

# The Change of CPX Scores according to Repeated CPXs

Yoon Hee Lee, Jae Hyun Park, Jin Kyung Ko and Hyo Bin Yoo

Department of Medical Education, Kyung Hee University School of Medicine, Seoul, Korea

## CPX 반복 시행에 따른 영역별 성적의 변화

경희대학교 의학전문대학원 의학교육학교실

이윤희, 박재현, 고진경, 유효빈

**Purpose:** Most medical schools have held clinical skills training programs recently. Despite these educational endeavors, few studies have attempted to address the effect of clinical skills assessments on clinical performance. This study investigated whether repeated experiences with the examination improved medical students' history-taking, physical exams, and patient-physician interactions (PPIs).

**Methods:** The subjects of the study were 101 4th-year medical students who participated in the clinical performance examination (CPX) 3 times. They completed their core clerkship before acquiring the first CPX scores; we tracked down the scores of three sets of CPX for 3 subdomains (history taking, physical exam, and patient-physician interaction) and investigated the changes in these scores. Additionally, we classified the research subjects into 3 groups by total CPX score-higher (upper 30%, n=30), intermediate (medium 40%, n=40), and lower (lower 30%, n=30)-and compared the curves for each group.

**Results:** Significant improvements were made on history taking and physical exam ( $F=130.786/237.358$ ,  $p<0.01$ ), while proficiency on the PPI declined ( $F=17.621$ ,  $p<0.01$ ). Additionally, scores in all levels improved continuously on history taking and physical exam, while students of the high and low levels experienced a sharp decline on the PPI ( $F=11.628$ ,  $p<0.01$ ).

**Conclusion:** Improvement in the history-taking score reflects an accumulation of clinical knowledge and clinical exposure. Improvement on the physical exam score is affected by repeated practice on similar or identical cases and receipt of feedback. That PPI can deteriorate might be an effect of one's negative experience in a clinical clerkship.

**Key Words:** Clinical performance, Clinical performance examination, Patient-physician interaction

## 서론

많은 의과대학에서 일차 진료의사로서의 핵심 역량을 향상시키기 위해 다양한 교육과 평가를 하고 있고, 그 중 하나가

임상수행능력시험(clinical performance examination, CPX)이다. CPX는 표준화 환자를 진료하는 동안 학생의사의 전반적인 능력과 기술을 평가함으로써 교육과정을 통해 획득한 역량들을 통합하고 실제 임상상황에 적응하는 능력을 볼 수 있다[1]. 또 CPX는 과거에 제대로 평가되지 못했던 병력청취

Received: April 19, 2011 • Revised: June 3, 2011 • Accepted: June 24, 2011

Corresponding Author: Jae Hyun Park

Department of Medical Education, Kyung Hee University School of Medicine, 1 Hoegi-dong, Dongdaemun-gu, Seoul 130-701, Korea

Tel: +82.2.961.9102 Fax: +82.2.969.6958 e-mail: fool@khu.ac.kr

A part of this paper was presented at the 6th Congress of the Asian Medical Education Association (AMEA) in March 2011.

Korean J Med Educ 2011 Sep; 23(3): 193-202.

doi: 10.3946/kjme.2011.23.3.193.

pISSN: 2005-727X eISSN: 2005-7288

© The Korean Society of Medical Education. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

와 신체진찰 영역을 측정하는 데 있어서 표준화 환자를 활용함으로써 좀 더 실제적이고 객관적인 평가를 가능하게 한다 [2]. 마지막으로 학생들이 의사로서 직면하게 될 의료 환경에 맞추어 임상수행능력을 갖출 수 있도록 교육하고, 그에 따른 교육의 효과를 평가할 수 있다. 이처럼 CPX는 평가의 형식을 갖지만, 임상실습 중 각 과에서 교육의 일환으로 활용하기도 하는 등 교육의 역할도 포함하고 있다.

우리나라에서는 1990년대부터 객관적-구조화 임상시험 (objective structured clinical examination, OSCE)이 알려지면서 임상수행평가에 관심을 갖기 시작했다. 도입 초기에는 Ronald Harden이 제안한 것처럼 OSCE가 임상수행평가 전체를 지칭하는 용어로 일반화되어 오다가, 현재는 OSCE와 CPX를 구분하여 사용하고 있다. OSCE는 표준화 환자 또는 모형을 통한 임상수기의 수행 여부가 중심이기 때문에 환자와의 면담을 통한 평가나 환자와의 의사소통에 관한 교육이 어렵다. 이에 반해 CPX는 실제 임상 상황과 비슷한 환경에서 학생들이 표준화 환자를 직접 대하는 기회를 갖게 함으로써 의사소통능력, 환자교육, 정보통합능력 등 OSCE보다 더 포괄적으로 임상수행능력을 측정하고 그에 따른 교육을 가능하게 한다.

임상수행능력을 평가하기 위해 여러 국가에서 다양한 방식의 임상수행평가를 의사면허나 의과대학 졸업자격을 부여하기 위한 총합 평가의 일환으로 도입하는 경우가 점차 늘어나고 있다[2]. 국내에서도 2010년도부터 의사면허시험에서 임상실기시험이 도입되어 시행되고 있고, 이로 인해 많은 의과대학들은 임상수행능력을 높이기 위해 여러 가지 형태의 교육과 평가를 실시하고 있다. 의사의 실제 직무 상황에 맞는 임상수행능력을 평가하기 위해서 가장 중요하게 간주되어야 하는 것이 환자와의 만남이고[3], 환자와의 만남을 통해 능력을 평가하는 방법 중 하나가 CPX이다.

하지만, 많은 학교에서 임상실습을 마친 후에 임상실습의 결과로 평가를 하거나 의사면허 시험에 임박해서 CPX를 시행하는 경우가 많다. 시행 시기에 있어서 쏠림 현상이 있다는 것은 CPX를 임상수행능력을 갖출 수 있도록 하는 교육의 일환으로 보기보다는 단순히 의사면허시험을 준비하기 위한 평가의 측면만을 강조하여 CPX가 갖고 있는 교육의 효과성이나 잠재성이 사장되고 있는 듯하다. 또한, 이렇게 CPX의 평

가 측면만이 강조되다 보니 CPX에 관한 국내 연구의 대부분은 평가와 관련된 신뢰도 및 타당도를 검증하는 연구들이거나 이에 영향을 미칠 수 있는 요인들에 대한 연구가 주로 이루어져 왔다. 국외 의학교육의 경우, 평가에 관한 연구 외에 CPX가 갖고 있는 교육의 효과성을 보기 위한 CPX 반복 시행에 따른 학생들의 임상수행능력 향상에 관한 연구 등[4] CPX를 교육적으로 활용하기 위한 노력에 적극적이다.

본 연구는 CPX를 여러 차례 경험한 의대생들의 병력청취, 신체진찰, 환자-의사관계(patient-physician interaction, PPI) 점수가 어떻게 변화되는지 분석하여 CPX가 갖고 있는 교육의 효과성에 대해 연구하고자 한다. 이를 위해 서로 연관성이 높으면서도 각기 다른 능력을 필요로 하는 병력청취, 신체진찰, PPI 영역의 점수 변화를 관찰하여 현재 학생 교육에서 취약한 영역을 확인하고, 성적이 높거나 낮은 학생들의 각 영역별 점수 변화 추이를 비교, 분석하여 더 전략적인 훈련방법에 대해 고찰하고자 한다.

연구 문제는 다음과 같다. 1) 3회에 걸친 CPX 시험에서 각 영역별 점수는 시간경과에 따라 변화 양상이 어떻게 다른가? 2) 학생들의 진료수행능력 수준을 상/중/하 3개의 그룹으로 구분했을 때, 각 그룹의 영역 점수의 변화 양상에 차이가 있는가?

## 대상 및 방법

### 1. 연구 대상 및 설계

연구 대상은 2010학년도 서울 소재 K대학교에 재학하는 의과대학/의학전문대학원 4학년 학생 113명이었다. 2009년 12월부터 2010년 9월까지 약 9개월에 걸쳐 시행된 3차례의 CPX 성적 변화를 추적하는 종단연구로 설계되었다. 이들 중 연구 참여에 동의하고, 3회의 CPX를 모두 치른 101명의 학생들의 성적을 분석하였다. 연구 대상자 중 여학생이 39명(38.6%)이고, 남학생이 62명(61.4%)이다. 의학과 재학생은 53명(52.5%)이었고 의학전문대학원 재학생은 48명(47.5%)이었다. 이들은 1차 CPX 이전에 3학년 임상실습과정을 마친 상태로 내과, 외과, 산부인과, 소아과, 정신과에서 임상실습을

Table 1. Cases Used in CPX

	1st CPX	2nd CPX	3rd CPX
Cases used	Dizziness Anxiety Weakness Menstrual irregularity Acute abdominal pain Chest pain Low back pain Fatigue Chest discomfort	Polyuria Hematuria Fever & Chill Headache Chest pain Heartburn Dyspnea Neck pain Abdominal pain	Diarrhea Hematemesis Hand tremor
Excluded cases	Vaccination Domestic violence Bad news delivery	Insomnia Vaccination Smoking cessation counseling	Insomnia
Total	9/12 cases	9/12 cases	3/4 cases

CPX, Clinical performance examination.

하였으며, 임상실습 기간에 내과와 정신과에서 교육과 형성 평가의 목적으로 CPX를 경험하기도 하였다. 이 외에도 3학년 1학기에 의료커뮤니케이션 수업에서 표준화 환자를 대상으로 의사소통법을 교육받았으며 CPX를 경험하였다. 능력에 따른 CPX 성적 추이 분석을 위해 연구대상을 세 그룹으로 분류하였다. 그룹의 구분은 3회의 CPX 총점을 기준으로 상(상위 30%, n=31), 중(중위 40%, n=40), 하(하위 30%, n=30)로 하였다.

1차 CPX는 2009년 12월에 실시되었고, 2차는 2010년 7월에, 3차는 2010년 9월에 실시되었다. 3회의 시험 사이에는 각각 7개월과 2개월의 시간차가 있었다. 1차와 2차 시험 사이 7개월 동안 학생들은 4학년 과정의 임상실습에 참여하였고, 2차 시험 직전에 2일간의 복습과정을 갖고 교수의 피드백을 받았다. 또 학생들은 2차와 3차 시험 사이 2개월 동안에는 의사 국가실기시험을 대비한 5일간의 '여름 방학 특별 교육과정 (summer skills lab)'에 그룹 혹은 개인단위로 참여하였다. 요약하면 본 연구는 자료 수집 시기를 등간 배치하지 않았고, 자료에 영향을 줄 수 있는 교육 등의 요소들을 엄격하게 통제하지 않았다.

## 2. CPX

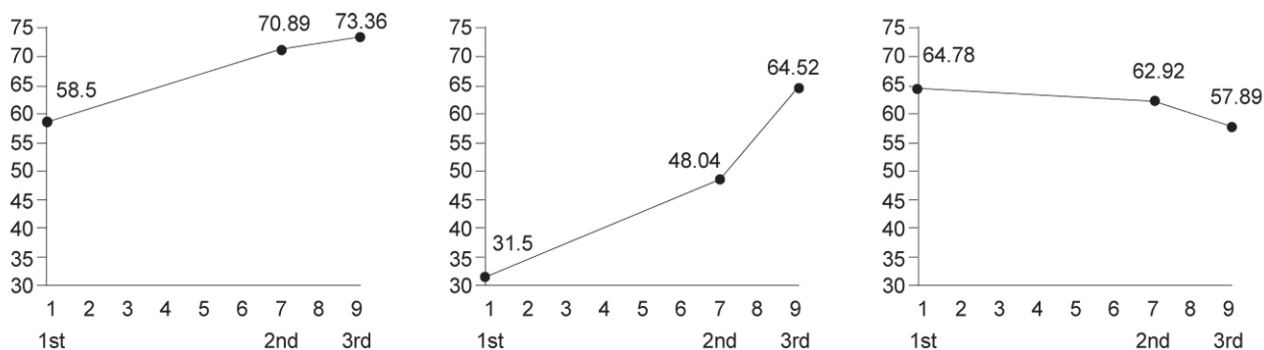
1회차 CPX는 2009년 12월 28일부터 3일 동안 실시하였고, 2회차는 2010년 7월 12일부터 5일 동안, 그리고 3회차는

2010년 9월 1일부터 5일간 시행하였다. 이 중 1, 2회차 시험의 준비와 운영은 서울·경기 CPX 컨소시엄이 주관하였으며, 3회차 시험은 컨소시엄의 방법과 기준에 준하여 K대학이 주관하여 증례를 개발하고 시험을 진행하였다.

시험의 진행은 증례별로 상황소개 숙지 1분, 표준화 환자 대면 10분, 사이 시험 5분이라는 컨소시엄의 방식에 따라 시행하였으며, 표준화 환자의 행동지침이나 채점 등 세부적인 사항도 컨소시엄의 방식을 따랐다. 한 가지 차이점이 있다면 컨소시엄이 주관한 1, 2차 CPX에서는 표준화 환자가 학생의 수행을 평가하고, 학생들의 태도에 국한하여 서면 피드백을 주는 것과 달리 대학이 주관한 3차 CPX에서는 표준화 환자와 더불어 임상교수가 평가에 참여하여 매 스테이션을 마친 후 2분간 학생들의 전반적인 수행에 대한 구두 피드백을 제공하였다.

CPX는 증례에 따라 22~38개의 문항으로 구성된 채점표를 기준으로 평가가 이루어진다. 채점표는 전반적 평가, 병력청취, 신체진찰, 임상 의사 예절, PPI 등의 기본적인 하위 영역을 포함하며, 증례에 따라 환자교육, 나쁜 소식 전하기, 정신과 평가 등의 영역이 추가되기도 한다. 본 연구는 병력청취, 신체진찰, PPI 3개의 하위 영역의 점수에 초점을 맞추고 있어 시험에 사용된 증례 중 위의 3개 하위 영역을 포함하지 않는 증례를 분석에서 제외하였다. 각각 12개 증례로 구성된 1차, 2차 CPX에서는 9개 증례를, 3차 CPX에서는 3개 증례를 분석

Fig. 1. Scores on History Taking, Physical Exam, and PPI



PPI, Patient-physician interaction.

하였다(Table 1).

각 하위 영역들은 각기 다른 평정척도로 채점하였다. 병력 청취 문항은 예(1)/아니오(0)의 2점 척도, 신체진찰 문항은 제대로 했음(2)/제대로 못했음(1)/하지 않았음(0)의 3점 척도, 그리고 PPI 영역의 문항은 최우수(5)부터 수준 미달(0)까지 6 점 척도상에 점수를 기록하였다. 병력청취와 신체진찰은 증례에 따라 완전히 다른 문항들로 구성되나 PPI 문항들은 모든 증례에 7개의 동일한 문항을 사용한다.

### 3. 자료 분석

본 연구에서 사용된 점수는 표준화 환자에 의해 평정된 점수이며, 표준화 환자와 임상교수가 이원 채점한 3회 차 시험의 경우, 점수의 표준화를 위해 표준화 환자의 점수만을 사용하였다. 하위 영역 중 병력청취와 신체진찰의 경우 증례별로 문항수가 다르므로 표준화를 위해 각 증례별 하위 영역 점수를 100점 만점으로 환산하였으며, 각 회차에 포함된 증례들을 통합하여 각 하위 영역 점수의 평균을 구한 값을 최종 회차별 CPX 점수로 분석하였다.

3개 시점의 자료를 분석하기 위해 반복측정 분산분석(repeated ANOVA)을 하였다. 분석 결과 Mauchly 구형성을 만족할 경우 개체 내 효과를 참조하고, Mauchly 구형성을 만족하지 않을 경우 다변량 검정(Wilks's lambda)값을 참조하였다. 연구 문제에 따라 하위 영역과 회차 간의 상호작용 효과와 그룹과 하위 영역과 회차 간 상호작용 효과를 분석하였다.

통계적 유의수준은 0.01에서 이루어졌으며 모든 자료의 분석은 SPSS version 12.0 (SPSS Inc., Chicago, USA)을 사용하였다.

## 결과

### 1. CPX 반복에 따른 영역별 점수의 변화

병력청취 및 신체진찰 영역은 CPX를 여러 차례 경험하면서 점수가 점차 향상 되었다(Fig. 1). 병력청취 영역의 경우, Mauchly의 구형성 검정결과( $\chi^2=13.197$ ,  $p<0.01$ )에 따라 참조한 Wilks's lambda값을 보면 3개의 시점에서의 영역점수에는 통계적으로 유의한 차이가 있었다( $F=130.786$ ,  $p<0.01$ ) (Table 2).

신체진찰 영역의 경우에는 Mauchly의 구형성 검정결과( $\chi^2=7.410$ ,  $p>0.01$ )에 의해 개체 내 효과 검정값을 참조하였다. 신체진찰 영역 역시 3개 시점의 점수가 통계적으로 유의하게 증가하였다( $F=237.358$ ,  $p<0.01$ ) (Table 2, Fig. 1).

PPI 영역은 병력청취나 신체진찰 영역과는 달리 시험이 반복될수록 점수가 점차 낮아졌다(Fig. 1). 통계적 검증은 Wilks's lambda값을 참조하였는데, 3개 시점의 PPI 점수가 유의하게 달랐다( $F=17.621$ ,  $p<0.01$ ) (Table 2).

Table 2. Results of Repeated Measures ANOVA and History Taking, Physical Exam, PPI with Scores of Mauchly's Test

Domain	Source	Mauchly's test of sphericity (p-value)	Mauchly's test of sphericity ( $\chi^2$ )	Wilks's lambda value (F-value)	Within-subject effect (df)	Within-subject effect (mean square)	Within-subject effect (F-value)	Within-subject effect (p-value)
History taking	No. of CPX	0.001	13.197	130.786				
	No. of CPX × Student level				4	54.695	1.472	0.212
	Error (No. of CPX)				196	37.157		
Physical exam	No. of CPX	0.025	7.410				237.358	0.000
	No. of CPX × Student level				4	244.331	2.199	0.071
	Error (No. of CPX)				196	111.117		
PPI	No. of CPX	0.000	80.856	17.621				
	No. of CPX × Student level				2.555	739.780	11.628	0.000
	Error (No. of CPX)				125.199	63.620		

PPI, Patient-physician interaction; CPX, Clinical performance examination.

## 2. 학생 수준과 시간 경과에 따른 영역별 점수의 변화 양상

### 1) 병력청취와 신체진찰 영역

병력청취의 경우 시간 경과에 따라 3회의 평균점수의 차이가 있지만( $F=130.786$ ,  $p<0.01$ ), CPX 차수와 학생 수준 간에는 유의한 상호작용이 없었다(Wilks's lambda value:  $F=1.472$ ,  $p>0.01$ ) (Table 2).

신체진찰 영역도 반복측정 시점에 따라 3회의 평균점수에 차이가 있지만( $F=237.358$ ,  $p<0.01$ ), CPX 차수와 학생 수준 간에는 유의한 상호작용이 없었다(개체 내 효과 검정:  $F=2.199$ ,  $p>0.01$ ) (Table 2).

요약하면, 병력청취 영역과 신체진찰 영역에서는 학생 수준에 따른 차이는 없고, CPX 반복측정에 따른 변화가 있음을 알 수 있다.

### 2) PPI 영역

PPI 영역에서는 반복측정 시점에 따라 3회의 평균 점수에 차이가 있고( $F=17.621$ ,  $p<0.01$ ), CPX 차수와 학생 수준 간 상호작용(교호작용)의 효과도 유의미하게 나타났다(Wilks's

lambda value:  $F=11.628$ ,  $df=2.555$ ,  $p<0.01$ ). 이는 임상수행능력의 높고 낮음에 따라 학생들의 PPI 영역 점수가 다른 양상으로 변화했음을 의미한다(Table 2).

Fig. 2를 살펴보면 상위 그룹은 최종 CPX에서 상승하는 추세를 보여주는 반면, 중·하위 그룹은 하락하는 양상을 보였으며, 그 중 하위 그룹의 하락폭이 더 컸다.

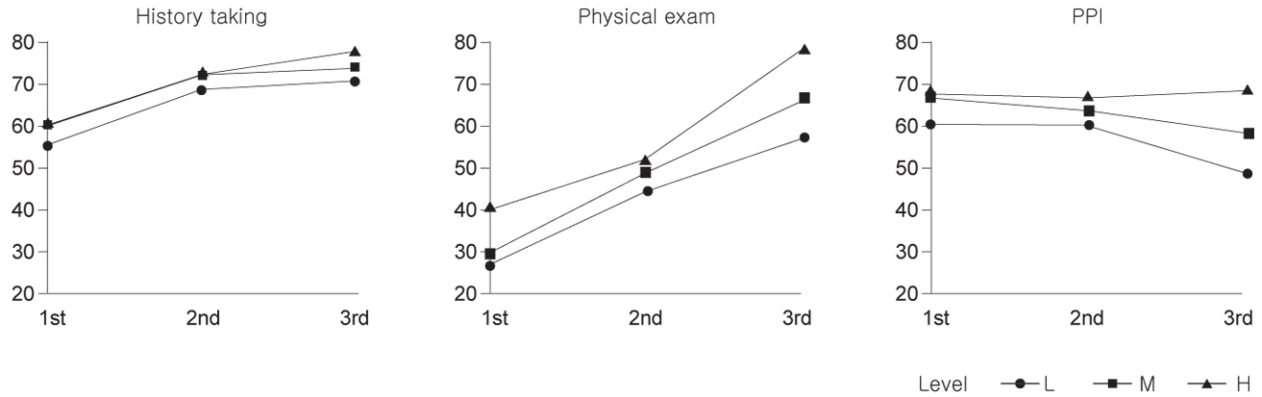
## 3. 시간 간격을 조정한 그래프 분석 결과

본 연구는 9개월 동안 3회에 걸쳐 시행된 CPX 결과를 이용하여 반복측정 분산분석 방법을 사용하여 분석하였다. 반복측정 분산분석은 본 연구에서처럼 반복 측정된 시점의 간격이 일정하지 않을 경우 간격의 차이에 따른 효과를 포함하는 결과를 주지 않는 단점이 있다.

이러한 한계점은 위계적 선형모형(hierarchical linear model, HLM)을 사용하여 보완할 수 있으나 HLM은 표본수가 충분히 많지 않은 연구에는 적용하기 어려운 또 다른 문제점을 가지고 있다. 연구자는 표본수 부족으로 HLM 분석방법을 적용하지 못했으나, 시점 간 간격 차이가 가지는 의미와 의사면허시험에 임박한 3번째 시점의 점수의 의미를 해석하기

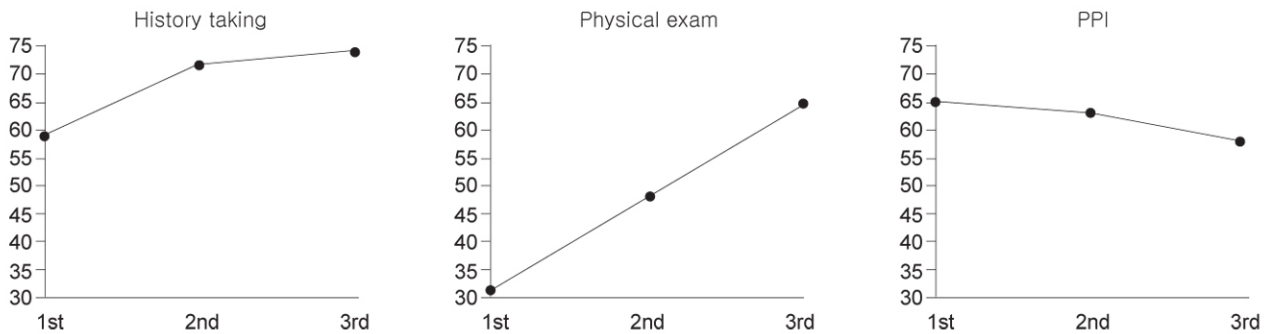


Fig. 2. Patterns of Scores by Ability Level



PPI, Patient-physician interaction.

Fig. 3. Adjusted Plots of CPX Scores



PPI, Patient-physician interaction.

위해 시점의 간격차를 조정한 그래프를 분석하였다.

Fig. 3에 따르면 병력청취는 9개월간 변화 폭에 차이가 없이 완만하게 점수가 향상되었다. 신체진찰은 1회차의 눈에 띄게 낮은 점수가 큰 폭으로 향상되는 변화 양상을 보이는데 특히 2회차와 3회차 사이의 2개월이라는 짧은 기간 동안 급격하게 점수가 향상되었다. 마지막으로 PPI의 경우에는 시간이 경과함에 따라 점차 점수가 낮아지는 경향을 보였으며, 특히 2회차와 3회차 사이에 하락폭이 컸다.

## 고찰

본 연구는 반복측정 분산분석을 이용하여 CPX 반복시행에

따른 학생들의 성적 변화를 분석하였다. 병력청취, 신체진찰, PPI 중 병력청취와 신체진찰은 CPX 시행 횟수가 늘어남에 따라 점수가 상승하였고, PPI는 떨어지는 양상을 나타냈다. 또한 학생의 능력을 기준으로 그룹을 나누어 상호작용 효과를 분석한 결과, PPI 영역에서 학생의 능력수준과 시간경과에 따라 성적 변화 양상에 차이가 있었다. 각 영역점수의 변화 양상은 Prislín et al. [4]의 연구와 거의 같은 결과를 보여주었다. 이러한 연구 결과에 대해서 다음과 같이 고찰하였다.

첫째, 병력청취 영역은 종료시점인 3회차의 점수가 세 영역 중 가장 높았고, 점수도 전체적으로 완만하게 상승하였다. 또한 시험의 반복이나 시험 사이의 시간차와 상관없이 회차 간 점수 변화의 폭에 변화가 없었다. 이러한 결과를 통해 학생들이 이 첫 CPX를 경험하기 전 여러 교육과정을 통해 병력청취에

필요한 기본 능력을 갖추었고, 이러한 능력이 3회차의 CPX를 통해 정교화 되었음을 알 수 있다. 이는 K의대에서 진행하고 있는 교육이 학생들의 병력청취능력 향상에 효과가 있었음을 추정할 수 있고, 또 다른 측면에서는 병력청취능력은 단기간에 습득되기 어려우며, 교육과 경험을 통한 내재화를 필요로 하는 영역인 것을 알 수 있다.

한편, Pfeiffer et al. [5]의 연구에서는 사회병력(social history)청취 문항의 점수가 2학년까지는 높아졌다가 그 이후에 하락하는 양상을 보였다. 이는 본 연구와 일치하지 않은 선행연구로서, 두 연구에서 사용된 평가문항이 측정하는 속성이 다르기 때문인 것으로 보인다. 선행연구의 문항들은 환자의 사회 환경 및 행동적 특성에 대한 질문이나 의사소통능력에 대한 평가를 주로 한 반면, 본 연구의 병력청취 문항은 환자의 증상에 대한 의학적 지식을 근거로 질병을 판단하는 질문들을 많이 포함하고 있다. 이에 대해 연구자들은 CPX에서 각기 다른 영역으로 측정되고 있는 병력청취와 PPI가 서로 공통적으로 포함하고 있는 역량이 현재 충분히 다루어지고 있는지에 대해 전반적인 재고가 필요하다고 논의하였다.

둘째, 신체진찰 영역도 병력청취와 마찬가지로 CPX가 반복 시행됨에 따라 점수가 향상되었으나 출발점과 점수 변화의 양상이 병력청취와는 달랐다. 신체진찰의 1회차 출발점 점수가 매우 낮았고, 각 회차 간 변화폭에 큰 차이가 있었다. 1회차 점수가 낮은 것은 연구 대상자들이 3학년 말이었던 당시 시점에서 환자에게 필요로 하는 신체진찰을 행하는 데 필요한 준비가 미흡했음을 의미하고, 이러한 결과는 현재 K의대의 교육과정인 신체진찰 수행능력을 기르는데 충분히 효율적이지 않았다는 추측을 하게 한다. 이와 더불어 학생들의 점수가 2회차와 3회차 사이에 급격하게 높아지는 현상은 자연스러운 학습곡선이 아니므로 추가적인 해석이 필요하다.

2회차와 3회차 사이의 2달 동안 K의대에서는 의사면허실기시험을 대비한 여러 가지 훈련을 실시하였다. 이 기간 동안 학생들은 평소보다 많은 훈련을 통해 각 신체진찰에 해당하는 수기를 익히고, 시험에 출제될 증례들을 학습하였다. 단기간 동안의 많은 훈련을 통해 점수가 크게 향상된 결과는 시간 대비 효율성이 높다는 측면에서 긍정적이나, 시험을 대비하여 그 시험에 맞게 훈련하고 연습한 것이 충분히 내면화되어 실제 임상수행능력으로 표출될 것인지의 문제에 대해서는 회

의적이라 할 수 있다.

Wilkerson & Lee [6]는 CPX에서의 신체진찰은 OSCE에서의 신체진찰과 측정요소가 다름을 주장하였다. 이들은 같은 장기에 대한 신체진찰능력을 OSCE와 CPX로 각각 평가하여 비교하였는데, CPX 점수가 OSCE보다 크게 낮았으며 상관관계를 이루지 않았다. CPX의 신체진찰은 단순히 특정 검사를 정확하게 시행하는 것뿐 아니라, 주어진 상황에 필요한 검사를 판단하고, 환자를 대면하고 있는 특정 맥락에서 이 검사를 적절하게 수행하는 종합적인 능력을 필요로 한다. 이러한 CPX 신체진찰의 종합적 속성을 고려할 때 본 연구에서 나타난 단기간에 이룬 높은 성적 향상은 통합적인 신체진찰 능력이 향상된 것이 아니라 의사면허실기시험에 포함된 제한된 증례에 필요한 신체진찰을 집중 연습하여 얻은 단기적 성과로 보는 것이 합당하다. 현재의 의사면허실기시험을 대비한 집중 훈련은 통합성과 맥락성을 강조하는 CPX 본래의 교육목적과 강점을 오히려 약화시키는 결과를 초래할 수도 있다는 것에 주의를 할 필요가 있다. 특정 시기에 집중되어 총합 평가(summative assessment)의 기능으로 활용되는 CPX를 의대 교육과정 전체에 골고루 분포시켜 반복적인 형성평가(formative assessment)로 활용함으로써 학생들이 임상수행능력을 꾸준히 향상시킬 수 있는 기회가 되도록 할 필요가 있다.

셋째, PPI 영역은 다른 영역과는 달리 CPX 경험이 많아질수록 점수가 하락하는 양상이 나타났으며 후반부에 하락 정도가 더 커졌다. 훈련의 양이 증가할수록 학생의 능력이 감소하는 현상에 대해서는 추가 연구가 필요하다.

Pfeiffer et al. [5]의 연구에 따르면 면담능력(interview skill), 의사소통 및 관계형성(communication & rapport)점수는 3학년까지 증가하다 4학년 때 감소하였다. 연구자들은 이러한 점수의 하락을 의과대학에 존재하는 숨겨진 교육과정(hidden curriculum)의 영향에 의한 것으로 설명하였고, 학생들이 임상실습을 하는 동안 적절한 PPI를 보여주는 역할 모델(role model)을 발견하지 못하는 현실에 대해 지적하였다[5]. 국내 연구 중에서도 Ahn et al. [7]은 전공의의 임상경험 부족과 부적절한 태도가 임상실습 중인 의대생들에게 부정적인 영향을 주며, 낮은 PPI 점수로 나타난다고 하였다.

본 연구에서 나타난 PPI 성적의 변화도 위의 선행연구와

같은 맥락의 해석이 가능하다. K의대 학생들도 임상실습에 노출될수록 PPI에 대한 부정적인 인식을 가질 기회가 많았고, 모범적 멘토(mentor)의 역할을 하기에는 시간이나 경험이 부족한 전공의로부터 훈련을 받으며 점차 숨겨진 교육과정에 익숙해진 결과가 PPI 점수의 하락으로 나타난 것으로 보인다.

한편, 선행연구 중에는 신체진찰 점수와 PPI 점수 사이에 부적 상관을 밝힌 흥미로운 연구가 있다[8]. 즉, 신체진찰을 정확하고 능숙하게 잘 하는 학생의사일수록 환자를 대하는 태도에 대해서는 낮은 점수를 받은 것인데 이는 유능한 의사에 대한 학생들의 왜곡된 인식 때문인 것 같다. 학생들은 질병을 능숙하게 다루는 능력을 환자를 편안하게 대하는 능력보다 우위에 두고 있어, 전자를 높이기 위해 후자를 소홀히 하는 경우가 많다. 특히 의사면허시험이 가까워 옴에 따라 환자와의 의사소통과 같은 비인지적 요소보다는 의학적 지식과 기술과 같은 인지적 요소에 더 초점을 맞추게 됨으로써 PPI에 대한 중요성을 간과하기 쉽다.

넷째, 학생들을 CPX 총점에 따라 상, 중, 하, 세 그룹으로 구분하고, 그룹과 성적 변화 간 상호작용을 분석한 결과에서는 PPI 영역에서만 그룹에 따라 변화 양상이 다른 것으로 나타났다. PPI는 신체진찰과 병력청취 영역과는 달리 비인지적 영역에 해당한다. 이러한 비인지적 영역은 일반적으로 다른 영역에 비해 잘 변하지 않는 특성이 있고[9], PPI 점수의 차이는 개인차에 기인한 부분이 크기 때문에[10] 시험반복과 학생수준에 따른 상호작용에서 PPI만 유의미한 결과를 나타낸 것으로 보인다. 상위그룹은 성적이 향상된 반면, 중·하위 그룹은 하락하였고, 특히 하위 그룹이 눈에 띄게 하락폭이 컸다. 이러한 결과는 PPI 점수가 전체 임상수행능력과 높은 상관을 나타낸다는 Han et al.[10]의 연구와 같은 결과로 PPI가 임상수행능력을 대표하는 중요한 영역임을 확인하였다. 그러나 상위그룹과 하위그룹 학생들의 어떤 특성이 후반부의 PPI 성적에 영향을 미친 것인지에 대해서는 설명하기 어려웠다. Han et al. [10]의 연구에 따르면 PPI 점수는 신체진찰에 비해 병력청취와 상관성이 높았다. 병력청취를 '환자가 자유롭게 하고 싶은 말을 하게하고 의사는 경청자의 역할을 함으로써 환자가 진정으로 원하고 필요로 하는 것이나 선호하는 것을 환자 수준에서 이해해 주고 적절히 반응해주기 위한 대화라고 정의하는 것에서[11] 보듯이 병력청취는 환자와의 의사소

통과 관계 형성을 중요한 요소로 포함한다. 그러므로 구조화된 평가표의 평가항목이 비인지적 측면에 초점을 맞춘 PPI 영역과 인지적인 측면에 초점을 맞춘 병력청취가 의사소통능력이라는 통합적인 능력을 적절히 평가하고 있는지에 대해서 고려해 볼 필요가 있음을 보여준다.

본 연구의 제한점은 다음과 같다. 첫째, 본 연구의 CPX 성적은 시험 반복에 의한 영향뿐 아니라 그 이외의 다른 교육, 훈련의 영향을 받았다. 이는 연구의 객관성을 저해할 수 있는 요소이나 현장성이 강한 교육연구에서 불가피하게 수용해야 하는 요소이기도 하다. 연구자는 이러한 연구의 특성을 인지하고, 연구 결과에 영향을 줄 수 있는 현장적인 요소들을 다음과 같이 밝힌다. 먼저, 총 3회의 시험 시행에 있어서 1, 2차 시험은 서울·경기 CPX 컨소시엄에서 주관하였고, 3차 시험은 K대학에서 자체적으로 주관하였다. 또한 3회의 시험 시행 간격이 7개월과 2개월로 일정하지 않았고, 2회차와 3회차 사이에는 의사면허실기시험을 앞두고 개별 혹은 그룹 단위의 다양한 밀도 있는 훈련 프로그램이 제공되었다.

둘째, 자료 분석에 사용한 반복측정 분산분석의 한계이다. 반복측정은 자료를 수집한 시점 간의 시간차를 고려하지 않는다. 반복측정의 이러한 한계는 HLM을 사용하여 보완할 수 있지만, 적은 표본수를 대상으로 하는 연구에는 적절치 않다 [12]. Kreft의 30/30 법칙에 따르면 HLM에서 0.90의 파워를 얻기 위해서는 적어도 30개의 그룹을 요하며 각 그룹당 30명은 있어야 하고, 수준 간 상호작용을 보고자 한다면 각 그룹당 20명의 개인이 있는 50개의 그룹이 있어야 한다고 하였다 [13]. 이러한 이유로 본 연구에서는 반복측정 분산분석을 사용하였으나 시험시점 간 시간차가 학생들의 성적 변화 양상에 영향을 미치므로 그래프의 X축 척도에 시간차를 포함시켜 그래프를 조정하여 전체적인 변화 양상을 분석하였다(Fig. 3).

셋째, 시험에 사용된 사례와 그에 따른 난이도 및 표준화 환자에 대한 영향을 고려하지 않았다. 본 연구에서는 총 21개 사례에서 표준화 환자에 의해 평정된 점수만을 사용하였는데, 각 사례별로 표준화 환자의 나이나 성별이 다르고, 이는 학생들을 평가하는 데 있어서 영향을 줄 수 있을 것이다. Kim et al. [14]은 진료수행평가에 사용되는 증례의 특성과 표준화 환자의 성별은 평가 결과에 영향을 미칠 수 있다고 하였고,



Lee et al.[15]은 PPI 측면에서 학생들이 '잘하는 것'과 '취약한 점'은 CPX 사례에 따라 달라질 수 있다고 하였다. 또한 Park et al. [16]은 시나리오와 평가표의 난이도가 문제에 따라 다를 수 있기 때문에 이러한 차이들이 채점 정확도에서 영향을 줄 수 있다고 하였다.

본 연구를 통해 병력청취는 통합적 훈련효과가 점진적으로 나타나는 영역이며, 신체진찰은 훈련의 맥락성과 적응능력이 중요하고, PPI는 현재의 교육환경에서 교육적 투입이 높아질수록 성도가 낮아지는 역설적인 결과를 내는 영역이라는 점과 개인특성의 반영이 높은 영역이라는 점을 밝혔다. 각 영역에서 발견된 이러한 교육적 요구들은 학생들이 지식의 습득과 적용을 분리하지 않고, 의대 교육과정을 통해 학습한 것을 임상현장에서 적절히 활용할 수 있는 교육환경을 마련함으로써 만족될 수 있을 것이다.

**Acknowledgements:** None.

**Funding:** None.

**Conflicts of interest:** None.

## REFERENCES

1. Chang A, Boscardin C, Chou CL, Loeser H, Hauer KE. Predicting failing performance on a standardized patient clinical performance examination: the importance of communication and professionalism skills deficits. *Acad Med* 2009; 84: 101-104.
2. MacRae HM, Vu NV, Graham B, Word-Sims M, Colliver JA, Robbs RS. Comparing checklists and databases with physicians' ratings as measures of students' history and physical-examination skills. *Acad Med* 1995; 70: 313-317.
3. Park JH. A psychometric evaluation of CPX in relation to validity and reliability [dissertation]. [Seoul, Korea]: Ewha Womans University; 2008.
4. Prislin MD, Giglio M, Lewis EM, Ahearn S, Radecki S. Assessing the acquisition of core clinical skills through the use of serial standardized patient assessments. *Acad Med* 2000; 75: 480-483.
5. Pfeiffer C, Madray H, Ardolino A, Willms J. The rise and fall of students' skill in obtaining a medical history. *Med Educ* 1998; 32: 283-288.
6. Wilkerson L, Lee M. Assessing physical examination skills of senior medical students: knowing how versus knowing when. *Acad Med* 2003; 78(10 Suppl): S30-S32.
7. Ahn S, Lee YM, Ahn DS. A study on how young doctors and patients perceive the doctor-patient relationship. *Korean J Med Educ* 2006; 18: 279-287.
8. Stratton TD, Elam CL, Murphy-Spencer AE, Quinlivan SL. Emotional intelligence and clinical skills: preliminary results from a comprehensive clinical performance examination. *Acad Med* 2005; 80(10 Suppl): S34-S37.
9. Kim D, Yoon H, Lee M, Ahn J, Lee S, Suh Y. Comparison of patient-physician interaction (PPI) evaluation between different grade medical students. *Korean J Med Educ* 2010; 22: 169-176.
10. Han JJ, Lee MJ, Im H. Comprehension of patient-physician interaction through analysis of relationships between domains in clinical performance examination. *Korean J Med Educ* 2010; 22: 177-184.
11. Smith RC. Patient centered interviewing: an evidence-based method. Philadelphia, USA: Lippincott Williams & Wilkins; 2002.
12. Pollack BN. Hierarchical linear modeling and the "unit of analysis" problem: a solution for analyzing responses of intact group members. *Group Dyn* 1998; 2: 299-312.
13. Kreft IG. Are multilevel techniques necessary? An overview, including simulation studies. Los Angeles, USA: California State University; 1996.
14. Kim J, Lee K, Yoo D, Yang E. Effects of case type and standardized patient gender on student performance in a clinical performance examination. *Korean J Med Educ*

- 2007; 19: 23-30.
15. Lee YH, Lee YM, Kim BS. Content analysis of standardized-patients' descriptive feedback on student performance on the CPX. *Korean J Med Educ* 2010; 22: 291-301.
16. Park WB, Kim AR, Lee SA, Kim EA, Chang SY, Kim YS, Kim SW, Shin JS, Lee YS. The accuracy of standardized patients' rating according to the order of examinees in clinical performance examination. *Korean J Med Educ* 2007; 19: 39-45.