

임상 실습 종합 평가로서 시행한 객관구조화진료시험 (OSCE)

- 이화 의대의 경험 -

이화여자대학교 의과대학 흉부외과학교실, 예방의학교실¹, 응급의학교실², 소아과학교실³,
비뇨기과학교실⁴, 진단검사의학교실⁵

한재진 · 박혜숙¹ · 어은경² · 유경하³ · 이동현⁴ · 정화순⁵

= Abstract =

An OSCE for Summative Assessment after Clinical Clerkship - Experience in Ewha Medical School -

Jae Jin Han, MD, Hyesook Park¹, MD, Eunkyung Eo², MD, Kyungha Yoo³, MD,
Donghyeon Lee⁴, MD, Wha Soon Jung⁵, MD

*Department of Thoracic & Cardiovascular Surgery, Preventive Medicine¹,
Emergency Medicine², Pediatrics³, Urology⁴, Laboratory Medicine⁵,
Ewha Womans University, Medical College, Seoul, Korea*

Purpose: To report the experience of OSCE performed as a summative assessment of the medical students in year 4 at Ewha Womans University.

Methods: Ninety-eight students took the twenty stations including 3 pairs of linked stations. We conducted duplicated OSCEs and each station runs 4 minutes 30 seconds. We analyzed the validity of stations, the reliability of duplicated OSCEs and the difference of scores according to the OSCE tools and categories.

Results: Mean OSCE score was 64.45±5.49 and Cronbach-α of total stations was 0.74. The scores between duplicated examinee groups and between duplicated evaluator groups were not different: the score of preceding examinee group was 63.3 and the later 65.7 (p=0.13), and the score of the first evaluator group was 64.6 and the second was 63.6 (p=0.16). However according to the individual analysis, 9 stations showed statistically discrepancies (p<0.05) between duplicated evaluator groups. Regarding OSCE tools and categories, physical examination and skill categories did not show statistically significant differences between duplicated evaluator groups. OSCE scores showed moderate correlation with the cumulative performance grade in general (r=0.53, p<0.01).

Conclusions: OSCE could be regarded as a reasonable evaluating tool for the summative assessment after two years of clinical clerkship of the fourth year medical students. Analysis of the OSCE stations which was done individually and according to OSCE tools/categories helps to improve the OSCE more complete and the duplicated stations more coincident.

Key Words: OSCE, Summative assessment, Clinical Clerkship

교신저자: 한재진, 이화여자대학교 의과대학 흉부외과학교실
서울시 양천구 목 6동 911-1
Tel: 02)2650-5629, Fax: 02)2649-4930, E-mail: jjhan@ewha.ac.kr

서 론

의과대학 교육에서 의과대학 졸업자가 ‘얼마나 많은 의학 지식을 갖고 있는가?’ 뿐만이 아니라 실제 임상에서 ‘독자적으로 의료행위를 얼마나 수행할 수 있는가?’를 목표로 할 때 이는 의학지식을 바탕으로 정보를 수집하고 머릿속에 저장하였던 지식을 끄집어내어 상황을 판단하고 필요한 술기를 이용하여 실행에 옮길 수 있는 능력을 갖추는 것을 의미한다고 할 수 있다. 이전의 의과대학에서 주로 이루어졌던 필기시험과 구두시험만으로는 이와 같은 수행능력을 평가할 수 없었으나 의학교육 (medical education)과 실제의료 (medical practice)간의 간격을 메우기 위한 의학 교육 분야 개혁의 일환으로 객관 구조화 진료시험 (OSCE: Objective Structured Clinical Examination)이 소개되어 1994년에 서울의대에서 처음 시행하였으며 2002년까지 26개 의과 대학에서 시행하였다 (이영미, 2003).

이화여대에서도 1999년에 OSCE를 도입하여 매년 의학과 4학년 전체를 대상으로 한 종합 평가의 한 방편으로 시작하여 지금에 이르고 있는데 2003년도 4학년 졸업시험의 일부로서 시행한 OSCE를 대상으로 OSCE 문항 자체의 신뢰도, 문항의 각 분야에 따른 차이점, 시험장 복제에 따른 문제점 및 OSCE 성적과 다른 학업 성적과의 차이점 등을 분석하여 의대 학생들의 임상 실습에 대한 평가로서 OSCE의 가치를 평가하고자 하였다.

대상 및 방법

2004년도 이화 의대 졸업 예정자 98명을 대상으로 모든 임상 과목의 실습이 종료된 직, 후에 졸업 시험의 일환으로 시행하였으며 평가 결과는 ‘임상 종합학’이라는 과목의 점수에 반영하였다.

가. 문제 출제 절차

시행 6 개월 전에 각 임상 교실과 예방의학 교실에서 교실별로 OSCE 출제 교수를 추천받아 OSCE 위원회를 구성하였으며 이 가운데 7명으로 준비위

원회를 만들었고 모든 출제 위원들은 해당 과목의 문제를 1~2문제 씩 출제하도록 하였다. 세 차례의 OSCE 워크숍을 가졌는데 1차 워크숍에서는 OSCE에 처음 참여하는 교수들을 위한 OSCE 설명회 및 이미 출제되었던 문제들에 관한 검토를 하였고 각자 출제할 주제는 미리 만든 문제 구성 행렬표 (matrix)에 맞추어 전체 협의를 통하여 정한 후 각자의 문항을 2차 워크숍까지 만들어 오도록 하였다. 2차 워크숍에서는 1박 2일의 합숙을 통하여, 각자 만들어 온 문항을 완성하였으며 그 가운데 이번 해에 출제할 문제를 최종적으로 정하였고 교수들이 직접 리허설을 실시하여 문항의 완성도를 높였다. 3차 워크숍에서는 문항의 복제 (duplication)를 위하여 제 2평가자에 대한 교육이 이루어졌고 그 동안 섭외한 전문연기자 표준화 환자 (standardized patients, SP) 및 일반 SP를 각 문항에 배정하였으며 모형과 준비물을 점검하였다. 이 때 전문연기자 SP가 시범을 보이는 시간을 가짐으로써 일반 SP들에게 실제 도움이 되도록 하였다. 각 문항 별로 전체 리허설 시간 전까지 3~4차례의 SP 훈련 및 제 1, 2평가자 간의 평가 일치를 위한 의견 조율 시간을 갖도록 하였으며 각 문항의 체크리스트는 준비위원회에서 점검하여 형식의 통일성을 높였다.

나. 대상 및 시행 방법

2003년도에는 응시생이 98명이므로 20개의 문항을 선정하였다 (Table I). 응시 학생 98명을 토요일 오후에 모두 시험을 치르게 하고자 시험장의 복제 (duplication)를 하였으며 또한 전체 학생을 선발팀과 후발팀으로 나누어 시험 시간 동안 격리 조치를 하였다. 20 문항 중에는 3쌍의 연속 문항이 있어서 휴식 스테이션을 합하여 총 28개의 스테이션을 만들었다. 각 스테이션에는 4분 30초간 머물며 문제를 풀고 30초 동안에 이동하도록 하였다. 그리고 가능하면 각 스테이션에서 시험 시작 후 4분이 지나면 종료 예고를 하도록 하며 퇴역임을 시행하도록 하였다.

다. 분석 방법

시험의 신뢰도를 구하기 위하여 각 문항 별로

Table I. OSCE Stations and Tools/Categories

OSCE Stations	Tools			OSCE Categories			
	Plastic MODEL	General SP	Special SP	History Taking	Physical Exam	Skills	Education
Breast 1	○				○		
Breast 2		○					○
Back Pain		○			○		
Headache 1			○	○			
Headache 2		○			○		
Acute Abdomen			○		○		
Diabetes		○		○			
Pregnancy	○				○		
Vaginal Spot			○	○			
Hemoptysis 1			○	○			
Hemoptysis 2		○					○
Foley	○					○	
Dyspnea		○			○		
Airway	○					○	
Eye Pain		○		○			
Pediatric Heart	○				○		
Mental Illness			○	○			
Transfusion		○				○	
Occupation		○		○			
Suture	○					○	

Cronbach- α 를 구하였다. 학생 선, 후발팀 및 평가자에 따른 OSCE 사용 도구 및 기술 분야 별 비교는 모든 학생들이 모든 도구와 기술 분야 항목을 시행하였기 때문에 반복 측정 분산 분석 (repeated measures ANOVA)으로 비교 평가하였으며 이들의 전체 학교 성적과의 상관 분석을 하였다. 또한 각 항목 별 점수와 전체 학교 성적과의 상관관계를 분석하여 실제 각 임상 실습 항목의 차이를 보고자 하였다. 사용한 통계적 검정법은 모두 SPSS version 11.0 통계 패키지를 이용하였다.

결 과

가. 문항신뢰도 및 문항별 성적 분포

전문연기자 SP가 환자의 역할을 한 문항은 5개, 일반 SP는 9개였으며 모형을 이용한 문항은 6개였으며 측정 분야별로는 문진이 7문항, 이학적 검사

가 7문항, 처치 및 기술이 4문항, 교육이 2문항이었다 (Table I). 전체 문항의 표준화 Cronbach- α 는 0.74로 비교적 강한 내적일치율을 보였다.

학생들의 전체 평균은 64.45 \pm 5.49점 (100점 만점 기준)이고 문항별 점수는 Fig. 1과 같다. 학생들의 OSCE 총점 분포는 Fig. 2와 같다.

나. 시험장 복제에 따른 성적

시험은 선발팀 48명과 후발팀 50명으로 나누어 시행되었고 각 팀 안에서 20명의 제 1평가자가 평가하는 시험장과, 20명의 제 2평가자가 평가하는 시험장으로 복제되었다. 선발팀이 시험을 치르는 동안 후발팀은 격리되어 모여 기다렸으며 후발팀이 시작할 때 선발팀은 분리되어 시험장을 떠나도록 하였는데 선발팀의 전체 평균은 63.3 (범위, 53.2~90.4) 점, 후발팀은 65.7 (범위, 50.3~73.3)점으로서 양 팀 간에 차이는 없었으며 (p=0.13) 문항별로도 유의한

임상 실습 종합 평가로서 시행한 객관구조화진료시험 (OSCE)

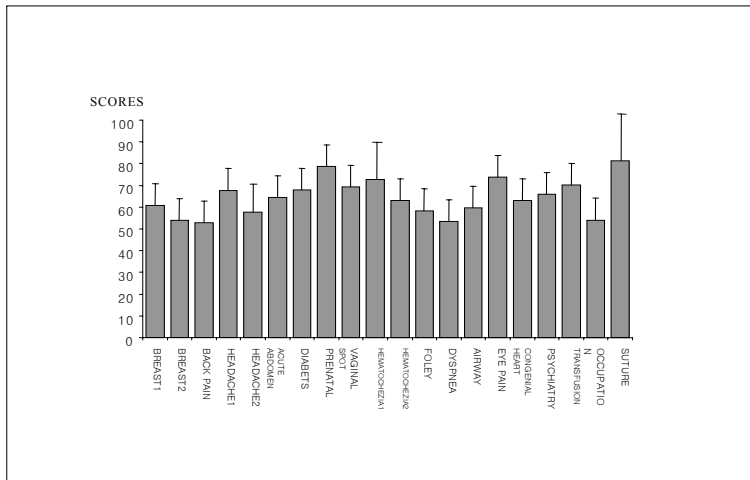


Fig. 1. OSCE Station Scores

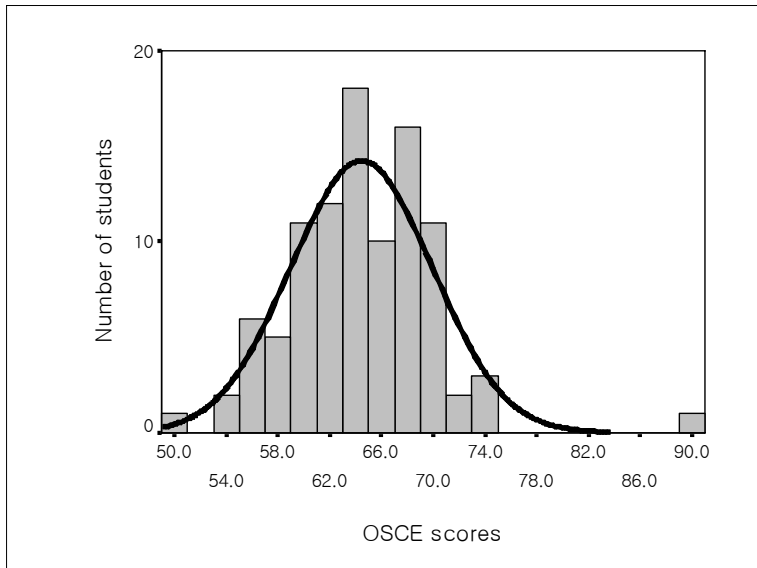


Fig. 2. Distribution of OSCE Scores

차이가 없었다 (Fig. 3).

출제 교수와 동료교수 또는 전임의가 평가를 시행하였으며 시험장을 복제하였기 때문에 제 1평가자와 제 2평가자로 나뉘어 평가를 시행하였는데, 문항 출제교수는 주로 제 1평가자로서 평가를 하였고 동료교수 혹은 전임의가 제 2평가자로서 평가를 하

였다. 제 1평가자 시험장 학생들의 OSCE 평균점수는 64.6 (범위, 55.4-90.4)점이고 제 2평가자 시험장 학생들의 OSCE 평균 점수는 63.6 (범위, 50.3~73.3)점으로서 유의한 차이는 없었다 ($p=0.16$). 그러나 문항별로 살펴보면 (Fig. 4) 전체 20문항 중 9개 문항에서 유의한 차이를 보였다 ($p<0.05$).

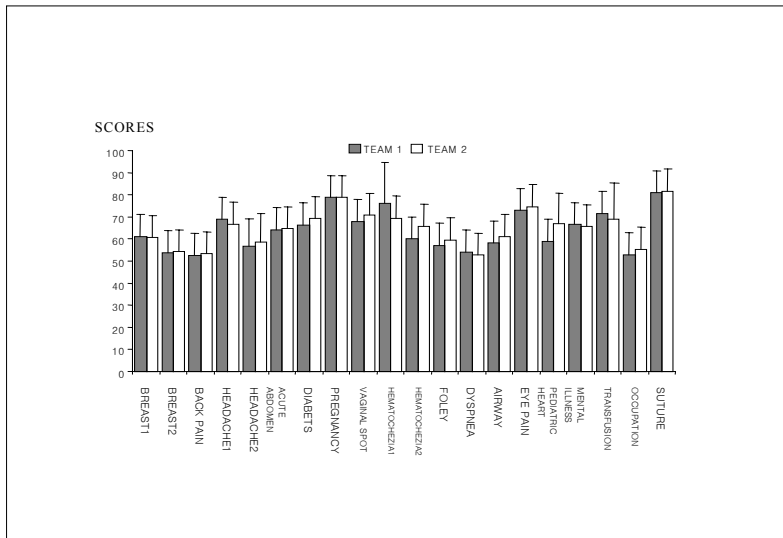


Fig. 3. OSCE Station Scores by Duplicated Examinee Teams

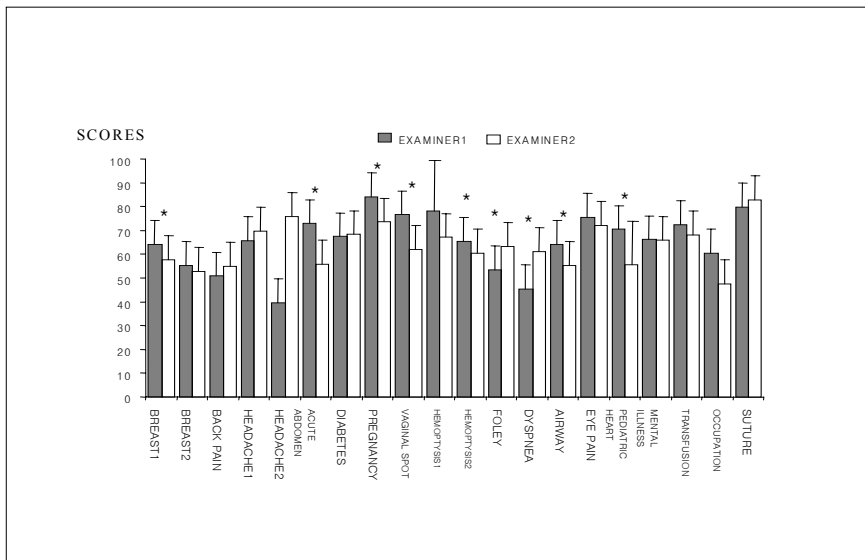


Fig. 4. OSCE Station Scores by Duplicated Evaluators. (* p<0.05)

다. 사용도구 및 측정 분야별 성적

학생 선·후발팀 여부와 평가자에 따른 차이를 통제하고 OSCE 사용도구 및 측정 분야별로 점수가 어떻게 다른지를 알아보기 위하여 반복측정 분산분

석을 시행하였다. 사용도구는 모형, 일반 SP, 전문 연기자 SP로 나누어 분석하였고 측정분야는 문진, 이학적 검사, 처치 및 술기, 교육으로 나누어 분석을 하였다 (Table II). 사용도구에 따른 전체 학생의 점수로 볼 때 일반SP의 점수가 다른 경우에 비해 낮

Table II. OSCE Scores according to the Tools and Categories by Duplicated Teams and Examiners mean(standard deviation)

	Total	Team		Evaluator	
		Team 1 (n=48)	Team 2 (n=50)	Examiner 1 (n=49)	Examiner 2 (n=49)
Tools*					
Plastic Model	67.0 (7.4)	65.9 (6.1)	68.1 (8.3)	69.3 (6.4)	64.7 (7.6)*
General SP	60.8 (5.9)	60.1 (6.2)	61.4 (5.5)	59.2 (4.9)	62.4 (6.3)*
Special SP	66.9 (6.1)	66.5 (6.6)	67.3 (5.5)	69.7 (4.3)	64.1 (6.3)*
Categories*					
History Taking	66.6 (5.7)	65.8 (6.1)	67.3 (5.1)	68.5 (3.8)	64.7 (6.6)*
Physical Exam	61.6 (6.8)	60.9 (6.9)	62.2 (6.6)	61.1 (6.6)	62.1 (6.9)
Skill	67.4 (7.6)	66.9 (6.7)	67.9 (8.4)	67.4 (7.3)	67.4 (8.1)
Education	58.5 (8.3)	56.9 (8.9)	60.0 (7.6)	60.3 (6.1)	56.7 (9.8)*

* p<0.05 by repeated measures ANOVA

Table III. Correlation Coefficients between the OSCE Score by Tools or Categories and the Performance Grade

	Lecture Score	
	Cumulative	Year 4
Total	0.531	0.583
OSCE Tools		
Plastic Model	0.412	0.333
General SP	0.494	0.512
Specialized SP	0.433	0.431
OSCE Categories		
History Taking	0.428	0.476
Physical Exam	0.493	0.434
Skill	0.326	0.256
Education	0.429	0.455

All items are at p<0.05 by pearson correlation

았으며 측정 분야별로는 이학적 검사와 교육 부분이 다른 두 분야에 비해 점수가 낮았다. 선·후발팀의 OSCE 도구 및 측정 분야별 차이는 없었으나 평가자에 따른 차이는 OSCE 도구 전체와 측정 분야 중 문진과 교육에서 유의한 차이를 보였고 이학적 검사와 처치 술기 분야는 유의한 차이가 없었다.

라. OSCE 점수와 학교성적과의 상관성

전체학교 성적과 OSCE 사용 도구 및 측정 분야와의 상관성을 알아본 결과 대부분 중등도의 양의

Table IV. Correlation Coefficients between the OSCE Station Scores and the Performance Grade

OSCE Stations	Lecture Score	
	Cumulative	Year 4
Breast 1	0.287*	0.313*
Breast 2	0.214*	0.255*
Back Pain	0.292*	0.328*
Headache 1	0.244*	0.208*
Headache 2	0.102	0.073
Acute Abdomen	0.298*	0.252*
Diabetes	0.212*	0.322*
Pregnancy	0.264*	0.146
Vaginal Spot	0.218*	0.226*
Hemoptysis 1	0.146	0.226*
Hemoptysis 2	0.415*	0.420*
Foley	0.113	0.101
Dyspnea	0.012	-0.015
Airway	0.180	0.117
Eye Pain	0.332*	0.298*
Pediatric Heart	0.276*	0.256*
Mental Illness	0.187	0.182
Transfusion	0.308*	0.303*
Occupation	0.146	0.203*
Suture	0.121	-0.065

* p<0.05

상관성이 있었으나 전학년 누적 성적과는 처치 기술 분야에서, 4학년 성적과는 모형과 처치 기술분야에서 상관성이 낮게 나타났다 (Table III).

전체 학교 성적을 OSCE 문항별로 분석해 본 결과 유방 진찰, 요통 진찰, 안통 병력 청취, 수혈 방법, 폐암 CT 소견 등에서 중등도의 상관관계를 보였고 요도관 삽입, 정신 질환 병력 청취, 뇌질환 방사선 검사의 판독, 기관 삽입, 봉합술 등에서 낮은 상관성을 보였다 (Table IV).

고 찰

OSCE 문항 출제를 위하여 내과 3명, 일반외과 2명, 소아과 2명, 산부인과 2명, 그 외 정신과, 신경과, 가정의학과, 예방의학과, 비뇨기과학, 응급의학과, 진단검사의학과, 마취과학, 안과학, 신경외과학, 방사선과학, 성형외과학, 정형외과학, 이비인후과학, 흉부외과학에서 각 1명 씩의 교수가 위촉되었다. 시험 대상 학생들이 모든 임상 실습을 다 마친 상황이므로 Major 과목 뿐 아니라 Minor 과목의 임상 실습의 문항을 전부 내도록 하였으며 측정 분야는 Harden 등 (1979)이 제안한 비율에 맞도록 전체 회의를 통하여 정하였다. 최종적으로 20개 문항의 선정은 문항의 완성도, 증례의 빈도수 및 수행 수월성 등을 고려하여 정했는데 의과 대학을 졸업하는 학생들이 필수적으로 수행할 수 있어야 하는 소위 'exit objectives'의 임상 형태와 기술 범주에 맞추어 각 문항의 완성도를 높였으며 (Scott, 1991) 황건 등(2001)이 OSCE 예비연구에서 제안한 42개 항목에서 크게 벗어나지 않음을 발견하게 되었다. 문항의 신뢰도를 구하기 위하여 Cronbach- α 를 구하였는데 전체 평균점이 0.74로 나와 문항 전체로서는 어느 정도 신뢰성을 보였다 (박훈기 등, 1998; Colliver와 Williams, 1993; Vu 등, 1992).

이전에는 SP로서 모집한 병원 직원들을 훈련시켜 시행하였으나 올해에는 한양의대에서 SP로 훈련된 전문 연기자를 5문항에 투입하였으며 (복제이므로 총 10명) Table II에서 보이듯이 이들과 일반 SP (간호사, 조교, 기사, 용역 직원 등)가 투입된 9개 문항 및 모형용 사용한 6개 문항과 비교하여 학생들의 점수에 미치는 영향을 알아본 결과, 모형 및 전문 연기자 SP를 사용한 문항에 비해 일반 SP의 문항이 점수가 낮았다 ($p < 0.001$). 물론 전문 연기자 SP의 문

항은 7개의 문진 문항 중에 4개였고 다른 하나는 이학적 검사 (급성 복통) 문항이었으며 일반 SP는 나머지 2개의 문진 문항과 다른 분야가 골고루 들어있었으므로 (Table I) 직접적인 비교는 무리가 있지만 문진 문항이 SP의 전문성에 더 의존함을 고려할 때 전문 연기자 SP의 문항이 일반 SP 문항에 비해 좋은 점수를 받은 점 및 시험 후의 학생들의 설문에서 만족도가 더 높은 점 등을 비교할 때 다른 문헌들의 주장과 마찬가지로 전문 SP의 필요성을 보였다고 할 수 있다 (김병수 등, 2001; 이병국, 2000; 최윤호 등, 2000; Barrows, 1993). 측정 분야 별로는 처치 술기가 가장 높았고 이학적 검사가 비교적 낮았으며 환자나 보호자에 대한 교육 혹은 의사소통을 측정하는 문항에서 제일 점수가 낮아서 실제 임상 실습 및 교육에서도 이학적 검사나 의사소통 훈련에 비해서 문진이나 도구를 이용하여 특정한 처치 술기의 훈련에 치우치는 것으로도 사료되었다 (Table II).

총 98명을 대상으로 토요일 오후에 만나절에 걸쳐 OSCE 시험을 다 치루게 하기 위한 설계로 시험 장소를 두 군데로 나누어 복제를 하였고 20 문항을 채택하였는데 6개의 문항을 3쌍의 연관 문항 (Linked Station)으로 만들었기 때문에 휴식 스테이션이 불가피하게 늘어서 총 28개의 스테이션으로 운영을 하여야 했다. 연관 문항 스테이션을 운영해 본 결과, 4~5분 정도 길이의 OSCE 문항 특성 상 측정하고자 하는 분야의 확실한 제시가 있는 점은 각기 독립적으로 문항을 구성하더라도 학생들에게 문항을 이해시키기에 충분한 듯 하다. 즉 단순히 같은 질병이나 주소이기 때문에 연관 문항으로 만들어 측정분야를 달리 하는 것이 꼭 학생들의 이해도, 문항 준비의 수월성, 혹은 문항의 완성도를 높이는 것은 아니라고 생각한다. 더구나 연관 문항으로 만들면 두 문항을 연결시키기 위하여 SP상황 및 시험 조건들을 일치시키는 노력이 필요하고 또한 필연적으로 휴식 스테이션이 늘기 때문에 전체 시험 시간이 길어지는 단점을 고려할 때 특별히 이로운 점은 없는 것 같았다. 시험장의 복제와 두 개의 팀으로 운영하여야 하는 이유로 평가자와 측정 도구의 복제가 필요했는데 선, 후발에 따른 점수의 차이는 유의한 차이가 없었으며 (Fig. 3) 복제 시험장에서의 문제점

중에 평가자의 일치도가 제일 어려운데(박훈기 등, 1999) 본 시험의 결과에서도 사용 도구별로 구분하였을 때 모든 도구에서 제 1 및 제 2평가자의 점수를 비교할 때 유의한 차이를 보였다. 다만 측정 분야별로 비교할 때 문진과 교육은 유의한 차이를 보였으나 이학적 검사와 처치술기 분야에서는 유사한 평가를 한 것을 볼 때 평가의 기준이 보다 명백히 눈에 보일 수 있는 분야인 이학적 검사나 처치 술기의 문항보다 문진이나 교육 등 주로 언어 소통과 내용을 평가해야 하는 분야에 보다 평가자의 일치를 위한 훈련과 연습이 필요하다고 할 수 있다. 한편 평가자에 따른 두 복제 장소간에 전체 점수에는 유의한 차이가 없었으나 문항 별로 분석을 하면 유의한 차이를 보였다. 따라서 다른 문항과 비교하여 평가자의 일치를 위한 노력이 더 필요한 문항을 객관적으로 제시해 줄 수 있는 이런 문항별 분석표나 그래프가 복제 스테이션의 운영 후 평가에 매우 유용할 것으로 사료된다 (Fig. 4).

임상능력을 측정하기 위하여 사용하는 평가도구는 타당성이 높고 신뢰성 및 실용성을 갖추어야 하며 평가로만 끝나지 않고 학습을 증진시키는 교육적 의미를 지녀야 한다고 할 때 의학 교육 과정 중의 여러 단계에서, 또한 각 교실이나 임상과 별로 다양한 형태의 OSCE를 시행함으로써 교육적 효과를 높일 수 있다고 사료된다 (Davis, 2003).

2003년도에 모든 임상 실습을 종료한 본과 4학년 학생들을 대상으로 이화의대에서 실시한 OSCE 시험 방법은 시험장 복제에 따른 평가자의 일치도에 문제가 있었지만 학생 및 교수에게 커다란 동기를 부여하고 임상 실습을 포함한 의학 교육 과정 및 방법에 있어서 새로운 발전의 계기로 삼을 수 있었다.

참 고 문 헌

김병수, 이영미, 안덕선, 박정률(2001). 임상의학입문(Introduction to Clinical Medicine) 평가를 위한 객관적임상실기시험(Objective Structured Clinical Examination)경험. *한국의학교육*, 13, 289-298
박훈기, 김동원, 김덕언, 최호순, 김경태(1998). 의학

과 4학년 종합평가로서의 객관적-구조적 임상능력평가(OSCE)의 경험. *한국의학교육*, 10, 43-56.
박훈기, 이정권, 김승룡, 김경태, 박해영(1999). 시험장 복제(Duplication)가 객관적 구조적 임상 시험(OSCE)의 신뢰도에 미치는 영향. *한국의학교육*, 11, 37-52
이병국(2000). 임상수행평가의 세계적 추세에 따른 표준화 환자 프로그램의 토착화 방안 연구. *한국의학교육*, 12, 377-392.
이영미(2003). 의학교육에서 OSCE의 적용 현황. 보건의료인 국가시험의 발전방향 모색; 임상실기평가. *제5회 학술세미나*, 한국보건의료인국가시험원
최윤희, 손희정, 이준혁 등(2000). 임상실습 교육을 위한 표준화 환자 프로그램의 개발 및 적용. *한국의학교육*, 12, 353-361.
황건, 이영미, 백상호(2001). 의사시험에 임상수기수행평가를 도입하기 위한 예비연구. *한국의학교육*, 13, 277-287.
Barrows HS(1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Acad med*, 68, 443-451.
Colliver JA, Williams RG(1993). Technical issues: test application. *Acad Med*, 68, 454-463.
Davis MH(2003). OSCE: the Dundee experience. *Medical Teacher*, 25, 255-261.
Harden RM, Gleeson FA(1979). ASME Medical Education Booklet No. 8: Assessment of clinical competence using an objective structured clinical examination (OSCE). Dundee.
Scott CS, Barrows HS, Brock DM, Hunt DD(1991). Clinical behaviors and skills that faculty from 12 institutions judged were essential for medical students to acquire. *Acad. Med*, 66, 106-111.
Vu NV, Barrows HS, Marcy ML, et al(1992). Six years of comprehensive, clinical, performance-based assessment using standardized patients at the Southern Illinois university school of medicine. *Acad Med*, 67, 42-49.