ORIGINAL ARTICLE

Using Video Recording in Evaluating Skills of Medical Students in the Performance of the Orthopedic Examination

Jose Ma D. Bautista, MD,[1] Peter B. Bernardo, MD[1] and Mark Anthony R. Ruanto, MD[2]

[1]*Department of Orthopedics, College of Medicine and Philippine General Hospital, University of the Philippines Manila*
[2]*Department of Orthopedics, Philippine General Hospital, University of the Philippines Manila*

ABSTRACT

Objective. The study aims to assess the similarity between the results of the evaluation of students during an Objective Structured Clinical Examination (OSCE) and a video recording of the same OSCE (VOSCE).

Methods. All Orthopedic surgeon preceptors in the actual OSCE were recruited to the study. Video recordings of the students taking the OSCE were collected and later reviewed and re-evaluated by the same preceptor after at least four weeks. The grades of actual OSCE and VOSCE were collected and analyzed using Cohen's kappa coefficient.

Results. High variability of intra-rater reliability was observed in different preceptors and station (slight agreement to perfect agreement). Overall intra-rater reliability between actual and video OSCE showed moderate agreement with Cohen's kappa coefficient equal to 0.43 (n-219).

Conclusion. Video OSCE is a reliable tool in assessing student clinical skills and knowledge in the musculoskeletal examination. Some factors have been suggested to further improve reliability.

Key Words: video-recording, skills evaluation, OSCE

INTRODUCTION

Musculoskeletal conditions are a major problem that affects individuals and their activities. Developing good clinical skills to perform correct physical examination is essential to recognize and diagnose these conditions. However, some evidence suggests that the majority of medical students lack these skills.[1,2]

The GALS (Gait-Arms-Legs-Spine) Examination is a useful screening tool for musculoskeletal problems widely used in training and clinical practice. It is a reliable measure to function in a variety of musculoskeletal conditions.[3] Studies showed that this examination improves the diagnosis of musculoskeletal problems and improves the confidence of medical students and family physicians in the examination of the musculoskeletal system.[4,5,6] Even though there are still several specialists that do not use GALS as a clinical tool,[7] this examination was proven to be valid and reliable particularly when executed by a specialist (rheumatologist or orthopedic surgeons).[3,8]

Ensuring that medical students, even during their pre-clinical phase, can competently evaluate a patient with a musculoskeletal condition is therefore of prime importance. This competence may be measured by using the Objective

Corresponding author: Jose Ma Bautista, MD
Department of Orthopedics
Philippine General Hospital
College of Medicine
University of the Philippines Manila
Taft Avenue, Manila 1000, Philippines
Email: jdbautista2@up.edu.ph

Structured Clinical Examination (OSCE) which was introduced as a tool to evaluate medical students and trainees of their clinical skills and knowledge to improve the curriculum of training institutions.[9]

A systematic review which evaluated the validity and reliability of methods for objective skills assessment of surgeons identified the Objective Structured Assessment of Technical Skills (OSATS) as the "gold standard" for objective skills assessment.[10]

Despite its frequent use due to its many advantages, the need for a significant amount of faculty to conduct the examination, as well as its actual cost is seen as significant disadvantages.[11,12] The use of video recording to evaluate students' performance in the OSCE is an attractive alternative.

In another study, two raters were tasked to evaluate the students on their skill to secure informed consent for a surgical procedure using a single-station video-recorded objective structured clinical examination (VOSCE). The study suggests that VOSCE was a feasible, efficient, and reliable assessment method for medical students' skills showing a high inter-rater agreement.[13]

Using video recordings and live OSCE to compare students' performance for shoulder or knee examinations, a study showed moderate reliability between OSCE and VOSCE checklist scores.[14]

Driscoll et al. compared OSCE and VOSCE among surgical trainees. They studied the surgeon's tissue handling skills and showed the feasibility, validity, and reliability of video assessment of skills. The study suggests that it can improve skill assessment in training surgeons.[15]

Barratt also used VOSCE to evaluate student nurse's competence and safety in the performance of commonly used advanced clinical practice skills. The study proves that simulated OSCE video-recordings are a reliable tool to improve nurse practitioner training and education.[16]

Bautista showed excellent reliability between OSCE and VOSCE when used to evaluate students who took a mock examination as regards the performance of the GALS examination.[17]

This study was done to determine if evaluating a student's performance in an actual OSCE is the same whether the evaluator is physically present or if done by viewing a video recording. The results of this study could be used to address the disadvantages presented by the standard OSCE.

MATERIALS AND METHODS

All Orthopedic surgeon preceptors during the end-rotation Musculoskeletal OSCE of Year Level 4 students of the College of Medicine in the University of the Philippines-Manila were recruited to participate in the study. Before the exam, the study was explained to the students and written consents were secured. This study did not in any way interfere or influence changes on how the traditional OSCE was done nor did it influence how the students were graded.

The OSCE used an ordinal type of grading scheme using very satisfactory (VS), satisfactory (S), and unsatisfactory (US).

The OSCE consisted of 7 stations (6 exam stations and 1 rest station). Preceptors that participated in the study were assigned to 6 different exam stations. The students were grouped into 7 students per group for a total of 13 groups with one extra student who was included in the last group (thus the last group has 8 students). The groups were further divided into two batches (7 groups in batch A and 6 groups in batch B). Two groups, one from A and one from B, took the OSCE at a given time.

During the OSCE, the students were evaluated by the preceptors as they went around the stations using the standard evaluation tools. There usually was no interaction between the students and preceptors aside from the instruction to begin and to ensure that the student had completed the required tasks. The OSCE was recorded and the recordings were compiled per station per preceptor. The evaluation of each student was collected and tabulated.

At least four weeks after the actual OSCE, a compilation of the video recording of the students performing the OSCE in each station was given to the same preceptor who was assigned to that particular station. Using the same grading scheme, the preceptor re-evaluated the students using the video recording. The preceptors could go through the video recording at their leisure and were initially given 1 month to submit the grades. These grades were collected and compared to the grades given during the actual OSCE.

The analysis of intra-rater reliability was done using Cohen's kappa coefficient. Cohen kappa coefficient is interpreted as follows (based on Landis and Koch, 1977)[18]:

0	agreement equivalent to chance.
0.01 – 0.20	slight agreement.
0.21 – 0.40	fair agreement.
0.41 – 0.60	moderate agreement.
0.61 – 0.80	substantial agreement.
0.81 – 0.99	near perfect agreement
1	perfect agreement.

RESULTS

Ninety-one out of the 92 YL4 students who took the exam consented to have their OSCE video recorded. Seven out of eight orthopedic surgeons participated in our study and were assigned to six stations (3A, 4A, 4B, 5A, 5B, and 6B). Only these same seven preceptors who participated in the study reevaluated the video recording after at least four weeks from the date of the actual OSCE.

Initially, only four preceptors participating in the study started the exam (Consultants A, B, E & F). After the first batch of 7 students, the preceptor in Station 5A left and was replaced by another orthopedic surgeon who participated in the study (Consultant C). Then after four batches, the preceptor in station 5B (Consultant E) left and

Table 1. Consultant and Student assignment

Group	Station	Consultant	Number of Students who took the exam	Number of students who have VOSCE
A	3	A	50	50
	4	B	50	45
	5	C	43	43
B	4	D	21	21
	5	E & G	14	14
	6	F	41	37

Table 2. Intra-rater reliability analysis per consultant/rater

Rater	Subjects	Coefficient	Interpretation
A	50	0.0449	Slight agreement
B	50	0.1450	Slight agreement
C	43	0.2120	Fair agreement
D	21	0.1010	Slight agreement
E	14	0.4170	Moderate agreement
F	41	1	Perfect agreement

Table 3. Intra-rater reliability analysis per station

Station	Subjects	Coefficient	Interpretation
3	50	0.0449	Slight agreement
4	71	0.158	Slight agreement
5	57	0.275	Fair agreement
6	41	1	Perfect agreement

was replaced by another orthopedic surgeon (Consultant G) who also agreed to participate in the study. The preceptor in station 4B was initially not an orthopedic surgeon but was replaced by an orthopedic surgeon (Consultant D) who also participated in the study. This explains the difference in the number of students per station per preceptor. (Table 1)

Table 2 shows individual intra-rater reliability per rater per station. Results show a wide range of the Cohen Kappa coefficient (0.0449 to 1) from slight agreement to perfect agreement. Three preceptor ratings showed slight agreement on OSCE and VOSCE while the other three preceptor ratings showed fair, moderate, and perfect agreement.

Table 3 shows intra-rater reliability per station of the OSCE. Based on the Cohen Kappa Coefficient, the results show slight to perfect agreement. Two stations (Station 3 and 4) have slight agreement while Station 5 has a fair agreement between OSCE and VOSCE grades. One station (which has only one preceptor who participated – Rater F) showed a perfect agreement.

When all preceptor ratings for all stations were combined, overall intra-rater reliability for the OSCE (combining analysis of all the consultant; n=219) shows a moderate agreement with a Cohen kappa coefficient of 0.43.

DISCUSSION

Our study showed variable intra-rater reliability per station and rater. Comparing the students' scores during the actual OSCE and after at least 4 weeks by watching a video recording of their exams showed highly variable results. It did show that overall intra-rater reliability between actual and video OSCE was moderate (Cohens' kappa: 0.43). These results are quite different compared to those of the previously cited studies.

Excellent intra-observer reliability was seen in the study of Bautista when the evaluators used the same video recording of the students on separate dates.[17] This is as opposed to the present study's evaluation of students during the actual OSCE and using a video recording at a later date. Having to observe and grade 50 consecutive students over five hours without taking breaks will take its toll on an evaluator. A tired and hungry faculty member cannot be expected to be truly objective anymore. This is as opposed to when one has the option to limit the amount of time and number of students being evaluated at any given time. One can also choose to pause and rewind the recording if an evaluator has doubts as to the student's answers and skills.

Previous studies comparing the evaluation of students and trainees using real-time video vs recordings of the same video showed higher reliability than the present study. The studies by Driscoll, Sturpe, and Kiehl all showed excellent reliability.[13,15,19] Although done in real-time, none of the studies seemed to require their evaluators to stay for five hours. Another factor could be the difference in what the evaluator can see and observe during the actual OSCE and what is caught in the recording by the camera. The possibility that a difference in what is observed may lead to lower reliability is also seen in Bautista's study.[17] Although involving only a limited number of examinees (eleven students), there was higher intra-rater reliability between those who used the same video recording at two separate times to grade the students taking the mock OSCE as opposed to those who first evaluated the actual OSCE and then the video recording.

Actual interaction between the preceptor and the student during a live OSCE can also affect the reliability of the test. The preceptor can seek clarifications from a student that can affect their grade. Observing an examinee's reaction might likewise sway an examiner's evaluation. This interaction is not present during the VOSCE. In this study, the recordings stopped after the time allotted for the student to take the exam expired. However, in the actual OSCE, preceptors were able to allow the student to finish the exam he/she is performing before giving the grade. Kiehl showed more failing grades given during the evaluation of video recordings.[13] It was not in this study's scope to determine in which format were the given grades higher.

One factor identified during the study that possibly affected the result is the video and audio quality. Construction was ongoing in a nearby building while the examination was taking place. Other technical problems experienced, like changing cameras, as well as the changing position of the patient and student might have affected the area of focus during the examination. As was shown in the study by Bautista during the mock OSCE, multiple cameras per

station (2 cameras with different viewpoints) could be used to improve the recording, and the reliability too.[17] This however would be an additional expense.

Another factor that could have affected the reliability of the test was the evaluation tool used to grade the OSCE. The final score was the preceptor's subjective assessment (very satisfactory, satisfactory, unsatisfactory). During the OSCE, each preceptor was provided with a guide on how to grade the student using a checklist. It however did not provide a basis to compute for the final grade. Previous studies, regardless of whatever combination of OSCE, real-time VOSCE, or recorded VOSCE, which used more objective tools showed excellent intra-rater reliability. Driscoll used both the Toronto scale and the Edinburgh Basic Surgical Trainee Assessment Form.[15] Bautista's study used the University of Toronto's Clinical Skills Assessment and Feedback Tool.[17] Vivekanada-Schmidt's study, which used both a 14-point checklist and a global rating scale showed higher reliability with the checklist (excellent vs moderate).[14]

CONCLUSION

Using a video recording of an OSCE is a promising tool in assessing a student's clinical skills and knowledge in the musculoskeletal examination. The results of this study add to the increasing evidence that the OSCE need not be done with all stations manned by an actual staff member. The use of an objective grading tool, having two (or more) cameras to do the recording and allowing evaluators their own pace as to accomplishing the evaluation of the video recordings seem to further improve reliability.

Statement of Authorship

All authors participated in data collection and analysis, and approved the final version submitted.

Author Disclosure

All authors declared no conflicts of interest.

Funding Source

No funding was provided for this study.

REFERENCES

1. Patel V, Patel P, Jeffery R, Taylor J, Thomas H. Examination of the musculoskeletal system: junior doctors' perceptions of the usefulness of the Gait, Arms, Legs and Spine (GALS) technique. Postgrad Med J. 2015;91(1078):418-22. doi: 10.1136/postgradmedj-2015-133340.
2. Baker KF, Jandial S, Thompson B, Walker D, Taylor K, Foster HE. Use of structured musculoskeletal examination routines in undergraduate medical education and postgraduate clinical practice – a UK survey. BMC Medical Education. 2016;16:277. doi: 10.1186/s12909-016-0799-6.
3. Doherty M, Dacre J, Dieppe, P, Snaith M. The "GALS' locomotor screen. Ann Rheum Dis. 1992. 51(10):1165-1169. doi: 10.1136/ard.51.10.1165
4. Fox, RA, Dacre JE, Ingham Clark CL, Scotland AD. Impact on medical students of incorporating GALS screen teaching into the medical school curriculum. Ann Rheum Dis. 2000.59(9):668–71.
5. Beattie, KA, Bobba R, Bayoumi I, Chan D, Schabort I, Boulos, P, et al. Validation of the GALS musculoskeletal screening exam for use in primary care: a pilot study. BMC Musculoskeletal Disorders. 2008. 9:115. doi: 10.1186/1471-24784-9-115.
6. Lillicrap MS, Byrne E, Speed CE. Musculoskeletal assessment of general medical in-patients – joints still crying out for attention. Rheumatology (Oxford). 2003.42(8):951-954. doi: 10.1093/rheumatology/keg259.
7. Blake T. Teaching musculoskeletal examination skills to UK medical students: A comparative survey of Rheumatology and Orthopaedic education practice. BMC Medical Education. 2014. 14:62. doi: 10.1186/1472-6920-14-62.
8. Foster HE, Kay LJ, Friswell M, Coady D, Myers A. Musculoskeletal screening examination (pGALS) for school-age children based on the adult GALS screen. Arthritis Rheum. 2006. 55(5):709-16. doi: 10.1002/art.22230
9. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J. 1975.1(5955): 447-451. Doi: 10.1136/bmj.1.5955.447.
10. Van Hove PD, Tuijthof GJM, Verdaasdonk EGG, Stassen LPS, Dankelman J. Objective assessment of technical surgical skills. British Journal of Surgery. 2010 Jul. 97(7):972-87. Doi: 10.1002/bjs.7115
11. Hamann C, Volkan K, Fishman MB, Silvestri RC, Simon SR, Fletcher SW. How well do second-year students learn physical diagnosis? Observational study of an objective structured clinical examination (OSCE). BMC Med Educ. 2002; 2:1. doi:10.1186/1472-6920-2-1
12. Zayyan M. Objective Structured Clinical Examination: The Assessment of Choice. Oman Med J. 2011 Jul; 26(4): 219-222.
13. Kiehl C., Simmenroth-Nayda A, Goerlich Y, Entwistle A. Standardized and quality-assured video-recorded examination in undergraduate education: informed consent prior to surgery. Journal of Surgical Research.2014.191(1): 64-73. doi: 10.1016/j.jss.2014.01.048
14. Vivekanada-Schmidt P, Lewis M, Coady D, Morley C. Exploring the Use of Videotaped Objective Structured Clinical Examination in the Assessment of Joint Examination Skills of Medical Students. Arthritis & Rheumatism. 2007.57(5): 869-876. doi: 10.1002/art.22763
15. Driscoll, P, Paisley AM, Paterson-Brown S. Video assessment of basic surgical trainees' operative skills. The American Journal of Surgery. 2008 Aug.196(2): 265-272. doi: 10.1016/j.amjsurg.2007.09.044.
16. Barratt, J. A focus group study of the use of video-recorded simulated objective structured clinical examinations in nurse practitioner education. Nurse Education in Practice.2010 May.10(3):170–5. doi: 10.1016/jnepr.2009.06.004.
17. Bautista J, Manalastas R. Using Video recording in Evaluating Clinical Skills. Medical Science Educator. 2017. 27(4): 645-50.
18. Landis JR, Koch G. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977 March. 33(1): 159-74.
19. Sturpe DA, Huynh D, Haines ST. Scoring Objective Structured Clinical Examinations Using Video Monitors or Video Recordings. American Journal of Pharmaceutical Education 2010 April.74(3): 44. doi: 10.5688/aj740344.