

객관적 임상실기시험(Objective Structured Clinical Examination)을 이용한 가정의학과 임상실습 학생의 임상수행능력 평가

고려대학교 의과대학 가정의학교실, *의학교육학교실
김수현 · 이도행 · 이영미* · 최윤선 · 조경환 · 홍명호

요 약

연구배경: 의과대학 실습 학생의 임상수행능력은 체계적이고 객관적인 평가가 필수임에도 불구하고 필기시험 혹은 교원의 주관적인 기준에 의해 평가되어 온 것이 사실이다. 이에 가정의학과에서는 임상실습 학생을 대상으로 일차의료에 필요한 임상수행능력을 습득하였는지를 평가하는 데 객관적 임상실기시험(OSCE)을 시행하였고, 실시한 OSCE의 신뢰도와 타당도를 평가하였다. 아울러 운영상의 문제점을 파악함으로써 전반적인 임상실습 평가 도구로의 적용 가능성을 알아보려고 하였다.

방법: 39명의 가정의학과 임상실습 학생 중 지원자 26명을 대상으로 4주 간격으로 2회에 걸쳐 OSCE를 실시하였다. 시험은 10개의 시험실로 구성되었고 시험 시간 5분, 이동 시간 1분이었다. 시험실은 병력청취 3개 영역, 신체검사 3개, 기술적 술기 3개 및 면담기술 1개 영역으로 구성되었다.

결과: 1차 시기에 11명, 2차 시기에 15명의 학생이 응시하였다. OSCE 총점은 평균 65.6 ± 5.63 이었으며 전체 문항의 신뢰도 계수(Cronbach's α)는 0.615였다. OSCE 총점은 평가 시기별로는 유의한 차이를 나타내지 않았으나 평가자가 바뀐 두 시험실(발열 병력청취 및 혈압측정)에서는 유의한 차이가 나타났다. OSCE 총점과 기존의 2회의 필기시험 점수와의 상관은 0.56, 0.54로 통계적으로 유의하였지만, 각 평가 영역별 점수와 기존 필기시험 점수의 상관은 신체검사 영역에서만 0.42로 통계적으로 유의하였다. 시험 운영상 큰 문제점은 나타나지 않았으며, 참가 학생들은 임상수기 평가도구로서의 OSCE에 대해 긍정적인 반응을 보였다.

결론: 본 연구에서 시행하였던 객관적 임상실기시험(OSCE)은 가정의학과 임상실습에서의 수기능력 평가와 실습 교육에 대한 되먹임 효과가 있으나, 중등도의 신뢰도를 보여 전반적인 임상수행능력 평가를 위해서는 더 높은 신뢰도와 타당도의 확보를 위해 향후 지속적 연구가 필요할 것으로 생각된다. (가정의학회지 2002;23:583-592)

중심단어: 객관적 임상실기시험(OSCE), 임상수행능력, 신뢰도, 타당도, 가정의학과 임상실습

서 론

의과대학의 교육 목표는 학생들이 의과대학 졸업

교신저자: 홍명호

Tel: 02-828-6276, Fax: 02-837-3966

E-mail: gurof@netian.com

후 일차료를 수행할 수 있는 능력을 양성하는 데 있다.^{1,2)} 따라서 이러한 임상수행능력은 임상 전 교육에서 학습된 의학 지식을 바탕으로 임상실습을 통해 실질적인 경험을 쌓음으로써 습득이 가능하다. 그러나 현재 임상실습의 대부분은 3차 병원 중심의 입원 환자를 대상으로 시행되고 있어 일차의료에 적합한 임상수행능력을 학습하는 데는 부적절

하다. 따라서, 가정의학과 임상실습은 일차의료의 자질을 갖춘 졸업생 배양이라는 의과대학 교육 목표 달성에 있어서 실질적으로 가장 중요한 위치를 차지한다.

고려대학교 의과대학 임상실습과정에서 가정의학과는 선택 과목으로, 기간은 1주일(5일)을 차지한다. 가정의학과 의 임상실습에서는 전공의가 모델이 되어 직접 가르치는 흉·복부 진찰, 모형을 통한 유방 및 골반 신체검사, 모형을 이용한 창상봉합 및 정맥 채혈 등을 통해 수기를 교육하며, 표준화 환자를 이용하여 면담기술이나 병력청취, 신체 검사 등의 임상수기를 가르치고 평가하고 있다.

이러한 임상실습에서의 학생들의 수행능력에 대한 평가는 암기 위주의 지식을 평가하는 필기 시험 일변도의 관행을 탈피해야 한다는 것에 의학 교육자들이 동의하지만, 가정의학과뿐 아니라 대부분의 과에서 학생들의 임상수행능력을 객관적으로 평가하지 못한 것이 현실이다.

객관적 임상실기시험(Objective Structured Clinical Examination: 이하 OSCE로 약칭)은 여러 개의 임상상황을 가정하여 각 시험실에서 주어진 시간 내에 피험자들의 수행능력을 평가하는 실기시험이다. 이는 의과대학 학생들에 있어서 병력청취나 신체검사, 임상술기, 자료 해석 능력, 면담기술 등의 임상수행 전반에 걸친 능력을 객관적이고 체계적으로 평가하는 방법으로 인정받고 있다. Harden이 1975년 최초로 소개한 이후^{3,4)}, 그 신뢰도와 타당도가 검증되어 외국 여러 나라에서 의과대학생 혹은 전공의 평가에도 사용되고 있다.^{5,6)} 우리나라에서도 일부 대학에서 졸업시험의 하나로 OSCE를 시도한 경험은 있지만^{7,8)}, 가정의학과에서 임상실습 교육을 평가하는데 적용된 적은 없었다. 본 연구에서는 가정의학과 실습 학생들을 대상으로 임상수행능력을 평가하기 위하여 OSCE를 실시하여, 운영상의 문제점을 찾고 OSCE의 신뢰도와 타당도를 평가하였다. 또한 이러한 경험과 되먹임을 통해 학생들의 임상실습 교육의 질을 향상시키고, 전반적인 임상실습 평가 도구로서의 유용성을 알아보고자 하였다.

방 법

1. 문항개발 및 진행

1) OSCE 개발팀 구성 및 증례 개발: 2000년 3월 가정의학과 전문의 2명과 전공의 1명, 성형외과 의사 1명, 의학교육학 전문가 2명으로 구성된 OSCE 개발팀을 구성하여 5월 18일과 6월 15일 양일간에 걸쳐 시험을 시행하기로 결정하였다. 피험자는 1999년 2학기부터 2000년 1학기 사이에 가정의학과 임상실습을 선택한 39명의 학생 중 자발적으로 시험에 지원하도록 하였고, 시험 당시 본과 4학년 1학기 과정 중이었다. 지원자는 총 26명이었으며 2일의 시험 날짜 중 본인이 원하는 날짜를 선택하도록 하였다.

시험실(Station)의 수는 10개로, 각 시험실당 시험시간은 5분, 다음 시험실로 이동하기 위한 시간 1분을 주어 총 60분 동안 시험이 이루어지는 형태를 취하였다. 시험 내용은 선행 연구의 설문조사에서 일차 진료에 반드시 필요한 임상수기로 정한 항목 중에서 중요도가 상위 15위에 해당하는 수기 가운데 중등도의 난이도를 갖는 9개 항목에 면담기술을 평가하기 위해 가정의학과 의 특수성을 고려하여 말기 암 환자의 호스피스 진료에 대한 항목을 추가하였다.

증례 시나리오, 채점표, 채점표를 위한 가이드라인 등 증례 개발은 내과 계열 문제는 가정의학과 의사가 외과계열 문제는 성형외과 전문의가 담당하였다. 증례 시나리오는 주어진 주제에 관한 가상의 시나리오로서 학생들이 질문할 수 있는 상황을 시나리오화 한 것이며, 채점표는 문항에 대한 시행 여부를 주로 '예' 혹은 '아니오'의 형태로 표시할 수 있도록 만들었고, 가이드라인은 채점의 일관성을 위해 채점 시에 주의할 사항이나 규칙, 혼돈을 줄 수 있는 부분에 관한 설명 등을 포함하였다. 개발된 증례를 OSCE 개발팀이 아닌 외부 전문가(내과 전문의와 일반외과 전문의)에게 보내 자문을 구하여 수정, 보완하였다.

2) 표준화 환자 및 평가자 훈련: 고려대학교 부속 병원 가정의학과 전공의를 표준화 환자로 훈련시켰으며 동시에 평가자로 활용하였다. 각 시험실에 한 명의 전공의가 할당되어 총 10명(1차 및 2차 시험)이 참여하였다. 참여한 전공의들의 대부분은 과거 학생

실습 과정에서 표준화 환자 역할을 한 경험은 있었으나 OSCE에 대한 경험은 없는 상태였으며, 표준화 환자 훈련을 위해 총 2회에 걸친 훈련을 받았다. 첫번째 훈련에서는 OSCE의 목적, 구성 및 진행 등 전반적인 사항에 대하여 오리엔테이션을 시행하였고 증례 시나리오를 나누어주어 다음 훈련 모임 전까지 읽도록 하였다. 두번째 훈련은 증례를 개발했던 2명의 의사가 전공의를 일 대 일로 만나 개별적으로 시행하였다. 표준화 환자는 시나리오에 맞게 연기하며 학생들의 질문 이외에는 미리 말해주지 않도록 하였고, 시나리오에 나와있지 않은 질문에 대해서는 가능한 한 '모른다' 혹은 '아니다'로 대답하게 하였다. 또한 평가를 위해 채점표 및 채점을 위한 가이드라인을 나누어주고 채점 방법에 대하여 교육하였다.

8개 시험실은 1차와 2차 시험에 동일한 전공의가 참가하였으나 2개의 시험실(발열 병력청취, 혈압측정)은 평가자 간 차이가 시험에 미치는 영향을 평가하기 위해 2차 시험에서 다른 전공의가 표준화 환자와 평가자 역할을 하였다.

3) 모의시험: 시험 2일 전 OSCE 개발팀과 표준화 환자 및 평가자로 참여하게 될 전공의들이 모여 모의 시험을 시행하였다. 수련의 2명이 피험자가 되어 5개 시험실(창상봉합, 정맥채혈, 기침 병력청취, 복통 병력청취, 복통 신체검사)에 대하여 시험 당일과 동일한 조건에서 예비시험을 한 결과, 시간이나 진행 사항에 대해서는 큰 문제점이 발견되지 않았다. 문제점이 발견된 창상봉합, 정맥채혈, 복통 병력청취 등 시험실에 대하여 시나리오와 채점표를 일부 수정하였고 수정된 내용에 대하여 재교육을 시행하였다.

4) 시험 당일 준비: 시험 장소는 현실감과 순환의 용이성 등을 고려하여 외래 진료실을 이용하였고, 편의상 진료가 끝난 평일 저녁을 이용하였다. 시험 시작 30분 전까지 평가자와 학생들을 집합시키고 20분에 걸쳐 시험 목적, 방법, 주의사항 등에 대하여 교육하였다. 각 방에는 비디오 카메라를 설치하여 평가자가 놓치는 부분이나 객관성에 대하여 점검할 수 있게 하였다.

2. 통계 분석 방법

각 시험실 점수 및 총점, 기존 필기시험 점수를

100점 환산 점수로 변환하여 분석에 이용하였다. 각 문항의 신뢰도는 문항의 내적 일관성 측정을 위해 Cronbach's α 값을 구하였고, 1차와 2차 시험 시기간의 평균점수의 비교는 t검정을 통해 분석하였다. 또한 타당도 측정을 위해 OSCE 총점 및 각 시험실 점수와 기존 필기시험 점수(2학년 2학기 필기시험 점수, 3학년 1학기 필기시험 점수)와의 상관성을 분석하였다. 상관성 분석은 피어슨 상관분석을 통하였고, 유의수준은 $P < 0.05$ 로 하였다. 이상의 모든 통계적 처리는 Window용 SAS (8.1 version)를 사용하여 분석하였다.

결 과

1. 피험자의 선수 과목 이수 분포

피험자는 1999년 2학기부터 2000년 1학기 사이에 가정의학과 임상실습을 선택한 39명의 학생 중 지원자 26명이었다. 시험 문항의 내용 구성으로 보아 주요 임상과를 모든 이수한 후에 시행하는 것이 이상적이지만, 지원한 학생들에게 두 가지 시험 날짜 중 본인이 가능한 날짜를 선택하도록 하였기 때문에 선수 임상실습과목에 차이가 있었다. 주요 임상 과목인 내과, 외과, 산부인과, 소아과, 정신과 및 가정의학과를 모두 이수한 학생은 13명이었으며 가정의학과는 전체 중 76%인 19명만이 이수하였고, 소아과는 92%의 학생이 이수하였으며, 나머지 4과목은 80% 이상의 학생에서 이수한 것으로 나타났다.

2. 문항의 내용 및 측정 영역

병력청취 3개, 신체검사 3개, 기술적 숙기 3개, 면담기술 1개로 총 10개 시험실을 구성하였고 문항의 내용 및 측정 영역은 Table 1과 같다.

3. 문항별 성적

전체 문항의 100점 환산 평균 점수는 65.6 ± 5.63 이며 발열에 대한 병력청취 시험실의 평균점수가 가장 낮았고 유방 진찰을 하도록 한 시험실의 평균점수가 가장 높았다(Table 2).

Table 1. Contents and clinical skill areas of the each station.

Station	Content	Clinical skill area	Data	No. of check-list items
1. Measurement of the blood pressure	Measurement of the blood pressure	TS*	SP	10
2. Cough - history taking	Hx of 25-year-old male with cough	Hx [†]	SP	13
3. Abdominal pain - history taking	Hx of 27-year-old male with acute abdominal pain	Hx	SP	10
4. Breast physical exam	Palpation & description of the given breast mass	PEX [‡]	model	12
5. Venous puncture	Peripheral venous blood sampling	TS	model	10
6. Cough - physical exam	PEX of the patient with cough	PEX	SP	15
7. Fever - history taking	Hx of 26-year-old female with fever	Hx	SP	18
8. Hospice	Interview with 40-year-old female who has terminal breast cancer	IS [§]	SP	14
9. Primary closure	Aseptic technique & suture of the given wound	TS	model	11
10. Abdominal pain - physical exam	PEX of the patient with abdominal pain	PEX	SP	12

*TS: Technical skill

† Hx: History taking

‡ PEX: Physical examination

§ IS: Interview skill

^{||} SP: Standardized patient

Table 2. Mean scores of the each station.

		No. of students	Min.	Max.	Mean	SD
History taking	Cough	26	30.77	84.62	55.62	16.03
	Abdominal pain	26	50.00	90.00	74.62	11.40
	Fever	26	11.11	77.78	47.01	19.20
Physical examination	Cough	26	30.00	100.00	66.73	18.54
	Abdominal pain	26	55.00	95.00	76.15	9.93
	Breast mass	24	50.00	100.00	77.78	14.56
Technical skill	Blood pressure	26	30.00	90.00	55.00	20.05
	Primary closure	26	25.00	83.33	60.26	12.11
	Venous puncture	26	40.00	90.00	68.85	12.11
Interview	Hospice	26	42.85	92.85	71.43	13.70
Total		24	49.12	74.42	65.60	5.63

4. 문항의 신뢰도

OSCE 문항의 내적 일관성 계수(Cronbach's α)는 0.615로 나타났다. 각 영역별로 신뢰도를 살펴보면, 보통 병력청취 및 신체검사 영역에서는 Cronbach's α 가 0.000 및 0.020으로 매우 낮았다. 또한 정맥채혈 영역 역시 Cronbach's α 가 0.222로 신뢰도가 낮은 것으로 확인되었다(Table 3).

5. 시험 시기별 피험자의 평균점수 비교

2000년 5월 18일에 시험을 실시한 1차 평가 집단과 6월 15일에 시험을 실시한 2차 평가 집단 사이에 평균점수의 차이가 있는지를 t검정을 한 결과 OSCE 총 점에 있어서는 1차 평가 집단과 2차 평가 집단 간에 통계적으로 유의한 차이가 나타나지 않았다. 그러나, 발열 병력청취와 혈압측정의 경우 시험 시기에 따라서 통계적으로 유의한 차이가 나타났다($P < 0.01$) (Table 4). 구체적으로 살펴보면, 발열 병력청취 영역에서의 평균점수는 1차 평가 집단이 28.3, 2차 평가 집단은 60.7로 2차 평가 집단에서 더 높은 점수를 나타내었다. 또한 혈압측정 영역에서는 1차 평가 집단이 73.6, 2차 평가 집단은 41.3로, 2차 평가 집단이

더 낮은 점수를 보였다.

6. 평가 영역별 평균점수

평가 영역을 병력청취, 신체검사, 기술적 수기, 면담기술의 4개 영역으로 구분하였다. 병력청취 영역

Table 3. Reliability of the each station.

	Station	Cronbach's α
History taking	Cough	0.485
	Abdominal pain	0.000
	Fever	0.771
Physical examination	Cough	0.669
	Abdominal pain	0.020
	Breast mass	0.595
Technical skill	Blood pressure	0.687
	Primary closure	0.467
	Venous puncture	0.222
Interview	Hospice	0.582
Total		0.615

Reliability is measured by Cronbach's α .

Table 4. Comparison of the score between the first and second session.

		Mean (SD)		t-value	P-value
		Session 1 (n=11)	Session 2 (n=15)		
History taking	Cough	58.01 (18.32)	53.85 (14.54)	0.65	0.52
	Abdominal pain	73.64 (12.06)	75.33 (11.26)	-0.37	0.72
	Fever	28.28 (9.11)	60.74 (11.01)	-7.97	0.00
Physical examination	Cough	63.64 (16.14)	69.00 (20.37)	-0.72	0.48
	Abdominal pain	77.27 (11.70)	75.33 (8.76)	0.48	0.63
	Breast mass	73.11 (16.60)	82.05 (11.70)	-1.54	0.14
Technical skill	Blood pressure	73.64 (11.20)	41.33 (12.46)	6.81	0.00
	Primary closure	58.33 (19.36)	61.67 (17.48)	-0.46	0.65
	Venous puncture	73.64 (12.06)	65.33 (11.26)	1.80	0.08
Interview	Hospice	67.53 (14.42)	74.29 (12.89)	-1.26	0.22
Total		64.71 (6.72)	66.36 (4.67)	-0.71	0.49

by t-test.

에는 기침, 복통, 발열 병력청취의 3개 영역 점수를, 신체검사 영역에는 기침, 복통, 유방 진찰의 3개 영역 점수를 합산하였다. 또한 기술적 수기영역에는 혈압측정, 창상봉합, 정맥채혈의 3개 영역 점수를 합산하였고, 면담기술 영역의 점수는 호스피스 영역 점수를 포함시켰다.

각 영역별 평균점수는 신체검사 영역에서 74.0±9.48점으로 가장 높았고, 병력청취 영역에서 59.1±10.08점으로 가장 낮았다.

7. OSCE 점수와 기존 필기시험 점수와의 상관관계

OSCE 총점과 2학년 2학기 및 3학년 1학기 필기시험 점수와의 상관관계는 각각 0.56과 0.54로 나타났으며, 이는 유의수준 0.01에서 통계적으로 유의하였다. 또한 각 평가 영역별 점수 및 OSCE 총점과의 상관성을 살펴보면 OSCE 총점과 병력청취, 신체검사 영역 점수와의 상관관계는 각각 0.71, 0.82로 높게 나타났다($P <$

0.01). 기술적 수기 영역 및 면담기술 영역 점수와의 상관관계는 통계학적으로 유의하지 않았다. 각 평가 영역별 점수와 기존 필기시험 점수와의 상관관계는 신체검사 영역과 2학년 2학기 필기시험 점수와의 상관관계가 0.42로 통계학적으로 유의하였으나($P < 0.05$), 다른 영역 점수는 유의한 상관관계를 보이지 않았다(Table 5).

8. 설문 결과

시험이 끝난 직후 학생들에게 시험 내용의 적절성, 시험의 전반적 사항 및 진행에 대한 만족도 등에 대한 설문조사를 실시하였다.

1) 시험 내용의 적절성: 각 시험실의 내용과 임상실습 경험과의 연관성을 묻는 질문에 대하여 창상봉합술이 가장 낮은 점수를 보여 전 임상실습에서 많이 다루어지지 않았던 영역으로 파악되었고 다른 9개 항목은 중간 정도의 반응을 보였다. 시험실의 내용이 의사로서 반드시 습득해야 할 것인가에 대하여 학생

Table 5. Correlations between OSCE scores and the previous written exams.

Pearson correlation coefficient	History taking	Physical exam	Technical skill	Interview skill	Total OSCE score	Score of Year 2	Score of Year 3
P-value							
History taking	1.000						
Physical exam	.541 [†]	1.000					
Technical skill	-.344	-.068	1.000				
Interview skill	.262	.230	-.380	1.000			
Total OSCE score	.713 [†]	.828 [†]	.190	.363	1.000		
Score of Year 2 [‡]	<.0001	<.0001	.375	.082	.560 [†]	1.000	
Score of Year 3 [§]	.330	.416 [*]	.183	.254	.560 [†]	.939 [†]	1.000
	.100	.043	.372	.210	.005		
	.262	.390	.223	.239	.536 [†]	<.0001	
	.200	.060	.274	.239	.007		

* $P < .05$, by Pearson's correlation.

[†] $P < .01$, by Pearson's correlation.

[‡] Score of Year 2: Score of the 2nd semester of year 2.

[§] Score of Year 3: Score of the 1st semester of year 3.

들은 10개 항목 전부에 대하여 긍정적으로 평가하였다. 표준화 환자나 의학모형 등 자료에 대해서는 10개 시험실 모두에 긍정적인 반응을 보였고 대체적으로 표준화 환자를 사용한 시험실에 대한 만족도가 의학모형을 사용한 시험실에 비하여 더 높았다. 각 시험실에서 문제를 해결하는 데 필요한 시간에 대해서는 대체적으로 중간적인 반응을 보였다. 문제의 난이도에 대하여 혈압측정과 정맥채혈의 기술적 수기 영역이 다른 영역에 비하여 어렵지 않았던 영역으로 평가되었다.

2) 시험의 전반적인 사항 및 진행에 대한 만족도: 학생들은 시험 진행에 대하여 전반적으로 원활하였던 것으로 평가하였고 시험 소요시간, 전체적인 시험 문항의 구성에 대하여 긍정적으로 응답하였다. 61.6%의 학생이 다른 시험보다 정신적 부담이 컸다고 응답하였다. 이 시험을 대비하여 준비를 했는지의 여부에 있어서는 21명(80.8%)이 준비하지 않은 것으로 응답하였고 시험에 대한 사전 정보가 있었는지의 여부에 있어서는 '그렇다'와 '아니다'의 반응 비율이 비슷하였다. 전체 학생 중 85.8%가 시험이 흥미 있었다고 응답하였고 다른 과목으로의 확대에 대해서는 46.1%의 학생이 긍정적인 반응을 보였다. 시험이 졸업 후 의사 생활을 하는 데 도움이 될지에 대해서는 19명(73.1%)의 학생들이 도움이 될 것으로 응답하였다.

고 찰

1975년 Harden 등이 임상수행능력을 평가할 수 있는 시험의 형태로 개발한 OSCE는³⁴⁾ 그 신뢰성 및 타당도가 여러 연구에서 입증되어 세계 각국으로 전파되었으며, 미국에서는 의과대학생 및 전공의 임상수행 평가방법에도 널리 사용되고 있고⁵⁾, 캐나다에서는 MCCQE II (Medical Council of Canada Qualifying Examination part II)라는 명칭하에 국가 시험의 일부로서 시행하고 있으며⁶⁾, 유럽에서도 널리 이용되고 있다.

1960년대 중반 Barrow와 Abrahamson이 개발한 표준화 환자를 OSCE와 같은 다시험실 임상실기시험(multi-station clinical examination)에 도입함으로써

시험의 객관성과 신뢰성이 높아졌고, 표준화 환자를 통한 되먹임(feedback)은 구성평가(formative assessment)로 이용되어 학생들의 수기 교육 향상에 큰 역할을 하고 있다.

학생들의 임상수행능력은 여러 분야에 걸쳐 있어 한 가지의 평가 방법이나 평가자의 주관적 기준에 의해서는 적절하게 평가되기 어렵다. 하지만 아직까지 가정의학과뿐 아니라 대부분의 임상 과에서 학생들의 임상수행능력 평가는 객관적이고 체계적으로 이루어지지 않았다. 그러나, 최근 몇 년 사이에 일부 국내 의과대학에서 OSCE를 도입하여 임상수행능력을 평가하고 있다.^{7,8)} 고려대학교 의과대학 가정의학과에서는 임상실습 학생들에게 표준화 환자를 이용하여 병력청취, 신체검사, 면담기술 등을 평가하고 비디오 촬영을 통해 되먹임 교육함으로써 학생들에게 긍정적 반응을 보여왔으며, 본 연구에서는 이 경험을 바탕으로 학생들의 가정의학과 임상실습 평가에 OSCE를 시도하였다.

본 연구에서 10문항의 신뢰도는 Cronbach's α 값이 0.615로 국내의 한 연구에서 보고한 0.14⁷⁾보다는 높았으나 외국의 OSCE 경험자가 제시하는 신뢰도의 기준인 0.8⁹⁾보다 낮았다. 일반적으로 OSCE의 신뢰도에 영향을 미치는 요소는 시험 시간, 질, 평가자와 응시자에 대한 교육, 시험조건 등이다.⁹⁾ 본 연구에서 문항의 신뢰도가 낮은 이유로는 첫째, 시험 시간이 짧고 문제수가 적었던 것을 생각할 수 있다. 둘째, 보통 병력청취 및 신체검사 영역의 신뢰도가 거의 0에 가깝게 나온 것이 전체 시험의 신뢰도에 부정적인 영향을 주었다. 이 영역의 신뢰도가 낮은 이유는 채점표를 너무 단순화하였거나 평가자의 일관성이 부족했던 가능성 등을 생각할 수 있다. 이를 보완하기 위해서는 보다 체계화된 세부 문항의 개발과 표준화 환자 및 평가자에 대한 좀 더 세심하고 체계적인 훈련이 필요하다. 기존의 연구에서 OSCE를 졸업시험이나 국가시험과 같은 자격시험에 이용하기 위해서는 최소한 20개 시험실 또는 4~6기간의 시험 시간이 요구된다고 하였다.^{9,10)} 본 연구에서의 신뢰도가 학생들의 당락을 결정하는 자격 시험에 이용되기에는 부족한 수준이었지만 몇 가지 문제점들을 보완한다면 가정의학과 혹은 전체 임상실습의 성취도 평가 및 되

먹임 교육에 이용하기에는 충분한 정도라고 생각된다.

시험 시기에 따라 1차 평가 집단과 2차 평가 집단 간의 평균점수를 비교한 결과 총점에서는 양 군 간에 유의한 차이가 없었다. 그러나, 발열 병력청취 및 혈압측정 영역에서는 양 군 간에 유의한 차이가 나타났다. 동일한 전공의가 표준화 환자와 평가자 역할을 하였던 타시험실에서는 시험 시기에 따라 시험 성적에 유의한 차이가 없고 전공의가 바뀌었던 발열 병력청취와 혈압측정의 경우에서만 시기에 따른 평균점수의 차이가 있는 것으로 볼 때 표준화 환자 및 평가자의 표준화를 위한 체계화된 훈련이 필요함을 시사하는 소견이라고 할 수 있다. 시험점수의 변이 요인으로는 응시자, 평가자, 시험문항 자체(표준화 환자의 연기 포함)의 요인 등을 생각할 수 있다. 본 연구에서는 응시자의 시험 전 능력에 관해서는 분석을 시행하지 않았다. 하지만 김병수 등¹¹⁾의 연구에서 응시자의 시험 전 능력은 OSCE 점수 결과 차이에 영향을 미치지 않는 것으로 나타났고, 박훈기 등¹²⁾의 연구에서도 시험장 사이의 평균점수 차이에 대해 평가자 간의 차이를 가장 중요한 요인으로 분석하고 있다.

평가 영역별 평균점수의 비교에서는 신체검사 영역의 점수가 가장 높았다. 본 시험에서는 해당 시기 시행 여부에 따라 “예” 혹은 “아니오”로 구분하는 이분법적 채점표를 이용하였다. 이러한 이분법적 채점표 체계는 채점이 용이하고 변이가 적어 병력청취나 기술적 시기 영역 평가에서는 문항을 세분화하여 시행 여부만을 평가해도 충분히 객관적인 평가가 가능하였다. 그러나, 신체검사 영역의 경우는 시행만 하면 정확도에 관계없이 동일한 득점을 하게 되기 때문에 본 연구에서 신체검사 영역의 채점표를 너무 단순화하여 신체검사수기의 평균점수가 오히려 다른 영역보다 상회하는 것으로 나온 것으로 추정된다. 신체검사 영역의 채점표 제작 시에는 이분법적 척도보다는 ‘정확하게 시행’, ‘시행하였으나 부정확함’, ‘시행하지 않음’과 같이 채점표를 보다 세분화하는 것이 더 바람직할 것으로 생각된다.

OSCE의 타당도를 보는 방법에는 몇 가지가 있다. OSCE가 중요한 임상수행능력 평가 영역을 전반적으로 다루고 있는가를 평가하는 내용타당도(content validity), 임상수행능력을 측정하는 다른 평가 방법

이나 다른 영역에서의 성취도와 OSCE 성적을 비교하는 준거타당도(criterion-related validity), 학생 초기나 전공의 초기의 성적과 후기의 성적을 비교하거나 저학년과 고학년의 성적을 비교하는 구성타당도(construction validity) 등이 대표적인 방법이다.^{13,14)} 내용타당도의 경우 본 연구에서 OSCE는 선행 연구에서 일차의료에 필요한 임상수기 상위 15위에 해당하는 시기 가운데 9개를 선정하였으며, 임상수기능력 평가에 기존의 필기시험보다는 우수할 것으로 생각되지만, 문항수가 적어 임상실습 전반에 걸친 수행능력 평가에는 부족한 점이 있다. 준거타당도의 경우 미국의 가정의학과 및 소아과 임상실습 학생의 OSCE 성적과 교원의 주관적 점수, 필기시험과의 상관성이 매우 낮게 나온 몇몇 연구가 있다.^{13,15)} 본 연구에서는 학생들의 2학년 2학기 및 3학년 1학기 필기 시험 점수와 OSCE 점수와의 상관성을 조사하였는데 두 학기 성적 모두 OSCE 총점과는 통계적으로 유의한 상관성을 보였다. 각 영역별 점수는 신체검사 영역과 2학년 2학기 성적에서만 통계적으로 유의한 상관성을 보였고, 그 외의 영역에서는 상관계수가 0.4 이하로 매우 낮았다. 신체검사 영역에서의 유의한 상관성은 필기 시험 점수가 높은 학생이 신체검사 능력도 뛰어나다고 말할 수 있지만 이는 앞서 말한 바와 같이 신체검사 능력을 세부적으로 평가하지 못했다는 문제점이 있다. 결과적으로 기존의 필기시험 성적이 좋은 학생은 전체 OSCE 점수, 즉 전반적인 임상수행능력이 뛰어난 경향이 있다고 말할 수 있지만, 병력청취, 기술적 술기, 면담기술 등이 각각 뛰어나다고 말할 수 없다.

학생들의 설문 결과 OSCE 시행에 대해서 전반적으로 긍정적인 반응을 보였지만 시험 전 많은 부담을 느꼈으며, 국가시험 등에서의 도입에 관해서는 부정적인 반응을 보였다.

본 시험에는 1회 시험당 표준화 환자와 평가자 역할을 시행한 10명의 전공의, 전체적인 진행에 교수요원 3명이 참여하였고, 학생 오리엔테이션을 포함하여 26명의 학생이 치르는 데 총 130분이 소요되었다. 기술적 수기에 필요한 모형은 기존 임상실습 교육에 사용하였던 장비를 사용하였고 표준화 환자를 맡았던 전공의들에게 약간의 인건비를 지급하였다.

본 연구는 임상실습 후 학생 평가도구로서 객관적 임상실기시험(OSCE)의 적용 가능성을 검토하기 위한 예비연구로서 시행되었으며, 대상수가 적고 문항 제작 및 평가자 훈련이 미흡하였다는 문제점이 발견되었다. 이러한 문제점들을 수정, 보완한다면 기존의 인력과 자원을 활용하여 적은 비용과 시간으로 가정의학과뿐만 아니라 전체 임상실습 평가 및 졸업 전 임상종합평가 등에 확대 적용할 수 있으며 나아가 학생들과 교육자 모두에 대한 되먹임을 통해 임상실습 교육의 향상을 가져올 수 있을 것으로 생각한다.

참 고 문 헌

1. 박재복, 박정환. 한국의과대학의 교육목적과 교육 목표의 내용분석. *한국의학교육* 1996;8:145.
2. 홍창희, 배직현, 이정신. 새로운 의과대학 교육과정 시안. *한국의학교육* 1990;2(1):43-8.
3. Harden RM, Mary S, Wilson DW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ* 1975;1:447-51.
4. Harden RM, Gleeson FA. Assessment of clinical competence using objective structured examination. *Med Edu* 1979;13:41-54.
5. Amitai Z, Miriam FB, Alton IS, Nancy EG. Lessons learned from six years of international administrations of the ECFMG's SP-based clinical skills assessment. *Acad Med* 1998;73:84-91.
6. 안덕선. The Medical Council of Canada Qualifying Examination Part II 연구. *한국의학교육* 1997;9:65-72.
7. 박훈기, 김동원, 김덕연, 최호순, 김경태. 의학과 4학년 종합평가로서의 객관적-구조적 임상능력 평가(OSCE)의 경험. *한국의학교육* 1998;1:43-57.
8. 서보양, 이두진, 권광보, 강복수. 객관적으로 구조화된 임상시험의 시행 경험. *한국의학교육* 1998;2:363-81.
9. Jerry AC, Reed GW. Technical Issues: Test Application. *Acad Med* 1993;68:454-60.
10. Shatzer JH, Darosa D, Colliver JA, Barkmeier L. Station-length requirements for reliable performance-based examination scores. *Acad Med* 1993;68:224-9.
11. 김병수, 이영미, 안덕선, 박정률. 임상의학입문(Introduction to Clinical Medicine)평가를 위한 객관적임상실기시험(Objective structured Clinical Examination)경험. *한국의학교육* 2001;13(2):289-98.
12. 박훈기, 이정권, 김승룡, 김경태, 박해영. 시험장 복제(Duplication)가 객관적 구조적 임상 시험(OSCE)의 신뢰도에 미치는 영향. *한국의학교육* 1999;11(1):37-52.
13. Mastell DG, Wolfish NM, Hsu E. Reliability and validity of the objective structured clinical examination in paediatrics. *Med Edu* 1991;25:293-9.
14. Hilliard RI, Tallett SE. The use of an Objective Structured Clinical Examination With Postgraduate Residents in Pediatrics. *Arch Pediatr Adolesc Med* 1998;152(1):74-8.
15. Michael D, Camile F, Fitzpatrick NP, Desiree L, Mark G, Stephen R, et al. Use of an objective structured clinical examination in evaluating student performance. *Fam Med* 1998;30(5):338-44.

Abstract

Assessing the Clinical Competence of Family Medicine Clerkship Students Using the OSCE

Soo Hyun Kim, M.D., Do Haeng Lee, M.D., Young Mee Lee, M.D.*
Youn Seon Choi, M.D., Kyung Hwan Cho, M.D., and Myung Ho Hong, M.D.

Departments of Family Medicine and *Medical Education College of Medicine, Korea University

Background: Although medical students' clinical competence should be assessed systematically and objectively, written exams or faculty's subjective methods were used in the past. We assessed the clinical competence of the family medicine clerkship students using the OSCE and evaluated the reliability and the validity of the OSCE. The purpose of this study was to find the logistic problems and the feasibility of the OSCE to an entire clerkship evaluation.

Methods: Twenty six volunteers took part in the OSCE. Ten, 5-minute long stations, were divided into 4 areas: 3 for history taking, 3 for physical exam, 3 for technical skill, and 1 for interview skill areas.

Results: Eleven students were examined on first session and 15 students on second session 4 weeks later. The mean OSCE score was 65.5 ± 5.63 . The reliability of the total stations was 0.615. The total OSCE score was not different by the sessions, but there was a significant difference in the two stations where the examiners had been changed. There were statistically significant correlations between the total OSCE score and the two previous written exams ($r=0.56, 0.54, P<0.01$). There was no significant logistic problem and the examinees showed positive responses to the OSCE.

Conclusion: The OSCE was a useful tool for assessing the clinical competence of family medicine clerkship students. The reliability of the OSCE was moderate and in order to extend the OSCE to an entire clerkship evaluation, we need further study to acquire a higher reliability and validity. (*J Korean Acad Fam Med* 2002;23:583-592)

Key words: OSCE, clinical competence, reliability, validity, family medicine clerkship