

일개 의과대학에서 실시한 표준화환자 (SP)를 이용한 임상수행능력평가지험 (CPX) 결과 분석

순천향대학교 의과대학 산업의학과, 산부인과¹, 소아과²

김주자 · 이경재 · 최규연¹ · 이동환²

= Abstract =

Analysis of the Evaluation for Clinical Performance Examination using Standardized Patients in one Medical School

Joo Ja Kim, MD, Kyung Jae Lee¹, MD, Kyu Yeon Choi², MD, Dong Whan Lee³, MD

Department of Occupational Medicine¹, Obstetrics and Gynecology², Pediatrics³

Purpose: Analysis of the evaluation for clinical performance examination (CPX) using standardized patient (SP) in one medical school was done to compare the evaluations between the staffs and SPs, among the staffs, and among the SPs, and to investigate the correlation between the evaluation scores and the school record.

Methods: Clinical performance examination was done to seventy-nine 4th grade medical students. Checklist was modified from the existing reference and added more items. Three SPs who were non-physician volunteers were trained by authors about what is the disease and how to act the patient and how to evaluate the students. Three staffs and three SPs both filled up the checklist at the same time right after each student's performance.

Results: Agreements between the staffs and the SPs were relatively high enough in the items of physical examination and management but low in the history taking related items although they were statistically significant. In the evaluation between the staffs and the SPs, the mean scores in the management and overall attitude were not significantly different, but, mean scores in the history taking, physical examination and overall skill were significantly different. Among the staffs there were no significant differences in history taking, physical examination, overall attitude, and there were significant differences in patient management and overall skill. Among the SPs there were no significant differences in physical exam, patient management and overall skill, and there were significant differences in history taking and overall attitude. The correlation coefficients between the evaluation scores and school records, between the evaluation scores and written report were relatively low respectively.

Conclusion: To increase the agreement rates between the staffs and SPs, more training for the staffs and SPs are necessary

Key Words: Clinical performance examination, Standardized patients, Evaluation

교신저자: 이경재, 순천향대학교병원 산업의학과
서울 용산구 한남동 657
Tel: 02)709-9449, Fax: 02)796-9025
E-mail: leekj@hosp.sch.ac.kr

서 론

표준화환자란 환자 아닌 사람을 훈련시켜 실제 환자처럼 가장해서 의과 대학생이나 전공의를 면담케 하여 이들의 임상수행능력을 평가할 목적으로 개발된 것이다.

표준화환자 프로그램은 1960년대 미국 남가주대학교 의과대학 교수팀을 중심으로 연구 개발되었으며 (Barrows, 1964), Harden (1975)이 제안한 객관적-구조화 임상시험 (Objective Structured Clinical Examination)과 함께 학생수행 평가시험으로 많은 의과대학에서 이용하고 있다. 캐나다는 1992년도부터 본격적으로 국가고시에 도입하여 성공적으로 평가되고 있으며 (안덕선, 1997) 의과대학에서는 이 시험에 통과되어야만 졸업을 시키고 있다. ECFMG 등 (Sutnick, 1993; Ziv 1998)에서도 외국의과대학 졸업자의 임상수행능력을 평가하기 위하여 임상수행실기시험 (Clinical Skill Assessment)을 개발하여 임상수행능력을 평가하는데 적합함을 입증하였다. 또한 1998년부터는 USMLE (United State Medical Licensure Examination)에 지원한 외국 의과대학졸업 수험생에게도 이를 시행하고 있다. 일본에서도 1994년도에 새로운 형태의 교과과정을 사용하는 신진 의과대학들이 연합체를 구성하여 OSCE를 실시하였다고 한다 (Kachur, 1996).

우리나라에서는 1996년 서울대학교 의과대학 외과학교실에서 미국 남가주 대학교 의과대학으로부터의 권고를 받아들여, 객관적-구조화 임상시험 (The objective structured clinical examination, OSCE)에 표준화환자 프로그램을 도입하게 되었다 한다 (이병국, 2000; 이병국, 2001). 그 후 여러 의과 대학에서 표준화환자를 이용한 임상수행능력시험을 임상실습 평가의 중요한 방법으로 계속 시도하고 있다 (박훈기 등, 1998; 서보양 등, 1998; 최윤희 등, 2000). 하지만 아직도 모의환자 또는 표준화환자, 임상수행능력 또는 임상실행능력, 임상실기능력 등 용어통일에 있어서도 의견일치가 이루어지지 않은 상태이다. 또한 편으로는 우리나라 의사국가시험 혹은 임상의학 입문 평가시험 등에도 임상수행능력 측정을 도입하

기 위한 예비연구 (이영미 등, 2001; 황건 등, 2001; 김병수 등, 2001)들도 시도되고 있다. 그런데 제도의 도입 뿐 아니라, 이미 외국에서는 평가에 있어서도 표준화환자에 의한 평가의 정확도를 보고하는 많은 연구가 있다 (Elliot, 1987; Vu, 1987; Tamblyn¹, 1991; Tamblyn², 1991; Vu, 1992;) 그러나 아직 우리나라에서는 임상수행능력 시험의 평가에 대한 연구는 많이 이루어지지 못하고 있다.

본 연구는 일개 의과대학에서 임상실습 및 문제 중심학습과정 (PBL)이 끝난 본과 4학년생을 대상으로 표준화환자를 이용하여 임상수행능력 평가시험 (CPX)을 실시하여 그 평가에 관하여 분석하고자 하였다. 즉, 첫째, 교수평가와 표준화환자의 평가를 비교하고 또 평가에 참여한 교수간 혹은 표준화환자간의 내적 차이가 있는가를 보고자 하였으며, 둘째, 학생들의 필기성적으로 대표되는 학업성적과 임상수행능력 평가시험과의 관련성을 알아보고자 했으며, 또한 학생들이 표준화환자시험 후 스스로 작성한 환자진단 및 관리에 관한 서면 보고서에 대한 교수의 평가점수와 수행능력 평가점수와 비교하여 그 관련성을 조사하였다.

대상 및 방법

가. 연구대상

2003년 6월 의과대학 본과 4학년 1학기 학생 79명을 대상으로 하였으며, 전 학생을 3개의 소그룹으로 나누어 3명의 산부인과 교수진과 3명의 표준화환자가 각 그룹에 할당되었다.

나. 증례 및 점검표 개발

시험은 일차진료에 반드시 필요한 임상적 수기능력 (적절한 병력청취, 진찰, 진단 및 관리 등), 환자와의 면담태도, 전반적 진찰술기로 나누어 평가하였다. 증례 및 점검표는 Reteguiz와 Cornel-Avendano (1999)의 'Mastering the OSCE/CSA'를 참고로 하여 번역하였고, 일부 문항들은 우리나라 현실에 맞게 수정한 임상적 수기능력에 관련된 24개의 문항을 택하여 '했다/안했다'의 이분법으로 채점하였다.

또 전반적인 환자 면담태도와 전반적 진찰술기에 대한 평가 등 2가지 항목을 추가하였다.

학생에게 주어진 증례는 다음과 같다. 「26세된 여자 대학원생이 심한 하복통으로 급하게 외래에 방문하였다. 그녀의 체온은 39.3℃였고 나머지 생징후는 정상이었다. 이 환자에 대한 중점적인 병력청취와 진찰을 시행하고, 가능성이 있는 진단명, 시행해야 할 검사들과 치료 등에 대하여 환자에게 간단히 설명하시오. 10분 내로 미혼인 김복자 여성을 진찰해 보시오.」

다. 표준화환자 훈련

시험에 동원되는 인력을 줄이고 경비를 절감하기 위하여 병원에서 근무하는 간호보조원 1명과 원목실에 근무하는 직원 2명을 표준화환자로 훈련시켰다. 1일 2시간씩 2회에 걸친 총 4시간의 훈련과정이 있었으며 환자 역할 및 학생 평가에 관하여 훈련을 하였다. 첫 번째 모임에서 임상실기시험의 목적 및 구성 및 진행 등 전반적인 사항에 대하여 오리엔테이션을 시행하고 증례 시나리오와 그 질병에 대한 내용을 나누어 주고 다음 훈련모임까지 읽어오도록 하였다. 두 번째 모임에서는 그 질환에 대한 시나리오대로 환자역할을 할 수 있도록 훈련을 실시하였다. 또 학생 평가를 위해 점검표가 적힌 채점표 및 채점을 위한 가이드라인을 나누어 주고 채점 방법에 대하여 교육하였다.

라. 시험장소

시험장소는 병원의 외래 진료실을 활용하기 위하여 토요일 오후에 실시하였다. 3개의 방이 준비되었고 각 방에는 교수 1명과 표준화환자 1명이 배치되었으며, 전체 학생들은 3 그룹으로 나누어 한 개의 방에만 들어가도록 하였다. 시험시간은 학생당 15분씩 할당되었다. 즉, 3개의 방에서 교수와 표준화환자는 방에서 각 학생은 시험장소 입구에 부착되어 있는 증례를 읽고 10분간은 환자 면담시간으로, 5분간은 면담결과를 정리하여 서면보고서(평가답안지)에 기록하는 시간으로 하였다. 이 때 교수와 표준화환자는 각각 점검표에 채점하였다.

마. 연구방법

각 학생의 환자면담이 끝난 후 교수와 표준화환자가 점검표에 있는 모든 항목을 채점하도록 한 후 이를 취합하여 분석하였다. 학생들의 임상수행능력 평가를 위하여 만들어진 점검표는 임상적 수기능력에 관련된 24개의 문항은 ‘했다=1, 안했다=0’으로 점수를 주었으며, 전반적인 환자면담태도와 전반적 진찰술기는 0~6점 중에서 점수를 주었다. 서면보고서는 교수만이 평가를 하였으며 5점 척도로 채점하였다.

분석은 임상적 수기능력 24개 문항 중 병력청취 14개 문항, 진찰 5개 문항, 진단 및 관리 5개 문항 등 3개의 항목으로 묶어 분석하였고 전반적인 환자와의 면담태도, 전반적 진찰술기에 대한 평가로서 임상수행능력 평가를 하였는데 이들에 대한 교수평가와 표준화환자 평가결과를 분석하였다. 또한 학생의 3학년 말 성적을 참고로 하여 임상수행능력 평가 점수 및 서면보고서 점수와의 관련성을 분석하였다.

바. 분석

통계분석은 SAS 6.12 통계 프로그램을 이용하였다. 교수 및 표준화환자평가에서 점검표에 열거된 문항을 실시한 정답률을 χ^2 -test로 비교하였으며, 항목별로 평균치 비교는 t-test, 교수간 혹은 표준화환자간의 비교는 one-way ANOVA, 학교성적과의 관련성은 피어슨 상관계수를 이용하였다.

결 과

가. 교수와 표준화환자의 평가성적 비교

교수와 표준화환자의 점검표 각 문항에 대한 평가에서 그렇다고 점수를 준, 즉 학생들이 환자에게 병력청취를 옳게 실시하였다고 생각한 비율은 교수, 표준화환자 각각 다음과 같다 (Table I). 통증의 위치 확인은 모두 98.7%, 통증의 시발시점 확인은 각각 96.2%, 100.0%, 통증의 특성 확인은 각각 82.3%, 83.5%, 통증의 강도 확인은 각각 38.2%, 54.4%, 통증의 완화인자 확인은 각각 21.5%, 19.0%, 통증의

Table I. Distribution of Correct Answers in Each Checklist Items by Staffs and Standardized Patients (SPs)

Checklist items	Correct answer(%)	
	Staffs	SPs
No. of subjects	79	79
<i>History taking (Did he/she ask about ...)</i>		
1. the location of the pain?	98.7	98.7
2. the onset of the pain ?	96.2	100.0
3. the quality of the pain?	82.3	83.5
4. the intensity of the pain?	38.2	54.4
5. any alleviating factors?	21.5	19.0
6. any aggravating factors?	40.5	58.2
7. any association with nausea or vomiting?	55.7	74.7
8. a cage in bowel movements?	51.9	72.2
9. any urinary complaints?	35.4	40.5
10. vaginal discharge?	38.0	40.5
11. sexual history?	70.9	58.2
12. any sexually transmitted diseases?	5.1	1.3
13. history of pregnancy?	69.6	68.4
14. use of condoms?	27.8	12.7
<i>Physical examination (Did he/she ...)</i>		
15. listen to the patient's abdomen with a stethoscope?	54.4	65.8
16. palpate the patient's abdomen?	92.4	92.4
17. attempt to elicit rebound tenderness?	86.1	84.8
18. try to elicit costovertebral angle tenderness?	25.3	27.8
19. ask to perform pelvic examination?	51.9	55.8
<i>Diagnosis and management (Did he/she ...)</i>		
20. discuss the initial impression with patient?	84.8	81.0
21. discuss the work-up for this problem?	78.5	79.7
22. explain that this could lead to infertility if not treated?	20.3	20.3
23. explain that this was a sexually transmitted disease?	24.1	16.5
24. explain that the sexual partner should be examined and treated?	26.6	39.2

악화인자 확인 각각 40.5%, 58.2%, 오심이나 구토에 대한 확인은 각각 55.7%, 74.7%, 장운동 또는 배변 습관의 변화에 대한 확인은 각각 51.9%, 72.2%, 배뇨 시 불편함에 대한 확인은 각각 35.4%, 40.5%, 질분비물 양상에 대해 문진은 38.0%, 40.5%, 성교력에 대한 파악은 각각 70.9%, 58.2%, 성병에 대한 확인은 각각 5.1%, 1.3%, 임신력에 대해 파악은 각각 69.6%, 68.4%, 콘돔 사용에 대한 확인은 각각 27.8%, 12.7%로 나타났다. 진찰조건에 관한 문항에서 청진여부에

관하여는 각각 54.4%, 65.8%, 복부촉진 여부는 각각 92.4%, 92.4%, 반사통 확인은 각각 86.1%, 84.8%, CVA tenderness 확인은 각각 25.3%, 27.8%, 내진 시행여부는 각각 51.9%, 55.8%로 나타났다. 진단 및 관리에 관한 문항에서 골반 내 감염성질환에 대한 논의 여부는 각각 84.8%, 81.0%, 정밀검사 시행에 대한 논의 여부는 78.5%, 79.7%, 불임 가능성에 대한 설명여부는 각각 20.3%, 20.3%, 성병의 설명여부는 각각 24.1%, 16.5%, 성적 파트너에 관한 설명여

Table II-1. Agreement Between Staffs and Standardized Patients (SPs) by Checklist Items

Checklist items	Staffs	SPs		Agreement test	
		yes	no	(k)	95% CI
<i>History taking (Did he/she ask about...)</i>					
1. the location of the pain?	yes	77	1	-0.01*	-0.03~0.00
	no	1	0		
2. the onset of the pain?	yes	76	0		
	no	3	0		
3. the quality of the pain?	yes	58	7	0.33*	0.06~0.60
	no	8	6		
4. the intensity of the pain?	yes	17	13	0.03*	-0.18~0.24
	no	26	23		
5. any alleviating factors?	yes	8	9	0.37*	0.13~.62
	no	7	56		
6. any aggravating factors?	yes	23	9	0.21*	0.01~0.41
	no	23	24		
7. any association with nausea or vomiting?	yes	41	3	0.44*	0.25~0.62
	no	18	17		
8. a change in bowel movements?	yes	37	4	0.38*	0.19~0.57
	no	20	18		
9. any urinary complaints?	yes	19	9	0.41*	0.21~0.62
	no	13	38		
10. vaginal discharge?	yes	25	5	0.68*	0.52~0.85
	no	7	42		
11. sexual history?	yes	38	18	0.29*	0.08~0.50
	no	8	15		
12. any sexually transmitted diseases?	yes	1	3	0.39*	-0.15~0.92
	no	0	75		
13. history of pregnancy?	yes	45	10	0.44*	0.22~0.65
	no	9	15		
14. use of condoms?	yes	8	14	0.39*	0.17~0.62
	no	2	55		

Total numbers of subjects are 79.

* p<0.05

부는 각각 26.6%, 39.2%로 나타났다. 즉, 문항에 따라 정답률이 차이가 있었다.

Table II-1 및 Table II-2는 교수와 표준화환자간의 평가의 일치도를 조사한 것으로 각 문항의 kappa 값과 95% 신뢰구간은 다음과 같다. 즉, 통증의 위치 확인 -0.01 (-0.03-0.00), 통증의 특성 확인은 0.33 (0.06-0.60), 통증의 강도 확인은 0.03 (-0.18-0.24), 통증의 완화인자 확인은 0.37 (-0.13-0.62), 통증의 악화인자 확인 0.21 (-0.01-0.41), 오심이나 구토에 대한 확인은 0.44 (0.25-0.62), 장운동 또는 배변습관의 변화에 대한 확인은 0.38 (0.19-0.57), 배뇨 시 불

편함에 대한 확인은 0.41 (0.21-0.62), 질 분비물 양상에 대한 문진은 0.68 (0.52-0.85), 성교력에 대한 파악은 0.29 (0.08-0.50), 성병에 대한 확인은 0.39 (-0.15-0.92), 임신력에 대해 파악은 0.44 (0.22-0.65) 콘돔 사용에 대한 확인은 0.39 (0.17-0.62)이었다. 진찰소견에 관한 문항에서 청진여부에 관하여는 0.71 (0.56-0.87), 복부촉진 여부는 0.82 (0.58-1.06), 반사통 확인은 0.64 (0.40-0.89), CVA tenderness 확인은 0.87 (0.75-0.99), 내진 시행여부는 0.87 (0.77-0.98)이었다. 진단 및 관리에 관한 문항에서 골반 내 감염성질환에 대한 논의 여부는 0.78 (0.59-0.96), 정밀

Table II-2. Agreement Between Staffs and Standardized Patients(SPs) by Checklist Items

Checklist items	Staffs	SPs		Agreement test	
		yes	no	(k)	95% CI
<i>Physical examination (Did he/she ...)</i>					
15. listen to the abdomen with a stethoscope?	yes	42	1	0.71*	0.56~0.87
	no	10	26		
16. palpate the abdomen?	yes	72	1	0.82	0.58~1.06
	no	1	5		
17. attempt to elicit rebound tenderness?	yes	64	4	0.64*	0.40~0.89
	no	3	8		
18 try to elicit CVA tenderness?	yes	19	1	0.87*	0.75~0.99
	no	3	56		
19. ask to perform pelvic examination?	yes	40	1	0.87*	0.77~0.98
	no	4	34		
<i>Management (Did he/she ...)</i>					
20. discuss the initial impression with patient?	yes	63	4	0.78*	0.59~0.96
	no	1	11		
21. discuss the work-up for this problem?	yes	57	5	0.58*	0.36~0.80
	no	6	11		
22. explain that this could lead to infertility if not treated	yes	11	5	0.61*	0.39~0.83
	no	5	58		
23. explain that this was a sexually transmitted disease?	yes	10	9	0.53*	0.31~0.76
	no	3	57		
24. explain that the sexual partner should be examined and treated	yes	21	0	0.72*	0.56~0.87
	no	10	48		

Total numbers of subjects are 79.

* p<0.05

검사 시행에 대한 논의 여부는 0.58 (0.36-0.80), 불임 가능성에 대한 설명여부는 0.61 (0.39-0.83), 성병의 설명여부는 0.53 (0.31-0.76), 성적 파트너에 관한 설명여부는 0.72 (0.56-0.87)이었다. 전체적으로, 중간정도 이하의 kappa 값을 보이는 병력청취 관련항목보다는 진찰조건과 진단 및 관리에 관한 항목에서 중간정도 이상의 좋은 일치도를 보여 전반적으로 일치도가 높게 나타났으며, 복부촉진 부분을 제외하고는 모두 통계적으로 유의하였다 (Fleiss, 1981).

Table III은 교수의 평가와 표준화환자의 평가를 체크리스트의 각 항목별로 평가점수의 평균치를 비교하였다. 병력청취에 관한 14개의 문항 평균치는 교수 7.3±1.7, 표준화환자 7.8±2.2로 통계적으로 유의한 차이를 보였으며, 진찰조건에 관한 5개의 문항 평균치는 교수 3.1±1.1, 표준화환자 3.3±1.1로 통계

적으로 유의한 차이를 보였다 (각각 p<0.05). 진단 및 관리에 관한 5개의 문항 평균치는 교수 2.3±1.3, 표준화환자 2.4±1.4로 통계적으로 유의한 차이를 보이지 않았다. 학생의 질문 태도에 대한 평가로는 6점 만점에 교수 4.2±1.2, 표준화환자 4.1±1.2로 유의한 차이를 보이지 않았으나 학생의 전반적 진찰술기에 대한 평가로는 6점 만점에 교수 3.6±1.5, 표준화환자 4.1±1.3으로 통계적으로 유의한 차이를 보였다 (p<0.01).

나. 교수간, 표준화환자간 평가성적 비교

3명의 교수간 평가비교에서 병력청취 14개 문항의 평가점수는 각각 7.7±1.8, 7.1±1.5, 7.2±2.0으로 통계적 유의한 차이가 없었다. 진찰조건 5개 문항의 평가점수는 각각 3.3±1.0, 3.0±1.1, 2.5±1.3으로 통계

Table III. Comparison of Scores Between Staffs and Standardized Patients (SP) by Categories of the Checklist for Evaluation

Categories	Staffs		SPs	t-test
	M±SD	M±SD	M±SD	
No. of subjects	79		79	
History taking (14)	7.3±1.7		7.8±2.2	p<0.05
Physical examination (5)	3.1±1.1		3.3±1.1	p<0.05
Management (5)	2.3±1.3		2.4±1.4	ns
Subtotal score (24)	12.8±2.8		13.4±3.0	p<0.05
Overall attitude[6]	4.2±1.2		4.1±1.2	ns
Overall skill[6]	3.6±1.5		4.1±1.3	p<0.01
Total score{36}	20.6±4.8		21.7±4.9	p<0.05

() indicates number of items and [] indicates score ranged form 0 to 6.

Table IV. Comparison of Scores Among the Staffs and Among the Standardized Patients (SP) by Categories of the Checklist for Evaluation

Categories	Staffs				p	SPs			p
	Staff A	Staff B	Staff C	SP1		SP2	SP3		
	M±SD	M±SD	M±SD	M±SD		M±SD	M±SD		
No. subject	26	27	26		26	27	26		
History taking(14)	7.7±1.8	7.1±1.5	7.2±2.0	ns	7.3±1.2	7.2±2.3	9.0±2.4	*	
Physical exam.(5)	3.3±1.0	3.0±1.1	2.5±1.3	ns	3.5±0.9	3.3±1.1	3.0±1.3	ns	
Management(5)	2.7±1.0	1.8±1.6	2.5±1.3	*	2.6±1.2	2.0±1.5	2.5±1.4	ns	
Subtotal(24)	13.7±2.5	11.9±2.8	12.7±2.8	ns	13.4±2.2	12.4±2.8	14.5±3.6	*	
Overall attitude[6]	4.2±1.2	3.7±1.3	4.8±0.9	ns	3.8±1.0	3.9±1.0	4.6±1.4	*	
Overall skill[6]	4.2±1.2	2.3±1.3	4.3±0.9	**	4.2±0.8	3.9±1.0	4.3±1.8	ns	
Total{36}	22.2±4.6	17.9±4.9	21.7±3.7	**	21.3±3.5	20.3±4.8	23.5±5.9	ns	

() indicates number of items and [] indicates score range form 0 to 6.

* p<0.05 and by one-way ANOVA

** p<0.01 and by one-way ANOVA

적 유의한 차이가 없었다. 진단 및 관리 5개 문항의 평가점수는 각각 2.7±1.0, 1.8±1.6, 2.5±1.3으로 유의한 차이를 보였다(p<0.05). 학생의 질문 태도에 대한 평가는 6점 만점에 각각 4.2±1.2, 3.7±1.3, 4.8±0.9로 유의한 차를 보이지 않았다. 학생의 전반적 진찰술기에 대한 평가로는 6점 만점에 각각 4.2±1.2, 2.3±1.3, 4.3±0.9로 통계적으로 유의한 차이를 보였다(p<0.01). 3명의 표준화환자간 평가비교에서 병력 청취 14개 문항의 평가점수는 각각 7.3±1.2, 7.2±2.3, 9.0±2.4로서 통계적 유의한 차이를 보였다(p<0.05). 진찰소견 5개 문항의 평가점수는 각각 3.5±0.9, 3.3±1.1, 3.0±1.3으로 통계적 유의한 차이가 없었다. 진단 및

관리 5개 문항의 평가점수는 각각 2.6±1.2, 2.0±1.5, 2.5±1.4로서 유의한 차이가 없었다. 학생의 질문 태도에 대한 평가는 6점 만점에 각각 3.8±1.0, 3.9±1.0, 4.6±1.4로 유의한 차를 보였다(p<0.05). 학생의 전반적 진찰술기에 대한 평가로는 6점 만점에 각각 4.2±0.8, 3.9±1.0, 4.3±1.8로 통계적으로 유의한 차이를 보이지 않았다(Table IV).

다. 임상수행능력 평가결과와 학교성적 및 보고서 점수와의 관계

학교성적과 임상수행능력과의 관계는 교수평가의

Table V. Correlation Coefficients of the CPX Score with the School Record or Written Report

	School record		Written report ¹⁾
	Staffs	SP	Staffs
	r	r	r
History taking	0.07	0.14	-0.05
Physical exam.	0.16	0.07	0.34**
Management	0.30**	0.32**	0.40**
Subtotal	0.25*	0.27*	0.30**
Overall attitude	0.22*	0.33**	0.23*
Overall skill	0.28*	0.31**	0.33**
Total	0.28*	0.32**	0.33**

¹⁾ written report was scored only by the staffs.

* p<0.05

** p<0.01

경우 병력청취, 진찰조건에서는 유의한 상관관계를 보이지 않았으며, 진단과 관리에서 상관계수 0.3으로 유의한 상관관계를 보였다 (p<0.01). 학교성적과 전반적 질문태도와의 관계는 상관계수 0.22로 유의한 상관관계를 보였으며, 학교성적과 전반적 진찰술기와의 상관계수는 0.28로 유의한 상관관계를 보였다 (각각 p<0.05). 표준화환자평가의 경우 병력청취, 진찰조건에서는 유의한 상관관계를 보이지 않았으며, 진단과 관리에서 상관계수 0.32로 유의한 상관관계를 보였다 (p<0.01). 학교성적과 전반적 질문태도와의 관계는 상관계수 0.33으로 유의한 상관관계를 보였으며, 학교성적과 전반적 진찰술기와의 상관계수는 0.31로 유의한 상관관계를 보였다 (각각 p<0.01). 서면보고서의 경우는 표준화환자가 평가할 수 없어 교수만이 평가하였다. 서면보고서 평가점수와 임상수행능력 평가점수와 관계는 병력청취에서는 유의한 상관관계를 보이지 않았으나, 진찰조건에서는 상관계수 0.34로 유의한 상관관계를 보였으며, 진단과 관리에서는 상관계수 0.40으로 유의한 상관관계를 보였다 (각각 p<0.01). 전반적 질문태도와의 관계는 상관계수 0.23로 유의한 상관관계를 보였으며 (p<0.05), 전반적 진찰술기와의 상관계수도 0.33로 유의한 상관관계를 보였다 (p<0.01) (Table V). 전체적으로 통계적으로 유의한 상관관계를 보였으나 실

제로 상관계수는 0.23~0.40의 낮은 값을 보인다고 하겠다.

고 찰

의과 학생이 어떻게 지식을 습득하고 그 지식을 어떻게 잘 적용해서 쓰느냐 하는 것은 별개의 주제로서 각기 다른 훈련을 요하며 평가 또한 다르다. 지금까지는 학생들을 평가하는 데 있어서 지식 평가가 주를 이루어 왔다. 이를 위한 보완방법으로 실습과정이 있으나 실질적으로 이에 대한 평가는 일방적이며 역시 지식전달 및 평가의 범주에서 크게 벗어나지 못하였으며 학생들이 무엇인가를 할 수 있는 능력 혹은 실행할 수 있음을 예측하는 평가는 되지 못하였다. 즉, 학생들의 임상능력에 관한 평가는 하지 못한 셈이다.

이러한 문제점들을 해결하기 위하여 최근에 시도 되는 방법으로 객관화-구조화된 임상시험 (OSCE), 임상수행능력 평가시험 (CPX) 등이 있다. 임상수행능력 평가는 지식에 대한 평가만이 아니며 실제 환자를 대면하면서 얼마나 잘 환자에게서 정보를 얻고, 문제를 판단하고 해결해 주는가에 대한 평가이다. 즉 지식, 태도, 술기를 동시에 평가할 수 있는 시험의 형태이다. 대부분 4학년 학생들에게 기본적인

면담술, 신체 진찰법 및 종합적인 문제해결 능력을 보는 정도까지 확장시키고 있다. 그런데 이는 주관적인 관찰에 영향을 받을 가능성이 많으므로 이를 객관화하는 것이 필요하다. 이를 위하여 환자의 객관화가 우선인데 표준화된 표준화환자(SP)를 이용하는 방법이 우리나라에서도 1996년도부터 의과대학에 도입되어 시도되고 있다. 표준화환자란 환자가 아닌 사람에게 환자인 양 훈련을 시키고 또 학생들을 평가하도록 훈련을 시키는 것이다. 표준화환자를 이용한 학생평가는 여러 가지 장점이 많다. 학생들이 환자와 의사소통하는 것을 직접 관찰할 기회를 제공하며 이들의 임상수행능력에 대하여 객관적인 평가가 어느 정도 가능하게 한다. 그 밖에도 표준화환자에 의한 임상수행능력 평가는 교수들의 시간절약 등 여러 장점이 있다. Abrahamson (1998)은 표준화환자를 이용한 임상수행능력 평가시험에 있어서 교수는 환자개발을 하고 rating 방법만 결정하면 되고 평가를 할 필요는 없으므로 교수의 시간을 절약할 수 있다고 하였다. 또한 평가에 있어서 무엇을 평가할 것인지 즉, 지식을 평가할 것인지, 지식을 사용하는 능력을 평가할 것인지 등을 명확히 할 것이며, 적절한 평가방법을 선택해야 하며, 대단히 조심스럽게 평가해야 한다고 하였다. 본 연구에서는 학생의 임상수행 능력을 임상적 수기(병력청취, 진찰, 진단 및 처치계획), 임상적 수기외적 요소(전반적인 환자와의 접근방법 및 의사소통능력)로 나누어 평가하였다.

그러나 본 연구의 목적은 교수평가와 표준화환자 평가에 차이가 있는가를 보고자 했으며, 또 시험에 참여한 3명의 교수간, 3명의 표준화환자간에 평가의 차이가 있는가를 보고자 한 것이다. 표준화환자가 정확히 환자역할을 하고 정확하게 학생들을 평가하고 기록할 수 있는가는 중요한 주제로서 많은 외국 연구들(Vu 등, 1987; Vu 등, 1992; Abrahamson, 1998)이 사전에 훈련을 받은 평가자들은 정확도를 충분히 유지할 수 있다고 보고하고 있다. Garman (1988)은 표준화 환자를 잘 훈련하면 90.2~93.4%의 정확도로 거의 교수만큼 정확할 수 있으며 test-retest 시에도 상당히 일관성이 있었다고 하였다. Tambllyn

등(1991)은 표준화환자가 환자를 묘사하는 정확도는 90% 이상이었다고 하였으며 85명 중 43명의 표준화환자가 95~100% 범위의 정확도를 보였다고 하였다. 또 표준화환자간의 평가일치도는 82%로 그 범위는 63~98%라고 보고하였다(Tambllyn 등, 1991). Elliot와 Hackam (1987)도 의사가 아닌 훈련받은 표준화환자가 임상술기를 평가하는 정확도를 83%로 보고하고 있다. 이병국 (2000)의 연구에서도 표준화환자와 의사간에 특정 사례에 대한 채점차이는 통계적 유의성이 없었다고 하며 표준화환자-채점자가 채점평가의사를 대신하여 OSCE의 채점평가자가 될 수 있다고 보고하였다.

본 연구에서는 교수평가와 표준화환자 평가와의 일치도는 항목별로 병력청취 관련항목에서는 kappa 값이 전반적으로 낮았으나 진찰소견, 진단 및 관리 항목에서는 비교적 높은 kappa값을 보여 주고 있다. 이는 임상수행능력 시험이 처음이었기 때문일 수도 있고 표준화환자에 대한 훈련이 부족했던 것으로도 판단된다. 표준화환자의 훈련을 좀 더 강화하고 보완하여 표준화환자에게 학생평가를 맡길 수 있도록 많은 준비가 필요하다. 우리나라에서 표준화 환자를 채점자로서 활용이 가능하려면 10시간 이상의 체계화된 훈련을 해야 한다고 하였는데, 본 연구에서는 표준화환자를 4시간 훈련시켜서 실제 시험에 응하게 하였다. 이렇게 이들 표준화 환자의 확보 및 훈련의 어려움 때문에 연수원 혹은 학회를 중심으로 한 컨소시엄의 방법을 제안하고 있다(이병국, 2001). 또 교수간의 평가비교에서는 각 항목별로 다른 결과를 보여주는데 병력청취, 진찰, 전반적 태도에서는 유의한 차가 있었다. 환자관리 및 전반적인 술기 능력은 교수간 차이가 통계적으로 유의하지 않았으나 다른 항목들은 유의한 차이를 보였다. 표준화환자간의 평가비교에서도 마찬가지로 각 항목별로 다른 결과를 나타내고 있는데 진찰항목과 진단 및 관리 항목과 전반적인 술기평가에서는 유의한 차가 없었다. 병력청취와 전반적인 면담태도에서 유의한 차이를 보였다. 이 역시 시험실시전 교수 및 표준화환자의 훈련에서 고려되고 개선되어야 할 부분이라고 생각된다.

본 연구의 또 다른 목적은 학생들의 임상수행능력과 학교성적과의 관련성, 서면 보고서의 평가점수와 관련성을 보고자 하였다. 교수평가와 표준화환자평가에 있어서 모두 학교성적과 임상수행능력평가 점수와의 관련성은 환자진단 및 관리와 전반적인 환자 면담태도, 전반적인 임상술기에 있어서 비록 통계적인 유의성을 보여 주었지만 상관계수가 낮아(0.22~0.33) 이 결과로 학교성적이 좋은 학생일수록 임상수행능력의 평가점수도 높게 받았음을 알 수 있다고 받아들이기에는 무리가 있다. 서면 보고서는 환자 즉, 표준화환자와 직접 면담 후 스스로 작성한 것이므로 환자 면접에서 긴장감으로 실력 발휘를 할 수 없었던 학생에게 자신의 실력을 표현할 수 있는 기회라 할 수 있는데 역시 병력청취를 제외한 대부분의 항목에서 통계적 유의성은 보였으나 상관계수가 낮으므로 연관성을 일반적으로 받아들이기에는 무리가 있다. 학생들이 다각도로 평가받을 수 있는 방법의 개선이 더 연구되어야 할 것이다.

특히 Table I에서 병력청취, 진찰, 진단 및 관리 항목별로 1차 단계에 해당되는 문항에서는 대개 정답률이 높았으나 감별진단 혹은 환자교육 등의 좀 더 다음 단계의 고려가 필요한 문항들에서는 비교적 낮은 정답률을 보이고 있다. 개별항목별 정답률이 낮은 항목들을 검토해서 학생들에게 부족한 부분은 다시 feedback 해서 학생교육 및 실습에 반영되어야 할 것이다. 임상 실습 시, 현실적으로 학생들이 책임을 지고 환자관리를 맡아서 임상수기, 태도를 직접 실습할 기회가 많지 않다. 그러므로 학생들의 임상수행능력을 높이고, 평가 시에도 학생들이 자연스럽게 실력 발휘할 수 있기를 기대하려면 모든 임상실습과목에서 충분한 학습이 이루어지도록 잘 계획되고 훈련되어진 제도가 필요하다. 김경한(2001)의 연구에서는 학생들이 실습 특과과목을 선택할 때 임상실습계획서를 참조하는 경우는 6%에 불과했으며 대부분이 학교선배나 친구들의 조언을 참고로 선택하는 것으로 나타났다. 또 학생이 중요하다고 보는 임상실습 평가 항목으로 출석과 태도를 가장 높게 꼽았으며, 임상술기를 배울 기회가 부족하다고 문제점으로 지적하였다. 학생들은 평가기

준이 큰 부분에 중점을 두게 되므로 실습목표와 평가기준을 먼저 제시한다면 실습에서 요구하는 방향에 맞추어 구체적인 방법을 스스로 준비할 것이다. 출석이 가장 중요하다고 생각할 때 출석만으로도 실습점수를 받을 수 있다면 이는 소극적인 학생들을 만들뿐이며 실제로 중요한 임상수행능력에 대한 훈련은 그 목적을 달성할 수 없다.

본 연구의 점검표에서 ‘했다/안했다’로 이분하여 평가하였기 때문에 점수를 세분화하지 못한 제한점도 있는데 즉, 교수나 표준화환자나 평가에 있어서 대개 좋은 점수를 주려는 경향이 있으므로 만약 점수를 세분화할 경우 이분적 평가보다는 학생들의 실력이 좀 더 반영되는 평가를 할 수 있을지는 목적에 따라 다를 수 있으며 이에 대한 연구가 좀 더 필요하다.

결론으로, 본 연구는 본과 4학년 학생들에게 임상수행능력시험 실시 후 교수와 표준화환자가 평가한 것을 비교 분석한 것으로 학생들 임상수행능력을 평가하는 시험방법을 향상시키기 위한 제도의 보완 및 발전을 위하여 활용하고자 하였다. 본 연구결과 앞으로 임상수행능력 평가에 있어서 보완점으로, 교수평가와 표준화환자평가의 일치도를 좀 더 높이고 또 교수간 및 표준화환자간의 평가 일치도와 신뢰도를 좀 더 높이기 위하여 교수 및 표준화환자의 훈련에 좀 더 충분한 시간을 할애하여야 하며, 표준화환자단독 평가만으로도 신뢰할 수 있도록 하려면 표준화환자의 연기 뿐 아니라 평가방법에 있어서도 충분한 훈련과 연구가 필요하다. 또한 본 임상수행능력평가를 위한 점검표의 내용이 대부분 질병진단 및 관리에 중점을 두고 있는데 나아가 환자 교육 및 상담술기에 대한 평가가 좀 더 강화된다면 학생들에게 그에 대한 훈련이 함께 이루어질 것이므로 학생들은 좀 더 풍부한 임상수행능력을 갖추게 될 것이다.

참 고 문 헌

- 김경한(2001). 임상실습시 선택과목제 시행 경험. *한국의학교육*, 13(2), 221-230.
- 김병수, 이영미, 안덕선, 박정률(2001). 임상의학입문

- (Introduction to Clinical Medicine) 평가를 위한 객관적 임상실기시험 (Objective Structured Clinical Examination) 경험. *한국의학교육*, 13(2), 289-298.
- 안덕선(1997). The medical Council of Canada Qualifying Examination Part II 연구. *한국의학교육*, 9, 65-72.
- 이병국(2000). 임상수행평가의 세계적 추세에 따른 표준화환자 프로그램의 토착화 방안 연구. *한국의학교육*, 12(2), 377-392.
- 이병국(2001). 우리나라 표준화환자 프로그램 활용 현황과 개선 방안 - 관련 문헌 고찰 및 현장 연구를 중심으로. *한국의학교육*, 13(1), 27-39.
- 이영미, 황 건, 백상호, 홍명호, 최윤선, 김수현(2001). 의사국가시험에 임상수행능력측정의 도입을 위한 객관적 임상실기시험 (Objective Structured Clinical Examination) 경험. *한국의학교육*, 13(1), 47-57.
- 박훈기, 김동원, 김덕언, 최호순, 김경태(1998). 의학과 4학년 종합평가로서의 객관적-구조적 임상능력 평가 (OSCE)의 경험. *한국의학교육*, 1, 43-57.
- 서보양, 이두진, 권평보, 강복수(1998). 객관적으로 구조화된 임상시험의 시행 경험. *한국의학교육*, 2, 363-381.
- 최윤호, 손희정, 이준혁, 정진상, 홍경표, 안병헌, 채수응, 서정돈(2000). 임상실습 교육을 위한 표준화환자 프로그램의 개발 및 적용. *한국의학교육*, 12(2), 353-361.
- 황건, 이영미, 백상호(2001). 의사시험에 임상수기수행평가를 도입하기 위한 예비연구. *한국의학교육*, 13(2), 277-287.
- Abrahamson, S(1998). 표준화 환자를 이용한 임상실기능력평가 워크샵 I. How to assess clinical competence. *한국의학교육*, 10(1), 153-164.
- Barrows H, Abrahamson S(1964). The programmed patient. *J. of Med Ed*, 39, 802-805.
- Elliot DL, Hickam DH(1987). Evaluation of physical examination skills: Reliability of faculty observers and patient instructors. *JAMA*, 258, 3405-3408.
- Fleiss JL(1981). *Statistical methods for rates and proportions*. John Wiley & Sons: New York.
- Garman(1998). 표준화 환자를 이용한 임상실기능력평가 워크샵. V. 1. How are rating scales are deloped? 2. Statistical scoring. *한국의학교육*, 10(1), 235-253.
- Harden RM, Stevenson M, Downie WW, Wilson GM(1975). Assessment of clinical competence using objective structured examination. *British Medical Journal*, 1, 447-451.
- Kachur EK, Lischka M, Das M, Ahn DS, Ban N, Swartz M(1996). Notes from the 7th Ottawa International Conference on Medical Education and Assessment. *Medical Encounter*, 12, 9-12.
- Reteguiz JA, Cornel-Avendano B(1996). *Mastering the OSCE/CSA*. McGraw-Hill Co: Health Professions Division.
- Sutnick AL(1993). ECFMG Assessment of clinical competence of graduates of fireign medical schools. *JAMA*, 270, 1041.
- Tamblyn1 RM, Klass DJ, Schnabl GK, Kopelow ML(1991). Sources of unreliability and bias in standardized patients rating. *Teaching and Learning in Medicine*, 3, 74-85.
- Tamblyn2 RM, Klass DJ, Schnabl GK, Kopelow ML(1991). The accuracy of standardized-patient presentations. *Medical Education*, 25, 100-109.
- Vu NV, Steward DE, Marcy ML(1987). An assessment of the consistency and accuracy of standardized patients' simulations. *J. of Medcial Education*, 62, 65-57.
- Vu NV, Marcy ML, Colliver JA, Verhulst SJ, Travis TA, Barrows HS(1992). Standardized patients' accuracy in recording clinical performance checklist items. *Medical Education*, 26, 99-104.
- Ziv Am Ben-David MF, Sutnick AI, Gary NE(1998). Lessons learned form six years of international administrations of the ECFMG's SP-based clinical skills assessment. *Aca Med*, 73, 84.