Developing an Objective Structured Clinical Examination to Assess Clinical Competence

Naumann F.L.
Faculty of Medicine, School of Medical Sciences, Exercise Physiology . The University of New South Wales
Faculty of Medicine, Prince of Wales Clinical School , The University of New South Wales

Moore K.M.
Learning and Teaching Unit, The University of New South Wales
School of Health and Human Sciences, Southern Cross University

Keywords: Exercise Physiology, Objective Structured Clinical Examination, Work-simulated Assessment of clinical skills.

Abstract

The distinguishing features of skill-based learning situations are that they are inherently variable, unpredictable, sometimes brief, high risk learning events which are not replicable, and this presents challenges for the quality assurance requirements of the assessment process (Cooper, Orrell & Bowden; 2003, Hodges, 2011; Yorke, 2011). One potential method for assessing clinical competence in the university setting is the Objective Structured Clinical Examination (OSCE).

This paper describes an action research project, a three-phase pilot of an OSCE for Exercise Physiology undergraduate students. Fifty-six students participated. As a means of assessing validity, after each of the three pilots the following elements of the design and implementation reviewed: number of stations, time allocated to station, content of each station, descriptors’ of tasks, preparation and training of simulated patients and examiners and assessment criteria.

Based on a comparative analysis all station and domain items, in the first OSCE, the Cronbach alpha for the examination was 0.52 and in the third OSCE 0.86, indicating continued improvement. This suggests the results would be reliable for making a summative judgement of student performance of Exercise Physiology clinical skills.

Introduction

In order for Exercise Physiology students to be accredited to practice, they undertake university studies and are required to complete practicum experience in a range of settings and environments. It is generally expected that during the practicum students will be assessed on their application of theory to practice. It is expected that these activities strengthen the work readiness of graduates.

However, assessing students during a practicum can be highly variable in terms of the opportunities for assessment and the skills of the examiners. This means some students face more stringent examinations than other students which raises questions of the equity of in-situ exams. Universities are rethinking curriculum design and assessment practices related to work-integrated learning (WIL) (Cooper et al, 2010). Therefore researching the effectiveness of skill-based assessment, work-simulated assessment practices, is both timely and pertinent to the current climate.

One potential method for assessing clinical competence in the work-simulated setting is the Objective Structured Clinical Examination (OSCE). The OSCE was first described by Harden et al. (1975) as a means to assess the clinical skills of final year medical students. Since its development, the OSCE has gained acceptance as a benchmark for assessing clinical competence (Bartfay et al. 2004) and has been adopted by several health professions including radiography (Marshall and Harris, 2000), nursing (McKnight et al., 1987; Alinier, 2006),
physiotherapy (Nayer, 1999) and dentistry (Brown et al. 1999). As Exercise Physiology is a relatively new allied health profession, it is critical to demonstrate that exercise physiologists have proficient clinical skills and OSCE was regarded by the team as a viable option.

The OSCE aims to enhance the validity of clinical assessment by simulating realistic clinical scenarios to reflect real-life professional tasks. Typically, an OSCE consists of multiple stations at which students are required to perform a range of tasks to demonstrate competency in relevant skills. During OSCE students rotate through the various stations spending a predetermined time. At each station the students are instructed to complete specified tasks in the time allocated. Stations can involve actual patients or surrogate volunteers who are normally trained to present a standardised portrayal of the clinical problem. Each station is scored separately using either detailed checklists or global rating scales.

Establishing reliable assessment is critical to the success of the Exercise Physiology Program. The underlying premise is that standardised assessment procedures ensure objectivity and maximise reliability (Bartfay et al. 2004; Major 2005). The aim of this study was to assess both the validity and reliability of an OSCE for assessing clinical skills of final year Exercise Physiology students. That is, we explored if each station in the Exercise Physiology OSCE tested the construct that it was intended to assess.

**Methodology**

This action research project constituted three pilots of an OSCE in the Exercise Physiology OSCE during 2011. Fifty-six final-year students participated. In addition to analysis of data from test results, feedback from examiners, clinical education staff and students provided insight into how the OSCE were perceived, the learning value and what improvements might be made.

**The OSCE**

The content of the work-simulated examinations were based on the diversity of skills expected of an exercise physiologist. Careful design of stations was required to ensure systematic and sufficient samples of the skill-based competencies, a key element of validity. A range of clinical presentations, typically expected in daily practice, were selected. These included the screening of healthy persons as well as assessment of patients with metabolic, cardiovascular, neuromuscular and musculoskeletal disorders. Each station assessed the student’s competency in the following domains:

- communication skills;
- clinical and procedural skills; and
- technical skills.

The assessment criteria and grading system used at each station evolved as the series of pilots progressed. Initially, checklists for assessing station-specific tasks were used. These were replaced by a global rating scheme employing criteria aligned to three competency domains mentioned above. A numerical scale was used for statistical analyses of test results. Therefore, each station was graded as follows:

- F for failed performance (score 0);
- P- for borderline performance (score 1);
- P for good performance (score 2); and
- P+ for excellent performance (score 3).

To strengthen validity, after each pilot, elements of the design and implementation of the OSCE were scrutinised. The elements reviewed were the:

- number of stations;
- time allocated to station;
- content of each station;
- descriptors’ of tasks;
- preparation and training of simulated patients; and
- preparation and training of examiners and assessment criteria.
Statistical analysis

The reliability of the examinations were analysed using methods based on classical test theory. This included generalisability theory which is an estimate of the examination’s reliability to consistently rank students. This means, in order for a station to be considered reliable, it should assess various tasks across the three competency domains and, distinguish between the best and worst performing students.

The measure for internal consistency utilised Cronbach alpha. The results for all marking criteria as well as the aggregate results from stations and competency domains were analysed. We were looking for a value >0.7 from the Charonbach alpha and generalizability coefficients analysis. All statistical analyses were performed using SPSS v20 with significance set at p<0.05.

Discrimination index was ascertained using Pearson’s correlation

In addition, in order to determine the effect of varying the number of stations or competency domains, Decision (D) studies were performed. This showed the minimum number of stations and domains needed to produce a reliable exam result.

Content validity was measured by using feedback from clinical education experts, examiners and academic staff. Assessing the content validity ensured representation of the diversity of clinical conditions typically encountered in clinical practice and the clinical competencies required to assess and manage patients.

Results

Based on analysis of all station and domain items in the first OSCE, the Cronbach alpha for the examination increased from 0.52 in the first OSCE to 0.86 for the third OSCE, indicating improvement across time which infers the OSCE is reliable for making a summative judgement of student performance (Table 1).

Table 1: Cronbach alpha across all items of the 3 pilot OSCEs

<table>
<thead>
<tr>
<th></th>
<th>Cronbach's Alpha</th>
<th>N of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSCE 1</td>
<td>0.520</td>
<td>22</td>
</tr>
<tr>
<td>OSCE 2</td>
<td>0.740</td>
<td>20</td>
</tr>
<tr>
<td>OSCE 3</td>
<td>0.863</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 2 shows the results of analysis of the Discrimination index (measured by Pearson correlation) for each station. Significant discrimination index’s (the term, significant indicates that the ranking of students at these stations, correlate with the overall OSCE performance ranking which were as follows.

- OSCE 1, there were only two such stations: which were significant
- OSCE 2, the discrimination index was significant for four stations.
- OSCE 3, the discrimination index was significant for all stations indicating that the ranking of students in these stations correlated strongly with the overall ranking

The aim of a good exam set up should be to have good students perform well at all stations. If there is an anomaly, it indicates the station could be suspect. The terms, significant meant that the top students did well on all stations
Table 2: The mean, SE and discrimination index for all OSCE stations

<table>
<thead>
<tr>
<th>Station</th>
<th>OSCE 1</th>
<th></th>
<th>OSCE 2</th>
<th></th>
<th>OSCE 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean SE</td>
<td>R*</td>
<td>Mean SE</td>
<td>R*</td>
<td>Mean SE</td>
<td>R*</td>
</tr>
<tr>
<td>CV Oncology</td>
<td>0.65</td>
<td>0.19</td>
<td>0.66</td>
<td>0.04</td>
<td>0.66</td>
<td>1.93</td>
</tr>
<tr>
<td>VO2 interpretation</td>
<td>0.50</td>
<td>0.60</td>
<td>0.62</td>
<td>0.71</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td>ECG interpretation</td>
<td>0.51</td>
<td>0.64</td>
<td>0.38</td>
<td>0.72</td>
<td>0.05</td>
<td>0.51</td>
</tr>
<tr>
<td>ROM</td>
<td>0.54</td>
<td>0.53</td>
<td>0.60</td>
<td>Rest</td>
<td>Rest</td>
<td></td>
</tr>
<tr>
<td>Lung function</td>
<td>0.60</td>
<td>0.16</td>
<td>0.15</td>
<td>Falls</td>
<td>0.67</td>
<td>0.04</td>
</tr>
<tr>
<td>Falls</td>
<td>0.57</td>
<td>0.06</td>
<td>0.04</td>
<td>Falls</td>
<td>0.54</td>
<td>0.57</td>
</tr>
<tr>
<td>Health &amp; Fitness</td>
<td>0.60</td>
<td>0.55</td>
<td>0.86*</td>
<td>0.72</td>
<td>0.03</td>
<td>0.76*</td>
</tr>
<tr>
<td>Mus-Sk</td>
<td>0.48</td>
<td>0.70</td>
<td>0.06</td>
<td>0.82*</td>
<td>0.73</td>
<td>1.49</td>
</tr>
</tbody>
</table>

Key:
CV Oncology – Cardiovascular oncology; VO2 interpretation – maximum Oxygen consumption
ECG Interpretation – Electrocardiogram interpretation; ROM – Range of Motion;
Mus-Sk – Musculoskeletal assessment; * Pearson correlation with total scores. Significant at the 0.05 level (2-tailed).

The effect of varying the number of domains and stations was explored Decision (D) studies. Figure 1 shows increasing the number of domains from more than three had little effect on the generalizability coefficient >0.7 with only seven stations. Similarly increasing the number of stations beyond seven had little effect on improving generalizability.

![Figure 1: Generalizability Coefficient by domain and station number.](image-url)
As far as content validity is concerned, staff and Clinical Educator feedback enabled station task and assessment criteria refinement throughout the project. Analysis of student feedback indicated that overall the OSCE was well perceived as a good way to assess their clinical competency. The student cohort remarked that it is essential to fully brief them on the format of the exam and assessment criteria. Staff and Clinical Educator feedback fully supported to use of the OSCE to assess clinical competence.

Discussion and Conclusion

Statistical analysis allowed improvements to be made to the reliability of the OSCE, exceeding the accepted Cronbach alpha benchmark of 0.7. This is consistent with the reported reliabilities for an OSCE in the medical and other allied health fields. This would suggest that the results from this examination would be reliable for making a summative judgement of Exercise Physiology student’s performance. Improvements in exam reliability were achieved through better selection and training of volunteer patients and surrogates, better preparation of the students, improved assessment criteria for the examiners to work from and a refinement in task requirements to be performed within the allocated time.

It is critically to successful implementation of an OSCE to determine the optimal number of stations and domains required to provide a reliable assessment of student performance. Decision (D) studies aid this process. Based on generalizability theory, Decision (D) studies can estimate the effect of varying conditions on the generalizability of an examination. Data from OSCE 2 and OSCE 3 shows the combination of seven stations: Cardiovascular oncology, diabetes, SCG interpretation, Strength, Falls, Health and Fitness and Musculo-skeletal assessment, each with three domains: communication, clinical and procedural skills and technical skills, achieved satisfactory generalizability. Increasing the number of stations or domains would not improve the reliability of the examination significantly.

As far as administration is concerned, an OSCE is both labour and time intensive. Nevertheless, one of the many benefits was that this study showed through an iterative process involving analyses of staff and student feedback along with statistic data from three pilot examinations, an OSCE can evolve into a reliable and valid means of assessing, via simulation, clinical skills in Exercise Physiology students.

An additional benefit was that the process of developing the OSCE facilitated a review of the curriculum. This project enables a fresh look at all aspects of clinical competency requirements throughout the curriculum. In some institutions, an OSCE is conducted prior to students undertaking WIL experiences. In this instance however, we used the OSCE after the WIL experience. Given that assessing students skills in-situ, during real- time clinical consultations can be highly variable an OSCE facilitates greater efficacy and equality in the assessment of a range of clinical skills. An OSCE is a valuable tool in the overarching assessment strategy of the clinical component of the Exercise Physiology curriculum.

Limitations of the study

The limitations are that this study focused on a small student cohort at one University. The student cohort were inexperienced taking this sort of exam. In future, an OSCE will be conducted in second year, to get them ready for the 4th year OSCE?

Directions for future research

This project has been followed by a second study in which we compared examiner reliability for assessment of the exercise physiology objective structured clinical examinations using observation of filmed student performance. It is anticipated that this project will add further academic rigor to the assessment of the students skills prior to engaging in and after WIL.
References


Copyright © 2012 Naumann F.L, Moore K.M.

The author assign to the Australian Collaborative Education Network (ACEN Inc.) an educational non-profit institution, a nonexclusive licence to use this document for personal use and in courses of instruction, provided that the article is used in full and this copyright statement is reproduced. The author also grants a non-exclusive licence to the Australian Collaborative Education Network to publish this document on the ACEN website and in other formats for the Proceedings ACEN National Conference Melbourne / Geelong 2012. Any other use is prohibited without the express permission of the author.