Original Article

Reliability and Validity of Objective Structured Clinical Examination for Residents of Obstetrics and Gynecology at Kermanshah University of Medical Sciences

Nasrin Jalilian M.D.[1], Reza Pourmirza Kalhori M.Sc.[2]*, Mansour Rezaei Ph.D.[3], Nahid Jalilian M.Sc.[4]

1. Dept. of Obstetrics and Gynecology, School of Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran

2. Dept. of Emergency Medicine, School of Paramedics, Kermanshah University of Medical Sciences, Kermanshah, Iran

3. Dept. of Biostatistics, School of Public Health, Fertility and Infertility Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran

4. Education Development Center, Kermanshah University of Medical Sciences, Kermanshah, Iran

**Email: pourmirza_r@kums.ac.ir*

(Received: 15 Oct 2012 Accepted: 17 Dec 2012)

Abstract

Introduction: Objective structured clinical examination (OSCE) is used for the evaluation of the clinical competence in medicine for which it is essential to measure validity and reliability. This study aimed to investigate the validity and reliability of OSCE for residents of obstetrics and gynecology at Kermanshah University of Medical Sciences in 2011.

Methods: **A** descriptive-correlation study was designed and the data of OSCE for obstetrics and gynecology were collected via learning behavior checklists in method stations and multiple choice questions in question stations. The data were analyzed through Pearson correlation coefficient and Cronbach's alpha, using SPSS software (version 16). To determine the criterion validity, correlation of OSCE scores with scores of resident promotion test, direct observation of procedural skills, and theoretical knowledge was determined; for reliability, however, Cronbach's alpha was used. Total sample consisted of 25 participants taking part in 14 stations. P value of less than 0.05 was considered as significant.

Results: The mean OSCE scores was 22.66 (±6.85). Criterion validity of the stations with resident promotion theoretical test, first theoretical knowledge test, second theoretical knowledge, and direct observation of procedural skills (DOPS) was 0.97, 0.74, 0.49, and 0.79, respectively. In question stations, criterion validity was 0.15, and total validity of OSCE was 0.77.

Conclusion: Findings of the present study indicated acceptable validity and reliability of OSCE for residents of obstetrics and gynecology.

Keywords: Reliability, Validity, Evaluation, Residency, Gynecology

Citation: Jalilian N, Pourmirza Kalhori R, Rezaei M, Jalilian N. Reliability and validity of objective structured clinical examination for residents of obstetrics and gynecology at Kermanshah University of Medical Sciences. Educ Res Med Sci. 2012; 1(2): 69-73.

Introduction

The aim of education is changing the learners' behavior. It is necessary to evaluate the teaching process to determine learning. This evaluation which is called measurement is determined based on objective and learning circumstances (1). Test is a well-known tool for evaluation in the teaching process (2). Every test should have three characteristics; validity, reliability, and practicality (3). OSCE is one of the assessment methods in medicine. OSCE is a proper tool for the assessment of internship and learning of practical skills. In this method, the examinees are required to carry out particular clinical duties in scheduled stations. Scoring is done in each station based on intern's competence in a specific skill and predetermined objective (4). Harden et al. introduced OSCE in 1972 (5). It is used in medical universities worldwide, including Iran.

There are three stations in OSCE; method (practical skill), question, and rest stations. OSCE possess strengths such as high validity and objectivity, equal examination of all students, absence of luck, and learning assessment in psycho-motor domains. Its disadvantages, however, are high labor requirement, abundant supplies, checklist preparation, and excessive time requirement (6). This method of evaluation is used more than other methods owing to satisfaction of medical (7), pharmaceutical (8), and nursing (9) interns about OSCE. This assessment method was so much welcomed that OSCE was recognized as an essential test for board examination of 24 medical specialties in Iran from 2005 and became was operational consequently (10). It is recommended OSCE should have reliability and validity (5). The reliability and validity of OSCE in psychiatry and residency (11) for the evaluation of nursing students' clinical skills (12) have been determined. The reliability of 0.89 and validity of 0.96 have been reported for OSCE for psychiatry residents (13). Tudiver et al. (2009), in Tennessee University, USA, reported the construct validity of 0.96 for the six-station OSCE used for the assessment of evidence-based medical skills and reliability of 0.58 for Cronbach's alpha of OSCE (14).

Reliability and validity have been carried out in other majors, too. In Shiraz, the criterion validity (correlation of the means of the scores for nursing clinical and theoretical courses with OSCE scores, r=0.38) and internal consistency (correlation of the OSCE scores for all students with students' scores in each station, r=0.52) have been reported (12).

Since OSCE is held every year as part of residency promotion test in obstetrics and gynecology department and the reliability and validity of this test have not been determined, not only in obstetrics and gynecology department, but also in other departments at Kermanshah University of Medical Sciences (KUMS), the present study was conducted to investigate the reliability and validity of the OSCE for residents of obstetrics and gynecology at KUMS in 2011.

Methods

In this descriptive-correlation study we used the data from OSCE for residents of obstetrics and gynecology that was administered 2011-2012. Part of the clinical competence evaluation of each resident is measured through OSCE. The instruments used for data collection were learning behavior checklists in question stations and multiple choice tests in question stations. Face validity and content validity of the instrument was approved by faculty members of obstetrics and gynecology department. To administer OSCE, the topics of stations were firstly proposed in obstetrics and gynecology meetings and subsequently, the faculty members designed the OSCE to be administered in the center for clinical skills at Imam Reza Hospital. Then, consulting with literature references for the promotion test, the objectives of the examination were determined and the draft of the OSCE and questions' card were prepared. Stations were designed as 9 method stations, 4 question stations, and 1 rest station. After doing the final control to fix latest defects in obstetrics and gynecology department OSCE stations were arranged the day before OSCE administration.

In the final meeting of the department, we discussed how to observe the individuals' performance in stations and how to score checklists. In the day of the examination, residents referred to the center for clinical skills and after their identity was checked, they referred to the stations. After administering each question or method in each station, the resident referred to the subsequent stations in specific times under the control of the examiner of the station who was a faculty member of obstetrics and gynecology department. The scores of each resident in each station were collected and analyzed by descriptive and inferential statistics using Pearson correlation coefficient and Cranbach's alpha.

Correlation of OSCE scores and scores of Direct DOPS was measured to determine the criterion validity. Correlation of total scores of promotion test and scores of theoretical knowledge, which was held in two stages, was measured to determine the concurrent

validity. To determine reliability (internal consistency), Cronbach's alpha was used for all stations. The statistical society of the research consisted of residents of obstetrics and gynecology who had taken OSCE. The calculated sample size with the assumption of reliability of 0.38 and alpha equal to 5% was five participants, however, the total of 25 residents of obstetrics and gynecology participated in 14 stations. P value of <0.05 was considered as significant.

Results

The mean age of participants was 29.43 (±5.62) years. The mean of obtained scores in method and question stations were 22.66 (±6.85) out of 32 and 8.70 (±3.19) out of 12, respectively. The reliability of method station was 0.79; this was 0.15, however, for the question station. The reliability of total OSCE for residents of obstetrics and gynecology was 0.77. The highest level of reliability (r=0.80) in method stations belonged to uterus vascular anatomy and breast physical examination stations, and for question stations belonged to pharmacology (p=0.001) (Tables 1, 2).

To determine the concurrent validity, Pearson correlation coefficient was applied and correlation of the scores of the stations with direct DOPS scores was r=0.48. Correlation of the stations' scores with first theoretical written exam, second theoretical written exam, and resident promotion written exam were, 0.74, 0.49, and 0.97, respectively (p=0.001). The highest level of correlation for method stations with OSCE belonged to histology of uterine infections (r=0.73) and IUD placement (r=0.52). The highest level of correlation of question stations with OSCE was reported for pharmacology stations (r=0.75) (p=0.001).

Table 1: . Mean and standard deviation of the scores and Cronbachs' alpha in method stations

Method station	Cronbach's alpha in the case of item elimination	Score in station	Mean±SD
The history of the infertile couple	0.78	3	2.00±1.92
Diagnostic laparoscopy	0.77	3	2.64±0.77
Uterine artery anatomy	0.80	4	2.88±1.45
Shoulder dystocia	0.78	4	3.10±0.76
Physical examination of the breast	0.80	3	2.12±1.00
Histology of uterine infections	0.76	4	2.44±1.12
IUD placement	0.76	3	1.64±1.11
Infertility Treatment	0.74	4	3.04±1.45
Uterine tumors	0.72	4	2.80±1.73

Table 2. Mean and standard deviation of the scores and Cronbach's alpha in question stations

Question station	Cronbach's alpha in the case of item elimination	Score in station	Mean±SD
Medical Ethics	-0.14	5	1.48±1.32
Fetal health	0.17	4	2.98±0.96
Male infertility	-0.06	4	1.12±0.83
Pharmacology	0.37	5	3.12±1.58

Discussion

The findings of the present study revealed the achievement of educational objectives in terms of validity of OSCE for residents of obstetrics and gynecology with DOPS test (r=0.48), first theoretical exam (r= 0.74), second theoretical exam (r=0.49), and concurrent validity with promotion test scores r=0.97. The reliability of the total OSCE for residents of obstetrics and gynecology at KUMS was r=0.77, which is considered acceptable. This was higher than what was obtained in the study carried out on nursing students (12) in Shiraz in terms of reliability, and construct and criterion validity. Correlation coefficient between the mean of checklist scores and OSCE total score of residents of obstetrics and gynecology was r=0.73 which was less than that of OSCE for psychiatry residents in Shiraz with r=0.89 (13).

The construct validity and Cronbach's alpha of OSCE for the assessment of evidence-based medical skills in Tennessee University, USA were r=0.96 and r=0.58, respectively (14). In OSCE for transportation of injured patients in emergency medicine, the criterion validity of the exam was r=0.58 and construct validity was r=0.63 (15). Since the criterion validity between the OSCE for the residents of obstetrics and gynecology and the scores of promotion test as well as the construct validity of first written exam were acceptable, it is suggested that DOPS be improved and second theoretical written exam for residents be reviewed in the analysis of clinical competence of residents of obstetrics and gynecology.

In OSCE for the nursing students, the reliability of 0.61was obtained by dividing the stations (ten stations) into half and calculating the correlation coefficient between them (12), but in the present study, Cronbach's alpha for all stations was considered as reliability coefficient. The reliability of 0.53 reported for OSCE for transportation of the injured patients in emergency medicine was also determined by Cronbach's alpha, which was lower than that of OSCE for the residents of obstetrics and gynecology (15). The acceptable reliability of OSCE indicates that this OSCE can be repeated in the upcoming years. Although raters' reliability (correlation between the reported scores by two raters in each station separately) is recommended as a more proper method of estimating the OSCE reliability (12, 16). This method is suggested to be used for estimating the OSCE reliability of residents of obstetrics and gynecology in the future studies.

Examiner bias in scoring is probably one of the reasons of low reliability of OSCE in some studies (17). To enhance the reliability of OSCE, various suggestions such as eliminating stations' problems and using trained pseudo-patient have been proposed (18).

Although many resources in medical education have emphasized the use of OSCE for the measurement of learning, there are several reports on lack of coverage by OSCE in terms of learning measurement in knowledge, attitude, and skill domains (19). The findings of the present study confirm this problem in terms of reliability of multiple choice questions. Moreover, since only pharmacology station had a proper reliability in terms of correlation with total OSCE, it is recommended that questions at question stations for residents of obstetrics and gynecology be redesigned and reviewed.

Conclusion

The results of this study confirm the regular use of OSCE for the assessment of clinical skills of the residents of obstetrics and gynecology. Improvement of DOPS for the residents of obstetrics and gynecology, and more precise designing of multiple choice questions on theoretical knowledge and questions at question station are recommended.

Acknowledgement

This study was taken from the research proposal number 91185 approved by the education development center at KUMS.

References:

1. Shabani H. Instructional skills (methods and techniques of teaching). 8th ed. Tehran: SAMT publication; 2003. [Persian]

2. Saif AA. Educational measurements, assessment and evaluation. 3th ed. Tehran: Doran Publication; 2010. [Persian]

3. Newble DI, Cannon RA. A handbook for medical teachers. 4th ed. Boston: Kluwer academic publications; 2001.

4. Alinir G. Nursing students and lectures perspective of objective structured clinical examination incurporating simulation. Nurse Education Today. 2003; 23(6): 419-426.

5. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination. Med Educ. 2011; 45(12): 1181-1189.

6. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J. 1975; 1(5955): 447–451.

7. Jalili Z, Noohi E, Ahmadpour B. Investigation of medical clerkship and internship satisfaction on OSCE As a clinical skill evaluation method in Kerman University of Medical Sciences. Strides in Development of Medical Education. 2005; 2(1): 18-24. [Persian]

8. Austin Z, O'Byrne C, Pugsley J, Quero Munoz L. Development and validation processes for an objective structured clinical examination for entry-to-practice

certification in pharmacy: the Canadian experience. Am J Pharm Educ. 2003; 67(3): Article 76.

9. Rushforth HE. Objective structured clinical examination (OSCE): Review of literature and implication for nursing education. Nurse Education Today. 2007; 27 (5):481-490.

10. Momtazmanesh N, Einolahi B, Malekan Rad E, Ghafari H. Strategic report in using of OSCE in board exams in medical exams. Strides in Development of Medical Education. 8th congress of medical education; Kerman: 2007. [Persian]

11. Moattari M, Abdollah Zargar S, Mousavinasab M, Zare N, Beygi Marvdast P. Reliability and validity of OSCE in evaluating clinical skills of nursing students. Pejouhesh. 2007; 31(1): 55-59. [Persian]

12. Taghva A, Rasoulian M, Panaghi L, Bolhari J, Zarghami M, Nasresfahani M et al. Validity and reliability of the first objective structured clinical examination (OSCE) in psychiatry in Iran. IJPCP. 2007; 13(1): 17-24. [Persian]

13. Attari A, Mirsepassi G, Taghva A, Bolhari J, Aminoroaia M, Hasanzadeh A. Validity and reliability of an objective structured clinical examination in psychiatry: A guided survey. IJPCP. 2007; 13(1): 41-48. [Persian]

14. Tudiver F, Rose D, Banks B, Pfortmiller D. Reliability and validity testing of an evidence-based medicine OSCE station. Innovations in Family Medicine Education. 2009; 41(2):89-91.

15. Pourmirza Kalhori R, Sabour B, Naderipour A, Rezaei M. The application, validity and reliability of OSCE for evaluation of carrying and transferring injured ones in Kermanshah Paramedics' Faculty Emergency Medical Students in 2011. Quarterly scientific Journal of Rescue & Relief. 2012; 4(2): 36-42. [Persian]

16. Movafagi Z, Akbari Farmad S, Dastani M, Vahid H. Comparative of perception for faculty members and residents of cardiovascular from validity of OSCE. 10th congress of medical education; Shiraz: 2009. [Persian]

17. Guraya SY, Alzobydi AH, Salman Sh. Objective structured clinical examination: Examiners' bias and recommendations to improve its reliability. Journal of Medicine and Medical Science. 2010; 1(7): 269-272.

18. Auewarakul C, Downing SM. Praditsuwan R, Jaturatamrong U. Item analysis to improve reliability for an internal medicine undergraduate OSCE. Adv Health Sci Educ Theory Pract. 2005; 10(2): 105-113.

19. Zolfagari B, Adibi N, Drakhshonfor S, Tarsaz M, Karbasi A, Niromand D. Academic achievement tests. 1th ed. Esfahan: Neshat publication; 2001. [Persian]