

진료수행시험 평가자로서 표준화 환자 활용의 타당성

가톨릭대학교 의과대학 의학교육학교실, 내과학교실¹, 정신과학교실²

김 선 · 박성환¹ · 허예라 · 이수정²

= Abstract =

The Appropriateness of using Standardized Patients' (SPs) Assessment Scores in Clinical Performance Examination (CPX)

Sun Kim, PhD, Sung-Whan Park¹, MD, Yera Hur, MA, Soo-Jung Lee², MD

*Department of Medical Education, Department of Internal Medicine¹, Department of Psychology²,
College of Medicine, The Catholic University of Korea, Seoul, Korea*

Purpose: The purpose of the study is to analyse the appropriateness of using standardized patients' (SPs) assessment scores of medical students in clinical performance examination (CPX).

Methods: 110 fourth year medical students in year four were divided into two large groups (group A and group B). Each group of students performed four different cases and overall, eight cases were tested. The examination were done by professors and the SPs using the same examination paper. Test scores were analysed per station and per each test domain. The differences between the two examiners were studied using paired t-test.

Results: There were significant differences in the scores given out by the professors' and the SP's in four out of the eight stations. According to the examination results of each domain, four cases showed significant differences in the history taking category, three cases in the physical examination, one case in the information sharing, four cases in clinical courtesy, and two cases in physician-patient relationship category. Cronbach's alpha scores of all stations were over 0.6, reflecting that the test items were appropriate for the examination.

Conclusion: From the result of this study and according to many preceding studies, using SPs' evaluation scores in CPX is appropriate. Some limitation of this study and ideas for improvement in using SPs in CPX are suggested.

Key Words: Standardized Patients, Clinical Performance Examination, Assessment

교신저자: 박성환, 가톨릭대학교 의과대학 내과학교실, 서울시 서초구 반포동 505번지
Tel: 02)590-1662, Fax: 02)3476-2274
E-mail: rapark@catholic.ac.kr

서 론

의학교육에서 표준화 환자를 활용한 교육 또는 평가는 외국의 경우 그 역사가 이미 30년이 넘었다. 미국의 경우 거의 모든 의과대학에서 표준화 환자를 사용하고 있다고 해도 과언이 아닐 정도이다. 이는 표준화 환자를 의학교육에 활용할 때, 교수들은 학생들이 환자에게 부적절한 말을 하거나 신체검사를 제대로 하지 못할까봐 걱정할 필요가 없고, 학생들도 익숙하지 않은 면담과 술기를 시행하면서 당황하거나 부끄러워할 필요가 없다. 또한 표준화 환자는 실제 환자와는 달리 원하는 시간과 장소에 활용될 수 있으며, 학생들의 수준에 맞고 난이도가 일정한 표준화 환자를 필요한 수만큼 만들어 낼 수 있다는 장점이 있다. 특히, 표준화 환자는 다양한 학생들의 질문과 태도에 따라 다양한 신체반응과 감정의 표현을 할 수 있고, 학생들의 임상술기에 대해 객관적인 피드백을 줄 수 있는 등 많은 장점이 있다 (최운호 등, 2000; Blue et al, 2000).

이러한 배경에서 표준화 환자 (standardized patient, SP)를 이용한 임상수행능력평가가 많이 이루어지고 있고, 그 대표적인 예로 객관구조화진료시험 (objective structured clinical examinations, OSCEs), 외국 의과대학 졸업생의 임상술기능력을 평가하기 위해 고안된 진료기능사정 (clinical skills assessment, CSA), 그리고 표준화 환자를 이용한 시험인 standardized patient examinations (SPEs) 등이 있다. 본 연구에서 사용된 진료수행시험 (clinical performance examination, CPX)의 특징은 각 스테이션 별로 표준화 환자를 배치하여 보다 현장감 있는 임상상황을 제공할 수 있다는 것이 가장 큰 장점이다. 따라서 진료수행 시험에서는 표준화 환자의 평가가 매우 중요한 역할을 한다. 표준화 환자와 관련된 과거의 많은 연구들은 표준화 환자의 평가에 대한 신뢰도를 측정하였으나, 최근에는 한 걸음 더 나아가서 표준화 환자 평가의 영역을 어느 정도로 할 것인가에 대한 연구를 많이 하고 있다 (Guiton et al, 2004; John et al, 2002; Solomon et al, 2000; Prislín et al, 1998 등).

한편, 우리나라의 상황을 살펴보면, 표준화 환자

를 이용한 임상수행능력평가는 1990년대 중반부터 시작하였고, 2004년 12월에 임상수행능력평가를 위한 서울지역 8개 의과대학 컨소시엄의 주최로 ‘제1차 CPX 심포지엄’을 개최하게 되었다.¹⁾ 이는 표준화된 환자뿐만 아니라 표준화된 시험 문항개발과 표준화된 평가 과정을 통해 일정 수준의 임상수행능력을 갖춘 의과대학생들을 길러내고자 하는 것이다. 또한 임상수행능력평가의 신뢰도와 타당도를 높이고자 하는 의도도 있다.

임상수행능력평가나 진료수행시험의 선행연구는 외국의 경우 관련 연구논문을 쉽게 찾아볼 수 있다. 그러나 우리나라의 경우 객관구조화진료시험의 도입이 10년밖에 되지 않았으며 진료수행시험을 실시한 연구논문은 국내의학학술지의 초록을 검색할 수 있는 ‘한국의학논문데이터베이스 (KMbase)’를 통해 검색해 보아도 거의 찾아볼 수 없다. 키워드와 제목 검색어로서 ‘진료수행평가’와 ‘CPX’를 사용했을 때 김주자 등 (2004) 논문 1편만을 찾을 수 있었다. 또한 ‘임상수행평가’라는 검색어를 사용했을 경우에는 이병국 (2000)의 연구 1편만을 찾을 수 있었다. ‘표준화 환자’ 또는 ‘Standardized Patient’에 대한 연구 논문은 김주자 (2004)와 이병국 (2000)의 연구 외에 4편의 논문을 찾을 수 있었다 (유문숙 등, 2002; 이병국, 2001; 최운호 등, 2000; 홍영재, 1992). 우리나라 의과대학에서는 최근 표준화 환자를 활용할 수 행평가가 임상교육과정의 평가 방법으로 많이 활용되고 있으나, 이와 관련된 선행 연구논문이 매우 부족한 실정이며, 관련 연구자료를 마련해준다는 측면에서 본 연구의 중요성과 가치가 강조될 수 있다고 하겠다.

이와 같은 배경에서 본 연구는 진료수행시험 상황에서 학생평가자로서 표준화 환자 활용이 우리나라 현실에서 타당한가를 분석하는 데에 목적을 두었다. 그 타당성을 확인하는 방법에는 여러 가지 방법이 있겠으나, 본 연구에서는 교수평가결과와 표준화 환자의 평가결과를 비교분석함으로써 진료수행

1) ‘제1차 CPX 심포지엄’ 자료집. 2004년 12월 3일. 임상수행능력평가를 위한 서울지역 8개 의과대학 컨소시엄.

시험에서의 평가자로서의 표준화 환자의 활용이 타당함을 검증하고자 한다.

이상과 같은 연구목적의 달성으로 본 연구는 다음과 같은 의의를 갖는다. 첫째, 교수와 표준화 환자의 평가를 통해 의과대학생들의 진료수행시험의 타당성을 분석한다. 타당성 분석의 방법은 교수의 평가 결과와 비교분석을 통해서 실시한다. 둘째, 표준화 환자 평가를 이용한 진료수행시험의 개선점을 논의하고 보다 효과적인 진료수행시험 방법을 모색한다.

대상 및 방법

가. 연구대상 및 방법

연구대상은 일개 의과대학의 의학과 4학년 학생 110명이며, 학생을 전체 2개 군(A군, B군)으로 구분하여 시험을 실시하였다. 시험은 2004학년도 9월 18일에 실시되었으며, A군(54명)과 B군(46명)은 7~8명씩 한 조를 이루어 군별로 서로 다른 4개의 사례에 대한 시험을 보도록 하였다. 본 연구에 사용된 총 8개의 시험사례는 서울CPX컨소시엄에서 선정한 것이다. 또한 CPX 시험은 OSCE와 접목하여 실시하였으므로 각 군은 4개의 CPX 사례와 4개의 OSCE 사례를 접하도록 하였다. OSCE 시험에서는 표준화 환자를 이용하지 않고 단순한 임상술기 문제를 실시하였기 때문에 이 연구에서의 대상에서는 포함하지 않았다.

언급한 바와 같이 CPX 시험의 경우, A군과 B군에 적용된 4개의 각각 다른 사례를 평가하여 총 8개 사례의 시험 결과를 얻을 수 있었다. A군의 경우 ① 불면, ② 기침, ③ 간염보균자, ④ 호흡곤란을 주제로 평가되었고, B군의 경우 ① 나쁜 소식 전하기, ② 복통, ③ 두통 그리고 ④ 비순응을 주제로 평가되었다. 각 사례에 대한 평가 문항의 내용은 병력, 신체검사(정신상태검사/신경학적 검사 포함), 정보 나누기(환자교육 포함), 임상 의사 예절 그리고 환자/의사관계라는 큰 영역으로 구분되었다. 각 영역의 평가문항은 사례에 맞게 조절하였기 때문에 사례별 평가내용이나 평가문항 수는 각각 달랐으며, 그

수는 23개 문항에서 34개 사이였다. 평가의 형식은 각 시험장에서 28명의 교수 평가자와 28명의 표준화 환자가 동시에 평가하였다. 표준화 환자의 경우 CPX컨소시엄을 통해 훈련된 환자들을 활용하였다. 평가자는 교수와 표준화 환자 동일한 것을 사용하였다. 교수는 학생이 수행을 하는 동안 관찰과 동시에 즉각 평가를 하였으며, 표준화 환자는 학생이 시험을 치른 후 5분 동안 학생의 수행 정도를 기억을 되살려 평가하도록 하였다.

나. 분석방법

표준화 환자 평가의 타당성을 검증하기 위해 교수와 표준화 환자의 평가결과를 비교분석하였다. 이를 위하여 대응표본 t검정(paired samples t-test)를 실시하였으며, 각 스테이션 별로 교수와 표준화 환자의 평가 결과에 대하여 평균, 표준편차, t검정통계량 등을 구하였다. 또한 각 스테이션별 신뢰도를 검증하기 위해 Cronbach's alpha 계수를 구하였다. 유의도 검증은 $p < .05$ 와 $< .01$ 을 기준으로 하였고 모든 통계분석은 PC-SPSS version 11.5를 사용하였다.

결 과

가. 교수와 표준화 환자 평가의 비교

교수와 표준화 환자가 평가한 8개 사례의 시험 총점은 100점 만점으로 환산하여 비교하였다. 그 결과는 Fig. 1과 같이 나타났는데, 가장 높은 평균을 보인 사례는 나쁜 소식 전하기로 교수는 72.36점, 표준화 환자는 70.89점의 평균을 주었다. 반면, 가장 낮은 평균을 보인 사례는 '호흡곤란'으로 교수는 66.28점, 표준화 환자는 61.75점의 점수를 주었다. 8개 사례 중 5개 사례(불면, 기침, 간염보균자, 두통, 비순응)에서는 표준화 환자의 평가가 교수의 평가 점수보다 높았다.

검사도구의 총점이 평가자간에 다르게 측정되었는가를 알아보기 위해 대응표본 t검정을 실시한 결과, 통계적으로 유의한 차이가 있는 사례는 총 4개로 나타났다. 불면 사례의 경우, 교수와 표준화 환자

진료수행시험 평가자로서 표준화 환자 활용의 타당성

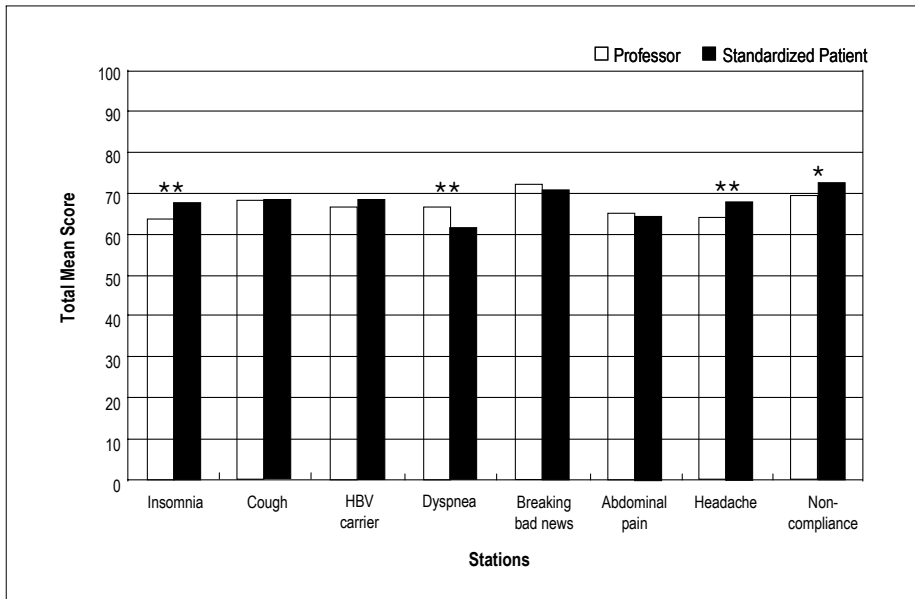


Fig. 1. Professors' and SPs' total mean scores of 8 stations (* p < .05, ** p < .01)

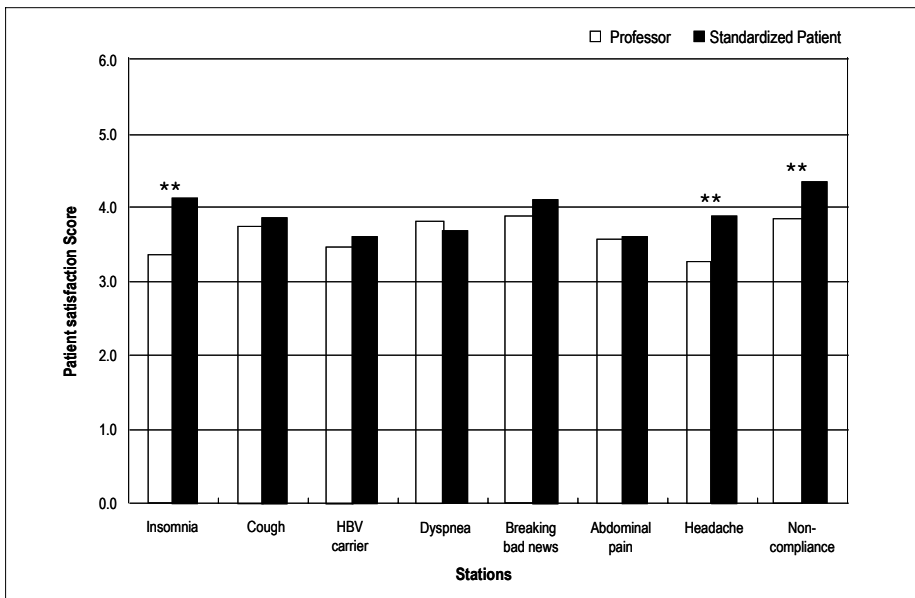


Fig. 2. Professors' and SPs' patient satisfaction scores (** p < .01)

접수의 차이의 평균과 표준편차는 각각 3.66, 7.494 이고, t통계값은 3.912, 자유도는 63이며 유의확률은 .000이다. 호흡곤란의 경우 교수와 표준화 환자 점

수의 차이의 평균과 표준편차는 각각 4.534, 4.702 이고, t통계값은 7.712, 자유도는 63이며 유의확률은 .000이다. 두통 사례의 경우 교수와 표준화 환자 점

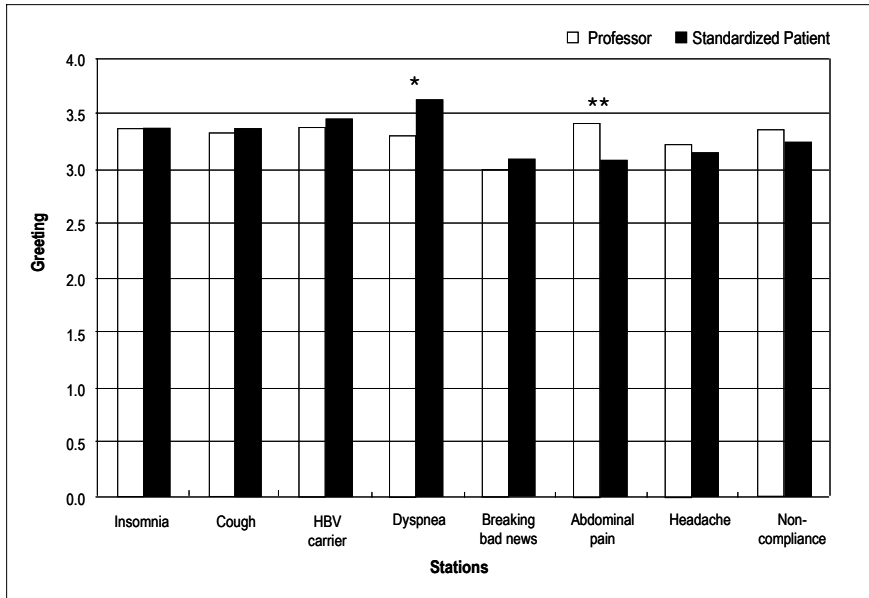


Fig. 3. Professors' and SPs' greetings scores (* p < .05, ** p < .01)

수의 차이의 평균과 표준편차는 각각 -3.696, 5.037 이고, t통계값은 -4.976, 자유도는 45이며 유의확률은 .000이다. 따라서 불면, 호흡곤란, 두통 사례는 유의수준 .01에서 유의한 차이가 있다. 비순응 사례의 경우, 교수와 표준화 환자 점수의 차이의 평균과 표준편차는 각각 -3.170, 9.189이고, t통계값은 -2.340, 자유도는 45이며 유의확률은 .024로써, 유의수준 .05에서 유의한 차이를 보였다.

나. 교수와 표준화 환자의 환자만족도 평가

교수와 표준화 환자의 환자만족도의 경우 평가의 척도가 최대값이 6, 최소값이 1이었는데 평가결과 는 Fig. 2와 같이 나타났다. 평균값이 4.0 이상인 스테이션 즉, '잘함' 이상의 평가를 받은 사례는 불면 (4.11), 나쁜소식 전하기 (4.09)와 비순응 (4.38)으로 표준화 환자 평가에만 해당되었다. 평균값은 최저 3.26점에서 최고 4.35점 사이였다. 호흡곤란 사례를 제외한 나머지 7개 사례에서 표준화 환자의 평가가 교수 평가 점수보다 높았다.

교수와 표준화 환자 평가의 환자만족도에 대한 대응표본 t검정 결과, 통계적으로 유의한 차이를 보

인 사례는 3개로 나타났다. 불면 사례의 경우, 교수와 표준화 환자 점수의 차이의 평균과 표준편차는 각각 -.730, .877이고, t통계값은 -6.697, 자유도는 63이며 유의확률은 .000이다. 두통 사례의 경우, 교수와 표준화 환자 점수의 차이의 평균과 표준편차는 각각 -.630, .645이고, t통계값은 -6.630, 자유도는 45이며 유의확률은 .000이다. 비순응 사례의 경우, 교수와 표준화 환자 점수의 차이의 평균과 표준편차는 각각 -.500, .937이고, t통계값은 -3.620, 자유도는 45이며 유의확률은 .001이다. 따라서 불면, 두통 그리고 비순응 사례는 유의수준 .01에서 유의한 차이를 보였다.

다. 교수와 표준화 환자의 인사하기(자기소개) 평가

인사하기 평가 항목의 경우 평가의 척도가 최대값이 4, 최소값이 1이었는데, Fig. 3을 보면 환자에게 인사하기 및 자기소개가 가장 잘 된 스테이션은 교수 평가의 경우 복통 (3.41) 사례였고 표준화 환자의 경우 호흡곤란 (3.60) 사례로 나타났다. 모든 스테이션에서 평균값이 3.0 이상이었고, 최저 3.00점에서 최고 3.60점이었다.

Table I. Paired T-test Scores of CPX Domain (Group A)

Station	Domain	Examiner	Mean	Full Marks	N	SD	T	P
Insomnia	History taking	Professor	13.547	18	64	1.436	-2.216	.030*
		SP	13.938		64	1.754		
	Mental exam	Professor	7.797	15	64	1.287	4.321	.000**
		SP	7.063		64	1.246		
	Information sharing	Professor	1.940	2	64	0.244	-.629	.531
		SP	1.970		64	0.355		
	PPI†	Professor	25.516	42	64	4.853	-4.245	.000**
		SP	28.297		64	4.527		
Cough	History taking	Professor	29.953	32	64	2.178	-5.405	.000**
		SP	25.266		64	2.018		
	Physical exam	Professor	7.297	12	64	1.094	1.271	.208
		SP	7.172		64	1.032		
	Clinical Courtesy	Professor	7.828	12	64	1.652	5.572	.000**
		SP	6.984		64	1.507		
	Information sharing	Professor	2.400	3	64	0.661	.659	.512
		SP	2.320		64	0.779		
	PPI	Professor	27.172	42	64	3.628	.752	.455
		SP	26.859		64	3.399		
HBV Carrier	History taking	Professor	14.767	18	64	1.514	.450	.654
		SP	14.734		64	1.348		
	Physical exam	Professor	5.078	9	64	1.440	.668	.506
		SP	4.953		64	1.227		
	Information sharing	Professor	9.234	12	64	1.678	1.449	.152
		SP	8.984		64	1.507		
	Clinical Courtesy	Professor	7.203	12	64	2.001	-8.396	.000**
		SP	9.047		64	1.750		
	PPI	Professor	25.719	42	64	3.073	-.194	.847
		SP	25.828		64	4.551		
Dyspnea	History taking	Professor	18.688	26	64	1.582	-1.655	.103
		SP	19.063		64	1.918		
	Physical exam	Professor	12.563	24	64	2.833	3.375	.001**
		SP	12.016		64	2.609		
	Clinical Courtesy	Professor	8.594	12	64	1.269	7.479	.000**
		SP	6.875		64	1.890		
	PPI	Professor	28.641	42	64	4.974	6.887	.000**
		SP	25.156		64	3.030		

† PPI denotes physician-patient interaction

* p<.05, ** p<.01

교수와 표준화 환자 평가의 인사하기 평가항목에 대한 대응표본 t검정 결과, 통계적으로 유의한 차이를 보인 사례는 2개로 호흡곤란과 복통으로 나타났다. 호흡곤란의 경우, 교수와 표준화 환자 점수의 차이의 평균과 표준편차는 각각 -.300, .927이고, t통계

값은 -2.582, 자유도는 62이며 유의확률은 .012로써, 유의수준 .05에서 유의한 차이가 있다. 복통 사례의 경우 교수와 표준화 환자 점수의 차이의 평균과 표준편차는 각각 .330, .790이고, t통계값은 2.798, 자유도는 45이며 유의확률은 .008로써, 유의수준 .01

Table II. Paired T-test Scores of CPX Domain (Group B)

Station	Domain	Examiner	Mean	Full Marks	N	SD	T	P	
Breaking Bad News	History taking	Professor	4.478	6	46	0.863	-1.401	.168	
		SP	4.696		46	0.940			
	Information sharing	Professor	19.739	24	46	3.672	1.506	.139	
		SP	18.978		46	1.915			
	PPI†	Professor	28.174	42	46	5.615	1.026	.310	
		SP	27.174		46	4.071			
Abdominal Pain	History taking	Professor	19.217	24	46	1.763	-1.093	.280	
		SP	19.435		46	1.846			
	Information sharing	Professor	4.304	6	46	1.030	3.078	.004**	
		SP	3.957		46	0.893			
	Physical exam	Professor	9.348	15	46	1.958	-3.30	.743	
		SP	9.413		46	1.869			
	Clinical Courtesy	Professor	7.435	15	46	1.785	1.430	.160	
		SP	7.000		46	2.022			
	PPI	Professor	25.457	42	46	2.553	.063	.950	
		SP	25.413		46	4.058			
	Headache	History taking	Professor	13.000	18	46	1.317	-2.374	.022*
			SP	13.565		46	1.455		
Neurological exam		Professor	11.065	18	46	1.831	-4.426	.000**	
		SP	12.739		46	2.824			
Clinical Courtesy		Professor	7.022	12	46	1.498	-5.818	.000**	
		SP	8.391		46	1.145			
PPI		Professor	26.696	42	46	2.874	.959	.343	
		SP	26.239		46	2.433			
Non-compliance	History taking	Professor	14.370	20	46	1.970	-3.574	.001**	
		SP	15.370		46	2.491			
	Information sharing	Professor	17.804	24	46	2.207	-1.771	.083	
		SP	18.761		46	3.497			
	PPI	Professor	27.348	42	46	4.138	-1.406	.166	
		SP	28.217		46	4.310			

† PPI denotes physician-patient interaction

* p<.05, ** p<.01

에서 유의한 차이를 보였다.

라. 각 사례의 영역별 평가

각 사례 별로 나타난 교수와 표준화 환자 평가의 차이에 대한 대응표본 T검정 결과는 Table I, II와 같이 정리하였다. 검정 결과에 따르면, 총 8개 사례의 33개 영역 중 통계적으로 차이를 보인 영역은 14개로 나타났다. 유의수준 .05에서 차이가 있는 영역은 2개로 나타났는데, 이는 불면과 두통 사례의 병

력 영역으로 t통계값이 각각 -2.216와 -2.374, 유의 확률은 각각 .030과 .022이다.

유의수준 .01에서 차이가 있는 12개의 영역은 불면 사례의 정신상태검사 (t값 4.321, 유의수준 .000)와 의사-환자관계 영역 (t값 -4.245, 유의수준 .000), 기침 사례의 병력 (t값 -1.313, 유의수준 .000)과 임상 의사 예절 영역 (t값 5.572, 유의수준 .000), 간염 보균자 사례의 임상 의사 예절 영역 (t값 -8.396, 유의수준 .000), 호흡곤란 사례의 신체진찰 (t값 3.375,

Table III. Cronbach's Alpha Scores of Each Station

Station	Examiner	N of Items	Alpha
Insomnia	Professor	24	.8157
	SP	24	.8183
Cough	Professor	32	.7712
	SP	32	.7456
HVB Carrier	Professor	31	.8026
	SP	31	.8319
Dyspnea	Professor	34	.8447
	SP	34	.8189
Breaking Bad News	Professor	23	.9310
	SP	23	.8516
Abdominal pain	Professor	33	.6725
	SP	33	.8023
Headache	Professor	28	.6259
	SP	28	.6710
Non-compliance	Professor	27	.7473
	SP	27	.9085

유의수준 .001), 임상 의사 예절 (t값 7.479, 유의수준 .000), 의사-환자관계 영역 (t값 6.887, 유의수준 .000), 복통 사례의 정보나누기 영역 (t값 3.078, 유의수준 .004), 두통 사례의 신경학적 검사 (t값 -4.426, 유의수준 .000)와 임상 의사 예절 영역 (t값 -5.818, 유의수준 .000), 그리고 비순응 사례의 병력 영역 (t값 -3.574, 유의수준 .001)이다.

사례별 신뢰도인 Cronbach's alpha를 살펴본 결과는 Table III과 같다. 기초연구분야에서 요구되는 0.8 이상의 값이 나온 사례는 불면, 간염보균자, 호흡곤란, 나쁜 소식 전하기, 그리고 복통과 비순응의 표준화 환자평가의 신뢰도가 이에 해당된다. 또한 모든 사례에서 신뢰도는 0.6 이상의 값을 보이고 있기 때문에 채점 문항들이 사례와 관련된 요소들을 잘 측정하고 있음을 알 수 있다.

이상의 결과를 종합해보면, 병력 영역에서 유의한 차이가 있는 사례는 불면, 기침, 두통과 비순응으로 총 4개였으며, 신체진찰(정신상태감사, 신경학적 검사 포함)에서 차이가 있는 사례는 불면, 호흡곤란, 두통으로 총 3개로 나타났다. 정보나누기 영역에서 통계적 차이가 있는 사례는 복통 사례 1개였으며, 임상 의사 예절 영역에서 차이가 있는 사례는 기침,

간염보균자, 호흡곤란, 두통으로 총 4개였다. 의사-환자 관계형성 영역에서 차이가 있는 사례는 불면, 호흡곤란으로 총 2개였다.

고 찰

외국의 많은 의과대학에서 표준화 환자를 이용한 임상수행능력평가를 의학과 과정 중에 여러 번 실시하는 것은 결과중심의 평가보다 과정중심의 평가를 중요시 여기기 때문일 것이다 (Prislin et al, 2000). 또한 표준화 환자를 활용했을 때 감정적인 부분이나 환자와의 의사소통에 대한 부분에서는 교수의 평가보다 표준화 환자가 평가했을 때 환자의 입장을 보다 잘 반영하는 측정이 될 수 있을 것이다. 이러한 배경에서 본 연구는 일개 의과대학의 진료수행시험에서 표준화 환자의 평가의 타당성을 검증하고자 하였다. 본 절에서는 이상의 연구결과를 바탕으로 진료수행시험에서 표준화 환자 평가의 개선점을 논의하고 보다 효과적인 진료수행시험을 위한 과제를 논의하고자 한다.

먼저, 본 연구의 결과를 통해 다음과 같은 결론을 얻을 수 있었다. 총 8개의 평가 스테이션 중 교수와

표준화 환자의 점수가 차이가 있는 사례는 4개 사례로 나타났다. 그러나 세부항목별 평가를 살펴보았을 때에는 환자만족도 평가의 경우 2개 사례에서만 차이를 보였고, 인사하기의 경우 3개 사례에서만 차이를 보였다. 또한 사례별 의사-환자관계형성 점수는 교수가 높게 부여하였으나, 전체적인 평균 점수는 표준화 환자가 더 높게 주었다. 단, 병력청취, 신체진찰, 정보나누기, 임상 의사 예절 등에 대한 문항은 객관적인 평가가 가능하여 교수와 표준화 환자간의 차이가 없을 것으로 예상하였으나, 두 평가간의 차이가 있었다. 이는 이 시험의 자체에 문제점이라기보다는 채점기준표의 문제점으로 지적된다. 채점기준표에 제시된 기준에 대해 교수와 표준화 환자가 서로 다르게 이해를 했을 가능성이 높다. 또한 교수 평가자의 경우, 보다 전문가적인 입장에서 평가했을 가능성이 높고, 이는 결국 동일한 평가항목에 대해 표준화 환자와의 차이를 가져다주었을 것으로 보인다. 그리고 교수는 학생이 수행을 하는 동안 관찰함과 동시에 평가를 한 반면, 표준화 환자는 기억에 의존하여 시험 후에 평가를 하였으므로 정확한 평가의 한계가 있음을 시사한다.

한편, 기존의 많은 선행연구들의 결과를 종합해 보았을 때, 표준화 환자와 전문의사의 채점 결과는 서로 다르지 않음을 시사하는 논문들에서 이 시험 자체의 한계가 아니라는 것을 알 수 있다(이병국, 2000). 표준화 환자들 간의 평가에 차이가 없는지 학생들을 표준화된 학생들로 훈련시켜 연구한 Worth-Dickstein et al (2005)의 최근 연구결과를 보아도 학생들의 개인적인 태도나 예의범절의 차이가 표준화 환자의 평가에 영향을 주지 않는다는 결과가 나왔다. 이는 훈련을 잘 받은 표준화 환자들의 평가를 신뢰할 수 있다는 증거이다. 따라서 진료수행시험에서 표준화 환자들의 평가 결과를 활용하는 데에는 큰 문제가 따르지 않는다고 볼 수 있다.

본 연구의 결과를 통해 생각해볼 수 있는 표준화 환자 평가의 개선점으로는 첫째, 보다 일관성 있는 채점을 위해 채점기준표와 평가항목을 개발하는 것이 필요하다. 둘째, 이번 연구에서는 교수와 표준화 환자의 평가 조건이 동일하지 않았기 때문에 향후

연구에서는 교수와 표준화 환자 모두 동일한 조건에서 평가 할 수 있도록 한 후에 다시 두 평가자간의 평가 결과를 비교하는 연구가 필요하다. 이를 통하여 교수와 표준화 환자간의 차이가 있는 부분이 채점기준표의 문제인지 또는 평가자의 문제인지를 보다 정확하게 파악해 낼 수 있을 것이다. 이러한 점에서 서로 다른 조건에서 평가결과를 비교한 것을 이번 연구의 한계점으로 지적될 수도 있다. 향후 보다 적절하고 타당한 연구 방법을 통해 표준화 환자를 CPX 시험에서 평가자로 활용하는데 문제가 없음이 밝혀지기를 기대한다.

참 고 문 헌

- 김 신(2004). “CPX의 타당도” 제1차 CPX 심포지엄 자료집. 한국의학교육학회/한국의학교육연수원. 51-57.
- 김주자 외(2004). 일개 의과대학에서 실시한 표준화 환자(SP)를 이용한 임상수행능력평가시험(CPX) 결과 분석. *한국의학교육*, 16(1), 51-61.
- 유문숙(2002). 표준화 환자 학습방법이 유치도뇨술, 의사소통능력, 학습동기에 미치는 효과. *기본간호학회지*, 9(1), 66-75.
- 이병국(2001). 우리 나라 표준화 환자 프로그램 활용 현황과 개선 방안 - 관련 문헌 고찰 및 현장연구를 중심으로. *한국의학교육*, 13(1), 27-39.
- 이병국(2000). 임상수행평가의 세계적 추세에 따른 표준화 환자 프로그램의 토착화 방안 연구. *한국 의학교육*, 12(2), 377-392.
- 최윤희 외(2000). 임상실습 교육을 위한 표준화 환자 프로그램의 개발 및 적용. *한국의학교육*, 12(2), 353-361.
- 한국의학교육학회(2003). *의학교육학용어집*. 아카데미아.
- 홍영재(1992). Standardized Patients를 이용한 임상 교육. *한국의학교육*, 4(2), 12-17.
- Blue AV, Chessman AW, Gilbert GE & Mainous AG(2000). Responding to patients emotions: important for standardized patient satisfaction.

- Family Medicine*, 32(5), 326-330.
- Boulet JR, McKinley DW, Norcini JJ & Whelan GP(2002). Assessing the comparability of standardized Patient and Physician Evaluations of Clinical Skills. *Advances in Health Sciences Education*, 7, 85-97.
- Boulet JR, Champlain A & McKinley DW(2003). Setting defensible performance standards on OSCEs and standardized patient examinations. *Medical Teacher*, 25(3), 245-249.
- Guillon G, Hodgson CS, Delandshere G & Luann Wilkerson(2004). Communication Skills in Standardized-Patient Assessment of Final-Year Medical Students: A Psychometric Study. *Advances in Health Sciences Education*, 9, 179-187.
- Prislin MD, Fitzpatrick CF, Lie D, Giglio M, Radecki S & Lewis E(1998). Use of an Objective Structured Clinical Examination in Evaluating Student Performance. *Family Medicine*, 30(5), 338-344.
- Prislin MD, Giglio M, Lewis EM & Ahearn S(2000). Assessing the acquisition of core clinical skills through the use of serial standardized patient assessments. *Academic Medicine*, 75(5), 480-483.
- Solomon DJ, Szauter K, Rosebraugh C & Callaway MR(2000). Global Ratings of Student Performance in a Standardized Patient Examination: Is the Whole More than the Sum of the Parts? *Advances in Health Sciences Education*, 5, 131-140.
- Worth-Dickstein H, Pangaro LN, MacMillan MK, Klass DJ & Shatzer JH(2005). Use of "standardized examinees" to screen for standardized-patient scoring bias in a clinical skills examination. *Teaching and Learning Medicine*, 17(1), 9-13.