

Association of the pre-internship objective structured clinical examination in final year medical students with comprehensive written examinations

Hasan Eftekhari^{1,2*}, Ali Labaf^{3,4}, Pasha Anvari⁵, Arsia Jamali⁵ and Farshad Sheybaee-Moghaddam⁵

¹Department of Health Promotion and Education, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran; ²Health Center for Community based Participatory Research, Tehran University of Medical Sciences, Tehran, Iran; ³Department of Emergency Medicine, Imam Khomeini Hospital, Tehran University of Medical Sciences, Tehran, Iran; ⁴Clinical Skills Development Center, Medical Faculty, Tehran University of Medical Sciences, Tehran, Iran; ⁵Students' Scientific Research Center (SSRC), Tehran University of Medical Sciences, Tehran, Iran

Aim: The purpose of this study is to evaluate the association of the pre-internship Objective Structured Clinical Examination (OSCE) in final year medical students with comprehensive written examinations.

Subjects and material: All medical students of October 2004 admission who took part in the October 2010 National Comprehensive Pre-internship Examination (NCPE) and pre-internship OSCE were included in the study ($n = 130$). OSCE and NCPE scores and medical grade point average (GPA) were collected.

Results: GPA was highly correlated with NCPE ($r = 0.76$ and $P < 0.001$) and moderately with OSCE ($r = 0.68$ and $P < 0.001$). Similarly a moderate correlation was observed between NCPE and OSCE scores ($r = 0.6$ and $P < 0.001$). Linear stepwise regression shows r^2 of a model applying GPA as predictor of OSCE score is 0.46 ($\beta = 0.68$ and $P < 0.001$), while addition of gender to the model increases r^2 to 0.59 ($\beta = 0.61$ and 0.36, for GPA and male gender, respectively and $P < 0.001$). Logistic forward regression models shows male gender and GPA are the only dependent predictors of high score in OSCE. OR of GPA and male gender for high OSCE score are 4.89 (95% CI = 2.37–10.06) and 6.95 (95% CI = 2.00–24.21), respectively ($P < 0.001$).

Discussion: Our findings indicate OSCE and examination which mainly evaluate knowledge, judged by GPA and NCPE are moderately to highly correlated. Our results illustrate the interwoven nature of knowledge and clinical skills. In other words, certain level of knowledge is crucial for appropriate clinical performance. Our findings suggest neither OSCE nor written forms of assessments can replace each other. They are complimentary and should also be combined by other evaluations to cover all attributes of clinical competence efficiently.

Keywords: OSCE; clinical skills; written examination; reliability; validity

Received: 17 December 2011; Revised: 28 February 2012; Accepted: 20 March 2012; Published: 24 April 2012

Medical education is an essential investment that involves much trainings and many clinical disciplines and programs. These programs are aimed at producing competent graduates who are able to perform appropriate history taking, comprehensive physical examinations, problem-solve, order and interpret essential paraclinic evaluations, arrive at a diagnosis, and outline a management plan.

Considering the significance of this investment, it is of vital importance to evaluate how this investment is paying off. In this regard, the evaluation of the medical students at certain training points plays a prominent role for the students and their trainers (1, 2).

Traditionally, multiple-choice questions, oral examinations, and tutor reports were frequently used in the evaluation of medical students. However, oral evaluations

and ward assessments lack practical reliability and validity (3–6). Although multiple-choice questions have an appropriate reliability, they are limited by the fact that they only assess one dimension of the students' competencies which is clinical knowledge (5–7). Therefore, evaluation of competencies and higher cognitive skills of undergraduate and graduate medical students, by means of OSCE has been more emphasized in recent years (8, 9). For example, the medical councils of Canada, Japan, and Korea employ OSCE in their licensing examination and the National Board of Medical Examiners incorporates use of OSCE into the US Medical Licensing Step 3 Examinations (10, 11) and nearly all medical schools in United States have reported use of OSCE in their regular evaluations (12).

Medical education in Iranian Universities is divided into four periods: basic sciences (five semesters), introduction to clinical medicine (two semesters), clinical clerkships (five semesters), and internship (18 months). Before being allowed to continue their medical studies, students are supposed to pass two national comprehensive examinations after finishing their basic sciences period (national comprehensive basic sciences examination) and clinical clerkships (national comprehensive pre-internship examinations). Both of these examinations are composed of multiple-choice questions and are conducted twice a year by the Ministry of Health and Medical Education. In 2009, Tehran University of Medical Sciences established pre-internship OSCE to evaluate the clinical skills of the medical students entering internship.

Despite appointing studies on the value of the OSCE in graduate and undergraduate settings, there are inconclusive findings regarding the relationship of the OSCE with other means of evaluation. More specifically it is not clearly known how different knowledge-based evaluations versus OSCE rank students. Similarly, it is not clear which kind of stations can assist OSCE in providing better divergent validity. The aim of the current study is to:

- (1) Evaluate the reliability and validity of the pre-internship OSCE.
- (2) Assess the association between traditional written examinations and OSCE scores in an Iranian medical school.
- (3) Determine the stations which mostly contribute to divergent validity of OSCE.

Subjects and materials

Study population

All medical students of October 2004 admission who took part in the October 2010 National Comprehensive Pre-Internship Examination (NCPE) and pre-internship OSCE were included in this study ($n = 130$).

Study measures

OSCE settings

The OSCE comprised 12 stations in four circuits, with 5 min at each station. The total examination took 8 hours. Each station evaluated one or more aspects of clinical competencies, including history taking, physical examination, communication skills, interpretation of laboratory findings, generating differential diagnosis, and management. Scoring was done by a single examiner at nine stations based on a prepared checklist (in history taking, psychiatric interview, neurologic examination, ophthalmologic examination, management of preterm rupture of membranes (PROM), arterial blood gas sampling, stitching, adult cardiopulmonary resuscitation, and orthopedic procedure of splinting). Three stations (chest X-ray findings in Mitral Stenosis [MS]), Treatment and prognosis of Kerion, and evaluation of child growth curve) were unmanned; the students recorded their findings at these stations, subsequently their recordings were evaluated by a single examiner. At five stations (history taking, psychiatric interview, neurologic examination, ophthalmologic examination, and orthopedic procedure of splinting) a standardized patient played a role. Communication skills and knowledge items in history taking station included establishing a rapport, acting respectfully, and identifying patient's concerns, involving the patient in decision-making process, and planning for management. In Kerion station, description of the lesion, making differential diagnosis, identifying most probable diagnosis, requesting appropriate laboratory tests, providing appropriate treatment were questioned. In chest X-ray finding station, examinees were asked about chest X-ray findings as well as complications and treatment of MS. In PROM station, students answered questions regarding management of PROM. In psychiatric interview examinees were supposed to inquire about risk factors of suicide including marital status, carrier, past medical and psychiatric history, and medications of a standardized patient attempting suicide. In ophthalmologic examination, examinees were supposed to assess visual acuity, pupillary light reflex, and choose appropriate treatment for a standardized patient with eye trauma. In other stations, related skills were examined. Examiners were trained prior to the exam and were not involved in the design of the station. Standardized patients were trained in three 45 min individual sessions and were equipped by written instructions 1 week prior to the examination. A description, explaining each station was written and placed at the door of each station and 1 min was assigned for reading them.

On the day of the OSCE, the examinees were given a 30 min orientation. In this session, the structure of the examination was reviewed and an opportunity to ask questions was provided.

Questionnaires to assess the content validity of the OSCE were distributed to the physician examiners (as lay experts), asking them to rate the stations they had observed in compliance with the importance of each station, adequacy of time, appropriateness of standardized patients' actions and capacity of check lists to assess stations' objectives.

National Comprehensive Pre-internship Examination (NCPE) scores and medical grade point average (GPA) NCPE scores and grade point averages (GPA) were kindly provided by the education deputy of the school of medicine. The highest attainable scores for NCPE and medical GPA were 200 and 100 respectively.

Statistical analysis

Data were analyzed with SPSS v13.0 for Windows (SPSS, INC., Chicago, IL, USA). Descriptive statistics, such as mean, median, count, range, and standard deviation (SD) were used to describe the characteristics, GPA, NCPE, and OSCE scores of the students. Reliability was calculated for the overall examination and each station by means of internal consistency statistic Cronbach's alpha. Independent T-test and ANOVA were used to assess the differences among groups. Pearson's correlations were employed to assess the associations between the NCPE score, GPA, and other quantitative variables with the OSCE scores. Variables correlated with OSCE score with $P < 0.2$ were selected for regression. NCPE scores, GPAs, gender, and age were employed as independent variables in linear and logistic regression models to predict the OSCE scores (dependent variable). For logistic models, the students were grouped to high and low OSCE scores ($n = 65$ and 65 , respectively) and further analysis was performed on these groups. Probability value of less than 0.05 was considered significant.

Ethical considerations

The study protocol was approved by TUMS Research Ethics Committee.

Results

Reliability and validity of OSCE

Of 130 students included in the study, 76 (58.46%) were female. Median age of the participants was 24 ranging from 23 to 26. Table 1 summarizes the internal consistency of the checklists at different stations of the OSCE and correlation between scores of each station of the OSCE and total OSCE score excluding that station from the total score.

The content validity of the OSCE is shown in Table 2. No significant difference was detected in the score of the students entering different circuits ($P = 0.08$).

Relationship of OSCE scores with comprehensive examinations

Mean OSCE score achieved by the students was 56 (22.44–80.08) out of total score of 120. Table 3 compares the OSCE and NCPE scores, as well as GPAs of the students according to gender. As this table presents, there was no significant difference in GPAs and NCPE scores of female and male students. However, male students performed better in the OSCE (Table 3).

GPA was highly correlated with NCPE ($r = 0.76$ and $P < 0.001$) and moderately with OSCE ($r = 0.68$ and $P < 0.001$). Similarly, a moderate correlation was observed between NCPE and OSCE ($r = 0.6$ and $P < 0.001$) (Data are not listed in the tables). Linear stepwise regression denoted r^2 of a model applying GPA as the predictor of the OSCE score was 0.46 ($\beta = 0.68$ and $P < 0.001$), while the addition of gender to the model increased r^2 to 0.59 ($\beta = 0.61$ and 0.36, for GPA and male gender, respectively, and $P < 0.001$). NCPE score was removed from both models due to high levels of collinearity. Logistic forward regression models showed that male gender and GPA were the only dependent predictors of a high score in the OSCE. OR of GPA and male gender for high OSCE scores were 4.89 (95% CI = 2.37–10.06) and 6.95 (95% CI = 2.00–24.21), respectively ($P < 0.001$).

Divergent validity of OSCE stations

Table 4 summarizes the association of each station of the OSCE with the NCPE score and medical GPA. The highest correlation between NCPE score and OSCE stations was observed in treatment and prognosis of Kerion, management of PROM, and chest X-ray findings in MS stations ($r = 0.5$, 0.38, and 0.35, respectively). Correspondingly, three stations of treatment and prognosis of Kerion, chest X-ray findings in MS, and management of PROM showed the highest correlation with GPA of the participants ($r = 0.46$, 0.37, and 0.36, respectively). No significant correlation was evident between NCPE score and arterial blood gas sampling, evaluation of child growth curve, stitching, and adult cardiopulmonary resuscitation stations (Table 4).

Discussion

This study illustrates that OSCE is a reliable and valid method for evaluation of medical students' clinical competencies. Previous studies have expressed a wide range of reliability from 0.12 to 0.89 for OSCE, indicating that there liability of the OSCE is setting dependent (3, 8, 9, 13–17). In our settings, the reliability of OSCE was similar to the standard of 0.8 which is comparable to written examinations (18, 19). In agreement with our findings, it is suggested that OSCEs with fewer than 10 stations might lack the ability to incorporate all necessary materials to even superficially cover reasonable measures of clinical competency, therefore reducing their validity

Table 1. Description of the OSCE stations and their reliability (number of participants = 130)

Station Description	Type of Station	Cronbach's alpha	P value	r	P value
History taking	H/C	0.91	<0.001*	0.47	<0.001*
Psychiatric interview	H/C	0.82	<0.001*	0.33	<0.001*
Management of preterm rupture of membranes	P/PS	0.88	<0.001*	0.60	<0.001*
Neurologic examination	P	0.89	<0.001*	0.51	<0.001*
Ophthalmologic examination	P/PS	0.83	<0.001*	0.31	<0.001*
Chest x-ray findings in mitral stenosis	PS	0.75	<0.001*	0.45	<0.001*
Treatment and prognosis of Kerion	P/PS	0.82	<0.001*	0.53	<0.001*
Evaluation of child growth curve	PS	0.75	<0.001*	0.34	<0.001*
Arterial blood gas sampling	S	0.82	<0.001*	0.26	0.002*
Stitching	S	0.80	<0.001*	0.52	<0.001*
Adult cardiopulmonary resuscitation	S	0.88	<0.001*	0.52	<0.001*
Orthopedic procedure of splinting	S	0.85	<0.001*	0.57	<0.001*

H, History taking; C, Communication skills; P, Physical examination; PS, Problem solving; S, Skill; r, item-total test score correlation. *Significant.

and reliability (20–22). Our findings justify the use of OSCE at particular points of medical education.

Our findings reveal that OSCE and examinations which mainly evaluate knowledge, judged by GPA and NCPE, are moderately to highly correlated. This exposes the interwoven nature of knowledge and clinical skills. In other words, a certain level of knowledge is crucial for appropriate clinical performance. In line with our findings, Simon et al have concluded that second-year medical students' OSCE scores were moderately correlated with USMLE Step 2 scores (23). Accordingly, Muller et al demonstrated a moderate correlation between clinical skills and USMLE Step 2 (24). A similar association was reported in dental students' or residents' OSCE and written examinations (9, 25, 26). On the other hand, the high correlations observed in the aforementioned studies indicate the low divergent validity of OSCE. In other words, clinical skills should be more weighted in OSCE to increase the divergent validity of this exam so that OSCE can provide some information which is not evaluated by knowledge-based written forms of examinations. In concordance with this finding, a detailed evaluation of OSCE stations pointed out that stations in which knowledge is more emphasized, such as

treatment and prognosis of Kerion, management of PROM, and chest X-ray findings in MS, showed a higher correlation with medical GPA and NCPE scores. In contrast, arterial blood gas sampling, stitching, adult cardiopulmonary resuscitation, and evaluation of child growth curve stations, which focused more on clinical skills, were the main sources of the divergent validity of OSCE. Notably, in this study, 5 min was considered for the students to accomplish their tasks in each station. This short time may also lead to reducing the divergent validity of OSCE. Prolonging the duration of the stations to 15 or 20 min can potentially provide the examinees with the opportunity of demonstrating their clinical skills in a superior quality, therefore improving the divergent validity of OSCE; however, there is some evidence which indicates this may not significantly change the scores of the students in knowledge-based stations (27).

Our findings show that although no significance difference is observed in the knowledge of the students with different gender, male students performed much better in the OSCE. This is in contrast to the findings of some studies that implied female students tend to have a better performance in clinical examinations (23, 28–30). Although most of these studies were conducted on

Table 2. Content validity of the OSCE (n = 31)

Question	Strongly agree%	Agree%	Neutral%	Disagree%	Strongly disagree%
Checklist items were precise and clear	16.1	64.5	16.1	3.2	0
Checklist items were in accordance with station objectives	20	63.3	13.3	0	3.3
Station objectives were prevalent in daily practice	48.4	35.5	12.9	3.2	0
Station objectives were imperative in daily practice	48.4	45.2	6.5	0	0
Time of the station was adequate	19.4	51.6	6.5	22.6	0
Standardized patients acted properly	25	55	20	0	0

Table 3. OSCE, NCPE scores, and GPAs of the students ($n = 130$)^a

OSCE Stations	Total	Gender		P value
		Female ($n = 76$)	Male ($n = 54$)	
History taking	5.7 ± 1.3	5.9 ± 1.3	5.5 ± 1.3	0.21
Psychiatric interview	6 ± 1	6.1 ± 1	5.9 ± 0.9	0.34
Management of preterm rupture of membranes	4.8 ± 2	4.1 ± 2	5.7 ± 1.5	<0.001*
Neurologic examination	5.3 ± 1.7	5 ± 1.7	5.7 ± 1.6	0.03*
Ophthalmologic examination	1.8 ± 2	1.7 ± 2.1	2 ± 2.1	0.65
Chest x-ray findings of mitral stenosis	2.4 ± 1.6	1.9 ± 1.4	3 ± 1.7	<0.001*
Treatment and prognosis of Kerion	3.4 ± 2	2.7 ± 1.9	4.5 ± 1.9	<0.001*
Evaluation of child growth curve	3.7 ± 1.7	3.9 ± 1.7	3.4 ± 1.9	0.18
Arterial blood gas sampling	6.6 ± 1.5	5.8 ± 1.2	7.3 ± 1.6	<0.001*
Stitching	6.5 ± 2.1	6.5 ± 2.1	6.3 ± 2.1	0.46
Adult cardiopulmonary resuscitation	6.8 ± 1.7	6.8 ± 1.9	6.9 ± 1.6	0.8
Orthopedic procedure of splinting	2.4 ± 2.5	1.3 ± 1.7	4 ± 2.5	<0.001*
Total OSCE score	56 ± 10.1	53.7 ± 9.8	59.3 ± 9.7	0.002*
NCPE	129.76 ± 21.52	127.72 ± 19.76	132.55 ± 23.62	0.21
GPA	82.3 ± 5.4	81.6 ± 5.2	83.55 ± 5.65	0.11

^aData presented as mean ± standard deviation.

*Significant.

written examinations, some studies have pointed out the superior performance of female students in clinical skills as well (31). In contrast, other studies have indicated no or minimal gender difference in the scores of students (and only in certain stations of OSCE) which may not seriously influence the performance of the students in reality (32, 33). We can postulate that better performance of male students in practical settings which is observed in this study can be attributed to lower levels of anxiety or higher levels of self-confidence or probable fewer social

interactions or communication skills of female students in our culture; however, further studies are paramount to support these speculative explanations.

An important beneficial aspect of the OSCE is providing students and faculty members with feedback. Reviewing the group performance of the students in this study revealed that the students performed poorly in orthopedic procedure skills, ophthalmologic examinations, and chest X-ray findings in MS. The significance of this finding is further enhanced by the evidence suggesting faculty

Table 4. Association of each OSCE station with medical GPA and NCPE score ($n = 130$)

	NCPE		GPA	
	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value
History taking	0.29	0.001*	0.28	0.008*
Psychiatric interview	0.23	0.008*	0.10	0.34
Management of preterm rupture of membranes	0.38	<0.001*	0.36	<0.001*
Neurologic examination	0.20	0.02*	0.16	0.12
Ophthalmologic examination	0.27	0.002*	0.21	0.04*
Chest x-ray findings of mitral stenosis	0.35	<0.001*	0.37	<0.001*
Treatment and prognosis of Kerion	0.50	<0.001*	0.46	<0.001*
Evaluation of child growth curve	0.14	0.11	0.29	0.006*
Arterial blood gas sampling	0.08	0.33	-0.02	0.83
Stitching	0.16	0.06	0.33	0.002*
Adult cardiopulmonary resuscitation	0.21	0.013	0.24	0.02*
Orthopedic procedure of splinting	0.33	<0.001*	0.43	<0.001*

*Significant.

expectations might lack correspondence to the actual competencies of the students (21) and will facilitate faculty members in developing a more competency-based curriculum. Despite the promising possibilities of the OSCE, one should also take into account that the OSCE is labor-intensive. It requires faculty time, standardized patient recruitment and training, administrative costs, quality control, security control, etc.

There were several limitations in current study: First the impact of standardized patients and examiners on students' performance was not evaluated correspondingly. Second, our study is limited to a single medical school. Therefore, our findings may not be generalized perfectly. Moreover, TUMS is ranked as the best medical school in Iran, therefore usually highly motivated and talented students with limited variation in their competencies are admitted into this university. This suggests that, generalization of our results should be done more cautiously. Third, in this OSCE, standardized patients, as well as various standard media and medical models were utilized, so our findings may not be applicable for the OSCEs that mainly use standardized patients. Fourth, although we assessed the academic performance of the students, it is widely acknowledged that a powerful academic background does not necessarily warrant a successful professional career. Furthermore, in interpreting our results, it is important to consider that OSCE has limited ability in measuring the real performance of the students in authentic situations (34, 35). Future studies should evaluate the role of the OSCE and other examinations on clinical performance of graduated doctors.

Our findings suggest neither the OSCE nor the written forms of assessments can replace each other. These examinations are complementary and they should be combined with other methods of evaluations to cover all attributes of clinical competence efficiently.

Conflict of interest and funding

The authors have not received any funding or benefits from industry or elsewhere to conduct this study.

References

- Petrusa ER, Blackwell TA, Rogers LP, Saydjari C, Parcel S, Guckian JC. An objective measure of clinical performance. *Am J Med* 1987; 83: 34–2.
- Newble DI, Entwistle NJ. Learning styles and approaches: implications for medical education. *Med Educ* 1986; 20: 162–75.
- Matsell DG, Wolfish NM, Hsu E. Reliability and validity of the objective structured clinical examination in paediatrics. *Med Educ* 1991; 25: 293–9.
- Maxim BR, Dielman TE. Dimensionality, internal consistency and interrater reliability of clinical performance ratings. *Med Educ* 1987; 21: 130–7.
- Sloan DA, Donnelly MB, Johnson SB, Schwartz RW, Strodel WE. Use of an Objective Structured Clinical Examination (OSCE) to measure improvement in clinical competence during the surgical internship. *Surgery* 1993; 114: 343–50; discussion 350–351.
- Schwartz RW, Donnelly MB, Sloan DA, Johnson SB, Strodel WE. The relationship between faculty ward evaluations, OSCE, and ABSITE as measures of surgical intern performance. *Am J Surg* 1995; 169: 414–7.
- Levine HG, McGuire CH, Nattress LW. The validity of multiple choice tests as measures of competence in medicine. *Am Educ Res J* 1970; 7: 69–83.
- Sloan DA, Donnelly MB, Schwartz RW, Strodel WE. The objective structured clinical examination. The new gold standard for evaluating postgraduate clinical performance. *Ann Surg* 1995; 222: 735–42.
- Carraccio C, Englander R. The objective structured clinical examination: a step in the direction of competency-based evaluation. *Arch Pediatr Adolesc Med* 2000; 154: 736–41.
- Huang CC, Chan CY, Wu CL, Chen YL, Yang HW, Chen CH, et al. Assessment of clinical competence of medical students using the objective structured clinical examination: first 2 years' experience in Taipei Veterans General Hospital. *J Chin Med Assoc* 2010; 73: 589–95.
- Lee YS. OSCE for the Medical Licensing Examination in Korea. *Kaohsiung J Med Sci* 2008; 24: 646–50.
- Barzansky B, Etzel SI. Educational programs in US medical schools, 2003–2004. *Jama* 2004; 292: 1025–31.
- Petrusa ER, Blackwell TA, Ainsworth MA. Reliability and validity of an objective structured clinical examination for assessing the clinical performance of residents. *Arch Intern Med* 1990; 150: 573–7.
- Cohen R, Reznick RK, Taylor BR, Provan J, Rothman A. Reliability and validity of the objective structured clinical examination in assessing surgical residents. *Am J Surg* 1990; 160: 302–5.
- Reznick R, Smee S, Rothman A, Chalmers A, Swanson D, Dufresne L, et al. An objective structured clinical examination for the licentiate: report of the pilot project of the Medical Council of Canada. *Acad Med* 1992; 67: 487–94.
- Grand'Maison P, Lescop J, Rainsberry P, Brailovsky CA. Large-scale use of an objective, structured clinical examination for licensing family physicians. *CMAJ* 1992; 146: 1735–40.
- Roberts J, Norman G. Reliability and learning from the objective structured clinical examination. *Med Educ* 1990; 24: 219–23.
- Grosse ME, Cruft GE, Blaisdell FW. The American Board of Surgery in-training examination. *Arch Surg* 1980; 115: 654–7.
- Langdon LO, Grosso LJ, Day SC, Norcini JJ, Kimball HR, Fletcher SW. A core component of the certification examination in internal medicine. *J Gen Intern Med* 1993; 8: 497–501.
- Wessel J, Williams R, Finch E, Gemus M. Reliability and validity of an objective structured clinical examination for physical therapy students. *J Allied Health* 2003; 32: 266–9.
- Malloy MH, Perkowski L, Callaway M, Speer A. The relationship between preceptor expectations and student performance on 2 pediatric objective structured clinical examination stations. *Arch Pediatr Adolesc Med* 1998; 152: 806–11.
- Hilliard RI, Tallett SE. The use of an objective structured clinical examination with postgraduate residents in pediatrics. *Arch Pediatr Adolesc Med* 1998; 152: 74–8.
- Simon SR, Bui A, Day S, Berti D, Volkan K. The relationship between second-year medical students' OSCE scores and USMLE Step 2 scores. *J Eval Clin Pract* 2007; 13: 901–5.
- Muller ES, Harik P, Margolis M, Clauser B, McKinley D, Boulet JR. An examination of the relationship between clinical skills examination performance and performance on USMLE Step 2. *Acad Med* 2003; 78: S27–S9.

25. Dennehy PC, Susarla SM, Karimbux NY. Relationship between dental students' performance on standardized multiple-choice examinations and OSCEs. *J Dent Educ* 2008; 72: 585-92.
26. Gerrow JD, Murphy HJ, Boyd MA, Scott DA. Concurrent validity of written and OSCE components of the Canadian dental certification examinations. *J Dent Educ* 2003; 67: 896-901.
27. Schoonheim-Klein M, Hoogstraten J, Habets L, Aartman I, Van der Vleuten C, Manogue M, et al. Language background and OSCE performance: a study of potential bias. *Eur J Dent Educ* 2007; 11: 222-9.
28. Haist SA, Witzke DB, Quinlivan S, Murphy-Spencer A, Wilson JF. Clinical skills as demonstrated by a comprehensive clinical performance examination: who performs better – men or women? *Adv Health Sci Educ Theory Pract* 2003; 8: 189-99.
29. Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *BMJ* 2002; 324: 952-7.
30. Lumb AB, Vail A. Comparison of academic, application form and social factors in predicting early performance on the medical course. *Med Educ* 2004; 38: 1002-5.
31. Haq I, Higham J, Morris R, Dacre J. Effect of ethnicity and gender on performance in undergraduate medical examinations. *Med Educ* 2005; 39: 1126-8.
32. Wisikin CM, Allan TF, Skelton JR. Gender as a variable in the assessment of final year degree-level communication skills. *Med Educ* 2004; 38: 129-37.
33. Fernandez A, Wang F, Braveman M, Finkas LK, Hauer KE. Impact of student ethnicity and primary childhood language on communication skill assessment in a clinical performance examination. *J Gen Intern Med* 2007; 22: 1155-60.
34. Williams RG, Barrows HS. Performance based assessment of clinical competence using clinical encounter multiple stations. In: Hart IR, Harden RM, Walton HJ, eds. *Further developments in assessing clinical competence*. Montreal: Canadian Health Publications; 1987, pp. 125-33.
35. Barman A. Critiques on the objective structured clinical examination. *Ann Acad Med Singapore* 2005; 34: 478-82.

***Hasan Eftekhari**

Department of Health Promotion and Education
 School of Public Health
 Tehran University of Medical Sciences
 Tehran
 Iran
 Tel: 00 98-21-88955888
 Fax: 00 98-21-66462267
 Email: eftkhara@sina.tums.ac.ir