Nurse Education Today xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Nurse Education Today

journal homepage: www.elsevier.com/nedt

Back to the future: An online OSCE Management Information System for nursing OSCEs

Pauline Meskell [a], Eimear Burke [a], Thomas J.B. Kropmans [b,*], Evelyn Byrne [a], Winny Setyonugroho [b], Kieran M. Kennedy [c]

[a] School of Nursing and Midwifery, College of Medicine, Nursing and Health Sciences, NUI Galway, Galway, Ireland
[b] School of Medicine, Medical Informatics & Medical Education, College of Medicine, Nursing and Health Sciences, NUI Galway, Galway, Ireland
[c] School of Medicine, College of Medicine, Nursing and Health Sciences, NUI Galway, Galway, Ireland

ARTICLE INFO

Article history:
Accepted 17 June 2015
Available online xxxx

Keywords:
OSCE
Nursing
OSCE management information system
OMIS
e-OSCE
Assessment
Generalizability theory

SUMMARY

Background: The Objective Structured Clinical Examination (OSCE) is an established tool in the repertoire of clinical assessment methods in nurse education. The use of OSCEs facilitates the assessment of psychomotor skills as well as knowledge and attitudes. Identified benefits of OSCE assessment include development of students' confidence in their clinical skills and preparation for clinical practice. However, a number of challenges exist with the traditional paper methodology, including documentation errors and inadequate student feedback.
Objectives: To explore electronic OSCE delivery and evaluate the benefits of using an electronic OSCE management system. To explore assessors' perceptions of and attitudes to the computer based package.
Design: This study was conducted using electronic software in the management of a four station OSCE assessment with a cohort of first year undergraduate nursing students delivered over two consecutive years (n = 203) in one higher education institution in Ireland. A quantitative descriptive survey methodology was used to obtain the views of the assessors on the process and outcome of using the software.
Methods: OSCE documentation was converted to electronic format. Assessors were trained in the use of the OSCE management software package and laptops were procured to facilitate electronic management of the OSCE assessment. Following the OSCE assessment, assessors were invited to evaluate the experience.
Results: Electronic software facilitated the storage and analysis of overall group and individual results thereby offering considerable time savings. Submission of electronic forms was allowed only when fully completed thus removing the potential for missing data. The feedback facility allowed the student to receive timely evaluation on their performance and to benchmark their performance against the class.
Conclusions: Assessors' satisfaction with the software was high. Analysis of assessment results can highlight issues around internal consistency being moderate and examiners variability. Regression analysis increases fairness of result calculations.

© 2015 Published by Elsevier Ltd.

Introduction

The Objective Structured Clinical Skills Examination (OSCE) is a well-established method of assessing skills and clinical competence among health practitioners including nurses (Oranye et al, 2012). The OSCE originated in the UK as an objective means to assess medical students' skills (Harden et al., 1975). The examination involves students progressing through a series of stations where they are assessed by an examiner with pre-determined marking criteria (Pugh et al., 2014).

Several authors have highlighted the importance of using OSCEs as an assessment method in nursing education (McWilliam and Botwinski, 2010; Baid, 2011; Oranye et al., 2012; Barry et al., 2012). The OSCE facilitates the assessment of students' competency with clinical skills in a controlled simulated environment instead of in the practice setting (Baid, 2011). According to McWilliam and Botwinski (2010), students recognize the value of the OSCE experience to their education.

A number of benefits have been attributed to the use of OSCEs including, the development of students confidence (Alinier, 2003), the preparation of students for clinical practice and the achievement of deeper more meaningful learning (Barry, et al., 2012). Importantly, the use of OSCEs facilitates the assessment of psychomotor skills as well as knowledge and attitudes (Baid, 2011). OSCEs have the potential to provide students with feedback on their clinical performance and facilitate the identification of strengths and weaknesses (McWilliam and Botwinski, 2010). The OSCE has been reviewed positively as an

* Corresponding author at: College of Medicine, Nursing & Health Sciences, School of Medicine/Medical Informatics & Education, National University of Ireland Galway, Clinical Science Institute/Comerford Building, Room 204, Newcastle road, Galway, Ireland. Tel.: +353 91495478; fax: +353 91495512.
E-mail address: thomas.kropmans@nuigalway.ie (T.J.B. Kropmans).

http://dx.doi.org/10.1016/j.nedt.2015.06.010
0260-6917/© 2015 Published by Elsevier Ltd.

Please cite this article as: Meskell, P., et al., Back to the future: An online OSCE Management Information System for nursing OSCEs, Nurse Educ. Today (2015), http://dx.doi.org/10.1016/j.nedt.2015.06.010

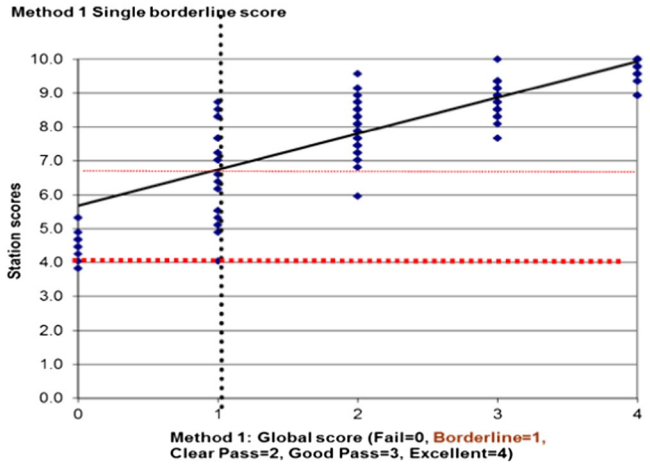


Fig. 1. Single borderline score regression analysis highlighting the difference between the static cut score (horizontal dotted line at 4.0 on the Y axis) and the dynamic cut score (solid line at 6.8 on the Y axis) (adjusted from John Patterson, honorary senior lecturer at the Centre for Medical Education of the Barts and London School of Medicine and Dentistry and Assessment Consultant).

assessment method for clinical competence and for responding to student diversity in education (Smith et al., 2012). However, there are a number of notable disadvantages associated with OSCEs. In particular, some students find them stressful and they are resource intensive in terms of staff, equipment and clinical skills laboratories (Baid, 2011). However, Alinier (2003) suggests that the educational benefits surpass the issues associated with resources (Alinier, 2003).

Traditionally OSCEs have been carried out using paper based methodology. However, a number of issues have been highlighted with this method including illegible handwriting, missing details (students' names and student numbers) and lost assessment sheets (Treadwell, 2006). Furthermore, it is known that manual calculation of results and entering them into a database are time-consuming and are subject to human errors and feedback is rarely provided to students on their performance after paper based assessments (Treadwell, 2006). Despite these issues there is a scarcity of literature regarding the use of computer or OSCE software and the assessment of OSCEs. Segall et al (2005) compared the usability of paper and pencil method and Personal Digital Assistant (PDA) based quizzes and found the PDA based quiz was more efficient and superior to the traditional based method (Segall et al., 2005). Similarly, Treadwell (2006) compared the conduction of a paper based OSCEs with an electronic method. The findings indicated that the electronic method was just as effective and more efficient (less time consuming) than the traditional paper based method. In addition, the electronic system was highly rated by the assessors, who found it less invasive and reported that they had more time to observe the students and permitted greater observation of the students when using the paper assessment. Schmitz et al. (2011) highlight a number of advantages to use an electronic handheld device to assess OSCEs including, speed of data gathering, simplicity of data evaluation and fast automatic feedback (Schmitz et al., 2011). Segall et al. (2005) support computer based assessment suggesting that grading is more accurate, feedback is immediate, security is enhanced and less time is spent by instructors on grading and data entry (Segall, et al., 2005). In the College of Medicine, Nursing and Health Sciences of the National University of Ireland in Galway (NUI Galway), an online assessment tool was developed and implemented from 2009 onwards. The OSCE Management Information System (OMIS) contains a station and item bank, an OSCE planning and student management tool plus a fully fledged results analysis tool (Kropmans, et al., 2012). We acknowledge, however, that other elements of OSCE planning, including use of simulators, props, actors/actresses and real patients, are not currently managed by OMIS. OMIS was originally developed with the primary objective of improving the quality assurance of our OSCE grading. The aim of this study was to explore the acceptability of an online OSCE Management Information System for the NUIG School of Nursing OSCEs, by means of a comparative cohort study.

Methods

We compared Objective Structured Clinical Examinations (OSCEs), delivered over two consecutive years of the first year nursing skills module using an online OSCE Management Information System (OMIS) (Cunningham et al., 2008). Each OSCE comprised of 4 individual stations. Both student cohorts (i.e. those from the 2012–2013 and 2013–2014 academic years) completed a hand washing, blood pressure measurement, manual handling and a documentation station, each of which was of 5 minute duration. Implementation of the OMIS software involved preparing both the system and the assessors. Assessment documentation was reviewed, reformulated, agreed by the module coordinators and uploaded to the OMIS system. Module coordinators and examiners were trained to use the electronic system and technical support was available at the time of the OSCE assessments. Student and assessor details were imported into the system and schedules finalized. In the first year of using OMIS in the School of Nursing & Midwifery, a User Acceptance Test (survey) was completed to explore the extent to which examiners accepted this online OSCE management solution. The questionnaire was developed following a comprehensive literature review around electronic OSCE management, expert review (Olsen, 2010) and consensus of the educationalists involved in OSCE planning. The 25 item questionnaire was divided into three sections evaluating 1. the OSCE Software user evaluation (3 Items); 2. usage of the electronic OSCE package (10 Items); and 3. the OSCE assessment process itself (12 Items). Last but not least examiners provided an overall impression of their experience with OMIS. The possible range of scores for the 3 sections ranged from 0 to 105, and the overall impression could be scored from 0 to 100%.

The total number of first year students that completed the OSCE was 203. The 2012–2013 cohort comprised 101 students, whereas the 2013–2014 cohort comprised 102 students. The station checklists for both OSCEs were identical. The novel online OSCE Management Information System, which was developed "in-house" at the National University of Ireland Galway, was used to administer both examinations (Cunningham et al., 2008; Kropmans et al., 2012). OMIS retrieves, stores and analyses assessment data electronically. Student feedback can be sent to students electronically using the Student Feedback Email System. We used item checklists to assess student competency with each task. The number of items per assessment form varied from 5, for the Documentation station, to 15, for the Blood Pressure station, with a maximum score of 30 marks for all three clinical stations and 10

Raw Data | **Summary** | Detailed | BL Groups | Graphs | Export | All

	Nr. of Results	Stations Passed	Stations Failed	Mean Result	Mean Result (%)	SD	SD (%)	Min	Max	Range	Mid Range	Station Items Cronbach Alpha (α)
Evaluation Questionnaire	18	18	0	89.9 / 105	85.6%	9.8	9.3%	75.0	101.0	26.0	88.0	0.139
Exam Totals & Mean %	18	18	0		85.6%							0.139
Students Passed | Failed		18 (100%)	0 (0.0%)									

Fig. 2. The outcome of the User Acceptance Questionnaire (n = 18) with an overall mean (min, max) score of 89.9 (75, 101) out of a total score of 105. The OSCE Management Information System was used to complete the User Acceptance Questionnaire (2012–2013).

Please cite this article as: Meskell, P., et al., Back to the future: An online OSCE Management Information System for nursing OSCEs, Nurse Educ. Today (2015), http://dx.doi.org/10.1016/j.nedt.2015.06.010

ARTICLE IN PRESS

Table 1
User Acceptance Test results.

User Acceptance Test (Section A) Using the electronic OSCE package (OMIS) n = 18	Strongly agree	Agree	Disagree or strongly disagree	No judgment
1.The training sessions prepared me sufficiently to use the package	4	12	1	1
2.It was easy to access the computer package	5	13	0	0
3.The screen layout and instructions were clear	6	12	0	0
4.The computer package was easy to use	5	13	0	0
5.I felt confident using the computer package	6	11	0	1
6.Drop down menus were easy to navigate	6	12	0	0
7.There was good technical support provided to help me with any problems	2	16	0	0
8.Using the package facilitated me to complete my work as an examiner efficiently	4	13	0	1
9.It was easy to exit from the programme	6	12	0	0
10.I would like to use the electronic system for future OSCES	4	14	0	0
The OSCE assessment process (Section B) Using the electronic OSCE package (OMIS)	**Strongly agree**	**Agree**	**Disagree or strongly disagree**	**No judgment**
1.Using the computer package facilitated me in assessing the skill	6	10	1	1
2.The outlined steps of the skill were appropriate to the skill being assessed	5	11	0	2
3.Having to use a laptop during the assessment was easier than using a paper form	5	10	1	2
4.Absent students were easy to manage	3	10	5	0
5.Late arrival students were easy to manage	2	5	10	1
6. The rating options were appropriate	7	9	0	2
7. Being able to input comments on student's performance was useful	5	9	1	2
8. I felt that I had to include a comment on students' performance for each student	3	2	2	10
9. I included a comment when I judged that a student had not completed an element of the skill appropriately	6	6	3	3
10. Not being able to see the students overall score was good	7	7	2	2
11. Being able to give a global rating for each student was good	7	10	0	1
12.Not being able to see score for each criterion was good	5	7	3	3
Overall level of satisfaction with using the computer package (Section C)				
Please move the slider on the Visual Analogue Scale to indicate your overall level of satisfaction with using the computer package to assess the OSCEs as you have experienced it.	88.40% (min = 74; max = 100)			

marks for the documentation station. The overall professional impression of the examiners was rated on a 5 item Likert Global Rating Scale (GRS) which included the following options: Fail (0), Borderline (1), Pass (2), Good (3) and Excellent (4). The numerical values of the GRS options were not incorporated in the final student scores, but were instead used for standard setting using an online Borderline Regression Analysis function that is built in to the OMIS (1). The static predetermined cut-off score for nursing studies is 40% (<40% means a fail score; >40% means a pass score).

Statistical Analysis

OMIS produces an online analysis of items and overall total (raw) scores and adjusted (raw) scores using standard setting of student performance after regression analysis. The mean result, standard deviation (SD), minimum and maximum and range and mid-range are produced instantly, in real time, during the examination. Internal consistency (Cronbach's alpha) is used to provide insight into the consistency of items in each station predicting the overall score of the student. Borderline Regression Analysis (Borderline Group Average versus Borderline Regression Method) calculates a 'flexible cut-off score' complementary to the general static cut score of 40% for each individual station. The overall average regression cut-score is used to adjust the average overall raw score of the students. Borderline Group Average, which is based upon calculation of the average mark of those students that were globally rated by their examiners as 'borderline', is the most simplistic method to use. A complete Borderline Regression Analysis, which is performed over all item marks matched with all of the global ratings (from fail to excellent), can also be used. The flexible cut-off score is calculated using the BRM Cut score (Intercept + 1 × Slope)—since borderline = 1 using FORECAST method (Fig. 1). All analysis reports and data were exported to Excel to facilitate further detailed analysis.

Dynamic Cut-off Score

Data were exported to perform a Generalizability Coefficient analysis using a G- and D-study with EduG software. The G-study generates information about whether the outcome can be generalized to other nursing OSCEs. The D-study provides information on how the generalizability of results can be improved.

Results

User Acceptance Test for Examiners

The internal consistency (Cronbach's alpha) of the questionnaire was 0.14. The overall mean score (min, max) was 89.9 (75, 101) out of a total of 105 for all three sections of the questionnaire (Fig. 2).

Setup | Raw Data | **Summary** | Detailed | Graphs | Export | All

	Nr. of Results	Stations Passed	Stations Failed	Mean Result	Mean Result (%)	SD	SD (%)	Min	Max	Range	Mid Range	Station Items Cronbach Alpha (α)
1 - Handwashing	102	102	0	27.1 / 30	90.3%	3.3	10.8%	15.5	30.0	14.5	22.8	0.663
2 - Blood Pressure	102	101	1	24.6 / 30	82.1%	4.5	15.1%	11.0	30.0	19.0	20.5	0.610
3 - Manual Handling	101	99	2	23.9 / 30	79.5%	4.8	15.9%	6.5	30.0	23.5	18.3	0.721
4 - Documentation	102	95	7	8.8 / 10	88.1%	1.6	16.2%	4.5	10.0	5.5	7.3	0.667
Exam Totals & Mean %	**407**	**397**	**10**		**84.4%**							**0.665**
Total Students Passed | Failed		**102 (100%)**	**0 (0.0%)**									

Fig. 3. Screenshot OMIS: OSCE Results summary table 2012–13 (OMIS 1.8.5).

Please cite this article as: Meskell, P., et al., Back to the future: An online OSCE Management Information System for nursing OSCEs, Nurse Educ. Today (2015), http://dx.doi.org/10.1016/j.nedt.2015.06.010

Setup | Raw Data | **Summary** | Detailed | Graphs | Export | All

	Nr. of Results	Stations Passed	Stations Failed	Mean Result	Mean Result (%)	SD	SD (%)	Min	Max	Range	Mid Range	Station Items Cronbach Alpha (α)
1 - Handwashing	101	99	2	26.1 / 30	87.0%	5.1	17.1%	0.0	30.0	30.0	15.0	0.787
2 - Blood Pressure	101	101	0	26.9 / 30	89.7%	3.3	10.9%	16.0	30.0	14.0	23.0	0.463
3 - Manual Handling	101	101	0	25.4 / 30	84.5%	4.6	15.3%	13.0	30.0	17.0	21.5	0.817
4 - Documentation	101	97	4	9.1 / 10	90.6%	1.6	16.4%	2.5	10.0	7.5	6.3	0.708
Exam Totals & Mean %	404	398	6		87.4%							0.694
Total Students Passed | Failed		101 (100%)	0 (0.0%)									

Fig. 4. Screenshot OMIS: OSCE Results summary table 2013–14 (OMIS 1.8.5).

Almost all examiners (17 out of 18) were experienced in assessing students in low technology (paper) based OSCEs. Two examiners used the online OSCE software before and were familiar with OMIS. Most examiners (16 out of 18) were academic lecturers in Nursing, whilst the remaining 2 were practicing Nurses/Midwives. Results of the User Acceptance Test are displayed in Table 1.

The majority of assessors (92%) indicated that using the computer package facilitated them in assessing the skill. Over 76% of assessors found using the laptop was easier than using a paper form; the remainder either were unable to make a judgment (12%) or found the paper form easier (12%).

Summary of Results

The summary of results for the 2012–2013 cohort (n = 102), as produced by the OMIS, demonstrated an overall internal consistency of 0.66, where the Cronbach's alpha per individual station varied between 0.61 for the blood pressure station and 0.72 for the manual handling station. The overall Cronbach's alpha of the examination was 0.67. In classical psychometric terms internal consistency was moderate. The overall average student performance for the clinical stations was 84.4%, with a minimum score of 21% (6.5 out of 30) and a maximum of 100% (30 out of 30). The overall average student performance for the documentation station was 88.1%, with a minimum score of 45% (4.5 out 10,) and a maximum score of 100% (10 out of 10) (Fig. 3).

The summary results for the 2013–2014 cohort (n = 101), as produced by OMIS, demonstrated an overall internal consistency of 0.69, whereby the Cronbach's alpha per station varied between 0.46 for the blood pressure station and 0.81 for the manual handling station. The overall Cronbach's alpha of the examination was 0.69. In classical psychometrics terms internal consistency was moderate. The overall average performance of the students was 87.4% for the clinical stations, with a minimum score of 0% (0 out of 30) and a maximum score of 100% (30 out of 30). The overall average performance of students was 90.6% for the documentation station, with a minimum score of 18% (2.5 out of 10) and a maximum score 100% (10 out of 10). The overall internal consistency was 0.67 and the Cronbach's alpha per station ranged from 0.61 for the blood pressure station and 0.72 for the manual handling station. In classical psychometric terms internal consistency was moderate. The overall average performance of the students was 84.4% with a minimum score of 21% (6.5 out of 30,) and a maximum of 100% (30 out of 30) for the clinical stations. The overall average performance of students was 88.1% for the documentation station, with a minimum score of 45% (4.5 out 10) and a maximum score 100% (10 out of 10) (Fig. 4).

Borderline Regression Analysis

Borderline Group Analysis is a simple way of calculating the average cut-off score of those students that were addressed as 'borderline performers' (i.e. examiners not being sure whether the student performance should be marked as fail or pass). Where there are a small number of students in this category, then Borderline Group Average estimates may be very unreliable as shown in the 2012–2013 cohort (Fig. 5), where only 5 student performances were marked as 'borderline' (i.e. 4 borderline scores for the blood pressure station and 1 for the documentation station). The 'average' score of these students was 67.5% (4 students) for the 'Blood Pressure' station and 55% (1 student) in the 'Documentation' station. Using Borderline Group Average in this cohort would not provide similar information for those stations where *no* students were marked as borderline performers. A fully fledged Borderline Regression Analysis would however provide this information due to the inclusion of all Global Ratings from Fail to Excellent (Fig. 1).

A similar situation arises with the Borderline Group Analysis of the 2013–2014 cohort (Fig. 6), within which 9 student performances were regarded as borderline (i.e. 3 for the hand washing station, 2 for the blood pressure station, 1 for the manual handling station and 3 for the documentation station). Due to the small numbers of students, cut-off scores of 58.9%, 69.2%, 43.3% and 41.7% may be unreliable.

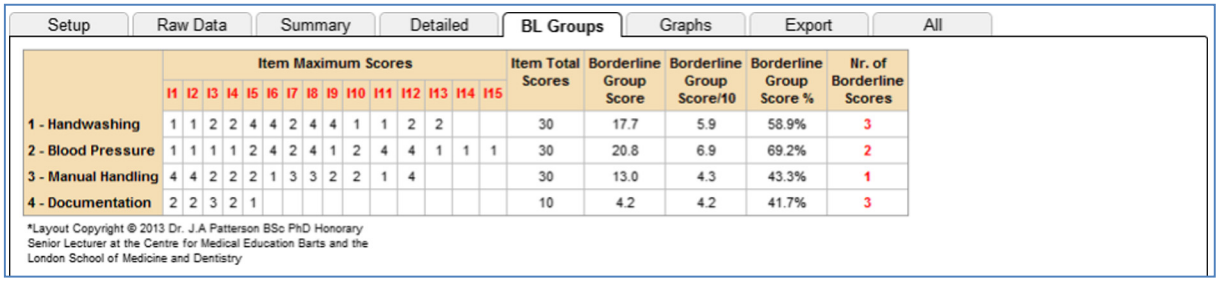
A fully fledged Borderline Regression Analysis is embedded into the OSCE Management Information System software whereby the forecast method is used to calculate new cut-off scores for each station taking into account the 'difficulty of the station' and the 'hawk and dove effect' of different examiners involved in the OSCE. With Borderline Regression Method 1, cut-off scores are calculated for all stations based upon analysing item scores and Global Rating Scores of all students varying

Setup | Raw Data | Summary | Detailed | **BL Groups** | Graphs | Export | All

	Item Maximum Scores															Item Total Scores	Borderline Group Score	Borderline Group Score/10	Borderline Group Score %	Nr. of Borderline Scores
	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15					
1 - Handwashing	1	1	2	2	4	4	2	4	4	1	1	2	2			30	N/A	N/A	N/A	0
2 - Blood Pressure	1	1	1	1	2	4	2	4	1	2	4	4	1	1	1	30	20.3	6.8	67.5%	4
3 - Manual Handling	4	4	2	2	2	1	3	3	2	2	1	4				30	N/A	N/A	N/A	0
4 - Documentation	2	2	3	2	1											10	5.5	5.5	55.0%	1

*Layout Copyright © 2013 Dr. J.A Patterson BSc PhD Honorary Senior Lecturer at the Centre for Medical Education Barts and the London School of Medicine and Dentistry

Fig. 5. Screenshot OMIS: Borderline Group Average analysis in OMIS for cohort 2012–13.



Setup | Raw Data | Summary | Detailed | BL Groups | Graphs | Export | All

	Item Maximum Scores															Item Total Scores	Borderline Group Score	Borderline Group Score/10	Borderline Group Score %	Nr. of Borderline Scores
	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15					
1 - Handwashing	1	1	2	2	4	4	2	4	4	1	1	2	2			30	17.7	5.9	58.9%	3
2 - Blood Pressure	1	1	1	1	2	4	2	4	1	2	4	4	1	1	1	30	20.8	6.9	69.2%	2
3 - Manual Handling	4	4	2	2	2	1	3	3	2	2	1	4				30	13.0	4.3	43.3%	1
4 - Documentation	2	2	3	2	1											10	4.2	4.2	41.7%	3

*Layout Copyright © 2013 Dr. J.A Patterson BSc PhD Honorary Senior Lecturer at the Centre for Medical Education Barts and the London School of Medicine and Dentistry

Fig. 6. Screenshot OMIS: Borderline Group Average analysis in OMIS for cohort 2013–14.

respectively from 48% in station 4 (was 55% in case of a 'group' average) to 66% for station 1 (no group average available).

Generalizability Theory Analysis

Exported data were analysed using Edu-G software (Swiss Society for Research in Education Working Group. Edumetrics—Quality of measurement in education). This analysis demonstrated poor generalizability of the results achieved in this 4 station OSCE. G-coefficient for the 2012–13 cohort was 0.29 and for the 2013–2014 cohort, it was 0.41.

Total error variation was high, respectively 91.9 and 83.7% for Observer (O), Station Observer and Circuit (SO:C) and Circuit and Observer (CO). Most error variance was observed between examiners of which 68 respectively 69% is due to interactions between Students, Observers and different Circuits. The outcome in terms of generalizability results could be improved by increasing the number of stations to between 5 and 10. In this way, the G coefficient, a more modern measure of reliability as compared to classic psychometric measures (i.e. correlation coefficients or Intra Class Correlation coefficients), would improve to 0.51.

Discussion

The overall outcome of the User Acceptance Test was good, with more than 80% of the examiners having agreed that functionalities did make a lot of sense and that they accepted this online OSCE solution. Its use was evaluated as very satisfactory. Items like managing the OSCE Assessment process i.e. '*5. Late arrival students were easy to manage*'; or *8.* '*I felt that I had to include a comment on student performance for each student*' or 9. '*I included a comment when I judged that a student had not completed an element of the skill appropriately*' were difficult to interpret according to the participating examiners, and therefore varied in user acceptance. Possible explanations for this include that no students arrived late in the study OSCEs and, secondly, no direct instructions were provided to examiners in relation to the inclusion of feedback, whereby the latter was left to individual interpretation. Optional 'written feedback' is a functionality that can be set as obligatory or not. The available 'student feedback email feature' was used for only the second cohort of students. Using 'open written feedback' in an OSCE Management Information System, examiners need to be trained in how to provide adequate, safe and achievable feedback (Black and Harden, 1986). The summary of results section provided instant information about the progress of each individual student in this OSCE. Although the results are quite high in both cohorts 10 students in the first and 6 in second cohort failed in one or two of the consecutive stations using a 'static pass mark' set prior to the start of the exam at 40%. Due to the availability of a Global Rating Score facility, and an appropriate number of students ($n \geq 100$), we could perform a Borderline Group Average analysis. The latter is based upon the overall professional impression of the examiner evaluating a student's performance. The examiner will mark this overall performance as a pass, borderline, fail, good, or excellent performance (Borderline Regression Method 1). In the borderline group feature (Figs. 3 and 4), the average performance of these 'borderline performing students' is substantially above the static pass mark of 40% in stations 1 and 4 of the first cohort of students and in stations 1 and 2 of the second cohort of students. Where N/A is indicated this means that no students were marked as Borderline performing students in the first cohort. 'Borderline performance' is an indicator of examiner uncertainty with regard to whether or not a student should pass or fail. Whether the regression outcome is high or low is an indicator of whether a station is 'easy' or 'difficult' to pass respectively. Where the 'Angoff Method' is a standard setting method used prior to an examination, Borderline Regression Analysis is a standard setting method used *after* the examination has taken place and is based upon the professional impression of the examiner evaluating students' performances according to a Global Rating Scale (Kaufman et al., 2000). In addition to the Borderline Group Average, OMIS provides a fully fledged borderline regression analysis that takes all scores into account and matches those with the professional impression of the examiners using a regression analysis (Fig. 1). We used the simple forecast method, in an Excel template, using all item total marks and the Global Rating Scale in the regression equation (Fig. 7) (Kaufman, et al., 2000). All station dynamic cut-off scores were above the 40% cut-off score indicating that stations were too easy to pass according to the professional impression of the examiners.

The OSCE design used for this nursing skills examination demonstrated poor generalizability of results in this 4 station OSCE. The generalizability would improve by introducing more stations (e.g. an OSCE with 5–10 stations). In this way, the G coefficient would improve from 0.34 to 0.51. However, these coefficients do not achieve the standards suggested in other research literature on the subject e.g. OSCEs (Newble and Swanson, 1988; Verhoeven et al., 2000; Wass et al., 2001a; Wass et al., 2001b; Guiton et al., 2004). The generalizability of results is only appropriate in OSCEs with a minimum of 15–18 stations (Iramaneerat and Yudkowsky, 2007; Iramaneerat et al., 2008; Iramaneerat et al., 2009).

Error variance is the 'error' that occurs around the 'true score' (i.e. the 'observed score' is the 'true score' plus 'error'). The Generalizability Theory is introduced to provide more insight into the error around the

Station		Item Maximum Scores																																				Total	Station	Station	Borderline	BRA	BR	BR
No:	Title	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20	I21	I22	I23	I24	I25	I26	I27	I28	I29	I30												
1	*Handwashing*	1	1	2	2	4	4	2	4	4	1	1	2	2																		30	9.0	1.1	0		*6.6*	7.2
2	*Blood Pressure*	1	1	1	1	2	4	2	4	1	2	4	4	1	1	1																30	8.1	1.6	5	6.1	*6.1*	6.7
3	*Manual Handling*	4	4	2	2	2	1	3	3	2	2	1	4																			30	8.0	1.6	0		*5.4*	6.1
4	*Documentation*	2	2	3	2	1																										10	8.8	1.6	1	5.5	*4.8*	5.7

Fig. 7. Screenshot OMIS: Individual weighting of checklist items for each station with total scores, station mean (SD), number of borderline scores, borderline group average and borderline regression analyses methods 1 and 2.

Please cite this article as: Meskell, P., et al., Back to the future: An online OSCE Management Information System for nursing OSCEs, Nurse Educ. Today (2015), http://dx.doi.org/10.1016/j.nedt.2015.06.010

ARTICLE IN PRESS

'true score' and to further improve the assessment procedure. In our design it appeared that most error around the 'true score' was caused by the examiners (Mitchell, 1979). Student interaction with examiners during different circuits caused approximately 80–90% of error in this assessment design. The latter might be due to the fact that examiners were not 'blinded' (i.e. they knew their students, and students might perform 'skills' in a different manner if they are being watched by their examiners). Interaction effects might be 'confounded' by the personal impression of the examiner(s) rather than the 'pure' evaluation of the skill(s). Furthermore, the manual handling and documentation station are rather technical stations where it is more obvious whether the students do it 'right' or 'wrong'. Scores of 'borderline performing students' varied around the 'static cut-off score' in the latter stations; whereas in the 'handwashing' and 'blood pressure measurement' station, the scores of borderline performing students appeared to be much higher with consequently more variation in 'true and error scores'. Making examiners aware of these differences and training them on existing pre-recorded scenarios and using well described rubrics might reduce the amount of error and should be the focus of additional research. The benefit of student feedback allowing them an opportunity to benchmark themselves against the group and to get relevant timely feedback on their performance should be obligatory in future comparisons. Although not the subject of this study, the overall impact on time reduction in running the OSCEs electronically on preliminary exploration indicated that approximately thirty hours of administration time was saved about this aspect needs to be further researched. In contrast to our previous paper based approach, results and feedback could be released immediately after the exam was finished.

Limitations of the Study

We acknowledge that a major limitation to the generalizability of this study is not just the small number of stations ($n = 4$), but also, the relatively small number of examiners ($n = 18$). We also recognize that the station content in this study was limited to a small number of clinical tasks. We have not explored the use of this online tool with more complex OSCE stations that assess competencies such as breaking bad news, explaining a treatment plan or responding to an angry patient or family member. In this study, we chose a limited set of clinical skills to pilot this user acceptance test. The opportunity for more detailed studies exists.

Currently, we are not aware of possible negative effects of using this online assessment tool. These could include distraction of the examiners as a result of concentrating on the software rather than observing the students' performance. Further observational studies on examiners behavior during online marking are in progress.

Acknowledgement

We acknowledge the work of Dr John Patterson, honorary senior lecturer at the Centre for Medical Education of the Barts and London School of Medicine and Dentistry and Assessment Consultant. His work on Borderline Regression Analysis (Excel file) is used as an essential part in OMIS software.

References

Alinier, G., 2003. Nursing students' and lecturers' perspectives of objective structured clinical examination incorporating simulation. Nurse Educ Today 23 (6), 419–426.
Baid, H., 2011. The objective structured clinical examination within intensive care nursing education. Nurs Crit Care 16 (2), 99–105.
Barry, M., Noonan, M., Bradshaw, C., Murphy-Tighe, S., 2012. An exploration of student midwives' experiences of the Objective Structured Clinical Examination assessment process. Nurse Educ Today 32 (6), 690–694.
Black, N.M.I., Harden, R.M., 1986. Providing feedback to students on clinical skills by using the Objective Structured Clinical Examination. Medical Education 20 (1), 48–52.
Cunningham, D.M., Walsh, D., 2008. OSCE Management Information System. *1.8.5*, OSCE Management Information System. http://www.qpercom.com.
Guiton, G., Hodgson, C.S., Delandshere, G., Wilkerson, L., 2004. Communication skills in standardized-patient assessment of final-year medical students: a psychometric study. Advances in Health Sciences Education 9 (3), 179–187.
Harden, R.M., Stevenson, M., Downie, W.W., Wilson, G.M., 1975. Assessment of clinical competence using objective structured examination. Br Med J 1 (5955), 447–451.
Iramaneerat, C., Yudkowsky, R., 2007. Rater errors in a clinical skills assessment of medical students. Evaluation and the Health Professions 30 (3), 266–283.
Iramaneerat, C., Yudkowsky, R., Myford, C.M., Downing, S.M., 2008. Quality control of an OSCE using generalizability theory and many-faceted Rasch measurement. Advances in Health Sciences Education 13 (4), 479–493.
Iramaneerat, C., Myford, C.M., Yudkowsky, R., Lowenstein, T., 2009. Evaluating the effectiveness of rating instruments for a communication skills assessment of medical residents. Advances in Health Sciences Education 14 (4), 575–594.
Kaufman, D.M., Mann, K.V., Muijtjens, A.M.M., Van Der Vleuten, C.P.M., 2000. A comparison of standard-setting procedures for an OSCE in undergraduate medical education. Academic Medicine 75 (3), 267–271.
Kropmans, T.J., O'Donovan, B.G., Cunningham, D., Murphy, A.W., Flaherty, G., Nestel, D., et al., 2012. An online management information system for objective structured clinical examinations. Computer and Information Science 5 (1), 38–48.
McWilliam, P., Botwinski, C., 2010. Developing a successful nursing Objective Structured Clinical Examination. J Nurs Educ 49 (1), 36–41.
Mitchell, S.K., 1979. Interobserver agreement, reliability, and generalizability of data collected in observational studies. Psychological Bulletin 86 (2), 376–390.
Newble, D.I., Swanson, D.B., 1988. Psychometric characteristics of the objective structured clinical examination. Medical Education 22 (4), 325–334.
Olsen, K., 2010. An examination of questionnaire evaluation by expert reviewers. Field Methods. 22 (4), 295-18.
Oranye, N.O., Ahmad, C., Ahmad, N., Bakar, R.A., 2012. Assessing nursing clinical skills competence through objective structured clinical examination (OSCE) for open distance learning students in Open University Malaysia. Contemp Nurse 41 (2), 233–241.
Pugh, D., Touchie, C., Wood, T.J., Humphrey-Murto, S., 2014. Progress testing: is there a role for the OSCE? Medical Education 48 (6), 623–631.
Schmitz, F.M., Zimmermann, P.G., Gaunt, K., 2011. Electronic Rating of Objective Structured Clinical Examinations: Mobile Digital Forms Beat Paper and Pencil Checklist in a Comparative Study. Springer, Berlin.
Segall, N., Doolen, T.L., Porter, J.D., 2005. A usability comparison of PDA-based quizzes and paper-and-pencil quizzes. Computers & Education 45 (4), 417–432.
Smith, V., Muldoon, K., Biesty, L., 2012. The Objective Structured Clinical Examination (OSCE) as a strategy for assessing clinical competence in midwifery education in Ireland: a critical review. Nurse Educ Pract 12 (5), 242–247.
Treadwell, I., 2006. The usability of personal digital assistants (PDAs) for assessment of practical performance. Med Educ 40 (9), 855–861.
Verhoeven, B.H., Hamers, J.G.H., Scherpbier, C., Hoogenboom, A.J.J.A., R. J. I., & Van der Vleuten, C. P. M., 2000. The effect on reliability of adding a separate written assessment component to an objective structured clinical examination. Medical Education 34 (7), 525–529.
Wass, V., Jones, R., Van Vleuten, C.D., 2001a. Standardized or real patients to test clinical competence? The long case revisited. Medical Education 35 (4), 321–325.
Wass, V., McGibbon, D., Van Vleuten, C.D., 2001b. Composite undergraduate clinical examinations: how should the components be combined to maximize reliability? Medical Education 35 (4), 326–330.

Please cite this article as: Meskell, P., et al., Back to the future: An online OSCE Management Information System for nursing OSCEs, Nurse Educ. Today (2015), http://dx.doi.org/10.1016/j.nedt.2015.06.010