

IMPLEMENTATION OF AN OBJECTIVE STRUCTURED CLINICAL EXAMINATION (OSCE) IN A KINESIOLOGY BACHELOR DEGREE

Sarah-Caroline Poitras¹, Sara Bélanger², Philippe Corbeil², Andréane Lambert-Roy²,
& Adrien Cantat¹

¹Vice-Dean's Office for Pedagogy, Laval University, Medical Faculty (Canada)

²Kinesiology department, Laval University, Medical Faculty (Canada)

Abstract

For the past four years, the Laval University's kinesiology bachelor degree has been using an OSCE to assess their students' clinical competencies. This presentation will describe the format the OSCE and will discuss the quality improvement process implemented following psychometric analysis of the test and its nine stations. The psychometrics qualities were tested using the Cronbach's alpha, the stations' difficulty level and discrimination index. An ANOVA has also been realised to ensure that the students' results of a same station in a different circuit were equivalent regarding the examiners.

In the 2018 edition, the psychometrics qualities were under the standards, especially the Cronbach's alpha and the stations' discrimination index. In order to improve the reliability of the test, modifications were made to seven of nine scoring grids. A committee revised each competencies' component assessed and removed the ambiguous ones. The psychometrics qualities of the revised results improved accordingly.

To prepare the 2019 edition, the OSCE committee reviewed the nine stations and adjusted the scoring grids. It also designed three new clinical situations. The psychometrics qualities of the 2019 edition have shown an improvement of the Cronbach's alpha and the stations' discrimination index. It has also demonstrated no significant differences between the circuits' performance. The appreciation surveys administered following each edition revealed the quality of the support offered to the students, examiners and simulated patients and the authenticity of the clinical situations. We thus consider this OSCE to be a reliable method to assess students' competencies of our kinesiology program.

Keywords: *Objective structured clinical examination (OSCE), assessment, competencies, kinesiology, test score reliability.*

1. Introduction

1.1. Kinesiology bachelor degree program's revision

From 2013 to 2016, the bachelor's degree in kinesiology at Laval University's Faculty of Medicine underwent a complete overhaul of its program, now centred on the competency-based approach (Scallon, 2004). In order to foster these skills' development, three classes on clinical approach were integrated to the program. They bring students to solve clinical vignettes involving a variety of clients and kinesiology-related issues. Due to the complex nature of the acquired knowledge, the implementation of these courses has necessitated a revision of the evaluation methods traditionally used in this program. Among the changes that were made, we note the addition of an objective structured clinical examination (OSCE) that is worth 40% of the last clinical approach course.

1.2. Project's objectives

This presentation aims at evaluating the validity of the first and second editions of the kinesiology program's OSCE (2018 and 2019). It is hypothesized that the OSCE provides a valid assessment to validate students' skills development following the last clinical approach course and to identify those who may be experiencing academic difficulties.

2. Design of the objective structured clinical exam in Kinesiology

The kinesiology OSCE contains nine clinical vignettes, divided into eleven ten-minute stations, interspersed with a two-minute transition time (Figure 1). Two of these vignettes are made of two linked stations, meaning that they are based on the same case. Completing this eleven stations circuit takes 2h36 minutes. The exam is divided into two periods: in the morning and in the afternoon. To ensure that all

forty registered students complete their examination in the same day, two circuits are held concurrently in the morning (10 students per circuit) and two others in the afternoon.

For nine stations, students have to interact with a patient who is played by an actor. To guarantee that the actors' performances are realistic and stable from one student to the other, actors meet with content experts for each vignette so that their performance is standardized (Loye & Fontaine, 2018). In addition, there are two writing stations: one in which the students must analyze a sports performance shown on video and another one during which they have to review an exercise program.

The students' performance, in acted vignettes, is evaluated by an examiner who is present in the room. The examiner may be a content expert for this particular vignette, a teacher in the program or a clinician. The evaluation grids that are used are mixed (Harden et al., 2016; Regehr et al., 1998). They contain a list of tasks that are assessed with a binary scale (observed/not observed), in addition to an overall qualitative scale measuring, with four response options, the level of students' professionalism (superior, adequate, limit, unacceptable).

Figure 1. Kinesiology's OSCE Design Plan.



3. Method and results

As in other studies, the OSCE validity of both editions was assessed using four of the five sources of Messick's validity framework (1989; Yazbeck Karam et al., 2018). These are presented in Table 1. Appreciation surveys were also submitted to each participant at the end of the OSCE (students and examiners) to collect their opinion about some aspects of the event. A four-point Likert scale was used in each survey (totally agree, agree, disagree, and strongly disagree). 75% percent of the students responded to the 2018 survey, whereas 55% responded to the 2019 survey. For the examiners, the response rate is respectively 78% in 2018 and 72% in 2019.

Table 1. Sources of validity, description and measures used to assess OSCE psychometric qualities.

Source	Description	Measure
Content	Assessed competencies in the OSCE represent the competencies taught in the courses	Clinical expert (clinical station development) Examiners' opinion (survey) Table of specification (Blueprint)
Internal structure	Psychometric properties of the OSCE and the stations	Cronbach's alpha (OSCE results) Discrimination index (stations results) Level of difficulty (stations results) Success rate for each competency's assessed component (stations results)
Response process	Evidence of data coherence	Comprehension of the task (students' survey) Examiners' opinion (survey) Analysis of variance (ANOVA) between mean groups (am, pm) and circuits (OSCE results)
Relation with other variables	Alignment of results with similar tools measuring the same subject	Correlation between OSCE scores and third clinical course scores (60% other than OSCE ones).

Content validity was supported by three initiatives. First, every station was designed by clinical experts, accompanied by an assessment consultant. Blueprints of the OSCE and the third clinical course were elaborated and compared to assure that the proportion of competencies assessed was respected in both. In addition, examiners' opinion was collected and they all agreed that the stations were realistic, authentic and well developed (mean of 3,9/4 on Likert's agreement responses scale).

The results of the 2018 edition were globally good (mean = 74,5%, low score = 60,3%, high score = 85,7%, stations' difficulty levels were between 57,5% and 85,4%). All students succeeded the

test. In regards to internal structure, some results of the psychometric analysis were below the standards, especially the Cronbach's alpha ($\alpha = ,40$) under the adequate rate of ,66 (Brannick et al. 2011). In addition, four stations' discrimination indices were inadequate to discriminate the good from the inferior performances ($D_{station1} = ,041$; $D_{station3} = -,022$ $D_{station5} = -,025$ $D_{station7} = ,032$). Otherwise, validity by response process was supported with the nonsignificant analysis of variance between the means of the groups and circuits ($F_{groups} = ,068$, $p = ,854$; $F_{circuit} = ,035$, $p = ,854$). It was also supported by students' survey answers about the clearness and unambiguous instruction of the clinical stations (mean of 3,8/4), and the observers' opinion about the ease of use of score grids (mean of 3,2/4). Students also agreed that the OSCE should be maintained in the program (mean of 3,5/4). The validity was also supported by the significant and positive correlation between OSCE's and the third clinical course's results (Pearson $r = ,729$, $p = ,000$).

To improve the internal structure validity of the second edition, a committee, composed of clinical experts, kinesiology teachers and an assessment consultant, revised competency's components failed by 70% or more of the students. The committee also created two new clinical stations to renew the exam and to reduce the chances of plagiarism from one year to the next.

The results of the second edition were as good as the previous one (mean = 74,8%, low score = 55,7%, high score = 89,8%, stations' difficulty levels were between 71,6% and 84,4%), although three students failed (score under 60%). The internal structure has shown improvement. More specifically, the Cronbach's alpha was adequate ($\alpha = ,747$) and 8 out of 11 stations showed an acceptable discrimination index ($D > ,20$; Bernier & Pietruliewicz, 1997). No significant difference was observed between the means of the groups and circuits ($F_{groups} = 1,345$, $p = ,254$; $F_{circuit} = ,862$, $p = ,360$). The appreciation surveys revealed almost the same results as in the 2018 edition for students' perception of clearness and unambiguous instruction of the clinical station (mean of 3,9/4) and for the ease of use of score grids (mean of 3,5/4). 2019 edition's students were also favorable to maintain the OSCE for the next cohort (mean of 3,7/4). Finally, a significant and positive correlation between OSCE's and the third clinical course's remaining results (60% other than OSCE ones) was also found (Pearson $r = ,888$; $p = ,000$).

4. Discussion and conclusion

Based on the results of the validity sources, we consider the OSCE in the kinesiology program as a valid method to assess students' competencies at the end of the third clinical course. Both editions seem to measure representative clinical courses' learnings. The OSCE situations appear to be realistic and the score grids are considered well developed. Stations are not too difficult nor too easy and, especially for the 2019 edition, are effective to discriminate a good from an inferior performance. The standardisation of the actors looks effective, as well as the examiners' training.

Nevertheless, here are some limitations about this study. First, the OSCE results are only based on checklist grids whereas many studies showed that global rating should be more effective to assess competencies (Hodges et al., 1999; Reznick et al., 1998). A global score should be added to the next edition to compare this score to the checklist results. Finally, data should be collected about the feedback provided on students' performance at the OSCE.

References

- Brannick, M. T., Erol-Korkmaz, H. T., & Prewett, M. (2011). A systematic review of the reliability of objective structured clinical examination scores. *Medical education*, 45, 1181-1189.
- Harden, R., Lilley, P., & Patricio, M. (2016). *The Definitive Guide to the OSCE: The Objective Structured Clinical Examination as a performance assessment*. Edinburgh: Elsevier.
- Hodges, B., Regehr, G., McNaughton, N., Tiberius, R., & Hanson, M. (1999). OSCE checklists do not capture increasing levels of expertise. *Academic Medicine*, 74, 1129-1134.
- Loye, N., & Fontaine, S. (2018). S'instrumenter pour évaluer. *Pédagogie médicale*, 19, 95-107.
- Messick, S. (1989). Validity. In: R. L. Linn (Ed.). *Educational measurement* (3rd ed., pp. 13-104). New York: Council on Education; Macmillan Series on Higher Education.
- Regehr, G., MacRae, H., Reznick, R. K., & Szalay, D. (1998). Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Med.*, 73(9), 993-7.
- Reznick, R. K., Regehr, G., Yee, G., Rothman, A., Blackmore, D. & Dauphinee, D. (1998). Process-rating forms versus task-specific checklists in an OSCE for medical licensure. *Academic Medicine*, 73, 97-99.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Saint-Laurent: Éditions du Renouveau pédagogique.
- Yazbeck Karam, V., Park, Y. S., Tekian, A., & Youssef, N. (2018). Evaluating the validity evidence of an OSCE: results from a new medical school. *BMC Medical Education*, 18, 313-319.