



## Validity and Reliability of pre-internship Objective Structured Clinical Examination in Shiraz Medical School

NARGES VASEGHI<sup>1\*</sup>, MAHVASH ALIZADEH NAINI<sup>1</sup>, REZA LABAF GHASEMI<sup>2</sup>, SOHEILA AMIRI<sup>1</sup>

<sup>1</sup>Clinical Skill Lab Center, Shiraz University of Medical Sciences, Shiraz, Iran; <sup>2</sup>Education Development Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

### Abstract

**Introduction:** Objective Structured Clinical Examination (OSCE) is one of the most appropriate methods for assessment of clinical skills. Validity and reliability assurance is a mandatory factor for any assessment tool. In Shiraz University of Medical Sciences, medical students' clinical competences are evaluated by a pre-internship OSCE. This study is designed to examine the validity and reliability of this exam. Validity is the extent to which the test measures what it intends to measure. Reliability refers to the accuracy of measurement and the consistency of test results.

**Methods:** Content validity was evaluated by expert opinion about blueprinting and station checklists. To determine the construct validity, station scores correlation with the total OSCE score and inter station correlations were calculated. The inter examiner reliability was assessed by coefficient of correlation.

**Results:** Content validity was established by alignment between the curriculum and the blueprint using expert opinion. Correlation of the station scores with the total OSCE score were positive and statistically significant in all stations except the 16<sup>th</sup> station (suturing). Inter examiner reliability coefficients of correlations ranged 0.33 – 0.99, with an average of 0.83.

**Conclusions:** Our findings support the assumption that the pre-internship OSCE is valid, reliable and suitable to assess students' clinical competence. Validity and reliability studies should be performed for all new assessment tools, particularly in high-stakes assessments.

**Keywords:** OSCE, Assessment, Validity, Reliability

*Corresponding author:*  
Narges Vaseghi  
Address: Clinical Skill Lab  
Center, Faghihi Hospital, Zand  
Ave., Shiraz, Iran.  
Email: vaseghin@sums.ac.ir  
Tel: +98-711-2303878

*Please cite this paper as:*  
Vaseghi N, Alizadeh Naini M,  
Labaf Ghasemi R, Amiri S.  
Validity and Reliability of pre-  
internship Objective Structured  
Clinical Examination (OSCE)  
in Shiraz Medical School. J. Adv  
Med&Prof. 2013;1(3):85-88.

### Introduction

Clinical skills assessment is an important aspect of medical education curriculum in all medical schools (1). Validity, reliability, feasibility and cost effectiveness are the main criteria for choosing assessment methods. On the other hand, an assessment tool should be able to predict clinical performance. Ignoring its limitation and based on its capability to evaluate clinical competencies, OSCE is used in medical schools to measure and compare medical students' skill. A well-constructed OSCE provides important information about the quality of training and can predict the candidate's performance well (2). In a typical OSCE, different stations are developed, using either real or standard patients (3). The examinees' competencies are assessed by structured

checklists. Examiners are chosen from trained staff, general physicians, assistants and nurses. All 6<sup>th</sup> year medical students at Shiraz University of Medical Sciences have been assessed by a pre-internship OSCE since 2005. The exam is concomitant with comprehensive MCQ exam and is held before the beginning of the internship period. Some aspects of these exams were assessed in prior studies. This study was designed to evaluate the validity and reliability of this exam to assure its applicability in the medical curriculum. Validity is the extent to which the test measures what it intends to measure, i.e. validity can be referred to as the truthfulness of the measuring tool (4). There are different kinds of validity of which five are the most common. They are (a) Face validity which is whether the assessment feels right on the

face of it, it can be demonstrated by expert opinion, (b) Content validity which refers to how much the exam covers the areas of competency, (c) Construct validity which is whether the assessment produces the expected results, (d) Concurrent validity which refers to the degree to which scores on a test correlate with the scores on an established test administered at the same time, and (e) Predictive validity which relates to the certainty with which a test can predict future performance (5). Reliability refers to the accuracy of measurement and the consistency of the results. In other words, it measures the extent to which an indicator or instrument measures consistently and accurately (6). In this study, we investigated face, content and construct validity as well as reliability of pre-internship OSCE used in Shiraz University of Medical Sciences. Concurrent validity of this exam was established by another study.

## Methods

### OSCE Implementation

All 6<sup>th</sup> year medical students (116; male: 55, female: 61) took part in OSCE in two main groups with 6 subgroups simultaneously. The students were informed about OSCE (rules and method) in an orientation session accompanied by an information sheet.

Considering the number of students, time factors and the availability of appropriate space (three similar floors at Shahid Motahari Clinic), the exam was repeated twice with 30 min interval, each in 3 subgroups. An expert team designed twenty short stations (each 5 min) with appropriate checklists for each station. One station was rest. All examiners were informed about scoring. Based on the pattern of evaluation, all security aspects were considered so that there was no contact between the two groups during intervals. All students completed the OSCE without any absentees. At the end of the exam, more than two thousand checklists were collected and scoring was completed. SPSS version 11.5 was used for processing and data analysis.

### Assessing Validity and Reliability

Based on undergraduate medical student curriculum goals and objectives, an OSCE blueprint was prepared by an expert team. It listed the clinical skills such as History Taking, Physical Examination, Patient Management, Communication Skills, Procedural Skills, Problem Solving and Para Clinical Workup (appropriate choose and interpretation) in horizontal line. Specific domains of practice such as Internal Medicine, Surgery, Psychiatry, Obstetrics, Pediatrics, Emergency Medicine, Public Health, ENT, Ophthalmology and Neurology were outlined

in a vertical line. Based on the curriculum, content validity was assessed by comparing the content of each OSCE station with the variable aspects of clinical competencies. This was facilitated by constructing a grid of core curriculum subjects and competencies. Face validity was evaluated by expert opinion, done by interview and agreement more than 0.75 was accepted.

The construct validity was evaluated by correlating station scores with the total OSCE score and inter-station correlations.

For measuring the inter-rater reliability, students' competencies were assessed by two independent examiners simultaneously; both for written and interactive stations and the coefficient of correlation were calculated and categorized to excellent ( $> 0.9$ ), good (0.7-0.89), borderline (0.5-0.69), acceptable (3-0.49), and poor ( $< 3$ ).

## Results

Face and content validity were reviewed and confirmed by faculty members. Content validity was established by alignment between the curriculum and OSCE using a blueprint (Table 1).

To assess construct validity correlation between station scores and the OSCE total scores was positive and statistically significant in all stations except the 16<sup>th</sup> station (suturing). The highest was in the 18<sup>th</sup> station ( $r=0.550$ ,  $p<0.001$ ) and the lowest in the 4<sup>th</sup> station ( $r=0.217$ ,  $p<0.019$ ). The inter-station correlations were positive in only less than 12 percent of scores and the highest was seen between 2<sup>th</sup> and 18<sup>th</sup> stations ( $r=0.450$ ,  $p<0.001$ ) (Table 2).

The inter-examiner reliability was assessed by the coefficient of correlation, an average of 0.83 (rang: 0.33- 0.99) as shown in Table 3. It was positive for all stations and statistically significant in 15 stations (more than 0.85), borderline in two stations and weak but acceptable in stations 1 & 2 (0.404 and 0.330, respectively).

## Discussion

Assessment has a main role in any educational program. It not only measures and compares trainees' capabilities but also provides students and teachers with feedback about their performance (7). To choose an appropriate assessment method, five factors should be considered; Validity, Reliability, Feasibility, Acceptability and Educational Impact. Based on Millers pyramid, MCQ or writing exams, even if designed in a proper manner, could assess only cognitive domain of learning objectives. All assessment methods for clinical competence such as Global rating scores (GRS), Direct observation of procedural skills (DOPS), mini clinical exams (mini

**Table 1.** Pre – Internship OSCE Blueprint

Number of stations	Practice Domain	Clinical skills							
		History taking	Physical Examination	Communication Skills & Ethics	Differential diagnosis	Problem Solving	Doing Procedure	Para clinical workup	Patient Management
1	Endocrinology					+		+	+
2	Pulmonary disease		+	+					
3	Infectious disease				+	+			+
4	Ophthalmology		+		+				
5	Nephrology				+			+	
6	Cardiology					+			
7	Hematology			+		+		+	
8	Dermatology				+				+
9	Obstetrics		+				+		
10	ENT					+		+	+
11	Pediatrics						+		
12	Psychiatry	+							
13	Rest								
14	Urology						+		
15	Gastrointestinal disease	+			+	+			+
16	Surgery						+		
17	Emergency medicine		+		+	+			+
18	Rheumatology		+	+					
19	Neurology		+						
20	Public Health			+					

CEX) have known limitations such as high cost, time consuming, weak validity and reliability to be choose routinely especially for under graduate trainees. On the other hand the degree for performance prediction differs in all above exams. OSCE take place in a standard artificial situation which differs with real work place, but if be choose appropriately may overcome some above defects. To be success in OSCE, it is necessary to have a good clinical knowledge and be able to perform clinical skills appropriately.

DE Blakemore in medical council of Canada in

Mayo Graduate School of Medicine showed that assessing knowledge with multiple-choice or written examination is not a clear predictor of performance on clinical practices. Kelly Kirby Ortega and coworkers showed that the students' performance on the USMLE step 1 examination did not correlate with the students' performance on a 3<sup>rd</sup> year OSCE (8). But Wilkinson, et al. explained that knowledge leads to good OSCE performance (9).

In Shiraz University of Medical Sciences, OSCE is used for students' clinical assessment, one is

**Table 2.** Correlation station scores with the total OSCE score

Stations	1	2	3	4	5	6	7	8	9	10
Correlation with Total score	0.276	0.475	0.550	0.217	0.313	0.372	0.479	0.394	0.499	0.452
Stations	11	12	13	14	15	16	17	18	19	20
Correlation with Total score	0.395	0.263	-	0.288	0.321	0.158	0.531	0.555	0.355	0.361

**Table 3.** The inter-examiner reliability, the coefficient of correlation

Station	1	2	3	4	5	6	7	8	9	10
<b>Pearson Correlation</b>	0.404	0.330	0.982	0.793	0.998	0.999	0.794	0.968	0.879	0.976
<b>P</b>	.011	.040	.000	.000	.000	.000	.000	.000	.000	.000
Stations	11	12	13	14	15	16	17	18	19	20
<b>Pearson Correlation</b>	0.914	0.986	-	0.598	0.520	0.822	0.983	0.842	0.975	0.981
<b>P</b>	.000	.000	-	.000	.001	.000	.000	.000	.000	.000

performed at the end of History taking and Physical Examination course (mini OSCE with 6 stations). The other take place in pre internship and the last for competent assurance at the end of internship. For its implementation in under graduate curriculum it was mandatory to assure its validity and reliability.

In this study we investigated the validity and Reliability of the pre-internship OSCE at Shiraz University of Medical Sciences.

To evaluate construct validity, correlation of the station scores with the total OSCE scores and correlations of inter-station were calculated. This issue was previously supported by Wilkinson, et al. (10).

## Conclusion

Positive and meaningful outcomes of the correlation between stations' scores and the OSCE total scores pointed out that designed stations in this study measured unique behavioral domain except for one station (the 16<sup>th</sup> station). This exception could be the result of a wrong design of the station or the examiner's mistake.

To assess inter-rater reliability, the coefficient of correlation between two examiners was calculated to be between 0.33-0.99. It was proposed that results were positive and meaningful. Moatari, et al. showed similar results between 0.38-0.95 in a research about evaluation of clinical skills among the 4<sup>th</sup> year nursing students (11).

This study ascertained that the pre-internship OSCE at Shiraz University of Medical Sciences has face, content and construct validity as well as inter-examiner reliability.

## References

1. Troncon LE. Clinical skills assessment: limitations to the introduction of an "OSCE" (Objective Structured Clinical Examination) in a traditional Brazilian medical school. Sao Paulo Medical journal. 2004; 122:12-17. PubMed PMID: 15160521.
2. Tervo RC, Dimitrievich E, Trujillo AL, Whittle K, Redinius P, Wellman L. The Objective Structured Clinical Examination (OSCE) in the clinical clerkship: an overview. S D J Med. 1997 May; 50(5):153-6. PubMed PMID: 9155233.
3. Walters k, Osborn D, Raven P. The development, validity and

reliability of a multimodality objective structured clinical examination in psychiatry; Med Educ. 2005; 39: 292 – 8. PubMed PMID: 15733165.

4. Shehmar M, Cruikshank M, Finn C. A validity study of the national UK colposcopy objective structured clinical examination-is it a test fit for purpose? An International Journal of Obstetrics and Gynaecology. 2009; 1796-1800. Doi: 10.1111/j.1471-0528.2009.02389.x.
5. Dent JA, Harden RM. A practical guide for medical teachers: McAleer S, editor choosing assessment instruments. 2nded: Elsevier, Churchill living stone; 2005; P302-310.
6. EdCaN. Competency standards in professional nursing: a summary of literature published since 2000:2008; P18-21.
7. Jeffries WB, Huggett KN. An Introduction to Medical Teaching: Mavis B, editors. Assessing Student Performance. Springer Dordrecht Heidelberg London New York; 2010; Chapter 11; P143 – 178.
8. Ortega KK, Natt N, Tieg R. The relationship between performance on a third-year medical student OSCE and performance on the USMLE step 1 examination. Mayo Graduate School of Medicine; Mayo Clinic, Rochester, USA. 2003; Available from:www.amee.org/conf2003/Bern\_FULL.pdf.
9. Wilkinson TJ, Frampton CM. Comprehensive undergraduate medical assessments improve prediction of clinical performance. Med Educ. 2004; 38: 1111–16. PubMed PMID: 15461657
10. Wilkinson TJ, Newble DI, Wilson PD, Carter JM, Helms RM. Development of a three-centre simultaneous objective structured clinical examination. Med Educ. 2000; 34:798-807. PubMed PMID: 11012928.
11. Moattari M, Abdollah-zargar S, Mousavinasab M. Reliability and Validity of OSCE in Evaluating Clinical Skills of Nursing Students; journal of Medical Education. Summer 2009; 13(3); 79-85. Persian.