

CPX 성적과 기존의 실습성적 및 필기시험 성적과의 상관관계

서울대학교 의과대학 의학교육실, 내과학교실¹, 성형외과학교실²

박완범 · 이성아 · 김은아 · 김연수¹ · 김석화² · 신좌섭 · 이윤성

= Abstract =

Correlation of CPX Scores with the Scores of the Clinical Clerkship Assessments and Written Examinations

Wan Beom Park, MD, Sung A Lee, RN, Eun A Kim, Yon Su Kim¹, MD, PhD, Suk Wha Kim², MD, PhD, Jwa-Seop Shin, MD, EdD, Yoon Sung Lee, MD, PhD

Office of Medical Education, Departments of Internal Medicine¹ and Plastic Surgery², Seoul National University College of Medicine, Seoul, Republic of Korea

Purpose: Recently, the clinical performance examination (CPX) using standardized patients has been introduced to several Korean medical schools. The purpose of this paper is to evaluate the correlation of CPX scores with the scores of other clinical clerkship assessments and written examinations.

Methods: We evaluated the correlation of the scores of CPX, which was administered to third-year medical students (n=230) at Seoul National University College of Medicine in January 2005, with the scores of a slide examination in internal medicine, an objective structured clinical examination (OSCE) in pediatrics and obstetrics/gynecology, and the written examinations in internal medicine and general surgery.

Results: CPX scores showed a mild correlation with the scores of the slide examination in internal medicine and OSCE in pediatrics and obstetrics/gynecology ($r=0.19, 0.28, 0.30$). CPX scores were also mildly correlated with written examination scores in internal medicine but not associated with those in general surgery.

Conclusion: Considering the mild correlation of CPX scores with scores of other clinical clerkship assessments and written examinations, CPX should be affirmatively considered as an examination complementary to previous examinations.

Key Words: Clinical performance examination, Objective structured clinical examination, Written examination, Correlation

교신저자: 신좌섭, 서울대학교 의과대학 의학교육실, 서울시 종로구 연건동 28번지
Tel: 02)740-8175, Fax: 02)740-8072, E-mail: hismed1@snu.ac.kr

*본 연구의 일부 내용은 2005년 5월 제17차 의학교육합동학술대회에서 발표되었음.

서 론

표준화 환자 (standardized patient, SP)란 실제 환자와 비슷하게 그리고 일관성 있게 진료 상황을 재현해 낼 수 있는 의사가 아닌 일반인을 의미한다 (Barrows, 1993). 최근 국내외 많은 의과대학에서 임상수행능력의 평가 방법으로 또는 졸업시험의 일환으로 이러한 SP를 이용한 진료수행시험 (clinical performance examination, CPX)를 도입하고 있다 (Williams et al., 1987; Stillman et al., 1987).

기존의 임상실습 평가 방법인 필기시험, 슬라이드 시험, 객관적-구조화 임상시험 (objective structured clinical examination, OSCE)과 비교하여 CPX는 임상에 보다 가까운 상황에서 임상 지식의 적용 능력, 대인관계, 임상예절 등 다양한 측면의 평가가 가능하다는 장점을 가진다 (Anderson et al., 1991; Schnabl et al., 1991). 최근까지 임상 수기를 객관적으로 평가하기 위해서 OSCE가 많이 사용되었으며 OSCE에서도 SP를 사용할 수 있으나 (이병국, 2000) OSCE와 CPX는 여러 측면에서 차이가 있다. 가장 중요한 차이 중 하나는 OSCE는 특정 술기의 수행 능력을 단편적으로 확인하는 반면, CPX는 특정 문제를 가진 환자를 대상으로 완전한 임상 수행 능력을 평가한다는 것이다. 따라서 CPX에서는 상황에 맞게 적절한 우선 순위로 임상 술기를 이용할 수 있는 능력까지 요구하게 된다. Luann 등 (2003)은 의과대학생을 대상으로 같은 장기에 대한 신체검사 능력을 OSCE와 CPX로 각각 평가하였을 때 CPX에서 점수가 통계적으로 유의하게 낮음을 보임으로써 특정 술기를 할 수 있는 능력과 항목별 중요성을 인식하여 정해진 시간 내에 그것을 적용하는 능력은 다를 수 있음을 지적하였다.

SP를 이용한 평가의 정확도는 이미 외국의 많은 연구를 통해 입증되었으며 (Elliot & Hickam, 1987; Vu et al., 1992), 최근 국내에서도 몇몇 연구를 통해 SP에 의한 평가와 교수에 의한 평가 사이에 큰 차이가 없음이 알려지고 있다 (이병국, 2000; 박훈기, 2003).

연구자들이 소속된 대학에서는 2005년 1월에 3학년을 갓 마친 학생을 대상으로 '서울CPX컨소시엄'을 통해 SP를 활용하여 CPX를 시행하였다. 본 연구

자들은 일개 대학에서 기존의 실습평가 도구에 의한 성적 및 필기 시험 성적과 CPX 성적과의 상관관계를 살펴보고 CPX의 여러 평가 항목별로 기존 시험 성적과의 연관성을 알아보고자 하였다.

대상 및 방법

가. 연구 대상

2005년 1월 17일부터 20일까지 4일간 서울의대 본과 3학년 학생 230명을 대상으로 CPX를 시행하였다. 일차의료에서 흔히 볼 수 있는 의학적 문제 8개를 '서울CPX컨소시엄'을 통하여 제공 받았으며 '서울CPX컨소시엄'에서 훈련된 SP 총 37명을 동원하였다. 시험방 (station) 별 시간 배정은 상황소개 및 지침 숙지 시간 1분, 환자대면진료시간 12분, 사이 시험 시간을 5분으로 하였다. 본 연구는 환자대면진료시간에 SP에 의해 채점된 점수만을 분석 대상으로 하였으며 사이시험 성적은 분석에서 제외하였다.

CPX와 같은 학기에 같은 학생들을 대상으로 내과에서 실습평가의 일환으로 시행한 슬라이드 시험 성적, 그리고 소아과와 산부인과에서 시행한 OSCE 시험 성적과 CPX 성적과의 관련성을 조사하였다.

또한 지식 평가를 대표하는 내과와 외과의 필기 시험과 CPX 성적 그리고 3학년 전체 평점과 CPX 성적 사이의 관련성을 평가하였다. 그리고 CPX 성적을 각 평가 항목별로 세분하여 각 평가 항목별 성적과 기존의 실습시험 및 필기시험 성적과의 관련성을 보았다.

1) CPX의 평가 항목 및 점수 비중

CPX의 문제 별 점검표 (checklist) 문항수는 총 232개 (시험방 당 평균 29개, 범위 23~34)이었으며 CPX의 세부 평가 항목은 병력청취 (34%), 신체검진 (14%), 임상예절 (9%), 의사-환자 관계 (24%), 환자 교육 (13%), 총괄채점 등 기타 (7.6%)를 포함하였다.

2) 과목별 실습시험의 형태

a. 내과 슬라이드 시험

주로 실습 과정을 통해 습득한 임상적 지식 또는 문제 해결 능력을 평가하였으며 1시간에 50문항을 슬라이드로 제시하고 풀도록 하였다.

Table I. Correlations of Overall Scores of CPX with Scores of Slide Exam, OSCE, and Written Exams

	Slide exam	OSCE		Written exam		Final grade (end of school year)
	Internal medicine	Pediatrics	Obstetrics/ Gynecology	Internal medicine	General surgery	
r	0.19	0.28	0.30	0.28	0.13	0.32
p-value	0.004	<0.001	<0.001	<0.001	0.06	<0.001

r: Pearson correlation coefficient

b. 소아과 OSCE

소아과 OSCE 시험의 시간 배정은 문제당 4분이었고, 30초가 feedback 시간으로 주어졌다. 4분 30초가 지나면 바로 다음 방으로 이동하였고 원칙적으로 중간 휴식은 없었으며 레지던트가 모의환자 역할을 하였다.

총 10문제가 출제되었으며 항목별 문제 배정은 병력 30%, 신체 진찰 10%, 수기 10%, 문제해결 능력 40%, 기타 10%로 이루어졌다.

c. 산부인과 OSCE

산부인과 OSCE의 시간 배정은 문제당 4분, 중간 휴식 5분을 취하였으며 휴식 시 추가적인 사이시험은 없었다. 산부인과 전공의가 모의 환자 역할을 하였으며 산부인과적 검진이 필요한 경우에는 여성 생식기 모형에 갖춘 인형을 사용하였다.

총 10문제가 출제되었으며 항목별 문제 배정은 병력 40%, 신체검진 10%, 수기 10%, 지식 20%, 기타 (태도 등) 20%로 이루어졌다.

3) 3학년 전체 평점의 과목별 비중

3학년말 전체 평점은 10개 과목을 합산하여 주어졌으며 각 과목별 학점은 임상총론 1학점, 내과학 및 실습 12학점, 외과학 및 실습 4학점, 산부인과학 및 실습 4학점, 소아과학 및 실습 6학점, 정신과학 및 실습 4학점, 정형외과학 및 실습 3학점, 방사선과학 및 실습 3학점, 마취과학 1학점, 핵의학 1학점으로 구성되었다.

나. 분석

각 성적간의 관련성은 피어슨 (Pearson) 상관계수

로 평가하였고 상관계수가 0.1과 0.3 사이이면 약한 양적 선형관계, 0.3과 0.7 사이이면 뚜렷한 양적 선형관계, 0.7과 1.0 사이이면 강한 양적 선형관계로 간주하였다. 양측검정을 통해 $p < 0.01$ 을 의미 있는 것으로 인정하였으며 통계 프로그램은 SPSS version 12.0 (SPSS Inc, Chicago, IL, USA)를 사용하였다.

결 과

가. CPX 총점과 기존 성적과의 상관관계

CPX의 총점은 기존의 실습평가 방법인 내과의 슬라이드 시험, 소아과, 산부인과의 OSCE 성적과 통계적으로 유의한 상관관계가 있었으나 그 정도는 미약하였다 (Table I). CPX의 총점은 내과 필기 시험 성적과 약하나마 통계적으로 유의한 상관관계가 있었으나 외과 필기 시험 성적과는 무관하였으며 학년말 전체 평점과는 뚜렷한 선형 상관관계를 보였다.

나. CPX 각 평가항목 별 성적과 기존 성적과의 상관 관계

CPX 성적을 각 평가항목별로 분석하였을 때 신체 진찰은 기존의 실습평가 방법인 내과의 슬라이드 시험, 소아과, 산부인과의 OSCE 성적과 약하나마 모두 상관관계가 있었으며 병력청취는 산부인과 OSCE와, 환자-의사 관계는 소아과와 산부인과의 OSCE와 약한 상관관계가 있었다 (Table II). 나머지 평가항목별 성적은 기존의 실습평가 성적과 무관하였다.

환자만족도, 신체진찰, 환자의사관계, 나쁜 소식

Table II. Correlations of Domain Scores from the CPX with Scores of Slide Exam and OSCE

		Slide exam		OSCE	
		Internal medicine	Pediatrics	Obstetrics/Gynecology	
Patient satisfaction	r	0.18	0.34	0.27	
	p-value	0.008	<0.001	<0.001	
Greeting	r	0.11	0.14	0.10	
	p-value	0.10	0.04	0.134	
History taking	r	0.12	0.13	0.24	
	p-value	0.09	0.05	<0.001	
Physical exam	r	0.24	0.20	0.18	
	p-value	<0.001	0.002	0.007	
Clinical courtesy	r	-0.2	0.10	0.14	
	p-value	0.78	0.13	0.04	
Patient-physician interaction	r	0.15	0.30	0.27	
	p-value	0.03	<0.001	<0.001	
Patient education	r	0.07	0.13	0.16	
	p-value	0.30	0.05	0.02	
Bad news delivery	r	0.11	0.12	0.07	
	p-value	0.10	0.07	0.27	

r: Pearson correlation coefficient

전하기는 내과 필기 시험 성적과 약한 상관관계를 보였으나 인사, 병력청취, 임상예절, 환자 교육은 내과 필기 시험 성적과 무관하였다 (Table III). 한편, CPX의 각 평가항목별로 분석하였을 때에도 외과 필기 시험과 유의한 상관관계를 가지는 항목은 없었다.

CPX 평가 항목 중 환자 만족도 ($r=0.3, p<0.001$), 신체 진찰 ($r=0.27, p<0.001$), 환자-의사관계의 성적 ($r=0.26, p<0.001$)은 학년말 전체 평점과 약한 상관관계를 보였으며 인사, 병력청취, 임상예절, 환자교육, 나쁜 소식 전하기의 성적과 학년말 전체 평점은 무관하였다.

고 찰

본 연구를 통하여 연구자들은 기존의 여러 평가

도구에 의한 실습 평가 성적과 CPX 성적과의 관련성을 살펴봄으로써 CPX 도입의 의미를 조망해보고자 하였다. 이번 연구 결과 CPX 성적은 기존의 실습 평가 방법인 내과 슬라이드 시험, 소아과와 산부인과 OSCE 성적과 약한 정도의 상관관계를 보였다. CPX 성적을 각 평가 항목 별로 분석하였을 때 기존의 실습평가 성적은 환자 만족도, 신체 진찰, 환자-의사 관계와 약하나마 상관관계가 있었으나 인사, 임상 예절, 환자 교육이나 나쁜 소식 전하기 등의 성적과는 유의한 관련성이 없었다.

이러한 사실은 기존의 실습평가 방법을 통해서 인사, 임상 예절, 환자 교육, 나쁜 소식 전하기 등을 제대로 평가할 수 없었다는 해석을 할 수 있으며 이 영역들은 실제 임상과 유사한 상황에서 SP를 통해서 평가를 제대로 할 수 있다는 것을 의미한다. 즉

Table III. Correlations of Domain Scores from CPX with Scores of Written Examinations

		Written exams	
		Internal medicine	General surgery
Patient satisfaction	r	0.27	0.12
	p-value	<0.001	0.07
Greeting	r	0.12	0.09
	p-value	0.07	0.20
History taking	r	0.13	0.03
	p-value	0.05	0.64
Physical exam	r	0.25	0.11
	p-value	<0.001	0.10
Clinical courtesy	r	0.05	0.02
	p-value	0.45	0.80
Patient-physician interaction	r	0.25	0.09
	p-value	<0.001	0.17
Patient education	r	0.15	0.07
	p-value	0.02	0.32
Bad news delivery	r	0.19	0.13
	p-value	0.003	0.05

r: Pearson correlation coefficient

CPX는 이전 평가방식과 비교하여 고유한 평가 영역을 가지고 있고 학생의 수행 능력에 대한 보다 자세한 정보를 제공해 줄 수 확인할 수 있었다.

박혜숙 등 (2004)은 병력청취와 신체검진의 OSCE 점수가 전 과목성적과 상관성이 낮음을 보임으로써 학생평가에 있어 OSCE의 역할을 강조하였다. 본 연구 결과에 따르면 병력청취와 신체검진에서 OSCE와 CPX가 약하나마 연관성이 확인된 반면 내과 슬라이드 시험과 CPX는 연관성이 없어 OSCE와 CPX의 평가 특성상 유사성을 간접적으로 확인할 수 있었으며 지식과 문제해결능력을 평가하는 내과 슬라이드 시험과 CPX와의 평가 도구로서의 차이점을 확인할 수 있었다. 하지만 OSCE도 인사, 임상 예절, 환자 교육 및 정보 공유 등 환자 진료 시 태도 영역에서 CPX 성적과 연관성이 없어 OSCE로 대신할 수

없는 CPX의 고유 영역이 있음을 확인할 수 있었다.

임상의학에 대한 일반적 지식이 환자 진료에 영향을 주기 때문에 CPX 성적이 내과의 필기시험 성적과 유의한 상관관계를 보이는 것은 어느 정도 당연한 일이기도 하다. Stillman (1991) 등은 레지던트를 대상으로 SP를 이용하여 CPX를 시행하면서 지식을 평가하는 객관식 시험을 추가했을 때 객관식 시험 성적과 환자 면담은 0.29, 신체 검진 0.12, 감별 진단 0.28의 상관관계가 있음을 보고하였다.

SP를 이용한 평가가 널리 사용되고 있지만 SP를 이용한 CPX의 문제점을 제시하는 연구가 많이 있었던 것도 사실이다. Regehr 등 (1999)은 수행해야 하는 문항간에 다양성이 커서 믿을 수 있는 결과를 얻기 위해서는 많은 시험방이 필요함을 언급하였다. Carpenter 등 (1995)은 문제 출제부터 훈련, 시행에

이르기까지 많은 비용이 소모되며 특히 시험방이 많아지면 비용 지출이 과도하게 커질 수 있음을 경고하였다. 그에 대한 대책으로 실기 능력을 필기 시험으로 평가하고자 하는 연구도 많이 진행되었는데 (Van der Vleuten C et al., 1989; Remmen et al., 2001), Kramer 등 (2002)은 수기에 관한 지식을 묻는 필기 시험 성적과 OSCE 간에 0.37의 상관관계를 보이며 일반적인 지식과 OSCE 성적은 0.22의 상관관계가 있음을 보임으로써 필기 시험으로 임상 수기 평가를 대신할 것을 제안하였다. 본 연구의 결과도 내과필기시험 성적과 CPX의 전체 성적이 0.28의 상관관계를 보여 Kramer의 결과와 비슷하였다. 하지만 영역 별로 분석을 해보면 인사, 병력청취, 임상예절, 환자교육은 내과필기시험 성적과 무관하였다. 따라서 단순히 전체 성적의 비교만으로 필기 시험의 유용성을 주장하는 것은 구체적이고 필수적인 평가항목이 간과될 수 있음을 시사하였다.

CPX 성적이 내과의 필기시험 성적과 유의한 상관관계를 보이는 또 다른 이유로는 내과적 지식은 CPX에서 환자대면진료에서도 도구적 지식으로 활용된다는 점이다. CPX 성적이 외과 필기시험과 무관한 것은 CPX의 평가 항목이 내과적 지식에 치우쳐 있기 때문이기도 하다. 실제로 본 연구에서 시행된 CPX에서 8개의 시험방 중 6개가 내과 질환을 가지고 있는 환자를 대상으로 하였고 정신과 질환 한 명, 신경외과 질환이 한 명 포함되어 있었다.

이렇게 내과 환자에 편중된 이유 중 하나는 외과적 질환의 시뮬레이션 (simulation)이 어렵다는 점도 있었다. 다양한 과의 환자를 포함한 CPX를 진행하기 위해서는 외과적 질환 환자를 표현하기 위한 기술과 경험이 축적되어야 할 것으로 생각한다. 또한 시나리오 개발 단계에서 보다 다양한 과가 참여해야 문제의 장기 시스템별 균형을 유지할 수 있고 한 환자에서 여러 과의 질환을 동시에 감별진단할 수 있는 문제를 개발할 수 있을 것이다.

본 연구에서 CPX 성적을 전체 과목의 성적과 비교하지 않고 일부과목의 실습성적과만 비교한 이유는 다음과 같다. 첫째, 3학년 말에 CPX 시험을 시행하였는데 그 시점에서 학생들이 실습한 과목이 CPX station을 구성한 과목과 차이가 있었기 때문이다. 예를 들어 CPX station에 신경외과 문항이 있었는데

신경외과는 3학년 때 실습하는 과목이 아니었다. 둘째, 본 연구를 진행한 대학의 경우 실습시험으로 OSCE를 시행하는 과목은 소아과와 산부인과 밖에 없었다. CPX와 OSCE 성적과의 관련을 보기위해 소아과와 산부인과를 선택할 수 밖에 없었다. CPX에 출제된 문항의 과목과 비교하는 과목의 차이는 그 자체가 둘 사이의 상관성을 떨어뜨리는 요소가 될 수 있었을 것이다.

CPX 평가 항목 중 환자 교육이 지식 평가 위주의 필기 시험과 유의한 상관관계가 없다는 점은 환자와의 정보 공유를 위해서는 단순한 지식의 축적 이외에 다른 요소가 필요로 함을 반증하였다. 이러한 점을 고려할 때 CPX는 이전의 실습 시험 및 필기 시험을 보완할 수 있는 평가 방법으로 생각된다.

CPX 성적은 3학년말 전체 평점과 뚜렷한 상관관계가 있었고 이는 전체적인 학습능력이 뛰어난 학생이 CPX를 잘 볼 수 있음을 의미하며 CPX의 타당도를 뒷받침한다. 김주자 등 (2004)은 표준화 환자를 이용한 CPX 성적과 전체 학교 성적사이의 연관성을 보고 한 바 있다. CPX 총점과 학교 성적 사이의 연관성이 상관계수 0.28~0.32 정도로 본 연구와 흡사하였으며 병력 청취가 학교 성적과 연관성이 없는 것도 본 연구 결과와 같았다. 하지만 신체 진찰이 학교 성적과 무관한 것은 본 연구와 약간 차이가 있었는데 본 연구에서는 전체 학교 성적과 신체검진은 0.27의 상관관계를 보였다. 김주자 등 (2004)의 연구는 한 시험방에 대한 연구 결과인 반면 본 연구에서는 8개의 시험방을 사용하였고 그에 따라 보다 많은 신체 검진 항목이 있었다는 점에서 신체검진과 전체 성적사이의 상관관계를 밝히는데 민감도의 차이가 있었을 것으로 생각한다. 본 연구의 결과에 따르면 신체검진과 비교하여 병력 청취는 학교성적과 관련성이 보다 약하며 병력 청취의 숙달을 위해서는 지식 습득이외에 보다 많은 경험과 훈련이 필요함을 시사하였다.

결론적으로, 기존의 평가 도구와 상관관계가 낮으면서, 내용 타당도가 높은 시험일수록 도입할 가치가 높다는 점을 고려할 때 CPX의 적극적 도입은 긍정적으로 평가된다.

참 고 문 헌

김주자, 이경제, 최규연, 이동환(2004). 일개 의과대학에서 실시한 표준화환자(SP)를 이용한 임상수행능력평가시험(CPX) 결과 분석. *한국의학교육*, 16, 51-61.

박혜숙, 한재진, 박미혜, 오지영(2004). 임상실습 전 학생과 임상실습 후 학생에서 시행한 객관구조화진료시험(OSCE) 평가. *한국의학교육*, 16, 63-71.

박훈기, 이정권, 황환식, 이재웅, 최윤영, 김혁, 안동현. 객관구조화진료시험(OSCE)에서 교수와 표준화환자 사이의 점검표 채점의 일치도. *한국의학교육*, 15, 141-150.

이병국(2000). 임상수행평가의 세계적 추세에 따른 표준화환자 프로그램의 토착화 방안 연구. *한국의학교육*, 12, 377-392.

Anderson DC, Harris IB, Allen S, Satran L, Bland CJ, Davis-Feickert JA, Poland GA, Miller WJ (1991). Comparing students' feedback about clinical instruction with their performances. *Acad Med*, 66, 29-34.

Carpenter JL(1995). Cost analysis of objective structured clinical examinations. *Acad Med*, 70, 828-833.

Elliot DL, Hickam DH(1987). Evaluation of physical examination skills. Reliability of faculty observers and patient instructors. *JAMA*, 258, 3405-3408.

Kramer AWM, Jansen JJM, Zuithoff P, Dusman H, Tan LHC, Grol RPTM, van Der Vleuten CPM (2002). Predictive validity of a written knowledge test of skills for an OSCE in postgraduate training for general practice. *Med Edu*, 36, 812-819.

Regehr G, Freeman R, Hodges B, Russell L(1999). Assessing the generalizability of OSCE measures across content domains. *Acad Med*, 74, 1320-1322.

Remmen R, Scherpbier A, Denekens J, Derese A, Hermann I, Hoogenboom R, Van der Vleuten C, van Royen P, Bossaert L(2001). Correlation of a written test of skills and a performance based test: a study in two traditional medical schools. *Med Teach*, 23, 29-32.

Schnabl GK, Hassard TH, Kopelow ML(1991). The assessment of interpersonal skills using standardized patients. In Proceedings of the thirtieth conference on research in medical education. *Acad Med*, 66(Suppl), S34-36.

Stillman PL, Swanson DB(1987). Ensuring the clinical competence of medical school graduates through standardized patients. *Arch Intern Med*, 147, 1049-1052.

Van der Vleuten C, van Luyk S, Beckers A(1989). A written test as an alternative to performance testing. *Med Edu*, 23, 97-107.

Vu NV, Marcy ML, Colliver JA, Verhulst SJ, Travis TA, Barrows HS(1992). Standardized patients's accuracy in recording clinical performance checklist items. *Med Edu*, 26, 99-104.

Wilkerson L, Lee M(2003). Assessing physical examination skills of senior medical students: knowing how versus knowing when. *Acad Med*, 78 (Suppl), S30-S32.

Williams RG, Barrows HS, Vu NV, Verhulst SJ, Colliver JA, Marcy M, Steward D(1987). Direct, standardized assessment of clinical competence. *Med Edu*, 21, 482-489.