

Reliability and Validity of OSCE in Evaluating Clinical Skills of Nursing Students

Moattari M, PhD¹; Abdollah-zargar S, MSc²; Mousavinasab M, MD³; Zare N, PhD⁴; Beygi Marvdast P, MSc²

¹Associate professor, Faculty of Nursing, Shiraz University of Medical Sciences

²Instructor, Faculty of Nursing, Shiraz University of Medical Sciences

³Professor, Psychiatric Department, Shiraz University of Medical Sciences

⁴Associate professor, Department of Statistics, Shiraz University of Medical Sciences

Abstract

Background: Objective structured clinical examination (OSCE) is one of the most preferable means of evaluating cognitive, emotional and psychomotor aspects of nursing students. The present study was designed to determine the reliability and validity of OSCE in evaluating clinical skills of nursing students.

Materials and methods: For this descriptive evaluative study, 10 different skills of nursing students were selected. Then, 37 students performed these skills at 10 different OSCE stations, during which two experienced inspectors evaluated their performance. Finally, the correlation between OSCE scores and the mean theoretical and clinical performance scores of students was calculated. Meanwhile, the correlation between the total OSCE scores and OSCE score of each station was determined. Inspector's reliability (correlation between scores reported by inspectors at each station) was also calculated.

Results: The correlation coefficient of OSCE scores and mean theoretical and clinical performance scores were 0.38 ($p=0.031$) and 0.52 ($p=0.005$), respectively. Correlation coefficient of inspectors' reliability was in a range of 0.38-0.95.

Conclusion: OSCE is strongly suggested as a reliable and valid means of evaluating nursing students' clinical skills.

Key words: CLINICAL EVALUATION, OSCE, RELIABILITY, VALIDITY.

Journal of Medical Education Summer 2009; 13(3); 79-85

Introduction

The importance of clinical teaching as the most important part of nursing education has been recognized by all educational planners and educational program directors. Clinical evaluation, as an important pillar of nursing education, must incorporate measuring the

nursing competencies. Therefore the necessity of using different evaluation methods for assessment of different aspects of nursing functions looks quite warranted. Objective structured clinical examination (OSCE) as one of the best methods of assessment is capable of evaluating the achievement of the three domains of educational objectives: cognitive, attitude and psycho-motor. OSCE as a test for evaluation of clinical competence consists of a series of time-scheduled stations in which the examinees are required to carry out special clinical tasks at each station (1). So that students demonstrate their competence in a variety of simulated conditions (2)

Corresponding author: Dr Marzieh Moattari is an associate professor who also has an MSc degree in Medical education

Address: Faculty of Nursing and Midwifery, Shiraz University of Medical Sciences Po.Box: 71935-1314, Shiraz Iran

Phone Cell: 98 917 7170856 Work: 98 711 626 6931

E-mail: moattar@yahoo.com

Since 1975 when OSCE was introduced by Harden(3) and was published in the medical Education Journal(4), many articles were written in medical journals about its application supporting the evidence of its use across Europe and north America (5,6); in Canada it became compulsory for their national state board examination (5). Within a few years OSCE became extensively used for the assessment of other health professionals such as chiropractors, nurses, nurse practitioners, pharmacists and physiotherapists(7) Teachers of Mac Master nursing school used it for assessment of nursing skills in third year nursing students in 1984 and it was for the first time that OSCE was used to assess nursing skills in Canada (8) it was also used in nurse education in the united kingdom since the introduction of project 2000(9). Features of OSCE such as testing the process and the product (giving importance to individual competencies), covering a broad range of clinical skills much wider than a conventional examination, objectivity in scoring(since standards of competence are preset and agreed check lists are used for scoring), objectivity in response stations where questions are asked, using simulations for acute cases, existence of the scope for immediate feedback. ,high validity due to eliminating patient variability and examiner variability (10) has led educators to use it all over the world.

This wide spread use of OSCE is a response to the critiques made on Common traditional clinical evaluation methods such as lack of appropriate time allocation, the large number of those taking part in each exam, and lack of objectivity (1) subjectivity, bias and halo effect inherent within many continuous assessment of practice (2,11,12) ,considering knowledge as the focal point of clinical evaluation(13) and domination of written tests that have measured little more than student's ability to recall facts(know) whereas there should be grater emphasis on the later stages of "shows how" and "does"(14). Though OSCE has been recognized as a gold standard (15,16) it has not been used in

Iranian nursing educational programs. In Iran, too, the issue of clinical evaluation still remains an unresolved issue and lack of specific standards for work in clinical environment, lack of specific and well – developed objectives (with the characteristics of specificity, attainability, measurability, relevancy and being time scheduled) have influenced not just the process of learning by students (17) but also the quality of evaluation. However according to the evidence available, the number of studies being made on assessment by this method in Iranian Nursing curricula is much limited. The evidence available shows the existence of only one such study, inclusively related to specific skills in the Newborn's section (1) in which the measurement of validity and reliability is not considered. The importance and necessity for conducting a valid test for assessment of clinical skills of nursing and scarcity of research on this subject led into the present evaluative study to be made to examine the validity and the reliability of the OSCE in assessing clinical skills of the fourth year nursing students.

Materials and Methods

In this descriptive and evaluative research all final year nursing students who were candidates for participation in the study (n = 37) were involved. The reason for selecting such a group was to find a general estimation of the performance status of graduates.

After obtaining official permission for this research, the reference books on basic nursing skills (18) and review lists for carrying out basic nursing skills (19) were studied and a list of nursing procedures (63 procedures) was prepared. In the next stage, to ensure the validity and to prioritize the skills, the list was provided to selected experts (faculty members of the nursing school and authorized head nurses working in hospitals) and they were asked to rank them based on the importance of each skill and the feasibility of its application so as their

feedback be considered in selecting the OSCE test items.

Based on the mean rating given to each skill, the ten-top rated skills were selected to be incorporated in the OSCE as follows: Tracheotomy suction, cardio – pulmonary Resuscitation, tracheotomy care, washing the hands by surgical methods, insertion of a Foley catheter, subcutaneous injection , putting on gloves by surgical method, nasogastric tube feeding, intra muscular injection and insertion of the nasogastric tube. The choice of those items based on professional views of instructors and head nurses was just towards providing the content validity of the test.

After selecting the procedures, a briefing session was held and the prepared check lists for each of the selected functions was provided to the examiners (faculty members of the group n=20) and they were asked to review them and give their corrective views on them. The check list is a collection of question, words and/or sentences concerning each function for which the examiner, without knowing about the true value of his/her views, check marks in front of the question, word or sentence that is mostly representing the characteristics or traits intended by the examinee (1). In this study, the number of items present in the check lists varied from 14 to 51 and totally consisted of 234 items.

In the next stage, all the students under study attended a briefing session and were introduced to the objectives of the study and procedures for conducting the test. They were assured that their taking part in the test had no effect on their evaluation and the data provided by them would only be used for the analysis of the study results. The purpose of that session was taking the attention of the students and encouraging them to participate in the study.

OSCE was conducted in two consecutive days. On each of the days, 5 stations were designed for evaluation of students. To prevent sharing the information between the students regarding the exam, they were asked

to attend and stay in a class during the examination period. At first they received their code numbers as their ID in the test. The students entered the skill lab which was designed for the exam in the order of their code numbers. The first student started taking part in the exam from the first station and after passing the first station; the next student was called to start the first station. At each station, after reading the guide for the station, they carried out of the function required, while being evaluated by two observers based on the check lists. Each student left the skill lab after completing his/her last station. During the test, no student partaking in the test could meet the students waiting to be called for the test. In addition, to amuse students who waited to be called for, a video tape was played. Since the time required for carrying out the functions varied from 5 to 7 min., all functions designed for 5 min. were evaluated on the first day and those designed for 7 min. , were evaluated on the second day.

Analysis of data

For determining the criterion referenced validity of the test, a criterion test as an acceptable representation of the students' performance was needed. Therefore, the grade point average of the clinical nursing courses (19 Credits) and the grade point average of theoretical nursing courses (25 credits) were used as criteria for determining validity. The correlation of each of these means with the score obtained from the ten OSCE stations was computed. To determine internal consistency, the correlations of the total OSCE scores of students with that of students' scores at each station were computed. For determining the inter-rater reliability, the correlations between the scores reported by two observers at each station were computed individually. Using split half reliability, the correlation between scores of the even and odd stations was computed.

Results

The results of criterion validity analysis of the test are as follows: the correlation between OSCE scores and the grade point average of theoretical and clinical courses were significant ($r = 0.376$, $P < 0.035$) and ($r = 0.523$, $P < 0.005$), respectively.

The results of internal consistency by computing the correlation between each student's total score in OSCE with student's score at any of stations showed that the scores of nasogastric tube feeding station had the highest correlation (0.744) and the station of putting on gloves by the surgical method had the least correlation (0.053). The findings show that the P-Values of all stations, except one (that of putting on gloves by the surgical method) were significant.

Inter-rater reliability of the examiners (the correlation between scores given by two observers at each station has been shown in the following table:

According to the findings presented in the above table, correlations between observer scores in all stations (except one) were more than 0.7. The correlation in that special case was 0.37. Despite of this, the Pearson

correlation supported significance of the relation between observer scores in all stations.

Split half reliability (the correlation between the two test) was positive and significant ($r=0.605$ and $P<0.001$). Since the correlation between the two observers at the station of putting on gloves by the surgical method was low (0.379) and the correlation between the scores of the same station and the total test score was weak and insignificant, its scores were excluded in the re-computation of the this reliability. In this stage, the correlation coefficient between the two test increased again to 0.651 ($P<0.001$). Therefore, the split half reliability of the test is approved by more confidence.

The findings related to split half reliability of the exam shows that the correlation between the mean score of the odd stations and that of even stations as well as the correlation between these means after deleting the students' scores at the ninth station (for lacking correlation with the total test score) * were significant ($r=0.605$ $P<0.001$) and ($r = 0.651$ and $P<0.001$) respectively.

Table. The correlation between scores given by two observers at each station

OSCE Station	Correlation Between Scores of the two Observers	P-Value
Tracheotomy suctioning	0.815	<0.001
Tracheotomy care	0.883	<0.001
Foley catheterization	0.728	<0.001
Cardiopulmonary resuscitation	0.946	<0.001
Washing hands by the surgical method	0.874	<0.001
Nasogastric tube insertion	0.812	<0.001
Nasogastric feeding	0.901	<0.001
IM injection	0.803	<0.001
Putting on gloves by the surgical method	0.379	0.021
Subcutaneous injection	0.768	<0.001
Total score	0.961	<0.001

Discussion

Criterion referenced validity of the OSCE held in this study is confirmed based on the positive and significant correlation obtained between OSCE scores and grade point average of both theoretical and clinical courses. The grade point averages of theoretical and clinical courses were used as an appropriate reference because they reflect the students learning during their study. They also are related, judicious, accessible and reliable. It is stated that prior academic performance rather than preparatory studying time is a better predictor of OSCE outcome (20) the correlation found between OSCE scores and theoretical courses are more than what was found in other studies (8,21) and lesser than what Harden found in 1975(3). However the weak correlation of OSCE scores with the GPA of theoretical courses comparing with the same correlation with that of clinical scores may reflect the theory practice gap in our nursing curricula. Considering the fact that OSCE must evaluate the three domains of knowledge, skill and attitudes, finding a criterion that covers all the three, especially the attitudes domain directly, is actually impossible. However the written test consisting of MCQ and short answer questions (mostly used in our theoretical course exams) has been reported as a good scale for discriminating the graduates as regard to their competencies (22). Also, there are supportive evidence for the predication that there is a correlation between knowledge and clinical performance (23). Validity of OSCE, was confirmed in another study using the views of medical students and examiners (24). In one of Iranian medical schools the validity of OSCE has been confirmed (25).

considering the acceptable correlation found between OSCE scores and those for each station, one can conclude that of the stations designed in this study, all (except one in which the likelihood of presence of a problem in the station's direction, the station's design or the evaluators of the station could not be rejected.) measure the same behavioral field

and thus, the internal consistency of the test is at a desired level,. There is supporting evidence regarding the constructive validity of OSCE using the correlation between the station scores and the total OSCE scores. (26) The correlations between scores given by the first and the second observer (inter rater reliability) at different stations varied from 0.379 to 0.946 with a P-value between 0.001 and 0.021 which are acceptable.

These findings represent the desired reliability for test being conducted. The result is in accordance with another study in which, the reliability of the OSCE was reported to be 0.65 and the inter rater reliability was found to be between 0.66 with 0.99(23). The reliability of another OSCE designed for evaluation of communication skills were between 0.59 and 0.65 (27). It is noteworthy that using other methods for measuring the reliability such as the test retest method, is not so common due to its needing design, spending time and high cost relating to OSCE test. On the whole, one can approve validity and reliability of the OSCE test conducted at Shiraz school of Nursing and Midwifery. Using OSCE both in education and service has many implications, for example, in education one can mention assisting students in self evaluation, triggering motivation in them, enhancing different aspects of learning in them and assisting instructors and officials in discovering strengths and weaknesses of the programs in schools. Moreover, OSCE may be used in clinical settings as a suitable method for evaluation of newcomer nurses, evaluation of new skills and as a part of evaluation of the continuing education plan.

References

- 1.Zareie Dehbaghee F. Evaluation of OSCE used in clinical evaluation of nursing students [MSc Thesis]. Tehran: Shahid Beheshti University of Medical Sciences; 2000.
- 2.Watson R, Stimpson A, Porock D. Clinical competence assessment in nursing: a systematic review of the literature. *J Adv Nurs* 2002;39(5):421-31

3. Harden R, Stevenson M, Wilson Downie W, Wilson G. Assessment of clinical competence using an objective structured clinical examination. *BMJ* Feb 1975;447-51.
4. Harden R, Gleeson F. Assessment of clinical competence using an objective structured clinical examination. *Med Edu* 1979;13:41-54.
5. Rezenick R, Regehr G, Yee G, Rothman A, Blackmore D, Dauphinee D. Process rating forms versus task-specific checklists in an OSCE for medical licensure. *Acad Med* 1998;73(10):97-115
6. Regehr G, MacRae H, Rezenick R, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998; 73 (9):993-7
7. Grand Maison P, Lescop J, Rainsbury P, Brailovsky C. Large-scale use of an OSCE for family physicians. *Can Med Assoc J* 1992;146(10):1735-40
8. Ross M, Caekill G, Knight J, Chamberlain M, Fothergill Bourbonnais F, Linton J. Using the OSCE to measure clinical skills performance in nursing. *J Adv Nurs* 1998;13:45-56.
9. Nicol M, Freeth D. Assessment of clinical skills: a new approach to an old problem. *Nurs Edu Today*. 1998;18:601-9.
10. Ananthakrishnan N. Objective structured clinical/practical examination(OSCE/OSPE). *J Postgrad Med* 1993;39-42
11. Lankshear A. Failure to fail: the teachers dilemma. *Nurs Standard* 1990;4 (20), 35-7.
12. Calman L, Watson R, Norman N, Redfern S, Murrels T. Assessing practice of student nurses: methods, preparation of assessors and students views, *J Adv Nurs* 2002; 38,516-523.
13. Daelmans H, Hem-Storrs H, vander Hoogenboom RJI, Scherpbier A, Stehoover A, vander Vloten CM. Feasibility and reliability of an in-training assessment program in an undergraduate clerkship. *Med Edu* 2004; 38:1270-7
14. Wilkinson TJ, Frampton CM. Comprehensive undergraduate medical assessments improve prediction of clinical performance. *Med Edu* 2004; 38:1111-6.
15. Bartfay W, Rombough R, Howse E, Leblanc R. The OSCE approach in nursing education. *Can Nurs* 2004; 100(3),18-23.
16. Educating future physician for Ontario. How to run OSCE.[Video and manual] Toronto:University of Toronto; 1994.
17. Moattari M, Parsayecta Z. Clinical performance evaluation: a model .[projects on evaluation methods in PhD course]. Tabriz: Tabriz university of medial sciences; 2001.
18. Potter P, Perry AG. Fundamentals of nursing. 5th ed. St Louis: Mosby; 2001
19. Delaune SC, Ladner PK. Skills checklist to accompany fundamentals of nursing standard and practice. Second ed. Sydney: Delmar Co; 2002.
20. Mavis BE. Does studying for an objective structured clinical examination make a difference? *Med Edu* 2000; 34:808-12.
21. Roberts J, Brown B. Testing the OSCE: a reliable measurement of clinical nursing skills. *Can J Nurs Research* 1990;22(1):51-9.
22. Jenson JJM, Scheropler AS, Metz Jc, Grol, RP, vander Veluten CP, Rethans JJ. Performance based assessment in continuing medical education for general practitioners. *Med Edu* 1996; (130):339-344.
23. Moattari M. Evaluation of a composed continuing medical education program executed for general practitioners of Shiraz. [MSc Thesis]. Tehran: Shahid Beheshti University of Medical Sciences; 2001.
24. Singgelton A, Smith F, Harvis T, Ross-Harper R, Hilton S. an evaluation of team objective structured clinical examination (TOSCE). *Med Edu* 1999; 33(1):34-41.
25. Azizi, F. Learning and research methods in medial education. Tehran: Javan Co;1993.
26. Wilkinson TJ, Newble DL, Wilson Carter JM, Helms RM. Development of a three-center simultaneous objective structured clinical examination. *Med Edu* 2000; 34(10):498- 507.
27. Hodge B, Turnbull J, Cohen R, Bienenstock A, Norman G. Evaluating communication skills in Objective structured

clinical examination format: Reliability and generalizability. Med Edu 1996; 30(1):38-43.