

진료수행평가에서 학생들 사이의 정보공유가 평가결과에 미치는 영향

연세대학교 의과대학 마취통증의학교실 및 의학교육학과, 연세대학교 의과대학 마취통증의학교실¹,
연세대학교 의과대학 의학교육학과²

김종훈 · 이기영¹ · 유동미² · 양은배²

= Abstract =

The Effects of Information Sharing between Students on Results of Clinical Performance Examination

Jonghoon Kim, MD, Kiyoung Lee¹, MD, Dongmi Yoo², MA, Eunbae Yang², PhD

*Department of Anesthesiology and Pain Medicine and Department of Medical Education,
Department of Anesthesiology and Pain Medicine¹, Department of Medical Education²,
Yonsei University College of Medicine*

Purpose: This study was designed to investigate the effects of information sharing between students on results of clinical performance examination (CPX).

Methods: 143 third-year students, who completed their core clinical clerkship, were randomly assigned to one day of a 3-day, 6 station CPX. Station checklists, scored by the SPs, consisted of history taking, physical examination, information sharing, clinical courtesy and patient-physician interaction. We compared the CPX station scores, CPX domain scores, self assessment scores and GPA of the three groups tested on different days with ANOVA, and analyzed the differences in CPX scores, after controlling for GPA, with ANCOVA. The correlations among the CPX total and domain scores, self assessment and the GPA of the 3rd year students were calculated. Data regarding information sharing between students and self-assessment were obtained through student questionnaires.

Results: CPX total scores, self assessment scores and GPA of 3rd year students from day-2 were significantly higher than scores from day-1 or day-3. Checklist analysis showed that PE, CC and PPI scores were significantly higher on day-2, but Hx and IS scores were not significantly different. Results from questionnaires showed that 43% of students shared evaluation information. And CPX total scores had little correlation with self assessment scores. GPA of 3rd year students showed a similar pattern with the CPX totalscore.

Conclusion: There is information sharing occurring between students during clinical performance examinations. Although there are significant differences in CPXscores for the treatment group, it seemed that CPX scores are not affected by information sharing between students. Reliability of information sharing from other students is low. The generalization of this study should be interpreted carefully.

Key Words: Clinical performance examination, Information sharing

교신저자: 양은배, 연세대학교 의과대학 의학교육학과, 서울시 서대문구 신촌동 134번지
Tel: 02)2228-2511, Fax: 02)364-5450, E-mail: nara@yumc.yonsei.ac.kr

서 론

의학교육은 환자를 진단하고 치료하는 진료행위를 잘 수행할 수 있는 능력 있는 의사를 양성하는 것을 주목적으로 한다. 이를 위해 학생들이 실제 환자진료에 적용할 수 있는 임상능력을 갖고 있는지 확인하는 것은 의과대학의 중요한 책무이다. 의학교육에 있어서 학생들의 임상능력을 확인하기 위한 객관구조화진료시험(OSCE: Objective Structured Clinical Examination)이 소개되었으며(Harden & Gleeson, 1979), 국내에도 학생의 임상수행 능력을 평가하기 위한 컨소시엄이 구성되고, 의사면허국가시험에 임상수행능력 평가를 위한 실기시험 도입 등이 논의되고 있다(이윤성, 2005). 특히 표준화 환자를 이용한 객관구조화진료시험, 진료수행평가(CPX: Clinical Performance Examination) 등의 임상수행능력 평가는 국내 여러 의과대학에 도입되어 널리 활용되고 있다.

표준화 환자를 이용한 진료수행평가는 그 특성상 동일한 시험 내용이 여러 날에 걸쳐 실시된다. 특히 고학년 학생을 대상으로 임상실습을 마친 후에 종합시험의 형태로 실시하는 진료수행평가는 평가 증례당 10분 이상의 시간이 소요되고 여러 개의 증례를 사용해야 하므로 학생 수가 많은 대학의 경우 이를 하루에 시행하는 것은 현실적으로 매우 어렵다. 따라서 평가내용에 대한 보안 유지는 평가의 신뢰성 확보를 위한 중요한 문제가 되고 있다. 보안을 철저히 유지하여 평가점검표의 내용에 대해 기밀을 지키는 것은 완벽하게 통제 가능하다해도 먼저 평가를 받은 학생이 나중에 평가를 받는 학생에게 자신이 평가받은 내용에 대해 함구하는 것은 현실적으로 어려운 일이다(Rutala et al., 1991). 표준화 환자를 이용한 진료수행평가에서 평가내용에 대한 학생들의 정보공유와 관련하여 Rutala et al.(1991)의 연구, Stillman et al.(1991)의 연구, Colliver et al.(1991)의 연구, Swartz et al.(1993)의 연구에 의하면 학생 상호간의 정보공유와 평가 성적 사이에는 유의한 관계가 없다. 그러나 Niehous et al.(1996)의 연구, Cohen et al.(1993)의 연구에 의하면 학생 상호

간의 정보공유가 평가 성적에 영향을 미친다. 국내에서 실시된 진료수행평가에서 평가내용에 대한 정보공유가 진료수행평가 성적에 어떤 영향을 미치는가에 대해서는 박훈기·권오정(2005)의 연구에 의해 이루어졌다. 이 연구 결과에 의하면 진료수행평가에서 동일한 문제로 며칠에 걸쳐 시험을 시행하는 경우 수험생들 사이에 평가문제에 대한 정보교환이 일어나며, 이러한 정보교환은 진료수행평가 성적에 영향을 미친다.

외국의 선행연구들은 학생 상호간의 정보공유가 진료수행평가 성적에 미치는 영향에 대해서 서로 다른 결과들을 제시하고 있다. 국내에서는 진료수행평가에 대한 준비와 정보공유가 평가결과에 어떤 영향을 주었는지를 분석한 박훈기·권오정(2005)의 연구를 제외하고는 거의 이루어지지 않았다. 진료수행평가에서 학생 상호간의 정보공유가 평가결과에 미치는 영향에 대한 결론을 일반화하기 위해서는 선행연구와는 다른 평가대상, 평가내용 및 평가환경에서 학생 상호간의 정보공유가 진료수행평가 성적에 어떤 영향을 미치는지에 대한 연구 결과들이 축적되어야 한다. 이 연구는 진료수행평가에서 학생들 사이에 정보공유가 있었는지 알아보고, 평가일자별로 성적에 차이가 있다면 학생들 사이의 정보공유가 평가결과에 영향을 미쳤는지 분석하고자 하였다.

대상 및 방법

2005년 1월 5일부터 7일까지 서울소재 Y 의과대학 3학년 학생 143명을 대상으로 진료수행평가를 시행하였다. 평가대상 학생은 평가에 사용되는 증례수와 3일간의 평가일정을 고려하여 세 개의 집단으로 임의표집하여 배치하였다. 평가에 참여한 학생들은 내과, 외과, 산부인과, 소아과 및 정신과 등 핵심 과목 임상실습을 이수한 학생들이었다. 진료수행평가에 사용된 증례 및 증례별 평가영역은 Table I과 같다. 증례는 '서울CPX컨소시엄(현재는 '임상수행능력평가를 위한 서울-경기 컨소시엄'으로 명칭 변경)'을 통하여 제공받은 증례 가운데 일차 진료에서 흔히 접할 수 있는 증상을 증례로 선정하였으며,

Table I. Names of Stations and Domains and Weights of Checklist (%)

| Case name | Domains of checklist | History taking | Physical examination | Information sharing | Clinical courtesy | Patient-physician interaction | Total |
|------------|----------------------|----------------|----------------------|---------------------|-------------------|-------------------------------|-------|
| Headache | | 40 | 24 | · | 14 | 22 | 100 |
| Diabetes | | 25 | · | 50 | · | 25 | 100 |
| Dyspnea | | 42 | 24 | · | 12 | 22 | 100 |
| Hepatitis | | 25 | 10 | 33 | 10 | 22 | 100 |
| Depression | | 59 | · | 12 | · | 29 | 100 |
| Bad news | | · | · | 70 | · | 30 | 100 |

Blue Print Matrix를 이용하여 학생이 익혀야 할 기본적인 임상 기술을 균형 있게 측정하도록 개발하였다.

증례별 표준화 환자는 서울CPX컨소시엄의 표준화 환자 훈련 담당 교수, 증례 개발 교수, 표준화 환자 교육담당자에 의해 훈련된 환자였으며, 일주일에 2~3시간씩 총 4회의 훈련을 받았다. 본 연구에서는 서울CPX컨소시엄의 표준화 환자 교육 시점과 본 대학의 평가 시점이 시간적 간격이 있었다는 점과 표준화 환자 상호간의 신뢰성을 높이기 위해 진료 수행평가를 시행하기 전에 비디오훈련을 추가로 실시하였다. 표준화 환자 훈련 담당자는 다른 의과대학에서 시행된 진료수행평가 비디오 4~5개를 보면서 점검표의 표준답안을 만들었으며, 표준화 환자와 훈련담당자가 비디오를 함께 시청한 후 표준화 환자간의 점검표 작성 기준을 일치시켰다. 병력청취, 이학적 검사, 정보나누기, 임상예결과 같은 영역은 점검표 작성이 95% 이상 일치할 때까지 훈련하였으며 환자의사관계는 완전한 일치를 보이기 어려워 표준화 환자의 채점이 2점 척도 안에 모이도록 훈련하였다. 한편, 진료수행평가는 의과대학에서 전통적으로 이루어졌던 평가방법과는 차이가 있다는 점에서 진료수행평가 1개월 전에 학생 오리엔테이션을 실시하고, 평가의 취지와 방법, 성적처리 방법 등에 대한 내용을 설명하였으며, 평가 당일에도 오리엔테이션을 실시하였다.

진료수행평가는 증례당 12분씩 여섯 증례가 사용되었으며, 표준화 환자는 각 증례별 시험이 종료된

직후 학생의 수행 정도에 대한 점검표를 완성하였다. 각 증례 사이에 직전에 시행한 증례에서 학생이 파악한 객관적 주관적 정보, 진단명, 그리고 향후 치료계획을 기술하는 5분간의 사이시험을 시행하였다. 사이시험이 진료수행평가 성적에 미치는 영향을 통제하기 위하여 사이시험 문제에서 학생이 만난 표준화 환자가 어떤 환자인지 짐작 할 수 있는 부분은 모두 사전에 제거하였다. 시험 종료 직후에는 시험에 대한 소감, 시험내용에 대한 정보교류 여부 및 학생 스스로가 자신의 수행정도를 100점 만점으로 평가하는 설문을 익명으로 실시하였다. 시험내용에 대한 정보교류를 확인하는 문항은 정보를 제공하였거나 받은 경우를 구분하지 않고 사실 여부만 확인하였다.

진료수행평가 일자별 표준화 환자의 점검표점수, 학생의 자기평가 점수, 이러한 점수가 평가에 참여한 학생들의 이전 평량평균과 어떤 관계가 있는지 평가일자별, 평가에 사용된 증례별 차이를 일원변량 분석 하였다. 일원변량분석 후 평가일자별로 차이가 나는 경우 Tukey 사후검증을 실시하였다. 또한 진료수행평가 날짜별 학생 평량평균의 차이를 고려하기 위하여 평량평균을 공분산으로 처리하여 진료수행평가 성적의 차이를 분석하였다. 진료수행평가 성적과 자기평가 점수, 학생의 평량평균과의 관련성은 피어슨(Pearson) 상관계수를 산출하고, 진료수행평가 내용에 대한 학생 상호간의 정보교류 정도는 백분율로 산출하였다. 통계적 유의 수준은 .05와 .01수준에서 검증하였으며, 통계 프로그램은 SPSS ver-

진료수행평가에서 학생들 사이의 정보공유가 평가결과에 미치는 영향

Table II. Comparison of CPX Scores, GPAs and Self Assessments by Day of Examination

| Variables | 1st day | | 2nd day | | 3rd day | | F | Post hoc test |
|------------------------|---------|-------|---------|-------|---------|-------|-------------------|----------------|
| | Mean | SD | Mean | SD | Mean | SD | | |
| CPX [†] score | 315.19 | 36.8 | 331.66 | 28.52 | 313.89 | 29.3 | 4.64* | 2nd > 1st, 3rd |
| Self assessment | 63.16 | 23.31 | 78.91 | 16.62 | 79.07 | 15.83 | 9.43 [†] | 3rd, 2nd > 1st |
| GPA [§] | 2.82 | 0.66 | 3.11 | 0.63 | 2.72 | 0.7 | 4.45* | 2nd > 3rd |

* p<0.05, [†] p<0.01, [†] CPX: Clinical performance examination, [§] GPA: Grade point average

Table III. Comparison of CPX Station Scores by Day of Examination

| CPX [†] station | 1st day | | 2nd day | | 3rd day | | F | Post hoc test |
|--------------------------|---------|-------|---------|-------|---------|-------|-------|---------------|
| | Mean | SD | Mean | SD | Mean | SD | | |
| Diabetes | 52.07 | 9.99 | 47.95 | 8.6 | 47.31 | 9.71 | 3.56* | 1st > 3rd |
| Depression | 36.06 | 9.32 | 41.71 | 12.05 | 37.92 | 10.83 | 3.42* | 2nd > 1st |
| Hepatitis | 58.13 | 12.01 | 61.98 | 8.78 | 57.28 | 11.43 | 2.55 | |
| Dyspnea | 44.77 | 10.38 | 50.2 | 11.3 | 46.59 | 6.45 | 3.96* | 2nd > 1st |
| Headache | 48.21 | 12.56 | 51.68 | 7.57 | 47.46 | 11.07 | 2.16 | |
| Bad news | 75.97 | 12.58 | 78.13 | 8.91 | 77.33 | 10.11 | 0.50 | |

* p<0.05, [†] CPX: Clinical performance examination

Table IV. Comparison of CPX Domain Scores by Day of Examination

| CPX [†] domain | 1st day | | 2nd day | | 3rd day | | F | Post hoc test |
|-------------------------------|---------|-------|---------|-------|---------|-------|--------------------|----------------|
| | Mean | SD | Mean | SD | Mean | SD | | |
| History taking | 89.23 | 16.79 | 92.93 | 15.95 | 87.73 | 13.77 | 1.41 | |
| Physical examination | 14.33 | 5.68 | 19.57 | 5.54 | 16.84 | 5.39 | 10.78 [†] | 2nd > 1st, 3rd |
| Information sharing | 112.34 | 18.8 | 113.72 | 12.45 | 109.66 | 15.56 | 0.80 | |
| Clinical courtesy | 27.58 | 3.49 | 29.53 | 3.15 | 27.21 | 4.39 | 5.38 [†] | 2nd > 1st, 3rd |
| Patient-physician interaction | 71.72 | 8.19 | 75.91 | 6.15 | 72.45 | 7.67 | 4.40* | 2nd > 1st |

* p<0.05, [†] p<0.01, [†] CPX: Clinical performance examination

sion 12.0 (SPSS Inc, Chicago, IL, USA)을 사용하였다.

결 과

가. 평가 일자별 점검표, 자기평가 점수 및 평량평균 차이

Table II는 표준화 환자의 점검표 점수, 학생 자기평가 점수 및 학생들의 3학년 평량평균을 평가 일자별로 차이를 비교한 것이다. 둘째 날 평가를 받은

학생들이 첫째 날이나 셋째 날 평가를 받은 학생에 비해 점검표 점수가 유의하게 높았다. 그러나 진료수행평가에 대한 학생의 자기평가 점수는 첫째 날 평가를 받은 학생이 다른 날에 평가를 받은 학생에 비해 낮았다. 학생들의 3학년 평량평균은 둘째 날 시험을 본 학생의 성적이 마지막 날 시험을 본 학생에 비해 통계적으로 유의하게 높았다. 진료수행평가에 사용된 각 증례별 점검표 점수를 분석한 결과는 Table III과 같다. 둘째 날 평가를 받은 학생들이 첫

Table V. Analysis of Covariance to estimate the Impact of Covariate on GPA[†] of 3rd Years

| Parameter | Estimate | Standard error | t |
|--------------------|----------|----------------|--------|
| (day1+day3)/2-day2 | -11.27 | 5.42 | -2.08* |
| (day1+day2)/2-day3 | 5.25 | 5.38 | 0.98 |
| (day2+day3)/2-day1 | 6.01 | 5.28 | 1.14 |

* p<0.05, † GPA: Grade point average

Table VI. Correlation of CPX Scores, Self Assessments, inter-station scores and GPAs

| Variables | CPX [†] total scores | Self assessment | GPA [§] |
|--------------------------------|-------------------------------|-----------------|------------------|
| History taking | .71 [†] | -.13 | .23 [†] |
| Physical examination | .54 [†] | .18* | .20* |
| CPX domain Information sharing | .71 [†] | .18* | .26 [†] |
| Clinical courtesy | .30 [†] | .07 | .18* |
| Patient-physician interaction | .77 [†] | .10 | .43 [†] |
| Inter-station examination | .28 [†] | .12 | .22 [†] |
| CPX total scores | . | .09 | .40 [†] |
| Self assessment | . | . | -.00 |

* p<0.05, † p<0.01, ‡ CPX: Clinical performance examination, § GPA: Grade point average

째 날에 평가를 받은 학생에 비해 호흡곤란과 우울 증에서 유의하게 높은 점수를 받았으며, 당뇨병의 경우에는 첫째 날 평가를 받은 학생의 성적이 셋째 날 시험 본 학생의 성적보다 유의하게 높았다. 또한, 점검표 영역별로 평가점수를 분석한 Table IV를 살펴보면, 신체진찰, 임상예절에서는 둘째 날 평가를 받은 학생들이 다른 날 시험을 본 학생에 비해 높은 평가를 받은 것으로 나타났으며, 환자의사관계에서는 둘째 날 평가를 받은 학생들이 첫째 날 평가를 받은 학생에 비해 높은 평가를 받았다.

나. 평량평균의 효과를 제거한 후 일자별 점검표 점수 차이

Table V는 평가 일자별 학생 평량평균의 차이를 통제하기 위하여 평량평균을 공분산으로 처리하고, 둘째 날 표준화 환자의 점검표 점수가 다른 날의 점검표 점수와 차이가 있는지 대비계수를 이용하여 분석한 결과이다. 학생들의 3학년 평량평균의 효과

를 제거한 후의 평가 일자별 점검표 점수는 둘째 날 평가를 받은 학생이 다른 날 평가를 받은 학생보다 유의하게 높았다.

다. 점검표 점수, 자기평가 점수 및 평량평균의 상관관계

Table VI은 표준화 환자의 점검표 점수, 학생 자기평가 점수 및 평량평균과의 상관관계를 나타낸 것이다. 표준화 환자의 점검표 점수와 학생들의 3학년 평량평균은 중등도의 상관관계를 보였으며, 특히 점검표 점수 중에서 환자의사관계 영역이 평량평균과 상관관계가 높았다. 표준화 환자의 전체 점검표 점수와 학생의 자기평가 점수 사이에는 상관관계가 나타나지 않았다.

라. 평가 내용에 대한 정보공유

진료수행평가의 평가내용에 대해 다른 학생들에게 정보를 전달하지 않겠다는 서약을 진료수행평가

전에 받았지만, 평가 후 설문조사에서 전체 평가 학생의 43%가 동료 학생들로부터 평가내용에 대한 정보를 전달 받은 것으로 나타났다.

고 찰

진료수행평가는 그 특성상 동일한 시험 내용을 여러 날에 걸쳐 실시한다. 따라서 평가내용에 대한 보안 유지는 평가의 신뢰성 확보 차원에서 중요한 문제이다. 그러나 점검표 내용에 대해 보안을 유지하는 것이 가능하다해도 먼저 평가를 받은 학생이 나중에 평가를 받는 학생에게 자신이 평가받은 내용을 전달하지 못하도록 통제하는 것은 현실적으로 어려운 일이다 (Rutala et al., 1991). 본 평가에서는 진료수행평가 내용에 대해 다른 학생들에게 전달하지 않겠다는 서약을 학생들로부터 받았지만, 평가후 설문조사에서 43% 정도의 학생이 동료 학생들로부터 평가내용에 대한 정보를 전달받은 것으로 나타났다. 학생 상호간에 정보전달에 대한 설문조사는 진료수행평가가 종료된 직후 설문조사가 실시되었기 때문에 첫째 날 평가받은 학생은 정보전달 여부에 대한 설문조사가 의미가 없었다. 둘째 날 평가받은 학생은 첫째 날 평가받은 학생으로부터 정보를 전달받았는지를 설문조사한 것이다. 셋째 날 평가받은 학생은 첫째 또는 둘째 날 평가받은 학생들로부터 정보를 전달받은 것이다. 본 평가에서 학생 상호간의 이러한 정보전달은 나중에 평가를 받은 학생들의 진료수행평가 성적에 영향을 주었을 가능성이 있다. 표준화 환자를 이용하여 여러 날에 걸쳐 시험을 볼 경우 학생 상호간의 정보공유와 일자별 성적 변화에 관해서 많은 연구가 진행되어 왔는데, Rutala et al.(1991)에 의하면 임상실습 후에 실시한 객관구조화진료시험 성적을 분석한 결과 학생들 간의 정보교환이 있었으나 그에 따른 시험성적은 큰 변화가 없었다. Stillman et al.(1991)에 의하면 전공의를 대상으로 한 진료수행평가에서 평가 시기는 성적에 큰 영향을 주지 않았다. Colliver et al.(1991)도 5년간 시행된 진료수행평가 성적을 분석한 결과 늦게 평가를 받은 학생의 성적이 일정한 상승세를

나타내지 않았다고 하였다. 또한, Swartz et al. (1993)에 의하면 먼저 시험을 본 학생이 다른 날 시험을 보는 학생에게 아주 자세한 정보를 주어도 양군간의 성적 차이는 없었다. 그러나 Niehous et al.(1996)에 의하면 학생 임상실습 동안에 시행한 객관구조화진료시험의 결과 평가 시기에 따라 성적은 큰 변화가 없었으나 몇 개의 증례는 나중에 평가를 받은 학생의 점수가 상승하였다. Cohen et al.(1993)도 30개의 증례를 이용한 객관구조화진료시험의 결과 연결형 증례의 경우에는 늦게 평가를 받은 학생의 성적이 향상되었음을 보고하였다.

본 연구에서 평가일자별로 학생들이 성취한 성적을 비교해본 결과 진료수행평가 성적에서 둘째 날 평가를 받은 학생의 성적이 첫째 날이나 셋째 날에 평가를 받은 학생에 비해 유의하게 높았으며, 특정 증례 및 평가영역에서도 둘째 날 평가 받은 학생의 성적이 높았다. 둘째 날에 높은 점수를 받은 학생들에 대한 원인은 크게 세 가지 측면에서 고찰해 볼 수 있다. 첫째는 둘째날 평가를 받은 학생들의 평량평균과 관련성이다. 둘째는 각 증례가 끝난 뒤 실시된 사이시험의 영향 가능성이다. 셋째는 학생 상호간의 정보전달이 영향을 미쳤을 가능성이다.

먼저, 둘째 날에 높은 점수를 받은 학생들에 대한 원인을 분석하기 위하여 진료수행평가 날짜별로 평가에 응시한 학생의 3학년 평량평균의 차이를 분석하였다. 그 결과 진료수행평가에 참여하는 학생들을 무작위 배정하였음에도 둘째 날에 평가받은 학생들의 평량평균이 셋째 날 시험 본 학생에 비해 통계적으로 유의하게 높은 것으로 나타났으며, 통계적 유의성은 없었지만 첫 날 평가받은 학생들보다도 평량평균이 높았다. 또한, 3학년 학생의 평량평균의 차이를 통계적으로 통제한 후 진료수행평가 성적의 차이를 분석한 결과에서도 둘째 날 평가를 받은 학생의 성적이 통계적으로 유의하게 높았다. 그러므로 3학년 학생들의 평량평균이 진료수행평가 성적에 영향을 미쳤다고 보기는 어렵다. 이러한 결과는 선행 연구 결과와 일치하고 있는데, 박훈기·권오정 (2005)에 의하면 진료수행평가를 받은 학생집단의 사전 평량평균이 평가 날짜별로 차이가 없음에도

진료수행평가 성적은 차이가 있었다. 평량평균이 진료수행평가 성적에 영향을 미쳤다고 보기 어렵다는 점은 학생들의 평량평균과 진료수행평가 점수의 상관에서도 알 수 있는데, 본 연구에서의 상관은 .40로 나타났다. 평량평균은 여러 가지 요소가 복합되어 있어 단정적인 결론을 도출하기 어렵지만, 필기 시험에 의한 학생의 지식 정도를 측정하는 것이 상당 부분을 차지한다. 그러나 진료수행평가는 단순히 학생의 지식 정도를 평가하는 것이 아니므로 의학적 지식이 많은 학생이 반드시 진료수행평가에서 좋은 성적을 받을 것으로 단정할 수 없다. Clauser et al.(1993)에 의하면 미국의사국가 시험의 임상지식에 관한 선다형 시험 성적과 진료수행평가 성적의 상관관계는 0.3 정도에 불과하다. Edelstein et al.(2000)도 비슷한 결과를 제시했는데, 학생의 1·2학년 및 3·4학년의 성적과 진료수행평가 성적의 상관관계는 각각 0.33과 0.28이었다.

그렇다면 특정 증례에서 높은 성적을 보이는 원인이 증례 중간에 실시한 사이시험의 영향이 아닌지 생각해 볼 수 있다. 사이시험 내용이 증례와 관련되어 교과서에 나오는 지식을 묻는 것이라면 이를 통해 학생들이 정확하게 증례의 진단명을 알 수 있으며 나중에 평가를 받은 학생의 점수가 올라갈 것으로 예상된다. 선행 연구는 사이시험이 진료수행평가 성적 향상에 도움을 준다고 보고한다. 박훈기·권오정 (2005)은 사이시험과 진료수행평가 성적이 관련이 있다고 하였다. Cohen et al.(1993)도 하나의 방안 사용하는 증례보다 2개의 방안을 연결하고, 시험관의 구두시험을 포함한 증례에서 평가 성적의 향상이 있음을 보고하였는데, 그 원인으로 학생들이 구두시험을 통해 해당 증례의 진단명에 대한 단서(사이시험에서 얻는 단서와 유사)를 얻은 것에 기인한 것으로 추정하였다. 그러나 본 연구에서는 선행 연구 결과에 기초하여 사이시험의 내용을 학생들이 환자를 보고 얻은 정보로만 한정하여 사이시험을 통해 증례의 진단명을 유추할 수 있는 가능성을 사전에 차단하였다. Table VI의 결과를 살펴보면, 사이시험과 진료수행평가와의 상관은 나타나지 않았다.

다음으로 학생 상호간의 정보공유가 진료수행평가 성적에 영향을 미쳤는지 알아보기 위해서 일차별 성적, 증례별 성적 및 각 증례의 평가영역별 성적이 어떤 차이가 있는지를 고찰하였다. 평가기간이 3일에 불과하여 평가점수의 변화 추세를 분석하는 것은 어려움이 있으나 둘째 날의 학생 성적 향상이 학생 상호간의 정보 전달에 의한 것이라면, 셋째 날에 평가를 받은 학생의 점수도 첫째 날에 평가받은 학생의 점수에 비해 높아야 할 것이다. 그러나 셋째 날 평가 받은 학생의 성적은 첫째 날 평가 받은 학생의 성적보다 약간 낮았다. 이러한 결과는 둘째 날에 평가 받은 학생이 높은 점수를 받은 원인이 학생 상호간의 정보 전달이 아니라 다른 원인에 있을 것이라는 점을 시사한다. Table IV의 영역별 점수에서는 임상예절, 신체진찰 및 환자의사관계 영역에서 둘째 날에 평가받은 학생이 다른 날 평가를 받은 학생들보다 우수하였으며, 병력청취와 환자교육에서는 시험일자 별로 큰 차이를 보이지 않았다. 이러한 결과는 선행 연구결과와 차이가 있는데, 박훈기·권오정 (2005)에 의하면 환자교육, 임상예절, 병력청취 부문은 날짜에 따라 점수가 선형적으로 상승하는 경향이 있으며, 신체진찰, 의사환자관계는 선형 증가의 경향을 발견할 수 없었다. 본 연구와 선행 연구결과의 차이가 발생한 원인은 크게 두 가지 가능성을 고찰해 볼 수 있다. 첫째는 평가에 참여한 학생집단의 특성과 평가에 사용된 증례의 특성과의 관련성이다. 박훈기·권오정 (2005)의 선행연구에 참여한 학생들은 평가 날짜별로 평량평균이 집단간에 차이가 없었던 반면, 본 연구에 참여한 학생들은 둘째 날 평가 받은 집단이 다른 집단에 비해 평량평균이 높았다. 따라서 진료수행평가에 사용된 증례 가운데 평량평균과 높은 상관을 보이는 증례의 경우에는 선행 연구 결과와 차이가 있었을 것이다. 그러나 진료수행평가 영역과 평량평균과의 상관관계를 나타낸 Table VI의 결과를 살펴보면, 환자의사관계를 제외하고는 학생들의 평량평균과 진료수행평가 영역별 성적은 상관이 없는 것으로 나타났다. 둘째는 특정 학생이 다른 학생에게 진료수행평가 내용과 진단명에 대한 정보를 제공해 주었을 가능성이

다. 평가내용에 대한 정보를 제공받은 학생이 진단명을 바탕으로 추가 학습을 하여 그 증례에 대한 지식수준이 올라갔다면 이에 가장 큰 영향을 받는 점검표점수는 병력청취와 환자교육과 관련된 영역이다. Colliver et al.(1992)에 의하면, 진료수행평가에서 진단을 내리는 증례는 평가 시기가 늦어질수록 성적이 증가하는 경향을 보였다. 그러나 신체진찰 영역은 단순한 지식뿐만 아니라 실습을 통해 정확한 술기를 익혀야 높은 점수를 얻을 수 있으므로 진단을 알았다고 해서 단기간에 향상되기는 어렵다(박훈기·권오정, 2005). 또한 환자의사관계나 임상예절의 점수도 학생이 어떤 증례인지 알고 시험에 임했다고 해도 이 영역은 학생들이 자신의 태도를 단기간에 쉽게 교정할 수 있는 영역이 아니므로 병력이나 환자교육보다는 평가 결과에 미치는 영향이 작을 것이다. 평가에 사용되는 증례에 대한 학생들의 사전 인지가 자신감을 주어 성적에 영향을 미쳤을 가능성도 있으나, 다른 영역이 받는 영향에 비해서는 상대적으로 적을 것이다. 따라서 진료수행평가 내용에 대한 학생 상호간의 정보전달이 진료수행평가 성적에 영향을 미쳤다고 단정할 수 없다. 한편, 학생 상호간의 정보전달이 어느 정도 신뢰로운지는 진료수행평가 후 학생들의 자기평가 점수와 표준화 환자에 의한 점검표 점수를 비교해 보면 알 수 있는데, 본 연구에서 두 점수 간에는 유의한 상관관계가 없었다. 이러한 결과는 학생들이 평가내용에 대해 정확하게 알고 있지 못하며, 자신이 어느 정도 수준에 있는지 모른다는 것을 의미한다. 따라서 먼저 평가받은 학생들이 다른 학생에게 전달했을 정보는 그 신빙성이 낮다고 할 수 있다.

여러 날에 걸쳐 진료수행평가를 실시할 경우 평가내용에 대한 학생 상호간의 정보 전달이 일어나는 것을 막을 수는 없다. 그러나 본 연구의 결과에 따르면, 평가내용에 대한 학생 상호간의 정보전달은 진료수행평가 성적에 영향을 미치지 않았다. 또한 먼저 평가를 받은 학생들이 나중에 평가를 받는 학생들에게 전달한 정보내용의 신뢰성도 낮은 것으로 나타났다. 그러므로 여러 날에 걸쳐 진료수행평가가 이루어진다고 해도 학생의 정보 전달은 크게 문제

가 되지 않을 것으로 보인다. 본 연구 결과가 서울 소재 한 개 대학의 경험적 자료에 국한되어 있다는 점에서 일반화 가능성은 조심스럽게 해석되어야 할 것으로 보인다.

참 고 문 헌

- 박훈기·권오정(2005). 진료수행평가(CPX: Clinical Performance Examination)에서 정보공유가 시험 성적에 미치는 영향. *한국의학교육*, 17(2), 185-195.
- 이윤성(2005). *의사국가시험에서의 실기시험 실행 방안*. 의사실기시험 실행방안 공개토론회 자료집. 한국보건의료인국가시험원, 1-13.
- Clauser, B.R. et al.(1993). A comparison of pass/fail classifications made with scores from the NBME standardized-patient examination and part II examination. *Acad Med*, 68(10 Suppl), 7-9.
- Cohen, R. et al.(1993). Impact of repeated use of objective structured clinical examination stations. *Acad Med*, 68(10 Suppl), 73-75.
- Colliver, J.A. et al.(1991). Test security in examinations that use standardized-patient cases at one medical school. *Acad Med*, 66(5), 279-282.
- Colliver, J.A. et al.(1992). Test security in standardized-patient examinations: analysis with scores on working diagnosis and final diagnosis. *Acad Med*, 67(10 Suppl), 7-9.
- Edelstein, R.A. et al.(2000). A comparative study of measure to evaluate medical students' performances. *Acad Med*, 75(8), 825-833.
- Harden, R.M., & Gleeson, F.A.(1979). Assessment of clinical competence using on objective structured clinical examination. *Med Educ*, 13, 41-54.
- Niehaus, A.H. et al.(1996). Is test security a concern when OSCE stations are repeated across clerkship rotation? *Acad Med*, 71(3), 287-289.
- Rutala, P.J. et al.(1991). Sharing of information by students in an objective structured clinical exami-

nation. *Arch Intern Med*, 151, 541-544.

Stillman, P.L. et al.(1991). Is test security an issue in a multistation clinical assessment? A preliminary study. *Acad Med*, 66(9 *Suppl*), 25-27.

Swartz, M.H. et al.(1993). The effect of deliberate, excessive violations of test security on performance on a standardized-patient examination. *Acad Med*, 68(10 *Suppl*), 76-78.