00–74.07	0.91
	OSCE examiner training <3 times	33.33	25.93–55.56	
	OSCE examiner training 3–5 times	33.33	11.11–59.26	
	OSCE examiner training >5 times	33.33	0.00–74.07	

* $p < 0.05$ shows significant differences

Min–Max, minimum–maximum; OSCE, Objective Structured Clinical Examination

4 Discussions

This study showed that the several steps to create a video for assessment, which were also done in this study, were planning or pre-production, recording or production, and editing or post-production [17]. Planning is important to ensure that the next step of video development is as expected, and this study describes how to develop and validate a video script, for which one video describes CPR that is appropriate and one video reflecting CPR not done properly according to the guidelines [18]. The video recording step needs to be supervised by the scriptwriter and the shooting must be done by a professional, which was also done in this study. This step is important so that video recordings record all relevant and objective information, can be seen clearly, and prevent video viewers from losing important details [19]. Post-production steps are also important as a final filter before the video is watched by video viewers as we did in the development of this assessment video. Submission of a post-editing video to the expert as the first viewer is expected so that the expert can identify potential gaps that can affect the assessment of the video, and provide an opportunity to make adjustments before the video is implemented [18].

The validation analysis of the two CPR videos in this study showed that although there were variations in the examiner's background that allowed differences in cognitive processes and various examiners' behaviors that could affect the assessment results, they were still able to provide consistent decisions. This study could illustrate that the results of the assessment of the two videos in this study were only influenced by differences in the performance of the students themselves. These results were consistent with previous studies showing similar results [20, 21]. Examiners will tend to make judgments easier and will give good judgment with accuracy when judging excellent performance and failing low quality performances because the examiners base their assessment on quantitative checklists of clinical skill performance [22,

23]. The tendency to more easily give assessments to students who perform well following the assessment guidelines is because the examiners base their assessments on quantitative measurements of the student's performance, including counting the number of correct points, and the examiners do not place more attention on the global assessment of pass and fail, so the examiner judges based on the fulfillment of checklist components [23]. This tendency can also be explained by when the examiner assesses good performance, it is easier for the examiner to choose the highest checklist point [22]. A video-based assessment accompanied by specific assessment instruments based on the newest and the most detailed evidence can increase the assessment's reliability [24].

As a reflection in the future, it is easier for examiners to give an assessment of good performance and the reduction of assessment deviations can be done by using specific cases indicating that there is a learning process when they evaluate when they use specific cases [24]. A video-based assessment using specific cases will be more effective than using general cases.

To minimize or avoid examiner biases, this example of video-based assessment can foster achievement of the highest objectivity of OSCE by applying this project in OSCE role-play scoring training. Through this role-play scoring training, we hope there will be the same perception between examiners on using assessment tools, using references, and minimizing the effect of background variability. Examiners' knowledge regarding their assessment performance, including the availability of clear checklists, understanding of the scoring rubric, a clear global rating scale, and how to rate it, is understandable so that it can be targeted in the training of OSCE examiners to minimize bias [25, 26].

One of the limitations of this study was that the results of this study could not be applied to other cases such as communication skills and clinical reasoning that had more complicated cases because in both cases there were differences in the way of assessment compared to the procedural skill with more standardized cases such as CPR in this study. The generalization was also a drawback in this study because the examiners came from a single institution. However, the examiners have the same standardization and are comparable with examiners in other institutions, hence, this approach can be also applied in other institutions.

Future research may use other clinical skills such as communication skills and clinical reasoning skills. Both have different forms of assessment and are more complex than the procedural skills in this study so that they can be used to answer with more certainty the influence of the examiners' backgrounds in conducting clinical skills assessments.

5 Conclusions

There were no significant differences in scoring between OSCE examiners, except for clinical practice experience and educational background categories. Video-based assessment can foster the objectivity of OSCE, hence, it can be applied in OSCE

scoring assessor training. However, this study shows that there are still sources of examiner biases that academics need to be aware of and consider.

Ethics Approval and Consent to Participate This study was approved by the Health Research Ethics Committee Faculty of Medicine Universitas Kristen Duta Wacana (Reference No.1068/C.16/FK/2019).

Competing Interest The authors declare that there are no competing interests related to the study.

Acknowledgements The author would like to thank the staff of the Faculty of Medicine, Universitas Kristen Duta Wacana for supporting the research.

Authors' Contribution *Oscar Gilang Purnajati*—conceived the research, reviewed the literature, designed the study, acquired funding, data analysis, and wrote the manuscript.

Rachmadya Nur Hidayah—developing study framework, data analysis the data, and reviewing the final manuscript.

Gandes Retno Rahayu—analyzed the data, reviewing the final manuscript.

References

1. Reid K, Smallwood D, Collins M, Sutherland R, Dodds A (2016) Taking OSCE examiner training on the road: reaching the masses. *Med Educ Online* 21:1–5. <https://doi.org/10.3402/meo.v21.32389>
2. Schleicher I, KL, Juenger J, Moeltner A, Ruessler M, Bender B, Sterz J, Schuettler K-F, Koenig S, Kreuder JG (2017) Examiner effect on the objective structured clinical exam: a study at five medical schools. *BMC Med Educ* 17(71):1–7. <https://doi.org/10.1186/s12909-017-0908-1>
3. Pell G, Homer M, Fuller R (2015) Investigating disparity between global grades and checklist scores in OSCEs. *Med Teach* 32(17):1106–1113. <https://doi.org/10.3109/0142159X.2015.1009425>
4. Kanada Y, Sakurai H, Sugiura Y (2015) Difficulty levels of OSCE items related to examination and measurement skills. *J Phys Ther Sci* 27(3):715–718. <https://doi.org/10.1589/jpts.27.715>
5. Taylor S, Haywood M, Shulruf B (2019) Comparison of the effects of simulated patient clinical skill training and student roleplay on objective structured clinical examination performance among medical students in Australia. *J Educ Eval Health Prof* 16:3. <https://doi.org/10.3352/jeehp.2019.16.3>
6. Zabar S, Kachur EK, Kalet A, Hanley K (2013) Objective structured clinical examinations 10 steps to planning and implementing OSCEs and other standardized patient exercises. Springer Science+Business Media, New York
7. Chahine S, Holmes B, Kowalewski Z (2015) In the minds of OSCE examiners: uncovering hidden assumptions. *Adv Health Sci Educ*
8. Chong L, Taylor S, Haywood M, Adelstein B-A, Shulruf B (2018) Examiner seniority and experience are associated with bias when scoring communication, but not examination, skills in objective structured clinical examinations in Australia. *J Educ Eval Health Prof* 15(17). <https://doi.org/10.3352/jeehp.2018.15.17>
9. Sobh AH, M.I. MI, Diab MI, Pawluk SA, Austin Z, Wilby KJ (2017) Qualitative evaluation of a cumulative exit-from-degree objective structured clinical examination (OSCE) in a Gulf context. *Pharm Educ* 17(1):73–80
10. Yeates P, Moreau M, Eva (2015) Are examiners' judgments in OSCE-style assessments influenced by contrast effects? *Acad Med* 90(7):975–980. <https://doi.org/10.1097/ACM.0000000000000650>

11. Kogan JR, Conforti LN, Iobst WF, Holmboe ES (2014) Reconceptualizing variable rater assessments as both an educational and clinical care problem. *Acad Med* 89(5):721–727. <https://doi.org/10.1097/ACM.0000000000000221>
12. Mortsiefer A, Karger A, Rothhoff T, Raski B, Pentzek M (2017) Examiner characteristics and interrater reliability in a communication OSCE. *Patient Educ Couns* 100:1230–1234. <https://doi.org/10.1016/j.pec.2017.01.013>
13. Massey D, Byrne J, Higgins N, Weeks B, Shuker M-A, Coyne E, Mitchell M, Johnston ANB (2017) Enhancing OSCE preparedness with video exemplars in undergraduate nursing students: a mixed method study. *Nurse Educ Today* 54:56–61
14. Erdogan A, Dong Y, Chen X, Schmickl C, Berrios RAS, Arguello LYG, Kashyap R, OK, Pickering B, Gajic O et al (2016) Development and validation of clinical performance assessment in simulated medical emergencies: an observational study. *BMC Emerg Med* 16(4). <https://doi.org/10.1186/s12873-015-0066-x>
15. Merchant RM, Topjian AA, Panchal AR, Cheng A, Aziz K, Berg KM, Lavonas EJ, Magid DJ, Basic A, Advanced Life Support PB et al (2020) Part 1: executive summary: 2020 American heart association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care 142(16_Suppl_2):S337–S357. <https://doi.org/10.1161/CIR.0000000000000918>
16. Panchal AR, Bartos JA, Cabañas JG, Donnino MW, Drennan IR, Hirsch KG, Kudenchuk PJ, Kurz MC, Lavonas EJ, Morley PTJC (2020) Part 3: adult basic and advanced life support: 2020 American heart association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care 142(16_Suppl_2):S366–S468. <https://doi.org/10.1161/CIR.0000000000000916>
17. Fleming SE, Reynolds J (2009) Wallace BJNe: lights... camera... action! a guide for creating a DVD/video 34(3):118–121
18. Lopes JdL, Baptista RCN, Domingues TAM, Ohl RIB, Barros ALBLdJRL-AdE (2020) Development and validation of a video on bed baths. *Rev Lat Am Enfermagem* 28:e3329. <https://doi.org/10.1590/1518-8345.3655.3329>
19. Beese NO, Rodriguez FS, Spilski J, Lachmann TJFiph (2021) Development of a digital video-based occupational risk assessment method. *Front Public Health* 9:683850. <https://doi.org/10.3389/fpubh.2021.683850>
20. Govaerts MJB, van der Vleuten CPM, Schuwirth LWT, Muijtjens AMM (2007) Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. *Adv Health Sci Edu* 12(2):239–260. <https://doi.org/10.1007/s10459-006-9043-1>
21. Naumann FL, Marshall S, Shulruf B, Jones PD (2016) Exploring examiner judgement of professional competence in rater based assessment. *Adv Health Sci Edu* 21(4):775–788
22. Byrne A, Soskova T, Dawkins J, Coombes L (2016) A pilot study of marking accuracy and mental workload as measures of OSCE examiner performance. *BMC Med Educ* 16:191. <https://doi.org/10.1186/s12909-016-0708-z>
23. Ali M, Pawluk SA, Rainkie DC, Wilby KJ (2019) Pass-fail decisions for borderline performers after a summative objective structured clinical examination. *Am J Pharm Educ* 83(2):142–147. <https://doi.org/10.5688/ajpe6849>
24. Daniels VJ, Bordage G, Gierl MJ, Yudkowsky R (2014) Effect of clinically discriminating, evidence-based checklist items on the reliability of scores from an internal medicine residency OSCE. *Adv Health Sci Edu* 19(4):497–506
25. Kozato A, Patel N, Shikino K (2020) A randomised controlled pilot trial of the influence of non-native English accents on examiners' scores in OSCEs. *BMC Med Educ* 20(1):268. <https://doi.org/10.1186/s12909-020-02198-y>
26. Chong L, Taylor S, Haywood M, Adelstein BA, Shulruf B (2017) The sights and insights of examiners in objective structured clinical examinations. *J Educ Eval Health Prof* 14:34. <https://doi.org/10.3352/jehp.2017.14.34>