

# 진료수행 시험에서 채점자 및 시험 일자에 따른 결과 차이

## Difference in Results according to Scorer and Test Date in Clinical Practice Test

권소희, 김영전  
원광대학교 의과대학

So-Hee Kwon(nannav74@wku.ac.kr), Young-Jon Kim(youngjonkim77@wku.ac.kr)

### 요약

본 연구는 진료수행 시험 상황을 재연하여 실제 의사가 채점한 결과와 표준화 환자가 채점한 결과의 차이가 있는지, 첫째 날 응시한 학생과 정보가 노출된 둘째 날 응시한 학생의 점수 차이를 비교하고자 하였다.

의과대학 4학년 101명이 4가지 세트(본태성고혈압, 수면장애, 기분저하, 무월경)로 구성된 진료수행 시험에 응시하였다. 학생들은 첫째 날과 둘째 날 중 표준화환자 채점 세트와 의사 채점 세트로 구성된 평가 일정 하나에 무작위로 배정되었다. 각 진료수행의 평가지침은 병력청취, 신체진찰, 환자교육, 환자의사관계, 진찰 태도로 영역을 구분하여 구성하였다. 각 상황별, 영역별 성취점수를 표준점수로 변환하여 집단 간 차이를 비교하였다.

모든 영역에서 여학생의 성취점수가 남학생의 성취보다 유의하게 높았다( $P=.001$ ). 표준화 환자가 채점한 집단과 의사가 채점한 집단의 평균의 유의한 차이는 없었다( $p=.425$ ). 연속되는 시험일자 중 둘째 날 응시한 군이 병력청취와 신체진찰영역에서 유의한 차이를 보였다. 본 연구에서는 채점자의 신분에 따른 평가 결과에 차이가 없음을 다시 한 번 밝히면서 표준화 환자를 활용한 평가의 가능성을 확인하였다.

■ **중심어** : | 진료수행평가 | 의사채점자 | 표준화환자 채점자 | 시험일자 | 정보공유 |

### Abstract

The purpose of this study is to clarify the difference between the scoring results by scorer(doctors and standardization patients) and examination dates. A total of 101 students in the fourth grade of medical school participated in four clinical practice test. Students were randomly assigned to either day-1 or day-2, which was consisted of a standardized patient scoring set or a physician scoring set. Station checklists consisted of history taking, physical examination, patient education, physician-patient relationship and clinical courtesy. The achievement scores of each case and each domain were converted to the standard score, and the differences between groups were compared.

Female students' achievement scores were significantly higher than male students' achievement scores in all domains. There was no significant difference between means by the standardized patients' group and doctors group. Day-2 group was significantly higher than day-1 group in both of history taking and physical examination domains. If the principles of checklist are clearly defined, the scorer status (either physician or standardized patients) does not determine the difference of students' practice test scores.

■ **keyword** : | Clinical Practice Test | Doctor | Standardized Patient | Test Date | Information Sharing |

## I. 서론

의과대학의 술기 역량 평가는 특정 의료술기의 수행 능력을 보는 객관적-구조화 임상시험(objective structured clinical examination, OSCE)과 특정 문제를 가진 환자를 대상으로 임상 수행 능력의 진반을 확인하는 진료수행시험(clinical performance examination, CPX)으로 구성된다. 그 중 진료수행시험은 실제 임상 상황과 유사한 환경에서 포괄적인 임상 수행능력을 측정하여 환자-의사 관계, 정보통합 능력, 환자 교육 등 다양한 측면의 평가가 가능하다[1].

진료수행 시험을 대비한 교육은 학교에서 학생이 먼저 인체 모형을 이용하여 술기를 습득하고, 모의 환자를 활용하여 진료를 연습하는 방식이다. 임상실습에서는 실제 환자의 진료에 직간접적으로 참여하여 의사로서의 역량을 습득한다. 진료수행 역량 평가는 시기와 목적에 따라 다양한 방식을 사용하지만, 가장 일반화된 방법은 체크리스트를 활용하는 것이다. 특정 술기의 시행 여부와 정확성 및 숙련도를 점수화하여 체크하는 방식으로, 술기나 진료의 과정을 동일한 기준으로 채점한다는 점에서 평가자의 임의성을 배제한 객관성을 확보한다.

또한, 진료수행 시험을 실시하기 위해서는 여러 채점 교수들이 동원되어야 한다. 그런데 대부분 임상교수의 일정을 고려하면 현실적으로 수십 명의 교원을 동시에 채점자로 활용하기에는 용이하지 않다. 이러한 이유로 표준화환자를 채점자로 참여시키는 것이 일반적인 상황이다. 표준화 환자(standardized patient, SP)란 실제 환자와 비슷하게 그리고 일관성 있게 진료 상황을 재현해 낼 수 있는 의사가 아닌 일반인을 의미한다[2]. 이러한 상황에서 우려되는 문제는 표준화 환자와 교수 간의 평가의 일치도이다. 지금까지의 연구에서는 표준화 환자와 교수들이 평가의 차이가 없다는 결과와[3][4] 표준화 환자와 교수의 평가 문제에서 차이가 있다는 연구결과[5] 혼재한다. 결국 학생의 입장에서는 동일한 증례에 응시하면서 채점자가 표준화환자인지, 의사 교수인지에 따라 다른 점수를 받을 수 있다는 불안감이 존재한다.

진료수행 평가에서 발생하는 또 다른 문제는 완벽히 통제될 수 없는 시험 상황에서 동일한 임상수행이 세팅되어 평가를 진행할 때 나타날 수 있는 결과의 차이이다. 현재 우리나라의 경우 지역 별로 평가에 활용될 수 있는 표준화 환자가 많지 않을뿐더러, 증례가 개발되었기에 맞춰 표준화 환자를 교육하고 연기에 투입하는 것에는 많은 노력과 시간이 소요된다. 대학이 필요한 수만큼 표준화 환자를 활용하여 여러 세트의 진료수행을 수행하는 것에는 물리적 한계가 있다. 그래서 수험생이 많아지면 동일한 증례에 대해 표준화 환자와 스테이션을 필요한 만큼 여러 세트 복제하지 않는 한 며칠 동안 연속적으로 시험을 시행해야 하는 것이 현실이다[6].

이러한 상황에서 보안을 철저히 유지하여 체크리스트의 내용에 대해 기밀을 지키는 것은 통제가 가능하다고 해도 먼저 평가를 받은 학생이 나중에 평가를 받는 학생에게 자신이 평가받은 내용에 대해 함구하는 것은 현실적으로 어려운 일이다[7]. 이러한 물리적 한계 때문에 학생들은 시험날짜가 뒤로 갈수록 더 좋은 결과를 얻을 수 있을 것이라는 합리적인 의심으로 종종 평가의 공정성에 문제를 제기한다.

그러나 시험의 시점과 관련한 해외 연구 사례는 대부분 시험 날짜가 성적에 영향을 미치지 않거나 미비하다고 보고한다[7-10]. 국내 연구에서는 학생들의 정보전달은 진료수행 평가 성적에 영향을 미치지 않는다고 밝히면서, 먼저 평가를 받는 학생이 나중에 평가를 받는 학생들에게 전달하는 정보의 신뢰도가 낮다고 밝힌 바 있다[11]. 그러나 또 다른 국내 의과대학의 연구 사례에서는 정보 공유 상황에서 시험 일자가 뒤로 갈수록 직선으로 성적이 상승한다고 밝힌 바 있다[12].

진료수행시험에서 기존의 채점자의 신분과 시험 일자에 따른 채점 결과의 차이는 채점자간 점수의 일치도를 확인하거나, 시험에 관한 정보공유 상황에서의 결과의 차이를 밝히는 연구가 대부분이었다. 이러한 연구는 채점자의 요소나 시험 일자가 결과에 영향을 미칠 수 있다는 합리적인 가정에 근거한다. 그러나 실제 이러한 요소에 따른 학생 평가 결과에 차이가 있는 지 그 사실을 밝히는 연구는 부족한 형편이다.

본 연구는 진료수행 시험의 실시과정에서 물리적 환

경적 제한요소로 불가피하게 적용되고 있으나 공정성 문제가 꾸준히 제기되고 있는 두 가지 요소에 주목하여 집단 간 평가 결과에 차이가 있는 지 확인하는 것이 목적이다. 통상적으로 사용하는 모의시험 양상을 재현하여 채점자 요소에 따른 집단 간 성적 결과의 차이가 있었는지, 시험 날짜에 따른 집단 간 성적 결과의 차이가 있는지, 차이가 있다면 평가의 어떤 영역에 해당하는지를 확인하고자 한다.

## II. 연구방법

### 1. 대상

본 연구의 대상은 일개 의과대학에 재학 중인 본과 4학년 학생 101명이다. 평가 대상 학생들은 4개의 시험 구성 중 하나에 랜덤으로 배치되어 응시하도록 하였다. 진료수행시험의 문제 상황 4개는 지역컨소시엄을 통해 제공 받은 것으로 ‘본태성 고혈압’, ‘수면장애’, ‘기분저하’, ‘무월경’ 증례이다.

증례별 표준화 환자는 지역 컨소시엄에 소속되어 있는 12명이 연령과 성별에 맞추어 배치되어 교육받았다. 이중 8명의 표준화 환자는 4세트씩 2개로 이루어진 시험장에서 환자 연기를 수행하였다. 채점자는 표준화 환자 채점 군에 나머지 4명의 표준화 환자가 교육받은 진료상황에 배치되어 평가를 담당했고, 의사 채점 군에 4명의 대학병원에 근무하는 교원이 참여하였다. 모의시험 형식으로 치러진 시험은 1일차, 2일차로 나누어 시행되었는데, 학생들에게 시험 시작 전 정보통제에 대한 교육은 따로 시행하지 않았다.

#### 1.1 시험 일자에 따른 성별의 분포

의학과 4학년 총 101명이 시험에 응시하였다. 대상자의 평균 나이는 26.2세였고, 남학생은 75명(74.5%), 여학생 26명(25.7%)이다. 첫날 의사 채점자에게 평가를 받은 학생은 총 27명(49.1%), 표준화 환자에게 평가를 받은 학생은 총 28명(50.9%)이고, 둘째 날 의사 채점자에게 평가를 받은 학생은 총 24명(52.2%), 표준화 환자에게 평가를 받은 학생은 22명(47.8%) 이었다[표 1].

표 1. 채점 날짜와 채점자에 따른 성별의 분포 N(%)

	첫째날			둘째날		
	의사채점자	표준화환자 채점자	총합	의사채점자	표준화환자 채점자	총합
남학생	22 (81.5)	21 (75.0)	43 (78.2)	16 (66.7)	16 (72.7)	32 (69.6)
여학생	5 (18.5)	7 (25.0)	12 (21.8)	8 (33.3)	6 (27.3)	14 (30.4)
총합	27 (49.1)	28 (50.9)	55 (100)	24 (52.2)	22 (47.8)	46 (100)

### 2. 평가방법

진료수행평가는 8개의 방에 4가지씩 진료수행 상황을 차례로 배치하여 학생들이 실제 시험장에 입실하여 평가를 받는 형식으로 진행되었다. 각 상황별 응시시간은 문제 확인 1분, 환자 대면시간 10분이고 사이시험은 시행하지 않았다. 학생들의 평가는 채점자가 실시간 입실한 상태에서 채점이 바로 이루어져 시험이 종료된 후 모든 자료를 취합하여 점수를 합산하였다. 각 상황별 영역점수는 병력청취, 신체진찰, 환자교육, 환자의사관계, 신체진찰태도로 구분되어있고, 병력청취는 ‘예’가 3점, ‘아니오’는 0점이고, 신체진찰은 ‘체대로 함’ 3점, ‘체대로 못함’ 1점, ‘하지 않음’ 0점, 환자교육은 ‘우수’ 2점, ‘보통’ 1점, ‘미흡’ 0점, 환자의사관계는 ‘아주 우수’ 4점, ‘우수’ 3점, ‘보통’ 2점, ‘미흡’ 1점, 신체진찰태도는 ‘아주 우수’ 4점, ‘우수’ 3점, ‘보통’ 2점, ‘미흡’ 1점으로 구성되어 있다. 평가지의 구성과 비중은 표와 같다[표 2].

표 2. 상황별 체크리스트 구성 및 비중 N(%)

	본태성 고혈압	수면 장애	기분 저하	무월경	영역별 총점
병력청취	39 (41.1)	42 (44.7)	50 (52.6)	66 (70.2)	197 (52.1)
신체진찰	24 (25.3)	12 (12.8)	10 (10.5)	12 (12.8)	58 (15.3)
환자교육	8 (8.4)	16 (17.0)	15 (15.8)	12 (12.8)	51 (13.5)
환자의사관계	20 (21.1)	20 (21.3)	20 (21.1)	4 (4.3)	64 (16.9)
진찰태도	4 (4.2)	4 (4.3)			8 (2.1)
총점	95 (100)	94 (100)	95 (100)	94 (100)	378 (100)

### 3. 분석방법

응시한 학생들의 각 성취 내용은 기본정보인 성별을 포함하여 응시날짜, 채점자 군으로 구분하여 정리하였고, 각 학생들은 체크리스트에 배정된 점수에 따라 각 항목별로 평가 점수의 평균치와 표준편차를 작성하였다. 영역별 비교에서는 4개의 진료수행 상황에 배정된 총 점수에 얼마를 성취했는지 산출하여 성별, 시험 일자, 채점자 군에 따라 나누어 t-test를 통해 집단 간 차이를 비교하였다. 각 상황별, 영역별 비교에서는 각 평가영역을 표준점수로 변환 합산하여 t-test를 통해 집단 간 차이를 비교하였다. 결과 분석은 SPSS ver. 24 한글판을 이용하였으며, 유의성은 p-value < .05 이다.

## III. 연구결과

### 1. 진료 수행 총점에 따른 차이

#### 1.1 총점에 따른 결과의 비교

학습자 기본 특성과 시험 일자, 채점자간의 총점의 차이를 본 결과 성별에 따른 총점의 차이는 여학생이 311.12±34.10점, 남학생이 280.76±40.93점으로 여학생의 점수가 높아 유의한 차이를 보였고([그림 1], p=.001), 시험 일자에 따른 차이는 첫째날의 총점이 278.45±41.8 이었고, 둘째날이 300.7±37.7로 더 높게 나타나 통계적으로 유의한 차이가 있었다([그림 2], p=.007). 채점자간의 차이는 의사 채점자는 291.60±43.09점이고, 표준화환자 채점자는 285.37± 39.57점으로 통계적으로 유의한 차이는 없었다([그림 3], p=.452).

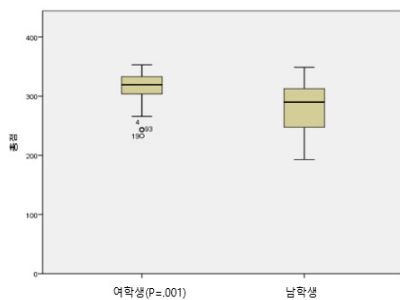


그림 1. 성별에 따른 총점의 차이

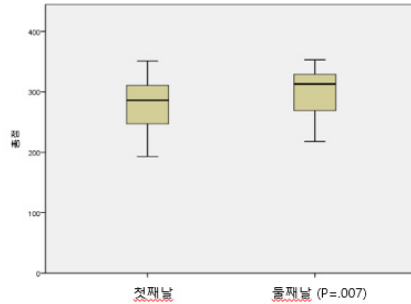


그림 2. 시험 일자에 따른 총점의 차이

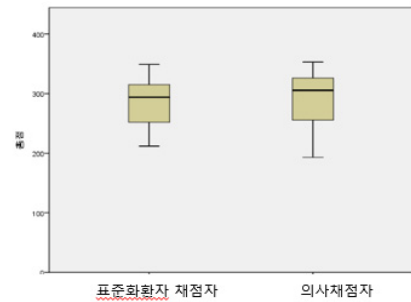


그림 3. 채점자간 총점의 차이

1.2 시험 일자에 따른 성별 및 채점자간 총점의 차이  
 성별에 따른 시험일자별 차이는 첫째날 여학생이 점수가 더 높았고(p=.008), 둘째날은 유의한 차이가 없었다. 채점자간 시험 일자별 총점은 유의한 차이가 없었다[표 3].

표 3. 시험일자에 따른 성별 및 채점자간 차이

		첫째날		둘째날	
		M±SD	P	M±SD	P
성별	여학생	306.1±35.3	.008	315.4±33.8	.079
	남학생	270.7±40.5		294.2±38.0	
채점자	표준화환자	278.2±39.9	.958	294.2±38.2	.272
	의사	278.8±44.3		306.6±37.1	

### 2. 진료수행시험의 상황별 결과의 비교

4가지 상황별로 총점을 비교해 본 결과 성별에 따른 차이는 여학생이 남학생보다 4가지 상황 모두 점수가 높게 나타나 통계적으로 유의하였고, 시험 일자에 따른 차이는 둘째날 시험을 본 학생들의 점수가 본태성고형

압, 수면장애, 기분저하의 세 가지 상황에서 첫째날 시험을 본 학생의 점수보다 높게 나타나 통계적으로 유의한 차이를 보였고, 채점자 간의 차이는 4가지 상황 모두에서 유의한 차이는 없는 것으로 나타났다[표 4].

표 4. 상황별 결과의 비교

	본태성 고혈압	수면장애	기분저하	무월경
남학생	48.6±10.1	48.0±9.9	48.4±10.4	48.8±10.9
여학생	54.2±8.6	56.0±7.9	54.8±7.1	53.5±9.1
p	.012	.000	.004	.041
첫째날	47.1±9.8	48.3±10.2	47.9±10.5	48.6±10.1
둘째날	53.5±9.4	52.7±8.3	52.4±9.0	51.6±9.8
p	.001	.023	.026	.139
의사채점자	51.2±9.8	51.9±9.5	48.5±11.3	51.3±10.5
표준화환자 채점자	48.7±10.3	48.6±9.5	51.6±8.4	48.6±9.5
p	.213	.078	.126	.175

M±SD

### 3. 진료수행시험 상황의 영역별 비교

진료수행 시험 상황 4가지의 각 영역별 표준화 점수를 합산한 결과를 비교하면 여학생의 점수가 병력청취 (p=.002), 신체진찰(p=.005), 환자교육(p=.018), 의사환자관계(.004), 진료태도(p=.036) 모두에서 남학생보다 점수가 높은 것으로 나타났고, 통계적으로도 유의한 차이가 있었다.

표 5. 영역별 결과의 비교

	병력 청취	신체 진찰	환자 교육	의사환자 관계	진료 태도
남학생	48.2±10	48.3±10	48.6±10	48.3±9	48.8±9
여학생	55.2±7	54.7±8	54.0±8	54.8±8	53.5±9
p	.002	.005	.018	.004	.036
첫째날	47.5±10	47.7±10	48.4±10	48.3±9	49.3±8
둘째날	53.0±8	52.7±9	51.9±9	52.1±10	50.9±11
p	.005	.012	.075	.059	.437
의사채점자	50.2±11	51.1±9	51.5±9	49.9±9	51.6±10
표준화환자 채점자	49.8±8	48.8±10	48.4±10	50.1±10	48.3±9
p	.851	.241	.110	.919	.089

M±SD

시험일자에 따른 차이는 병력청취(p=.005), 신체진찰 (p=.012), 두 가지 영역에서만 첫째날보다 둘째날의 점수가 높은 것으로 나타나 유의한 차이를 보였다. 채점자간의 차이는 의사 채점자에게 평가를 받은 학생의 점

수와 표준화환자 채점자에게 평가를 받은 학생들의 점수는 통계적으로 유의한 차이를 보이지는 않았다(p>0.5)[표 5].

## IV. 논의 및 고찰

진료수행 시험에서 학생의 성별에 따른 집단 간 차이를 발표한 연구는 대부분 증례의 성적 민감성이나 표준화 환자의 성별에 따른 차이를 밝히고 있으며[13][14], 국내연구에서 김종훈 등(2007)은 성(性) 적으로 민감한 증례와 중립적인 증례의 일부에서 남녀 학생 간의 평가 결과에 차이가 있음을 밝힌 바 있다. 여학생이 남학생에 비해 전반적으로 높은 점수를 획득하나, 평가 영역별로 고려할 때, 환자의사관계 영역만 유의미한 차이를 보인다는 것이다[15].

본 연구에서는 활용된 세 가지 중립 사례와 한 가지 민감 사례 모두 대부분 여학생의 총 성취 점수가 남학생 보다 높아 이전과는 다른 결과를 보였다. 평가 영역별로 볼 때도 병력청취, 신체진찰, 환자의사관계, 환자교육, 진료태도의 모든 영역에서 여학생이 남학생보다 높은 유의한 차이를 보였다.

여학생이 모든 영역에서 남학생보다 높다는 것은 진료수행 평가 이전의 학업성취도에서 여학생 집단의 성취도가 남학생 집단보다 높을 것을 시사한다. 진료수행이 도입된 초기에는 진료수행에 필요한 의학적 지식이나 술기의 수행이 성별로 차이가 없을 수 있다. 그러나 지난 10여 년간 진료수행과 관련한 교육과정의 체계화되고, 학생들이 평가를 준비하기 위한 전략들이 정착되면서 다른 과정과 마찬가지로 ‘진료수행’ 과정 또한 여학생의 성취가 높을 수 있다. 실제 국내 의과대학의 성별에 따른 성취도 차이를 연구한 내용을 보면, 대부분 여학생이 남학생과 비교하여 우수하다[16][17]. 이번 연구결과를 보면, 진료수행에 대한 평가 또한 일반적인 의과대학 교육과정의 학업성취도와 같은 맥락에서 이해될 수 있다.

채점자 집단에 따른 평가 결과는 총점의 유의한 차이는 없었다. 상황별로 보면 4가지 임상상황 중 3가지 상

황에서 의사 채점 집단의 평균점수가 표준화환자 채점 점수보다 약간 높게 나타났으나 유의미한 차이는 없었고, 영역별로는 병력청취, 신체진찰, 환자교육, 진료태도에서 의사 채점군이 표준화 환자 채점군보다 점수가 약간 높게 나왔으나 통계적으로 유의한 차이는 보이지 않았다.

2000년대 초 중반 국내외 연구에서는 의사 교원의 평가가 표준화 환자의 평가 보다 점수가 많이 낮았으며, 특히 병력청취 영역에서 표준화 환자 채점군이 의사 채점군보다 10점 이상의 차이를 보여 표준화 환자의 평가 교육이 논의된 바 있다. 표준화 환자가 학생을 배려하는 오류를 보이고, 전문적 지식이 미비하여 병력청취 등의 영역에서 부정확한 평가를 할 수 있다는 것이다 [18].

그러나 실제 본 연구에서는 각 상황과 영역별로 세부적인 요소에서 표준화 환자 채점 집단과 의사 채점 집단 간의 차이가 확인되기는 하나, 표준화 환자 채점군이나 의사 채점군의 평가 결과가 다른 집단에서 특정영역이나 사례에서 높다는 전형적인 내용이 확인되지 않는다. 이러한 결과는 진료수행시험 도입 후 10여년이 지나면서 평가 지침이 표준화되고, 표준화 환자 교육 또한 이전과 비교하여 높은 수준으로 수행되었음을 의미한다. 현재 진료 수행을 위한 평가는 행위에 초점을 맞춘 체크리스트가 개발되어 반영되고, 견소시험의 형태로 지속적인 표준화 환자 교육이 이루어지면서 평가의 임의적인 판단이 반영되기는 어려운 구조가 되었다. 본 연구에서는 표준화환자 채점자와 의사 채점자 특성에 따른 차이는 확인되지 않으며, 일부 특정 상황의 특정 영역에서 확인된 평가 점수의 차이는 채점자 개별 특성에 의한 것으로 추정된다. 이와 관련하여 본 연구에서 표준화환자와 의사 채점자간 일치도가 산출되지 않아 평가의 신뢰성을 보고하지 못한 것은 연구의 한계이다.

시험 날짜에 따른 결과를 비교하면 둘째날 시험을 본 학생들의 점수가 첫 날 시험을 본 학생보다 높았으나 통계적으로 유의한 차이를 보이는 것은 병력청취와 신체진찰 영역이다. 박훈기(2005)등에 의하면 평가 시기가 늦어질수록 성적이 증가하는 경향이 보이나, 사례에

따라 통계적으로 유의한 차이를 보이는 것은 없었으며, 병력청취, 환자교육, 예절 등의 영역에 한정한다[12]. 실제 학생이 시험에 관한 정보를 취득하더라도 신체진찰 등의 영역은 단순한 지식뿐 만 아니라 술기를 익혀야 하므로 단시간에 향상되기 어려우며, 환자-의사관계와 같은 영역은 단 시간에 교정이 되지 않아 별 차이가 없다는 것이다[15]. 더불어, 연구자들은 평가받는 학생이 다른 학생에게 정보를 전달했을 때 그 신빙성이 낮아 실제 학생들의 평가 결과에 미치는 영향은 미비하다고 지적한 바 있다.

진료수행에서 진행 시나리오, 평가 체크리스트 등이 완벽하게 통제된 상황이라면, 먼저 평가를 받는 학생이 다른 학생들에게 전달할 수 있는 정보는 환자가 처음 의사를 방문하는 목적에 해당하는 '주 증상'에 관한 정보가 전부이다. 사이 시험이 없는 상황에서 학생들이 주 증상에 대한 정보를 취득한다면 병력청취의 성취수준을 높일 수 있을 것이라 예상된다. 신체진찰과 관련하여서는 술기 역량을 강화한 교육체제의 변화를 짐작해볼 수 있다.

2000년대 초반과 비교하여 현재 우리나라 의과대학의 술기 교육은 진일보한 상황이다. 2009년부터 의사 술기 역량이 국가고시에 도입되면서 대부분의 의과대학은 술기센터를 통해 저학년부터 조기 교육을 시행하고 있다. 의학과 4학년 학생의 술기 역량은 기초적인 신체진찰 등의 술기 수행이 가능한 상태이다. 늦은 날짜에 시험을 본 학생이 다른 학생들로부터 평가에 대한 정보를 취득했다면, 신체진찰과 관련한 수행수준은 비교적 짧은 시간에 보완 가능한 상황이다. 주 증상에 대한 정보를 통해 신체진찰의 여러 요소 중 무엇을 어떻게 선택해서 수행할지 결정할 수 있어 그렇지 않은 집단에 비해 높은 점수를 획득할 수 있다.

이와는 다르게 응시날짜에 따른 환자교육, 환자-의사관계, 진료태도 영역 등은 점수 차이는 없었다. 진단에 대한 정보나 추정 없이는 환자교육이 불가능하고, 실제 학생들 사이에 전해지는 정보가 일반정보에 한정되어 있고, 그 내용 또한 신빙성이 낮았음을 짐작해 볼 수 있다. 환자-의사관계나 진료태도 영역은 학생 교유의 특성과 관련하거나, 장기간에 걸친 교육을 통해 축적되는

학습의 산물이다. 이러한 영역은 단 시간에 교정이 되는 것이 아니기 때문에 하루 이틀 간격에 따른 평가 결과의 차이는 없었을 것으로 생각된다.

우리나라는 2009년 의사 국가시험에 실기 시험이 도입되어 반영되면서, 진료수행 평가와 관련해서는 그 체계나 내용의 타당성이나 합리성이 확보된 상황이다. 진료수행에 관한 내용 또한 무엇을 어떻게 가르치고 평가 할지에 관한 세부적인 내용도 개선되고 교육에 필요한 다양한 도구도 확보된 상태이다. 그러나 일개 의과대학에서는 국가고시를 준비하면서 시행하게 되는 ‘진료수행’의 평가에서는 공간, 인력, 재정 등의 물리적인 제약 때문에 국가고시의 실기시험을 그대로 재현할 수 없는 한계가 있다. 이러한 상황에서 학교에서 학생들에게 시험의 공정성 측면에서 꾸준히 제기되는 것은 ‘채점자가 누구이며’, ‘응시날짜가 첫날인지, 둘째 날인지’에 관한 것들이었다. 본 연구에서는 채점자의 신분에서 따른 평가 결과에 차이가 없음을 다시 한 번 밝히면서 표준화 환자를 활용한 평가의 가능성을 확인하였다. 응시날짜에 따라 집단 간 차이가 유의미한 경우는 기존에 밝혀진 ‘병력청취’와 더불어 ‘신체진찰’ 영역도 포함되어 있음을 주목해야한다. 학생들 사이에 전해지는 정보를 완벽하게 통제 할 수 없는 상황에서 실제 학생들의 성취도를 정확하게 평가할 수 없는 한계를 확인하고, 공정성의 측면에서 이러한 점수 차이를 어떻게 반영할 것인지에 대한 연구가 필요하다.

#### 참 고 문 헌

- [1] Y. H. Lee, J. H. Park, J. K. Ko, and H. B. Yoo, "The Change of CPX Scores according to Repeated CPX Scores according to Repeated CPXs," *Korean J Med Educ.*, Vol.23, No.3, pp.193-202, 2011.
- [2] H. S. Barrows, "An overview of the uses of standardized patients for teaching and evaluating clinical skills," *Academic Medicine*, Vol.68, No.6, pp.443-451, 1993.
- [3] 이병국, "임상수행평가의 세계적 추세에 따른 표준화환자 프로그램의 토착화 방안 연구," *한국의 학교교육*, 제12권, 제2호, pp.141-150, 2000.
- [4] 박훈기, 이정권, 황환식, 이재웅, 최윤영, 김혁, 안동현, "객관구조화진료시험(OSCE)에서 교수와 표준화환자 사이의 점검표 채점 일치도," *한국의 학교교육*, 제15권, 제2호, pp.141-150, 2003.
- [5] 권복규, 김나진, 이순남, 어은경, 박혜숙, 이동현, 박미혜, 오지영, 한재진, 허정원, 유경하, "표준화환자를 이용한 임상수행능력평가시험(CPX)에서 교수와 표준화 환자의 결과 비교," *한국의학교교육*, 제17권, 제2호, pp.173-183, 2005.
- [6] 박훈기, 이정권, 김승룡, 김경태, 박혜영, "시험장 복제(Duplication)가 객관적 구조적 임상 시험(OSCE)의 신뢰도에 미치는 영향," *한국의학교교육*, 제11권, 제1호, pp.37-52, 1999.
- [7] P. J. Rutala, D. B. Witzke, E. O. Leko, J. V. Fulqiniti, and P. J. Taylor, "Sharing of information by students in an objective structured clinical examination," *Arch Intern Med*, Vol.151, pp.541-544, 1991.
- [8] P. L. Stillman, H. L. Haley, A. I. Sutnick, M. M. Philbin, S. R. Smith, J. O'Donnell, and H. Pohl, "Is test security an issue in a multistation clinical assessment?: A preliminary study," *Acad med*, Vol.66, No.9, pp.25-27, 1991.
- [9] N. V. Vu, M. L. Marcy, J. A. Colliver, S. J. Verhulst, T. A. Travis, and H. S. Barrows, "Standardized patient's accuracy in recording clinical performance check list items," *Med Edu.*, Vol.26, pp.99-104, 1992.
- [10] R. G. Williams, J. S. Lloyd, and D. K. Simonton, "Sources of OSCE examination information and perceived helpfulness: a study of the grafevine," In I. R. Hart, R. M. Harden, and J. D. Marchais(eds.), *Current developments in assessing clinical competence*, Montreal: Can-Heal Publications, pp.363-370, 1992.

- [11] 김종훈, 이기영, 유동미, 양은배, "진료수행평가에서 학생들 사이의 정보공유가 평가결과에 미치는 영향," 한국의학교육, 제18권, 제3호, pp.239-247, 2006.
- [12] 박훈기, 권오정, "진료수행평가(CPX: Clinical Performance Examination)에서 정보공유가 시험 성적에 미치는 영향," 한국의학교육, 제17권, 제2호, pp.185-195, 2005.
- [13] G. Furman, J. A. Colliver, and A. Galofré, "Effects of student gender, standardized-patient gender in a single case using a male and a female standarized patient," Acad Med, Vol.68, No.4, pp.301-303, 1993.
- [14] K. A. Chambers, J. R. Boulet, and G. E. Furman, "Are interpersonal skills ratings influenced by gender in a clinical skills assessmant using standardized patients?," Advances in Health Sciences Education, Vol.6, No.3, pp.231-241, 2001.
- [15] 김종훈, 이기영, 유동미, 양은배, "진료수행평가에서 증례 유형과 표준화 환자 성별이 학생 성적에 미치는 영향," 한국의학교육, 제19권, 제1호, pp.23-29, 2007.
- [16] 안덕선, 임형, 김경성, "예과 입학성적, 졸업 총평점, 의사 국가고시간의 상관관계 및 성별 비교에 관한연구," 한국의학교육, 제12권, 제1호, pp.91-96. 2000.
- [17] 김하원, 고희선, 이무송, 남주현, 김원동, 이재담, "일개 의과대학에서 입학 시 특성에 따른 의예과 학업성취도에 대한 연구," 한국의학교육, 제14권, 제2호, pp.257-268, 2002.
- [18] N. Heine, K. Garman, P. Wallace, R. Bartos, and A. Richards, "An analysis of standardized patient checklist errors and effect on student scores," Med Edu, Vol.37, No.2, pp.99-104, 2003.

저 자 소 개

권 소 희(So-Hee Kwon)

정회원



- 2005년 2월 : 원광대학교(박사)
- 2018년 3월 ~ 현재 : 원광대학교 의과대학 책임연구원

<관심분야> : 술기교육, 보건교육

김 영 전(Young-Jon Kim)

정회원



- 2014년 8월 : 서울대학교(박사)
- 2017년 3월 ~ 현재 : 원광대학교 의과대학 교수

<관심분야> : 의학교육, 교육과정개발 및 평가