Developing an objective structured clinical examination to assess work-integrated learning in exercise physiology

FIONA NAUMANN
KERI MOORE
SALLY MILDON
PHILIP JONES
The University of NSW, Sydney, Australia

This paper aims to develop a valid method to assess the key competencies of the exercise physiology profession acquired through work-integrated learning (WIL). In order to develop a competency-based assessment, the key professional tasks needed to be identified and the test designed so students’ competency in different tasks and settings could be assessed. The objective structured clinical examination (OSCE) was adopted for this purpose. Key competencies for practice were identified by clinical academics, practicing accredited exercise physiologists (AEPs), and by a review of the exercise physiology scope of practice document. Final year exercise physiology students who had completed professional placements participated in three pilot OSCEs. Content validity was evaluated following feedback from students and staff, and test results were analyzed to confirm reliability. Key competencies were grouped within the domains of communication, clinical and procedural skills, and technical proficiency. The results indicate that the OSCE is a valid and reliable method of assessment of the key professional competencies of exercise physiology students acquired through WIL. Asia-Pacific Journal of Cooperative Education, 2014, 15(2), 81-89

Keywords: Exercise physiology, key competencies, assessment, work-integrated learning

Exercise physiology is an emerging allied health profession that is undergoing continued and rapid growth in the design of clinical training curricula. In order for exercise physiology students to be accredited to practice, they are required to complete 500 hours of clinical practicum experience within a work-integrated learning (WIL) environment (Exercise and Sports Science Australia [ESSA], 2012). WIL is an umbrella term that refers to an experience where students learn through authentic engagement in a workplace. Students are provided with an opportunity to apply their knowledge and skills in a clinical practice setting (Council in Higher Education, 2011). Successful WIL activities are an essential component of competence-based degrees and contribute to the training of future professionals. While there is a growing abundance of literature on work-integrated learning, specifically, fieldwork, there is little on how work-integrated learning is assessed, the quality of the assessment process and ultimately, the outcomes for students’ (Ferns & Moore, 2012).

The expectation that university experience will include the application of theory in a practice-based setting and produce work-ready graduates has forced universities to rethink curriculum design and assessment practices (Cooper et al, 2010). The distinguishing features of work-based learning situations are that they are inherently variable, unpredictable, sometimes brief, high-risk learning events that are not replicable. This varying environment, combined with varying experience of supervisors to mentor the assessment, presents challenges for the quality assurance of the assessment process (Cooper, Orrell & Bowden; 2003, Hodges, 2011; Yorke, 2011). As a consequence, there is a need to simulate the learning environment as much as possible during assessment, in order to assess each student’s skills in an authentic manner. 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

1 Corresponding author: Fiona Naumann, f.naumann@unsw.edu.au
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).

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. Prior to developing an OSCE, it is necessary to identify the key competencies of the profession and develop OSCE stations that present a standardized portrayal of typical clinical scenarios. Once the stations were developed, establishing reliability of the assessment was critical. The underlying premise is that standardized assessment procedures ensure objectivity and maximize reliability (Bartfay et al. 2004; Major 2005). The aim of this study was twofold. Firstly to develop a series of OSCE stations that reflected the core competencies of a practicing Exercise Physiologist. The second aim was then to assess both the validity and reliability of an OSCE for assessing clinical skills of Exercise Physiology students.

METHODOLOGY

Prior to setting up the Exercise Physiology OSCE, key clinical academics and practicing accredited exercise physiologists (AEPs), developed a list of key competencies for exercise physiology clinical practice, based on the ESSA scope of practice document (ESSA, 2008) and professional experience. Once these key competencies were identified, a range of clinical presentations, typically expected in daily practice, were selected. These included the screening of healthy persons, assessment of patients with various chronic diseases, interpreting and explaining results to clients, prescribing and delivering exercise to manage their chronic condition, demonstrating competency with clinical reasoning and decision making, and providing health education and behavior change advice to support and enhance health and well-being. A decision was made to conduct the OSCE at the completion of the WIL, to ensure skill competency had been attained.

Once developed, the Exercise Physiology OSCE was piloted on three occasions in 2011. Students completing their final clinical practicum course were invited to participate, with all students providing consent. After each pilot, the test results were analyzed and examiners, clinical education staff and students were surveyed. The surveys with staff and students were conducted by an independent researcher from the University Learning and Teaching Unit. Examiners were asked to provide feedback from their station and suggest any changes that could improve the station for the next examination. Students were asked to complete a survey that captured their OSCE experience. Questions covered the suitability of the examination, whether it was a valuable learning experience, potential gaps in learning and the challenges students faced during the examination. Finally, an experienced clinical educator made observations of all stations and provided feedback for improvements. Test data and staff and student feedback were used to refine the examination for the next pilot. The trial examinations did not contribute to their course result.

For all examinations there was a single examiner in each station. For OSCE 1, examiners were academic staff only. Examiners for OSCE 2 and OSCE 3 included academic staff, practicing Accredited Exercise Physiologists (AEPs) and clinical supervisors. A briefing
session was held prior to the OSCE to review the assessment criteria. All stations involved either real patients or surrogate volunteers who were provided with detailed scripts and instructions on how to portray the information. After each pilot, the following elements of the design and implementation of the OSCE were reviewed as a means of assessing validity: 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.

The assessment criteria and grading system evolved as the series of pilots progressed. Initially checklists involving station-task specific criteria were used. Following OSCE 1, a global rating scheme was developed with criteria aligned to three competency domains: communication skills, clinical and procedural skills, and technical skills. The domains were developed from an extensive review of other allied health domains and the most appropriate domains selected. Each assessment criterion was graded as follows: F for failed performance (score 0), P for borderline performance (score 1) and P for good performance (score 2) and P+ for excellent performance (score 3). The numerical scale was used for statistical analyses of test results.

STATISTICAL ANALYSIS

We measured content validity using feedback from clinical education experts, examiners and academic staff. The reliability of the examination was analyzed using methods based on classical test theory including generalizability theory, which is an estimate of the examination’s reliability to consistently rank students. Estimates of internal consistency were calculated using Cronbach alpha based on results for all marking criteria and aggregate results for stations and competency domains. A reliable station should be able to assess various tasks across competency domains and distinguish between the best and worst performing students. Generalizability (G) studies were performed to estimate the variances related to different facets of the examination and to calculate generalizability coefficients. For both Cronbach alpha and generalizability coefficients, a value >0.7 was sought.

Using the above variance data, it was possible to calculate a generalizability coefficient (GC) which is an estimate of the examination’s reliability to consistently rank students. Using the GC, a study can be performed to show how varying the conditions of the examination may improve reliability. Decision (D) studies were performed to identify the effect of varying number of stations or competency domains on generalizability. This was done to determine the minimum number of stations and domains needed to produce a reliable exam result. Correlations of performance across stations and competency domains were estimated by Pearson product-moment coefficients. All statistical analyses were performed using SPSS v20 with significance set at $p<0.05$.

RESULTS

A total of 56 students completed the pilot examinations; eight in OSCE 1, 14 in OSCE 2 and 34 in OSCE 3. The design of the OSCE and the analyses of the test results are reported for each pilot followed by a comparison of data across all three pilots.

OSCE 1

OSCE 1 included eight stations with six minutes allocated to each station and a one-minute changeover between stations. Students were not provided with the task description prior to entering the station. Three of the stations were designed to assess knowledge rather
than skills, with students completing written tasks or being questioned by an examiner without any patient interaction. Station specific checklists were used and a single grade was awarded for the station. Some of the surrogate patients were members of academic staff. Feedback from the clinical educator was that some of the stations were too ambitious. It is essential to ensure that all skills can be demonstrated within the timeframe allocated. The examination should also assess the competence of the student’s performance and the length of the stations needs to be sufficient for this purpose. Issues identified by student feedback included students’ uncertainty about the examination format and insufficient time to complete expected tasks in some stations.

The overall Cronbach alpha (stations and domains) for the OSCE 1 was 0.52. This would suggest that the results from this examination would be unreliable for making a summative judgment of student performance. If a mark of 50% was considered a pass, the 95% confidence interval would be 42%-58%. Results from the stations which focused on assessing knowledge correlated poorly with the overall results and with the other stations. As students’ performances can be dependent on the content of the stations and the competency domains, the Cronbach alpha was also estimated by stations and domains. The inter-station alpha was 0.55 and the inter-domain alpha was 0.82.

OSCE 2

Based on the experience and results of OSCE 1, the design of OSCE 2 was modified. The number of stations was reduced from eight to seven with all stations assessing skills on real patients or volunteer surrogates (not staff). The time per station was increased from six to seven minutes, with the task requirements scaled down to ensure they could be completed within the time allocated. The changeover time also increased to two minutes to allow students to read “student instructions” prior to entering the station. Assessment criteria, based on station specific tasks, were aligned to three competency domains, with a grade awarded for each domain.

Feedback following this pilot was more positive with greater clarity of tasks reported by the students and adequate time to complete tasks, improved alignment of assessment criteria to tasks. The Cronbach alpha based on all assessment criteria increased to 0.82. The inter-station alpha was 0.74 and the inter-domain alpha was 0.86.

OSCE 3

Given the improved outcomes with OSCE 2, there were few refinements made for OSCE 3. The number of stations and domains were not changed. The alignment of assessment criteria to the competency domains was reviewed to improve this. Feedback from students and staff following OSCE 3 did not identify any major issues. The Cronbach alpha based on all assessment criteria was 0.86 (Table 1). The inter-station alpha was 0.77 and the inter-domain alpha was 0.83.

a. Comparison of Data Across the Three OSCE Pilots

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 and would suggest that the subsequent results would be reliable for making a summative judgment of student performance (Table 1).
TABLE 1. Cronbach alpha across the three pilot OSCEs, incorporating seven stations, on the three assessment domains of communication, exercise physiology procedural and technical skills.

<table>
<thead>
<tr>
<th></th>
<th>Cronbach’s alpha</th>
<th>Number 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>

b. Analysis by Stations

Discrimination index (measured by Pearson correlation) for each station is shown in Table 2. Significant discrimination indexes indicate that the ranking of students at these stations, correlate with the overall OSCE performance ranking. For OSCE 1, there were only two such stations. All stations were reviewed, knowledge only stations removed, task requirements tightened and examination criteria rewritten to improve the alignment of domains. For OSCE 2, the discrimination index was significant for four stations. For 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. Overall, this result indicated that OSCE 3 was an internally consistent measure of clinical competence, as defined by the competency at each station.

Decision (D) studies were performed to determine the effect of varying the number of stations and domains. There was a progressive increase in generalizability with increasing numbers of stations and domains. Increasing the number of domains to three resulted in a generalizability coefficient >0.7 with only seven stations. Further increasing the number of domains had little effect on generalizability. Similarly increasing the number of stations beyond seven had little effect on improving generalizability.
TABLE 2. The mean, SE and discrimination index for all pilot OSCE stations

<table>
<thead>
<tr>
<th>OSCE 1 Station</th>
<th>Mean</th>
<th>SE</th>
<th>R²</th>
<th>OSCE 2 Station</th>
<th>Mean</th>
<th>SE</th>
<th>R²</th>
<th>OSCE 3 Station</th>
<th>Mean</th>
<th>SE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 CV assess cancer</td>
<td>0.65</td>
<td>0.19</td>
<td>0.81*</td>
<td>CV assess cancer</td>
<td>0.66</td>
<td>0.04</td>
<td>0.96*</td>
<td>CV assess cancer</td>
<td>0.66</td>
<td>1.93</td>
<td>0.76*</td>
</tr>
<tr>
<td>2 VO₂ assess</td>
<td>0.50</td>
<td>0.60</td>
<td>0.62</td>
<td>Diabetes consult</td>
<td>0.71</td>
<td>0.06</td>
<td>0.12</td>
<td>Diabetes consult</td>
<td>0.66</td>
<td>1.44</td>
<td>0.58*</td>
</tr>
<tr>
<td>3 ECG Interpret</td>
<td>0.51</td>
<td>0.64</td>
<td>0.38</td>
<td>ECG set &amp; protocol</td>
<td>0.72</td>
<td>0.05</td>
<td>0.51</td>
<td>ECG set &amp; protocol</td>
<td>0.66</td>
<td>1.87</td>
<td>0.69*</td>
</tr>
<tr>
<td>4 ROM assess</td>
<td>0.54</td>
<td>0.53</td>
<td>0.60</td>
<td>Rest</td>
<td></td>
<td></td>
<td></td>
<td>Rest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Lung assess</td>
<td>0.60</td>
<td>0</td>
<td>NA</td>
<td>Lung assess</td>
<td>0.51</td>
<td>0.06</td>
<td>0.69*</td>
<td>Strength assess</td>
<td>0.68</td>
<td>1.64</td>
<td>0.65*</td>
</tr>
<tr>
<td>6 Falls assess</td>
<td>0.57</td>
<td>0.16</td>
<td>0.15</td>
<td>Falls assess</td>
<td>0.67</td>
<td>0.04</td>
<td>0.54</td>
<td>Falls assess</td>
<td>0.57</td>
<td>1.97</td>
<td>0.64*</td>
</tr>
<tr>
<td>7 Corporate health</td>
<td>0.60</td>
<td>0</td>
<td>NA</td>
<td>Corporate health</td>
<td>0.62</td>
<td>0.03</td>
<td>0.76*</td>
<td>Corporate health</td>
<td>0.64</td>
<td>1.41</td>
<td>0.60*</td>
</tr>
<tr>
<td>8 Mus-sk ex delivery</td>
<td>0.48</td>
<td>0.55</td>
<td>0.86*</td>
<td>Mus-sk ex delivery</td>
<td>0.70</td>
<td>0.06</td>
<td>0.82*</td>
<td>Mus-sk ex delivery</td>
<td>0.73</td>
<td>1.49</td>
<td>0.82*</td>
</tr>
</tbody>
</table>

* Pearson correlation with total OSCE scores. Significant at the 0.05 level (2-tailed)
DISCUSSION

The application of the OSCE at the completion of the WIL experience, proved a valid and reliable means to assess exercise physiology skill competency and readiness to accredit into the profession. As defined by Gonczi et al. (1990) a competent professional has the knowledge, skills, attitudes and general attributes necessary for job performance to the appropriate standard. This paper sought to design and assess the validity and reliability of a tool to assess this standard at the completion of 500 hours of WIL, clinical practice. While we chose to implement an OSCE at the completion of the WIL period, the OSCE could also have value mid-way through WIL. The feedback students receive on a mid WIL OSCE, could then provide valuable direction as to the skill competencies they have mastered and more importantly, the competencies that they need to practice at the subsequent clinical placements. Such a process embraces the concept of assessment for learning whereby the performance on the OSCE could provide information on what the student has achieved feedback used to shape further learning (Black & Wiliam, 1998).

Through an iterative process involving analyses of staff and student feedback and test data from three pilot examinations, an OSCE has been developed which is a reliable and valid means of assessing clinical skills in exercise physiology students. The content validity of the current examination was ensured by representing the diversity of clinical conditions that exercise physiologists encounter in clinical practice and the clinical competencies required to assess and manage patients (ESSA scope of practice, 2012). Within each station, students encountered common clinical scenarios and were required to demonstrate specific skills relevant to the content of the station. Clearly the short duration of each station would not allow an assessment of all relevant skills, but each station was designed to assess an aspect of the principal competencies. It was evident that some stations were initially too complex and were refined to ensure students could complete tasks within the allocated time frame. In the first pilot, the inclusion of stations which were primarily designed to assess knowledge significantly affected the reliability of the examination. As the OSCE is intended to assess clinical competencies, it is important that each station is designed with this purpose in mind. A review of the curriculum also ensured that all skill competencies that were assessed had been delivered throughout the curriculum and practiced during work-integrated learning.

Improvements were made to the reliability of the OSCE, exceeding the accepted Cronbach alpha benchmark of 0.7 and consistent with the reported reliabilities in the medical and other allied health fields (Sloan et al., 1995; Bland et al., 1997). This would suggest that the results from this examination would be reliable for making a summative judgment of student performance. Improvements in exam reliability were achieved through better training of volunteer patients and surrogates, whereby patients were provided the scripts ahead of the exam and time provided to ask additional questions regarding their case history. Improvements were also brought about by 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. The surrogates were better matched to the requirements of the stations and all volunteers were given sufficient time to learn their scripts and to discuss the script and expectations of their role with the examiner. All volunteers were instructed in how to respond to students’ questions and instructions in a standard way to ensure consistency in their behavior.

Initial feedback from the students indicated that they required more information about the format and structure of the examination. Subsequently students were provided with written
guidelines and attended a pre-examination information session. Written instructions were available to students before they entered the station which clearly outlined the requirements of the station and ensured that all students received the same instructions. Although the initial grading system, using detailed checklists, documented what examiners should be assessing, it was not clear how a station score was to be determined. The grading system was subsequently revised to align the assessment criteria based on the station’s task with three competency domains: communication skills, clinical and procedural skills and technical skills.

As the OSCE is labour and time intensive, it is important to determine the optimal number of stations and domains required to provide a reliable assessment of student performance. Decision (d) studies, based on generalizability theory, can estimate the effect of varying conditions on the generalizability of an examination. Using data from OSCE 2 and OSCE 3, it was shown that the combination of seven stations and three domains achieved satisfactory generalizability and that increasing the number of stations or domains would not improve the reliability of the examination significantly.

Beyond its use as a form of assessment, there is evidence that the OSCE can enhance the quality of health professional teaching and learning (Mitchell et al., 2009). It clearly provides a motivation for learning (Bartfray et al., 2004) and can provide feedback to students on their performance to identify where corrective training is required (Nicol and Freeth, 1998). In addition to assessing student skill competence, Alinier et al. (2006) found the OSCE a useful method for evaluating the curriculum, quality of teaching and profiling strengths and weaknesses of the student cohort, thus promoting course review and continuous quality improvement.

CONCLUSION

The evidence suggests that with careful attention to elements of design and implementation of an OSCE, followed by appropriate analyses of test results, the OSCE is a valid and reliable form of assessment of clinical competence in student exercise physiologists, at the completion of WIL suitable domains for assessment in exercise physiology incorporate communication skill competency, clinical and procedural skill competency and technical skill competency. To ensure the content validity of the examination, stations must represent the diversity of clinical situations that exercise physiologists encounter in clinical practice. Careful attention needs to be put into station design, student instructions and examiner assessment criteria to ensure reliability of each station. The OSCE also enables a careful review of the curriculum that ensures all aspects of clinical competency requirements are delivered throughout the curriculum and consolidated within the WIL environment. The OSCE may also identify global areas where students didn’t perform well and ensure a greater teaching focus into these competency areas. Personalized student feedback may also guide a student to aspects of practice that they need to work on within the WIL environment.

REFERENCES


About the Journal

The Asia-Pacific Journal of Cooperative Education publishes peer-reviewed original research, topical issues, and best practice articles from throughout the world dealing with Cooperative Education (Co-op) and Work Integrated Learning/Education (WIL).

In this Journal, Co-op/WIL is defined as an educational approach that uses relevant work-based projects that form an integrated and assessed part of an academic program of study (e.g., work placements, internships, practicum). These programs should have clear linkages with, or add to, the knowledge and skill base of the academic program. These programs can be described by a variety of names, such as work-based learning, workplace learning, professional training, industry-based learning, engaged industry learning, career and technical education, internships, experiential education, experiential learning, vocational education and training, fieldwork education, and service learning.

The Journal’s main aim is to allow specialists working in these areas to disseminate their findings and share their knowledge for the benefit of institutions, co-op/WIL practitioners, and researchers. The Journal desires to encourage quality research and explorative critical discussion that will lead to the advancement of effective practices, development of further understanding of co-op/WIL, and promote further research.

Submitting Manuscripts

Before submitting a manuscript, please ensure that the ‘instructions for authors’ has been followed (www.apjce.org/instructions-for-authors). All manuscripts are to be submitted for blind review directly to the Editor-in-Chief (editor@apjce.org) by way of email attachment. All submissions of manuscripts must be in MS Word format, with manuscript word counts between 3,000 and 5,000 words (excluding references).

All manuscripts, if deemed relevant to the Journal’s audience, will be double blind reviewed by two reviewers or more. Manuscripts submitted to the Journal with authors names included with have the authors’ names removed by the Editor-in-Chief before being reviewed to ensure anonymity.

Typically, authors receive the reviewers’ comments about a month after the submission of the manuscript. The Journal uses a constructive process for review and preparation of the manuscript, and encourages its reviewers to give supportive and extensive feedback on the requirements for improving the manuscript as well as guidance on how to make the amendments.

If the manuscript is deemed acceptable for publication, and reviewers’ comments have been satisfactorily addressed, the manuscript is prepared for publication by the Copy Editor. The Copy Editor may correspond with the authors to check details, if required. Final publication is by discretion of the Editor-in-Chief. Final published form of the manuscript is via the Journal website (www.apjce.org), authors will be notified and sent a PDF copy of the final manuscript. There is no charge for publishing in APJCE and the Journal allows free open access for its readers.

Types of Manuscripts Sought by the Journal

Types of manuscripts the Journal accepts are primarily of two forms: research reports describing research into aspects of Cooperative Education and Work Integrated Learning/Education, and topical discussion articles that review relevant literature and give critical explorative discussion around a topical issue.

The Journal does also accept best practice papers but only if it present a unique or innovative practice of a Co-op/WIL program that is likely to be of interest to the broader Co-op/WIL community. The Journal also accepts a limited number of Book Reviews of relevant and recently published books.

Research reports should contain; an introduction that describes relevant literature and sets the context of the inquiry, a description and justification for the methodology employed, a description of the research findings-tabulated as appropriate, a discussion of the importance of the findings including their significance for practitioners, and a conclusion preferably incorporating suggestions for further research.

Topical discussion articles should contain a clear statement of the topic or issue under discussion, reference to relevant literature, critical discussion of the importance of the issues, and implications for other researchers and practitioners.