

## The relationship between second-year medical students' OSCE scores and USMLE Step 2 scores

Steven R. Simon MD MPH<sup>1</sup>, Anh Bui BS<sup>2</sup>, Shelley Day BS<sup>2</sup>, David Berti BA<sup>3</sup>, and Kevin Volkan EdD PhD MPH<sup>4</sup>

<sup>1</sup>Assistant Professor of Ambulatory Care and Prevention, Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA, USA

<sup>2</sup>Medical Student, Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA, USA

<sup>3</sup>Registrar, New England School of Law, Boston, MA, USA

<sup>4</sup>Professor of Psychology, California State University Channel Islands, Camarillo, CA, USA

**Keywords:** assessment, clinical competency, medical education, OSCE

### Abstract

#### Rationale, aims and objectives

A valid tool to measure clinical competency early in medical school could identify students who may require special educational attention. The overall aim is to assess the relationship between students' scores on an objective structured clinical examination (OSCE) given in the second year of medical school and their subsequent performance on Step 2 of the United States Medical Licensing Examination (USMLE Step 2).

#### Methods

Participants were 390 second-year medical students participating in a required OSCE; complete data (Medical College Admission Test, OSCE, USMLE Step 1 and Step 2 scores) were available for 340 students (87%). Univariate correlations and linear regression analyses were performed.

#### Results

Total OSCE score was moderately correlated with USMLE Step 2 score ( $r = 0.395$ ,  $P < 0.001$ ), as were two skills subscores of the OSCE, differential diagnosis ( $r = 0.343$ ,  $P < 0.001$ ) and identification of abnormality ( $r = 0.322$ ,  $P < 0.001$ ). In linear regression analysis, neither OSCE total score nor any of the subscores independently predicted Step 2 scores; only Step 1 score ( $\beta = 0.687$ ,  $P < 0.001$ ) and female sex ( $\beta = 0.152$ ,  $P < 0.001$ ) remained independent correlates of Step 2 score.

#### Conclusion

OSCEs early in medical school can be useful in the early assessment of clinical competence.

## **Introduction**

During the past decade, objective structured clinical examinations (OSCE) and standardized patients (SPs) have become widespread for clinical skills evaluation in undergraduate and graduate medical education. For example, in 1993–1994 only 38 of 126 medical schools in the United States employed OSCE/SP examinations. By 2003–2004, nearly all (121 of 126) reported regular use of these examinations [1].

A prior study found a moderate correlation between OSCE scores given in the second year at Harvard Medical School and performance on Step 1 of the United States Medical Licensing Examination (USMLE), a test of preclinical basic science knowledge [2]. Those results suggested the predictive validity of the OSCE and supported its use in the early identification of students who might require remedial education. The association between this OSCE and USMLE Step 2 has not previously been studied. This relationship is of particular interest given that USMLE Step 2 tests clinical knowledge, rather than basic science knowledge. We therefore examined the relationship between OSCE and USMLE Step 2, taking into account their correlation with the Step 1 examination and other factors.

## **Methods**

### **Setting**

This study took place as part of the Patient–Doctor II course at Harvard Medical School as previously described [2–4]. In brief, approximately 130 second-year medical students participate in this 220-hour physical diagnosis course each year. Taught at nine clinical sites, the course includes listening to lectures, practicing on each other and then taking histories and performing physical examinations on selected patients. The Human Studies Committee of Harvard Medical School approved the study protocol.

### **USMLE Step 2**

The USMLE Step 2 Clinical Knowledge is part of the second step of a three-step complementary process used for the granting of US medical licenses. Results of the USMLE examinations are reported to individual state medical licensing authorities for use as a common evaluation system in granting the initial license to practice medicine. USMLE Step 1 assesses basic science knowledge, while Step 2 evaluates students' understanding and application of clinical knowledge for supervised postgraduate training. Step 3 tests medical knowledge and understanding of biomedical and clinical science essential for the unsupervised practice of medicine. A Clinical Skills component, which employs SPs, was recently added to the Step 2 examination, but was not in place at the time of the study. USMLE Step 1 is generally taken in the second year of medical school, while Step 2 is generally taken in the third or fourth year. Step 3 is typically taken during postgraduate (residency) training.

Administered in a computer-based format since 1999, the USMLE Step 2 Clinical Knowledge is an 8-hour, multiple-choice test that is organized into two dimensions: Normal Conditions and Disease, and Physician Tasks. Normal Conditions and Disease addresses normal growth and development, basic concepts, general principles and individual disorders. Physician Tasks are further subdivided into promoting preventive

medicine and health maintenance, understanding mechanisms of disease, establishing a diagnosis and applying principles of management [5].

## **Objective structured clinical examinations**

The design, implementation and scoring of the OSCE have been described previously [2,3]. In summary, 10 sessions of the OSCE examination were held during a 1-week period in April for all second-year students in 1998, 1999 and 2000. Students rotated through 16 clinical stations, in which nine different formats were used alone or in combination (question and answer, preceptor role play, SPs, actual patients, mechanical or structural models, 35-mm slides, audio tape, videotape and CD-ROM). Content of each station remained constant over the 3 years of the study. Questions and tasks in the 16 stations were designed to target one of the seven following clinical skills: (1) history taking; (2) physical examination technique; (3) physical examination knowledge; (4) identification of abnormality; (5) differential diagnosis; (6) patient interaction; (7) patient presentation. There were a mean number of 55 questions per skill (range 14–70).

## **Data collection and analysis**

Correct answers to all OSCE questions were recorded on checklists by preceptors and double entered by research staff into a database. Total OSCE score, station scores and skills scores were calculated as follows. Each task or question counted as one point, and the sum of tasks performed or questions answered correctly, divided by the total number of possible answers for the OSCE ( $n = 382$ ), gave the total OSCE score a percentage. The sum of tasks performed or questions answered correctly for each station, divided by the total possible answers for each station, yielded the station score for each student. Each task or question on the checklist for every station was uniquely designated as representing one of seven skills. The sum of tasks performed or questions answered correctly for each skill, divided by the total possible answers for each skill, produced the skill score. We computed the mean  $\pm$  one standard deviation for the total OSCE score, each of the 16 station scores, and each of the seven skill scores.

We linked OSCE scores from 1998, 1999 and 2000 with students' sex, race/ethnicity, Medical College Admission Test (MCAT) scores, USMLE Step 1 and Step 2 scores. Students' self-reported race/ethnicity categories included White, African-American, Hispanic/Latino, Asian-Pacific Islander and others; we dichotomized White versus non-White because of relatively small sample sizes within each non-White race/ethnicity group. We used Pearson or Spearman correlation coefficients, as appropriate, to examine the univariate relationship between predictor variables and USMLE Step 2 score.

We performed a linear regression analysis to assess the relationship between the OSCE, MCAT and demographic variables (age, sex and race) with USMLE Step 2. A second linear regression model was performed with the original OSCE, MCAT and demographic variables, with the addition of USMLE Step 1 scores, as predictors of USMLE Step 2 scores.  $\beta$ -coefficients and adjusted  $R^2$  for these models were calculated.

In the two-stage least squares analysis, residuals from the regression of the original independent variables (OSCE stations or skills scores, MCAT scores, demographics) onto Step 1 scores were subsequently regressed onto Step 2 scores. All  $P$ -values are two-tailed.

## Results

### Sample characteristics

The OSCE scores, MCAT scores, USMLE Step 1 and Step 2 scores were available for 340 students. Table 1 shows the demographic characteristics and standardized test performance for study participants and for matriculating students who applied to US medical schools in years 1996–1998.

**Table 1**

Characteristics of study participants and matriculating students who applied to US medical schools in years 1996–1998

	Study participants (n=390)*	US medical school matriculated students (n=16 179)*
Median age (years)	23	24
Sex		
Female	202(51.8%)	43.4%
Male	188(48.2%)	56.6%
Race/ethnicity		
White	203 (52.5%)	64.8%
Other	184(47.2%)	35.2%
MCATsubscores (mean ± SD)		
MCAT biological sciences	11.73 (1.52)	10.1 ±1.7
MCAT physical sciences	11.63(1,71)	9.8 ±2.0
MCAT verbal reasoning	10.82(1.47)	9.6 ±1.7
USMLE Step 1 score (mean ± SD)	234±17	210±20*
USMLE Step 2 score (mean ± SD)	229 ±21	218±23*

\*Of the 390 students who completed the OSCE, 340 students with complete data for all measures were included in the analyses.

\*Data from the Association of American Medical Colleges (source: <http://www.aamc.org>).

\*USMLE Step 1 and Step 2 scores were average scores among all test-takers in 1998-2002 (source: <http://www.USMLE.org>).

MCAT, Medical College Admission Test; SD, standard deviation; USMLE, the United States Medical Licensing Examination.

### Univariate analyses

The Pearson correlation coefficient for the total OSCE score with USMLE Step 2 score was 0.395 ( $P < 0.001$ ). In univariate analysis, scores on Step 1 and Step 2 were highly correlated ( $r = 0.723$ ,  $P < 0.001$ ). Five of the seven OSCE skills subcomponents (differential diagnosis, history taking, identification of abnormality, physical examination knowledge, physical examination technique, and presentation skills) were significantly correlated with Step 2 scores ( $P < 0.05$ ), with differential diagnosis ( $r = 0.343$ ,  $P < 0.001$ )

and identification of abnormality ( $r = 0.322$ ,  $P < 0.001$ ) manifesting the strongest correlations (Table 2).

**Table 2**  
Univariate correlates of USMLE Step 2 scores

	Correlation coefficient*	P-value
Demographics		
Female	0.131	0.016
White	0.233	<0.001
Total OSCE score	0.395	<0.001
OSCE station subscore		
Presentation	0.131	0.016
Skin	0.281	<0.001
Alcohol and abdominal	0.173	0.001
Hemoptysis	0.170	0.002
Rectal-prostate	0.0101	0.062
Heart	0.126	0.020
Knee	0.055	0.310
Ear	0.154	0.004
Headache	0.115	0.035
Breast	0.040	0.465
Arthritis	0.312	<0.001
Thyroid	0.045	0.414
Cranial nerve	0.085	0.411
Mental status	0.154	0.016
Calf pain	0.210	<0.001
Lung	0.179	0.001
Abdominal	0.188	0.001
OSCE skills subscore		
Differential diagnosis	0.343	<0.001
History taking	0.211	<0.001
Identification of abnormality	0.322	<0.001
Patient interaction	0.085	0.228
Physical examination knowledge	0.190	<0.001
Physical examination technique	0.214	<0.001
Presentation skills	0.110	0.042
MCAT subscores		
MCAT biological sciences	0.285	<0.001
MCAT physical sciences	0.256	<0.001
MCAT verbal reasoning	0.313	<0.001
MCAT writing	0.133	0.014
USMLE Step 1 score	0.723	<0.001

\*Pearson or Spearman correlation coefficient, where appropriate.

MCAT, Medical College Admission Test; OSCE, objective structured clinical examination; USMLE, the United States Medical Licensing Examination.

## Multivariate analyses

In a linear regression model that included only MCAT, race and sex as predictors of Step 2 score, the model  $R^2$  was 0.201. When OSCE skills scores were added to the model, the  $R^2$  was 0.306. This result indicates that the predictor variables accounted for approximately 30% of the variability in Step 2 scores. Three of the seven skill scores (history taking, differential diagnosis and identification of abnormality) remained independent predictors of Step 2 score.

When Step 1 scores were added to the linear regression model, the  $R^2$  became 0.575, indicating that the model accounted for more than half of the variability in Step 2 scores (Table 3). As expected, Step 1 score was the strongest predictor of Step 2 score ( $\beta = 0.687$ ,  $P < 0.001$ ). In this fully adjusted model, sex remained the only other significant independent predictor of Step 2 scores. In a model that included Step 1 scores, MCAT scores, race and sex but excluded OSCE scores, the  $R^2$  was 0.562. Thus, the OSCE scores accounted for a relatively small percentage ( $0.575 - 0.562 = 0.013$ , or about 1.3%) of the variance in Step 2 scores when all of the other variables were included in the model.

In a two-stage least squares analysis, the residuals for a regression of the original independent variables (OSCE stations or skills scores, MCAT scores, demographics) were regressed onto Step 1 scores. The residuals of this analysis were then subsequently regressed onto Step 2 scores, yielding an adjusted  $R^2$  of 0.324.

**Table 3** Multivariate predictors of USMLE Step 2 scores

Model Pr variables	Without Step 1 0.306		With Step 1 0.575	
	Standardized $\beta$ -coefficient	P-value	Standardized $\beta$ -coefficient	P-value
White	0.128	0.010	0.072	0.071
Female	0.178	<0.001	0.152	<0.001
OSCE total score	0.307	<0.001	0.078	0.063
OSCE skills subscores				
Differential diagnosis	0.176	0.002	0.050	0.284
History taking	0.133	0.011	0.081	0.053
Identification of abnormality	0.117	0.033	-0.006	0.897
Patient interaction	-0.074	0.141	-0.070	0.083
Physical examination knowledge	-0.039	0.528	-0.016	0.766
Physical examination technique	0.107	0.079	0.039	0.412
Presentation skills	0.022	0.651	0.015	0.701
MCAT subscores				
MCAT biological sciences	0.125	0.033	-0.047	0.331
MCAT physical sciences	0.136	0.021	-0.013	0.789
MCAT verbal reasoning	0.127	0.020	0.062	0.152
MCAT writing	0.005	0.917	0.019	0.636
USMLE Step 1 score			0.687	<0.001

MCAT Medical College Admission Test; OSCE, objective structured clinical examination; USMLE, the United States Medical Licensing Examination.

## **Discussion**

Assessing medical students' clinical competency is essential throughout undergraduate medical education. Identifying students early in their training whose developing clinical skills lag behind their peers would be particularly useful. We sought to measure the ability of an OSCE in the second year of undergraduate medical training to predict clinical competency at the end of medical school, as measured by Step 2 of the USMLE. The key finding of this study was that this OSCE does predict performance on Step 2, with noteworthy levels of correlation between certain OSCE skill scores (identification of abnormalities and differential diagnosis) and the USMLE Step 2. As expected, when the relationship between Step 1 and Step 2 of the USMLE is included in assessing the association between OSCE and Step 2, the strong correlation between the two USMLE Steps overwhelms the ability of the OSCE to predict Step 2 scores. Previous study had shown that this OSCE was moderately correlated with Step 1 of the USMLE, the focus of which may be classified as basic science knowledge. As the OSCE measures students' clinical competency and clinically necessary skills, its correlation with Step 2 of the USMLE, with its emphasis on clinical knowledge, is reassuring. The findings of this study are consistent with those of Muller *et al.*, who showed modest correlation between clinical skills examinations and the USMLE Step 2 [6].

The strong relationship between Step 1 and Step 2 of the USMLE is not surprising and should not diminish the importance of identifying a correlation between the OSCE and Step 2, even though the relationship between Step 1 and Step 2 statistically overwhelms the OSCE Step 2 correlation. Our study had scores for all three of these examinations for 340 students; however, in reality, faculty and students completing the OSCE in the second year of medical school have only the OSCE score to assess competency at that stage of training and to anticipate the need for focused or intensified education programmes. Although of less importance than predicting clinical competency per se, the ability of the OSCE to identify students early in medical school who may struggle with future standardized tests can help faculty and students prepare for these examinations.

Our analyses identified female sex as a robust predictor of Step 2 performance, even when adjusted for Step 1 scores. This finding is consistent with other studies, which have suggested that female medical students tend to perform better in clinical assessments [7–9]. Our finding that Whites outperformed non-Whites on the Step 2 examination, a test of clinical knowledge, is supported by previous study [10,11]. The relationship between students' sex and race/ethnicity and their performance on clinical knowledge assessments and clinical skills assessments deserves further study.

This study has several important limitations. This study was conducted at one medical school, and the results may not be generalizable to other institutions. The OSCE was performed at the second-year medical student level and is being compared with an examination typically taken in the fourth year. The examinations consist of two different formats: one composed of an array of situations including patient encounters and focusing on performance assessment in the use of clinical knowledge; the second, multiple-choice and focusing on factual tests and clinical-contextbased tests. Furthermore, the OSCE in this study consisted of interactions with SPs as well as other structured evaluation stations, utilizing various media as well as interaction with a faculty observer. The results of this study may not be applicable to other OSCEs that principally

or exclusively use SPs for evaluation. In addition, it is important to point out that the Harvard Patient–Doctor II OSCE has recently evolved to include SPs at every station.

Assessing the development of clinical competency in undergraduate medical education and predicting the future clinical competency of practising doctors remain of paramount importance to medical teachers. This study, demonstrating moderate correlation between a second-year OSCE and a national board examination taken near graduation, suggests that OSCEs early in medical school can be useful in both of these regards. The modest level of correlation between the OSCE and the Clinical Knowledge component of the USMLE Step 2 reinforces the observation that these examinations measure different competencies and provides indirect support for the USMLE's adoption of the Clinical Skills component for the Step 2 examination. As OSCEs become more widely incorporated in undergraduate medical education and in standardized assessment, future work should examine the potential for OSCEs at different stages of training to predict the clinical performance of practicing doctors.

### **Acknowledgements**

The authors thank Eric Diddlemeyer for assistance in manuscript preparation. Dr Simon had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. There was no extramural funding of this project.

### **References**

1. Barzansky, B. & Etzel, S. I. (2004) Educational programs in US medical schools, 2003–2004. *Journal of the American Medical Association*, 292, 1025–1031.
2. Simon, S. R., Volkan, K., Hamann, C., Duffey, C. & Fletcher, S. W. (2002) The relationship between second-year medical students' OSCE scores and USMLE Step 1 scores. *Medical Teacher*, 24 (5), 535–539.
3. Hamann, C., Volkan, K., Fishman, M. B., Silvestri, R. C., Simon, S. R. & Fletcher, S. W. (2002) How well do second-year students learn physical diagnosis? Observational study of an objective structured clinical examination (OSCE). *Biomed Central Medical Education*, 2 (1), 1.
4. Makoul, G., Curry, R. H. & Novack, D. H. (1998) The future of medical school courses in professional skills and perspectives. *Academic Medicine*, 73 (1), 48–51.
5. *The United States Medical Licensing Examination*. Available at: <http://www.usmle.org/> (accessed 25 November 2005).
6. Muller, E. S., Harik, P., Margolis, M. J., Clauser, B. E., McKinley, D. & Boulet, J. R. (2003) An examination of the relationship between clinical skills examination performance and performance on USMLE Step 2. *Academic Medicine*, 78 (10 Suppl.), S27–29.
7. Ferguson, E., James, D. & Madeley, L. (2002) Factors associated with success in medical school: systematic review of the literature. *British Medical Journal*, 324, 952–957.
8. Haist, S. A., Murphy-Spencer, A., Quinlivan, S., Wilson, J. F. & Witzke, D. B. (2003) Clinical skills as demonstrated by a comprehensive clinical performance examination: who performs better – men or women? *Advances in Health Science Education*, 8, 189–199.

9. Lumb, A. B. & Vail, A. (2004) Comparison of academic, application form, and social factors in prediction early performance on the medical course. *Medical Education*, 38, 1002–1005.
10. Wass, V., Roberts, C., Hoogenboom, R., Jones, R. & Van der Vleuten, C. (2003) Effect of ethnicity on performance in a final objective structured clinical examination: qualitative and quantitative study. *British Medical Journal*, 326 (7393), 800–803.
11. Liddell, M. J. & Koritsas, S. (2004) Effect of medical students' ethnicity on their attitudes towards consultation skills and final year examination performance. *Medical Education*, 38 (2), 187–198.