

# 진료수행시험 (CPX: Clinical Performance Examination)에서 정보공유가 시험성적에 미치는 영향

한양대학교 의과대학 의학교육학교실

박 훈 기 · 권 오 정

= Abstract =

## Sharing of Information among Students and Its Effect on the Scores of Clinical Performance Examination (CPX)

Hoon Ki Park, MD, Oh Jung Kwon, MD

*Department of Medical Education, Hanyang University College of Medicine*

**Purpose:** During the high-stake examinations such as OSCE (Objective structured clinical examination) or CPX (clinical performance exam), test security is generally accepted as a major concern for test validity. This study was conducted to investigate the effect on examinee's scores of repeated, serial administrations of essentially the same standardized patient (SP)-based performance exam.

**Methods:** A performance-based examination using eight SP cases was administered to 123 senior medical students at Hanyang University School of Medicine. Students were randomly assigned to one of 16 groups of 8 students each. Three groups were tested serially each day, requiring 5 days for the complete administration of the examination. We compared the mean scores of the five groups of the examinees tested on different days with ANOVA and linear trends with multiple regression analyses.

**Results:** For both checklist scores and written scores during the interstation work, the mean scores of the first day groups were significantly lower compared to subsequent groups. And, there were slight linear trends in the scores over the five days. Scores related to case-specific history taking, information sharing, and clinical courtesy were significantly affected by the sharing of information between students. Scores related to patient satisfaction, physical exam, and physician-patient interaction were not influenced by the same pattern of behaviour.

**Conclusion:** Test security may be violated during SP-based performance exams even though the checklists are not accessible to the examinees. It would be desirable for the test-givers to prepare alternative forms of cases for maintaining the validity of SP-based performance exams.

**Key Words:** Clinical competence, Performance, Evaluation, Standardized patients, Validity

교신저자: 박훈기, 한양대학교 의과대학 가정의학과, 서울시 성동구 행당동 17번지

Tel: 02)2290-8740, Fax: 02)2281-7279, E-mail: hoonkp@hanyang.ac.kr

\* 이 논문은 2004년 홍릉의과대학 주최 의학교육평가 심포지움에서 구연 발표하였음.

## 서 론

객관구조화진료시험 (Objective structured clinical examination: 이하 OSCE)은 지식, 태도, 수기 등을 동시에 측정할 수 있는 시험 형태로 처음 소개된 이후 의학교육에서 평가 뿐 아니라 술기교육 방법으로 발전해왔다 (Harden et al, 1975). 표준화 환자는 Barrows가 1963년 처음 창안한 이후 교육, 평가에 널리 활용되고 있다 (Barrows, 1993). 표준화 환자를 이용한 임상수행능력평가는 진료수행시험 (Clinical performance examination: 이하 CPX) 형태로 졸업 전 의학교육 뿐 아니라 졸업 후 의학교육, 평생교육에 이용되고 있으며 의사 자격고사에도 이 평가방법이 도입되었다. OSCE와 CPX를 시행할 때 수험생이 많아지면 동일한 증례에 대해 표준화 환자와 스테이션을 필요한 만큼 여러 세트 복제하지 않는 한 며칠 동안 연속적으로 시험을 시행해야 한다 (박훈기 등, 1999). 표준화 환자 이용시험 운영에서 시험의 보안유지정도는 시험의 공정성과 타당도에 중요한 영향을 준다. 특히 표준화 환자 이용 시험을 임상실습 순환교육 평가나 국가시험으로 시행할 경우 동일한 문제를 계속 사용하면 시험에 관한 정보공유는 시험의 공정성을 위협할 수 있다.

다양한 형식의 OSCE가 수험생의 당락을 결정하는 시험으로 이용되면서 시험의 보안유지 (test security)가 커다란 문제로 대두되었다. OSCE에서 수험생들 사이에 정보교환이 어느 정도 일어나긴 하지만 이러한 정보교환이 시험성적에 큰 영향을 주지는 않는다고 보고하고 있다 (Williams et al, 1992). 임상실습 OSCE 문제에 대한 학생들 사이의 정보교환과 이에 따르는 시험성적의 변화는 거의 없는 것으로 알려져 있다 (Rutala et al, 1991; Vu et al, 1992). 전공의 대상 CPX에서는 초반부에 시험을 치른 수험생이 후반부에 시험을 치른 수험생보다 유리했다는 증거를 발견하지 못했다 (Stillman et al, 1991). 남일리노이 의대생 70여명을 5조로 나누어 5주간에 걸쳐 5년 동안 시행한 CPX에서 시험 일자에 따라 체계적이고 일정한 성적향상 경향은 발견할 수 없었다 (Colliver et al, 1991). 하지만 이러한

결과는 의미를 둘 만큼 광범위하고 심각한 시험보안문제가 없었다는 것을 의미할 뿐이지 증례에 따라서는 시험 일자별로 의미 있는 성적향상경향이 있다는 사실을 부정하지는 못하고 있다. 외과임상실습교육 평가시험으로 시행한 15개 스테이션으로 구성된 OSCE에서 전반적으로는 시험일정에 따라 일정한 성적상승경향은 없었지만 3개 스테이션에서는 성적 상승경향을 보였다 (Niehaus et al, 1996). 캐나다 온타리오 주 인턴선발고사로 시행한 30개 스테이션으로 구성된 OSCE에서 스테이션의 반복 사용은 단일 스테이션보다 연결형 스테이션에서 성적향상을 가져올 수 있었다 (Cohen et al, 1993). 하지만 동일한 문제라도 적어도 1년 이상 간격을 두고 사용하면 성적에는 거의 영향을 주지 않았다. OSCE에서 환자대면 스테이션 보다 스테이션과 연결되어 있는 지필고사가 시험의 보안여부에 따라 성적영향을 많이 받는다고 알려져 있다 (Gledhill & Capatos, 1985).

표준화 환자를 이용한 시험에서 환자의 문제에 대해 정보를 어느 정도 알고 있더라도 점검표 (checklists)를 모르면 기능향상에 큰 도움이 되지 않는다고 알려져 있다. 형성평가로 실시한 OSCE에서 시험공부는 성적향상에 큰 도움을 주지 않고 오히려 이전까지 축적된 학업성적이 OSCE와 상관성을 보였다 (Mavis, 2000). 시험에 대한 피드백을 구체적으로 주는 것 자체가 시험성적에 큰 영향을 주지는 않았다 (Furman et al, 1997). 하지만 학생들 사이에서 자세한 정보교환이 일어났을 때 시험 성적에 어떤 효과가 나타나는 지 아직 명확히 밝혀진 바는 없다. 표준화 환자 이용 시험에서 처음 시험을 본 그룹이 다음 그룹에게 시험정보를 구체적으로 알려주었을 때 양군간의 시험성적 차이는 유의하지 않았다 (Swartz et al, 1993). OSCE의 문제 내용이 학생실습 학습목표와 일치한다면 사전실습교육시간은 OSCE성적에 큰 영향이 없다고 알려져 있다 (Thomas & Shatzer, 2000). 임상의학입문과정에서의 경험 역시 OSCE 성적과 상관성이 0.2이하로 매우 낮은 것으로 알려져 있다 (Jolly et al, 1996). 하지만 외과영역의 수기에 관한 경험과 피드백의 양은 OSCE

성적을 향상시킨다는 보고가 있다 (Chatenay et al, 1996). 수험생간 문제의 사전 유출을 해결하기 위해 OSCE와 슬라이드 시험 혹은 지필고사를 직렬적으로 연결하여 운영하기도 한다. 남알리노이의대에서는 매년 50~70%씩 새로운 문제로 바꾸어 임상실습 종합평가로서 표준화 환자를 이용한 수행시험을 시행하고 있다 (Vu et al, 1992).

표준화 환자를 이용한 시험인 CPX에서 시험의 보안유지여부가 시험성적에 어떤 영향을 주느냐에 대한 지금까지의 연구결과는 세 가지로 요약될 수 있다. 첫째는 시험의 내용에 대한 수험생간의 정보교환은 일어날 수 있지만 적어도 일자별로 심각한 수준으로 일정하고 체계적인 성적상승 경향은 볼 수 없다. 둘째, 환자를 진료하는 스테이션보다는 필기시험 형식의 스테이션이 더 많이 시험정보공유의 영향을 받을 수 있다. 셋째, 시험의 목적에 따라 정보공유의 정도가 달라질 수 있고 이에 따른 CPX 성적 차이를 가져올 수도 있다. CPX를 시행 경험이 짧은 우리나라 현실에서 시험의 보안유지가 CPX성적에 어떤 영향을 줄 수 있는지에 대한 연구는 아직 없다.

이 연구는 표준화 환자를 이용한 임상수행평가에서 시험날짜에 따라 학생 성적 차이가 날 수 있는지 알아보고 시험에 대한 준비와 정보공유가 진료수행시험에 어떤 영향을 주는 지 알아보고자 했다.

### 대상 및 방법

한양대학교 의과대학에서는 2004년 8월 16일부터 5일 동안 4학년 학생 123명을 대상으로 임상실습을 모두 마치고 선택실습을 하는 기간에 CPX를 총합 시험 (summative examination)으로 시행하였다.

CPX는 8개 시험방 (station)과 8개의 사이 시험방 (interstation)으로 구성되었다. 일차의료에서 흔히 볼 수 있는 의학적 문제 8개가 서울CPX컨소시엄을 통하여 선정되었으며 표준화 환자 훈련 역시 컨소시엄에서 담당하였다. 문제 당 점검표 (checklists) 문항수는 평균 29개 (범위 23~34), 총 232개였다. 문항분포는 병력청취 34%, 신체진찰 14%, 의사에절 9%, 정보공유 (환자교육) 13%, 환자-의사관계 24%,

기타 (총괄채점)문항이 6%를 차지하였다. 모든 학생은 5일 동안 동일한 8개 문제를 접하였다. 한 명의 표준화 환자가 하루에 12명 이하의 학생을 대면하였으며 총 32명의 표준화 환자가 참가했다. 학생들이 사이시험을 보는 5분 동안 표준화 환자는 기억에 의존하여 점검표를 작성하고 구체적인 피드백을 서면으로 작성하였다. 표준화 환자의 훈련은 증례에 따라 4~10회 정도 시행하였다. 사이시험은 문제중심의무기록작성법으로 사정 (assessment), 향후계획 (plan)만 작성하게 하였고 미리 만들어진 채점기준표에 따라 채점을 했다.

학생들은 시험 직전 오리엔테이션을 통하여 CPX의 의미와 시험 요령을 설명 듣고 시험 내용에 대한 비밀유지서약서에 서명을 하였다. 이번 CPX가 성적 반영 측면에서 중요한 비중을 차지하지 않지만 학생 자신의 의사로서의 성숙도와 장단점을 파악하는데 중요한 시험임을 감안하여 서로 정보의 공유를 하지 말 것을 학생들에게 강조했다. 학생들은 사이시험시간이나 휴식시간에 서로 대화를 할 수 없었으며 메모용지는 별도로 배부한 빈 용지만을 사용하게 했고 시험이 끝난 후 사용한 용지는 모두 회수하였다.

학생들은 한 조에 8명씩 총 16조로 나누었으며 마지막 조만 4명으로 구성되었다. 처음 4일 동안은 하루 3개조, 마지막 날에는 4개조가 연속적으로 시험을 보았다. 이전 학점을 포함한 학생들의 조별 특성을 없애기 위해 조배정은 무작위로 하였다.

시험방법 할당 시간은 상황 소개 및 지침 숙지시간 1분, 환자대면진료시간 12분 (10분 짜 예비 신호), 사이시험 시간 5분으로 18분이었다. 4개 시험방을 돈 후 8분의 휴식 시간이 있어 한 학생당 총 시험 시간은 휴식 시간을 포함하여 2시간 32분이었다. 학생들은 시험 일 주전에 CPX 학생오리엔테이션을 받았으며 시험 직전에 조별로 20분 동안 구체적인 시험요령에 대한 오리엔테이션을 받았다. 시험 성적은 2004학년도 4학년 임상실습 시작 후 4주 짜 시행한 OSCE (Objective structured clinical exam) 과목 1 학점에 50% 반영되었다. 시험 직후 문제의 내용에 대한 교수의 피드백이나 표준화 환자의 피드

**Table I.** Students' Preparation for clinical Performance Exam (CPX) over Days of Exam\*

Variables	Day 1	Day 2	Day 3	Day 4	Day 5	P-value
Past credits	3.1 (2.5~4.0)	3.2 (2.4~4.2)	3.2 (2.5~4.2)	3.3 (2.4~4.3)	3.2 (2.4~4.2)	0.73
Perception of exam Goals*	3.6 (2~5)	4.2 (2~5)	3.9 (1~5)	3.9 (2~5)	4.1 (1~5)	0.18
Preparation for CPX*	2.6 (1~5)	3.1 (1~5)	3.0 (1~4)	3.5 (2~5)	3.5 (1~5)	0.03

\* Numbers in parentheses are range (min-max) by 5-point Likert scale

백은 없었다.

CPX를 마친 학생은 시험 전체에 대한 피드백 설문지를 기명으로 작성하였다. 이 설문지에서 CPX 시험 목적 인지정도(“나는 이 시험의 목적에 대해서 잘 알고 있다”)와 시험 전 준비 정도(“나는 이 시험에 대비하여 준비(공부)를 했다”)를 5점 Likert 척도(정말 그렇다 5점, 전혀 그렇지 않다 1점)로 응답하게 하였다.

시험 종료 3주 후 스테이션 성적, 영역별 성적, 당락 판정, 표준화 환자 서면 피드백을 포함한 CPX 성적 보고서를 개인별로 통보해 주었다. 성적 통보 1개월 후 시험기간 동안 정보 공유에 대한 설문을 무기명으로 시행하였다. 질문은 “먼저 시험을 치른 학생으로부터 CPX 내용에 관한 정보를 들었는가?”, 정보를 들었다면 “CPX에 대한 사전정보가 성적향상에 도움이 되었는가?”, 시험공부 시간, 시험 준비 방법, 가장 효과적인 준비방법, 결과에 대한 서면 피드백이 도움이 되었는지 여부 등을 포함하였다.

CPX 성적 합산은 스테이션 성적, 영역별(domain) 성적을 따로 구하고 당락판정은 평균에서 표준편차 두 배 값을 뺀 수치를 기준으로 결정하였다. 각 스테이션 점수는 환자대면(patient encounter)진료시험 성적과 사이시험 성적을 3:1의 비율로 가중치를 두어 평균을 구하였다. 시험 일자별, 시험 조별 성적 평균치 비교는 ANOVA 검정을 시행하였으며 다중 비교법으로 Duncan법을 이용하였다. 시험 일자과 그룹별 시험순서와 선형관계는 피어슨 상관계수를 산출하였다. 시험일자에 따른 CPX성적의 선형

증가 경향은 성별, 연령, 학점을 보정하고 다중선형 회귀분석을 시행하였다. 유의수준  $\alpha$ 값 0.05를 기준으로 통계적 유의성을 판단하였다. 모든 통계 분석은 통계전문프로그램인 SAS<sup>9</sup> 8.2판을 이용하였다.

## 결 과

학생 한 명이 휴학을 하고 시험에 참여하지 않아 전체 연구 대상자는 123명이었으며 이 중 남학생이 96명으로 78%를 차지하였다. 대상군의 연령 평균은 25.4세, 범위는 23~33세였다. 표준화 환자 진료 문제 8개로 구성된 CPX의 내적일치신뢰도 Cronbach  $\alpha$ 값은 0.74였다.

### 가. CPX 성적 분포

CPX 전체 점수의 평균은 61.5점, 표준편차 6.0점, 최저 46점, 최고 73점이었다. 환자대면 진료시험 성적은 평균이 62.9점, 표준편차 6.0점, 최저점 43점, 최고점 78점이었다. 사이시험의 평균은 57.4점, 표준편차 9.2점, 최저점은 35점 최고점은 75점이었다. 평균-2×표준편차 값을 기준으로 한 낙제 학생 수는 스테이션별로 평균 4명(범위, 2~7명)이었고 전체 점수 기준으로 낙제생 수는 3명(2.4%)이었다.

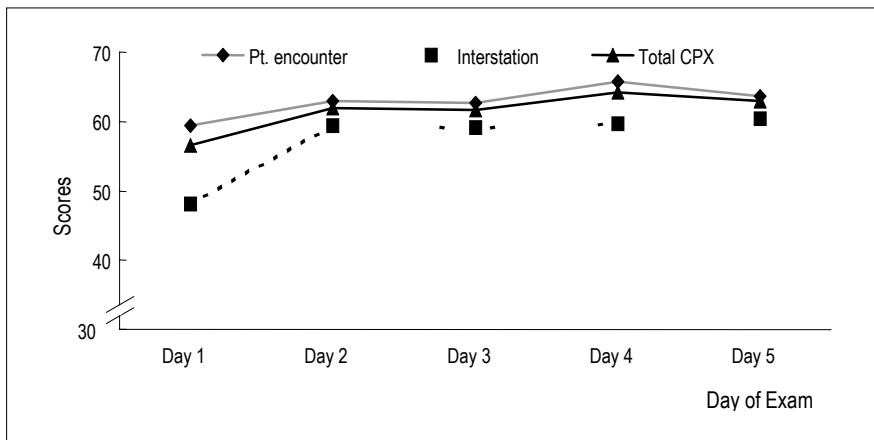
### 나. 시험 일자별 CPX 시험 준비

시험 일자별로 배정된 학생군의 1~3학년 학점 평균은 서로 차이가 없었다(Table I). 학생들의 CPX 목적에 대한 인식 정도도 시험 일자별로 서로 차이

**Table II.** Clinical Performance Examination (CPX) Station and Composite Scores by Day of Exam

Station	Day 1		Day 2		Day 3		Day 4		Day 5		P-value	Linear Trend*
	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range		
Dyspnea	46.4	26~59	55.4	41~70	49.4	39~63	58.8	46~75	53.3	29~70	0.00	0.01
HBV carrier	63.6	52~79	65.1	37~89	63.1	39~63	65.9	39~84	65.4	51~79	0.03	0.77
Non-compliance	73.9	53~95	71.4	49~94	74.0	44~80	80.1	57~97	77.8	59~100	0.04	0.08
Insomnia	46.6	26~59	53.6	34~74	55.8	56~83	54.7	22~81	55.1	36~71	0.02	0.01
Headache	59.8	42~76	62.6	44~85	59.2	22~75	62.1	44~78	59.6	38~85	0.71	0.65
Cough	59.8	44~78	64.8	45~81	65.6	32~78	71.1	61~83	61.4	46~77	0.00	0.28
Bad news	64.2	30~80	67.3	35~90	70.8	44~78	70.3	42~90	69.2	38~91	0.39	0.08
Abdominal pain	61.1	42~74	62.6	45~79	62.8	49~89	64.6	48~77	68.2	41~86	0.04	0.00
Patient Encounter	59.4	48~70	62.9	43~73	62.6	51~70	65.8	56~78	63.7	53~74	0.01	0.00
Interstation work	48.1	35~61	59.4	37~75	59.1	46~70	59.7	44~73	60.4	37~75	0.00	0.00
Total CPX scores	56.6	47~64	62.0	46~72	61.7	50~69	64.3	54~73	62.9	53~72	0.00	0.00

\* P-values of linear trends were obtained by multiple linear regression after adjusting for age, sex, and past credits.



**Fig. 1.** CPX scores by day of exam

가 없었다. CPX에 대한 시험 전 공부(준비) 정도는 첫 날이 가장 낮고 2~3일째, 4~5일째 순으로 Likert 척도 0.4~0.5 구간씩 높아졌다.

**다. 시험 일자별 CPX 스테이션 성적**

두통과 나쁜 소식 전하기 두 스테이션 성적은 일자별로 통계적으로 유의한 차이가 없었다 (Table II).

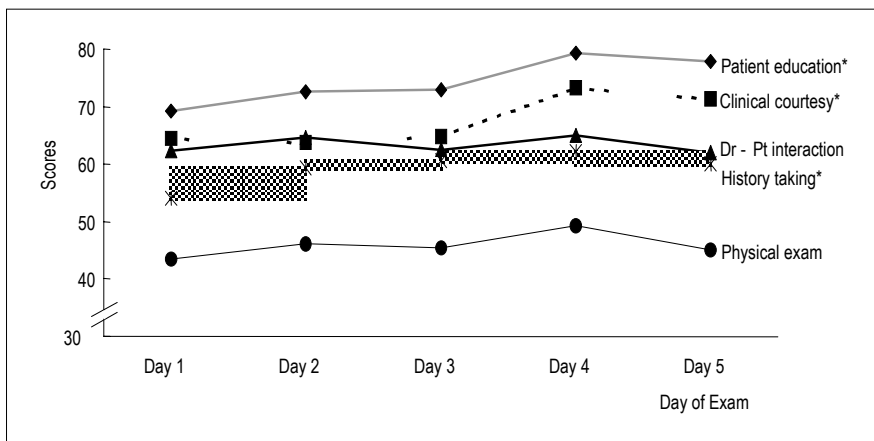
호흡곤란, 간염보균자, 비순응, 불면, 기침, 복통 스테이션 성적은 일자별로 통계적으로 유의하게 차이가 있었다. 환자대면진료성적 (p=0.0053), 사이시험 성적 (p=0.0001), CPX 총점 (p=0.0001)은 시험 일자별로 유의한 차이를 보였다. 첫째 날과 둘째 날의 성적 상승폭은 환자진료가 +3.5점, 사이시험이 +11.3점, CPX 총점에서 +5.4점이었다. 사이 시험성적과

**Table III.** Clinical Performance Examination (CPX) Domain Scores by Day of Exam

Station	Day 1		Day 2		Day 3		Day 4		Day 5		P-value	Linear Trend*
	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range		
Patient satisfaction	65.1	50~83	68.9	43~83	66.8	50~75	67.5	48~83	65.2	48~83	0.48	0.53
Greetings	91.7	42~100	97.4	83~100	95.1	54~100	98.3	88~100	96.6	58~100	0.16	0.12
History taking	54.1	39~73	59.4	44~72	60.6	51~71	62.2	52~73	60.1	52~75	0.00	0.00
Physical exam	43.5	32~65	46.1	18~66	45.3	24~69	49.2	32~73	45.1	29~63	0.42	0.64
Clinical courtesy	64.4	35~85	63.8	40~80	64.8	40~85	73.3	50~85	71.3	45~90	0.01	0.00
Information sharing	69.2	53~82	72.6	42~93	73.0	58~85	79.3	60~92	77.9	50~95	0.00	0.00
PPI†	62.4	54~73	64.7	46~78	62.5	49~69	64.9	55~73	62.0	46~76	0.36	0.62
Total scores	59.2	48~70	62.6	44~72	62.4	51~70	65.7	56~77	63.2	52~74	0.00	0.01

\* P-values of linear trends were obtained by multiple linear regression after adjusting for age, sex, and past credits.

† PPI denotes physician-patient interaction



**Fig. 2.** CPX domain scores by day of exam (\* p for linear trends < 0.01)

CPX 총점은 둘째 날 이후로는 일자별로 통계적으로 유의한 차이가 없었다. 환자대면진료성적은 첫날과 제 2, 4, 5 시험일 성적은 차이가 있고 둘째 날 이후로는 일자별로 성적 차이가 없었다 (Fig. 1). 시험일자와 피어슨 상관계수는 환자대면진료성적 0.27 ( $p=0.0027$ ), 사이시험성적 0.38 ( $p=0.0001$ ), CPX 총점 0.35 ( $p=0.0001$ )이었다. 시험일자 1일 경과에 따른 CPX 성적의 변화량은 성별, 연령, 학점을 보정한 후 환자대면진료성적 1.01 ( $p=0.0038$ ), 사이시험성적 2.20 ( $p=0.0001$ ), CPX 총점이 1.31 ( $p=0.0001$ )이었다.

### 라. 시험 일자별 CPX 영역별 성적

환자 만족도, 인사하기, 신체진찰, 환자-의사 관계 영역 점수는 일자별로 통계적으로 유의한 차이가 없었다 (Table III). 병력청취 ( $p=0.0014$ ), 의사예절 ( $p=0.0087$ ), 환자교육 ( $p=0.0029$ ) 영역 성적은 일자별로 통계적으로 유의하게 차이가 있었다. 의사예절 영역 점수는 제 3 시험일까지는 차이가 없다가 4일째 점수가 전날에 비해 8.5점 상승하였다 (Fig. 2). 병력청취 영역점수는 첫날에 비하여 둘째 날에 5.3점이 상승하고 그 이후로는 일자별로 통계적으로 유의한 차이가 없었다. 환자교육 (정보공유)영역점

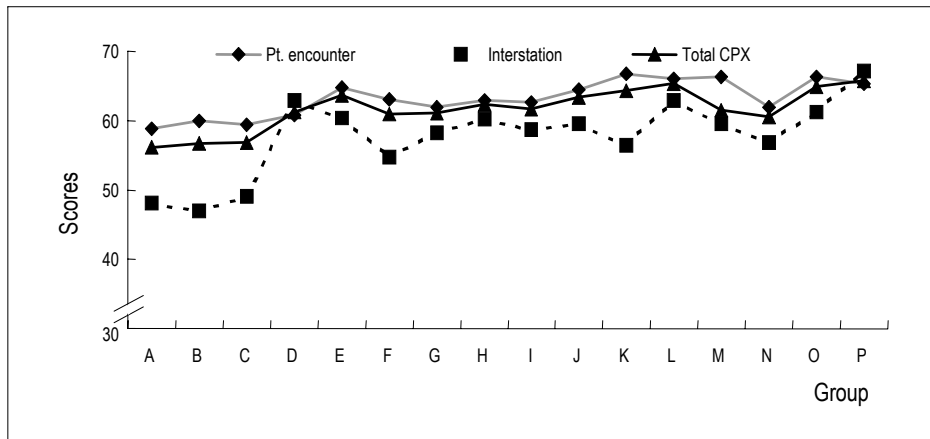


Fig. 3. CPX scores by serial order of students' groups

Table IV. Clinical Performance Examination (CPX) Scores by Students' Preparation for CPX

Variable	Very low		Low		Middle		High		Very High		P-value	Linear Trend*
	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range		
Past credits	3.0	2.53.7	3.4	2.44.3	3.1	2.44.2	3.2	2.44.3	3.0	2.44.0	0.10	0.26
Patient encounter	56.9	4867	62.3	4373	60.9	5471	65.5	5178	65.4	5374	0.00	0.07
Interstation work	50.2	4361	55.5	4375	55.9	3569	62.2	3774	54.7	4473	0.00	0.00
Total CPX scores	55.2	4764	60.6	4673	59.7	5368	64.7	4873	62.7	5372	0.00	0.01

\* P-values of linear trends were obtained by multiple linear regression after adjusting for age, sex, and past credits.

수는 첫날에 비해 둘째 날에 3.3점이 상승하고 넷째 날에는 전날에 비해 5.0점이 상승하여 최고점수를 보였다.

마. 시험 순서(그룹)별 CPX 성적 차이

환자대면진료성적은 시험 순서(그룹)에 따라 통계적으로 유의한 차이가 없었다. 사이시험 성적 (p=0.0001), CPX 총점 (0.0141)은 시험 순서별로 유의한 차이를 보였다. 제 1시험일 셋째 그룹과 제 2 시험일 첫째 그룹의 성적 상승폭은 환자진료가 +1.4 점, 사이시험이 +13.8점, CPX 총점에서 +4.5점이었다 (Fig. 3). 시험 순서(그룹)와 피어슨 상관계수는 환자대면진료성적 0.29 (p=0.0014), 사이시험성적 0.38 (p=0.0001), CPX 총점이 0.36 (p=0.0001)이었다. 시험순서가 하나씩 뒤로 갈 때 CPX 성적의 변화량은 성별, 연령, 학점을 보정한 후 환자대면진료성적

0.34점 (p=0.0021), 사이시험성적 0.70점 (p=0.0001), CPX 총점이 0.43점 (p=0.0001)이었다.

바. 시험 준비정도에 따른 CPX 성적 차이

시험 준비 정도에 따른 학생군별 1-3학년 학점은 서로 차이가 없었다 (Table IV). 시험 준비 정도에 따라 환자대면진료성적 (p=0.0001), 사이시험 성적 (p=0.0001), CPX 총점 (p=0.0001)은 유의한 차이를 보였다. 시험 준비정도와 피어슨 상관계수는 환자대면진료성적 0.34 (p=0.0001), 사이시험성적 0.25 (p=0.0046), CPX 총점이 0.36 (p=0.0001)이었다. 시험준비정도 1 구간 증가에 따른 CPX 성적 변화량은 성별, 연령, 학점, 시험일자를 보정한 후 환자대면진료성적 0.07점 (p=0.0675), 사이시험성적 0.16점 (p=0.0048), CPX 총점이 0.09점 (p=0.0104)이었다.

**Table V. Students' Preparatory Studying for CPX and Its Helpfulness\***

Method	User	Helpfulness
Textbook reviewing: Physical exam	45 (50.0)	20 (27.4)
Textbook reviewing: Medical interviewing	12 (13.3)	1 (1.4)
Lecture note reviewing	57 (63.3)	21 (28.8)
Other references	45 (50.0)	12 (16.4)
Practicing with other persons	16 (17.8)	6 (8.2)
Audiovisual aids	7 (7.8)	0 (0.0)
Internet surfing	21 (23.3)	7 (9.6)
Information from senior	22 (24.4)	4 (5.5)
Information from other students	2 (2.2)	2 (2.7)

\* Numbers in parentheses are percentage of ninety respondents.

### 사. 수험생간의 정보공유 효과 및 CPX 시험 준비 방법

전체 응답 학생 90명 중 53명 (58.9%)에서 시험에 관한 정보를 서로 공유했다고 응답했다. 이 중 시험 전정보공유의 성적향상효과에 대한 응답 분포는 “전혀 그렇지 않다” 4명 (7.5%), “약간 그렇지 않다” 3명 (5.7%), “그저 그렇다” 21명 (39.6%), “약간 그렇다” 21명 (39.6%), “매우 그렇다” 4명 (7.5%)이었다.

CPX 시험 준비를 한 학생은 전체의 90%를 차지하였고 평균 준비 시간은 6.3시간, 범위는 0.5~48시간이었다. 시험 준비 방법으로는 강의 노트 보기, 진단학 교과서 읽기, 선배도움, 인터넷 검색, 다른 사람대상 실습, 의학면담 책 읽기의 순이었다 (Table V). 이 중 강의노트, 진단학 책 읽기, 인터넷검색, 다른 사람 대상 실습 순으로 효과가 있었다고 응답했다.

### 고 찰

표준화 환자를 활용하는 진료수행시험에서 시험의 보안 유지는 공정성, 신뢰도 확보를 위해 꼭 필요하다. 이 연구는 하루 이상 연속적으로 시행하는 진료수행시험에서 시험 일자별로 동일한 문제를 사용할 경우 수험생들 사이에 정보교환이 어느 정도 일어나고 그 효과가 어떻게 나타나는 지 알아보았

다. 연구결과 학생들은 시험일자가 뒤로 갈수록 CPX에 대한 준비를 더하였으며 CPX성적은 첫날과 둘째 날에 차이가 많이 났다. 진료 후 환자 평가 및 향후계획 수립 능력을 측정하는 사이시험이 환자대면 성적보다 일자별로 더 큰 차이를 보였다. 첫날과 둘째 날의 CPX 총점 차이는 5.4점으로 문제에 대한 단서 제공은 1 × 표준편차 (6.0점)에 해당하는 정도의 성적향상 효과를 보여주고 있다. 이는 정보공유에 의한 시험 성적의 향상이 당락 결정에도 영향을 줄 수 있다는 점을 시사하고 있다.

기존의 연구에서는 CPX에서 일자별로 시험성적의 차이가 날 수도 있지만 일정한 경향을 나타내지는 않는다고 보고하고 있다 (Rutala et al, 1991; Stillman et al, 1991; Vu et al, 1992.; Williams et al, 1992). 그런데 이번 연구에서는 시험일자가 뒤로 갈수록 수험생의 성적은 직선적으로 향상됨을 알 수 있다. 전체 8 문제 중 6개 스테이션 (75%)에서 시험 일자에 따른 차이를 보여 주었으며 4개 스테이션 (50%)에서 시험 일자에 따라 직선적인 성적 상승 경향을 보였다. 영역별로 보면 병력청취, 정보공유, 임상 의사 예절 등의 세 영역에서 일자별 차이를 보이고 시험일자가 뒤로 갈수록 직선적으로 앞의 세 영역별 성적의 직선적 상승 경향을 확인할 수 있었다. 병력 청취와 정보 공유 기능은 증례 혹은 문제와 연관된 특이 항목 중 가장 필수적인 것으로 구성된다. 따라서 문제가 무엇인지 알고 꼭 필요한 내용



이 무엇인지를 미리 정리하여 준비하면 진료수행시험 성적을 올릴 수 있다(Niehaus et al, 1996). 하지만 점검표의 항목이 공개되지 않는 한 정보공유의 시험성적향상 효과는 그다지 크지는 않다고 보고하고 있다(Vu et al, 1992). 이와는 반대로 신체진찰이나 환자-의사관계, 전반적인 환자 만족도는 상호작용이나 기능의 성숙도를 보기 때문에 수험생이 문제의 내용을 어느 정도 알아도 정보공유가 성적향상에 별 도움이 되지 않는다. 이번 연구에서 특이한 것은 임상 의사 예절 영역에서 초반 3일보다 후반 2일에 시험을 본 학생들이 더 점수가 높았었는데 이 연구만으로는 왜 그런 현상을 보였는지 정확한 원인을 추정하기는 힘들고 임상 의사예절과 신체진찰 영역의 높은 관련성을 고려하면 더욱 더 해석하기 힘든 현상이다. 좀 더 미시적으로 각 조별 시험 순서에 따른 성적 차이를 보면 환자진료성적은 시험 순서에 따라 일정한 성적변화 양상을 볼 수 없었으나 사이시험에 있어서는 시험을 나중에 볼수록 성적향상에는 더 도움이 되었다. 사이시험은 환자진료 결과를 바탕으로 문제를 정의하고 가설을 설정하고 향후 계획을 세우는 의사의 판단능력을 평가하므로 문제에 대한 단서를 알고 들어가는 것이 성적향상에 도움을 많이 주었다고 해석할 수 있다(Gledhill & Capatos, 1987).

시험 준비는 가장 적게 한 군에서 성적이 가장 낮았다. 하지만 가장 많이 준비를 한다고 가장 높은 점수를 얻지는 않았으며 이는 이전의 연구결과와 비슷하다(Rutala et al, 1991; Swartz et al, 1993). 하지만 시험에 대한 준비가 CPX성적의 12%를 설명하고 있다. 실제로 학생들은 60% 정도에서 시험에 관한 정보를 서로 공유했다고 인정하였으며 이 중 47%는 실제로 정보를 알고 CPX 시험을 보면 성적이 오른다고 알고 있었다.

이번 연구에서처럼 CPX에서 시험 일자별 보안유지가 시험성적에 영향을 주는 경우 그 해결책은 두 가지 방향으로 제시할 수 있다. 첫째는 시험 일자별로 난이도가 비슷한 문제로 조합을 달리하여 시험 문항을 구성함으로써 정보공유효과를 최소로 만드는 방법이 있다(Vu et al, 1992). 미국의사 국가고

시에서 사용하는 방법으로 비슷한 난이도의 문제로 구성된 문제 은행에서 여러 가지 대등한 시험문항의 조합을 만들어 시험일 전후의 문제가 서로 차이가 나도록 조정하는 방법이다. 둘째 방법은 시험 일자를 가능한 한 하루 혹은 이틀에 집중하여 CPX를 치르는 방법이다. 이 방법은 CPX 형태의 시험에서는 물리적으로 불가능하고 OSCE형태의 시험에서는 가능한데 교사장을 여러 군데로 나누고 동일한 세트를 여러 개로 복제하여 동시에 많은 인원을 수용할 수 있는 방법이다. 이 경우 문제 보안유지를 위해 표준화 환자 훈련이나 채점관 훈련은 시험 당일에 시행한다.

우리나라 현실에서 CPX를 시행할 때 대등한 문제의 조합을 여러 개 운영하려면 표준화 환자의 훈련 부담이 증가하고 문제 개발 및 문제 은행 관리체계가 절실하게 필요하다(Ker et al, 2005). CPX 성적으로 학점을 부여할 때 이번 연구처럼 시험 일자별로 성적 차이가 나면 일자별로 성적분포를 고려하여 표준화점수를 갖고 학점을 부여하는 방법이 있다.

이 연구의 제한점은 첫째, CPX시험 성적에 영향을 줄 수 있는 요인으로 정보공유이외에도 다른 요인이 있는데 이를 제대로 고려하지 않았다는 점이다. 따라서 시험 일자별 성적 상승이 정보공유 한 가지 이유만으로 설명될 수는 없다. 하지만 본 연구에서는 성적에 영향을 줄 수 있는 다른 변수 즉, 연령, 성별, 과거 학점 등을 통제하기 위하여 시험조를 무작위로 배정하였으므로 CPX성적에 영향을 줄 수 있는 변수는 우선적으로 시험일자가 된다. 둘째, 정보공유와 시험 준비와는 서로 속성이 다를 수 있다는 점이다. 시험 준비에 대한 설문은 시험이 끝난 후에 실시를 하였고 정보공유여부에 대한 설문은 성적표 배부 후에 시행하였다. 따라서 앞에서 시험을 본 학생들이 알려 주는 문제에 관한 정보를 공유하는 것이 시험 준비이기도 하지만 또 다른 변수로 시험에 대한 일반적인 준비도 여기에 속한다. 학생들의 시험 준비 방법이 주로 강의 노트를 검토해 보거나, 진단학/의학면담 교과서 등을 검토하는 것이었다. 이전에 교육받은 자료를 검토하는 것이 성적

향상에 효과적이었던 점은 CPX 시험문제에 관한 정보교환 이외의 일반적인 시험 준비도 성적향상에 도움을 주었다고 해석할 수 있다. 셋째, 정보공유에 대한 동기 유발 정도는 시험 결과를 어떻게 성적에 반영하는가에 따라 달라질 수 있다(Rutala et al, 1991). 이번 CPX시험에서는 당락 결정보다 학점부여 쪽으로 성적을 활용하였기 때문에 학생들 사이에 정보공유 노력은 다른 때보다 더 많아졌을 것으로 본다. 따라서 이 연구의 결과를 시험성적 활용방법의 차이를 고려하지 않고 일반화시키는 것은 무리가 있다. 학점 부여, 당락 판정, 피드백 등 어떤 목적으로 이 CPX 성적이 활용되는 지에 따라 학생들의 정보공유 정도는 차이가 날 수 있다.

결론적으로 학점을 부여하는 진료수행시험에서 동일한 문제로 며칠에 걸쳐 시험을 시행하는 경우 수험생들 사이에 문제정보교환이 일어나며 이는 성적향상에 부분적으로 도움을 준다. 따라서 진료수행 시험(CPX)의 공정성을 위해 일자별로 난이도가 비슷한 대등한 문제 조합을 교대로 운영하는 방안을 적극적으로 고려해야 한다.

### 감사의 글

이 연구 논문 자료 제공에 도움을 준 서울 CPX 컨소시엄 집행위원회 및 평가 소위원회에 감사 드립니다.

### 참고 문헌

박훈기, 이정권, 김승룡, 김경태, 박해영(1999). 시험장 복제(Duplication)가 객관적 구조적 임상 시험(OSCE)의 신뢰도에 미치는 영향. *한국의학교육*, 11, 37-52.

Barrows HS(1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. *Acad Med*, 68:443-451; discussion 451-453.

Chatenay M, Maguire T, Skakun E, Chang G, Cook D, Warnock GL(1996). Does volume of clinical

experience affect performance of clinical clerks on surgery exit examinations? *Am J Surg*, 172, 366-372.

Cohen R, Rothman AI, Ross J, Poldre P(1993). Impact of repeated use of objective structured clinical examination stations. *Acad Med*, 68(10 Suppl), s73-s75.

Colliver JA, Barrows HS, Vu NV, Verhulst SJ, Mast TA, Travis T(1991). Test security in examinations that use standardized-patient cases at one medical school. *Acad Med*, 66, 279-282.

Furman GE, Colliver JA, Galofre A, Reaka MA, Robbs RS, King A(1997). The effect of formal feedback sessions on test security for a clinical practice examination using standardized patients. *Adv Health Sci Edu*, 2(1), 3-7.

Gledhill RF, Capatos D(1985). Factors affecting the reliability of an objective structured clinical examination (OSCE) test in neurology. *S Afr Med J*, 23;67(12), 463-467.

Harden RM, Stevenson M, Downie WW, Wilson GM(1975). Assessment of clinical competence using objective structured examination. *Br Med J*, 22, 447-451.

Jolly BC, Jones A, Dacre JE, Elzubeir M, Kopelman P, Hitman G(1996). Relationships between students' clinical experiences in introductory clinical courses and their performances on an objective structured clinical examination(OSCE). *Acad Med*, 71, 909-916.

Ker JS, Dowie A, Dowell J, Dewar G, Dent JA, Ramsay J, et al(2005). Twelve tips for developing and maintaining a simulated patient bank. *Medical Teacher*, 27(1), 4-9.

Mavis BE(2000). Does studying for an objective structured clinical examination make a difference? *Med Educ*, 34, 808-812.

Niehaus AH, DaRosa DA, Markwell SJ, Folse R(1996). Is test security a concern when OSCE

- stations are repeated across clerkship rotations? *Acad Med*, 71(3), 287-289.
- Rulata PJ, Witzke DB, Leko EO, Fulginiti JV, Taylor PJ(1991). Sharing of information by students in an objective structured clinical examination. *Arch Intern Med*, 151, 541-544.
- Stillman PL, Haley HL, Sutnick AI, Philbin MM, Smith SR, O'Donnell J, Pohl H(1991). Is test security an issue in a multistation clinical assessment? A preliminary study. *Acad Med*, 66(9 Suppl), S25-7.
- Swartz MH, Colliver JA, Cohen DS, Barrows HS(1993). The effect of deliberate, excessive violations of test security on performance on a standardized-patient examination. *Acad Med*, 68 (10 Suppl):S76-8.
- Thomas PA, Shatzer JH(2000). Standardized patient assessment of ambulatory clerks: effect of timing and order of the clerkship. *Teach Learn Med*, 12(4), 183-188.
- Vu NV, Barrows HS, Marcy ML, Verhulst SJ, Colliver JA, Travis T(1992). Six years of comprehensive, clinical, performance-based assessment using standardized patients at the Southern Illinois University School of Medicine. *Acad Med*, 67, 42-50.
- Williams RG, Lloyd JS, Simonton DK(1992). Sources of OSCE examination information and perceived helpfulness: a study of the grafevine. In: Hart IR, Harden RM, and Marchais JD. (eds.), *Current developments in assessing clinical competence*, pp. 363-370. Montreal: Can-Heal Publications.