

Clinical Performance Examination Utilizing Standardized Patients in Board Examination: Based on the Board Examination of Korean Neurological Association for Three Years

Beom Joon Kim^{1,*}, Jung-Joon Sung^{1,*}, Hoon Ki Park², Dae-Won Seo³, Chin-Sang Chung³ and Byung-Woo Yoon¹

¹Department of Neurology, Seoul National University Hospital, Seoul National University College of Medicine,

²Department of Family Medicine, Hanyang University Hospital, Hanyang University College of Medicine, and

³Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

전문의 고시에서 표준화 환자를 이용한 진료 수행 시험의 경험: 3개 연도 신경과 전문의 시험 결과 분석을 바탕으로

¹서울대학교병원 신경과, ²한양대학교병원 가정의학과, ³삼성서울병원 신경과

김범준^{1,*}, 성정준^{1,*}, 박훈기², 서대원³, 정진상³, 윤병우¹

Purpose: Evaluation of clinical skills and attitude including development of dynamic patient-doctor relationship is important in board examination (BE). Korean Neurological Association (KNA) has introduced clinical performance examination (CPX) utilizing standardized patients (SP) to BE in 2007. In this study, the authors describe the 3-year experience of CPX in BE through 2009.

Methods: To implement CPX session in BE, KNA developed CPX workshop for BE attendees and members of grading committee. CPX sessions in BE consisted of two model scenarios mimicking neurological patients in clinical practice. The total score and itemized scores of CPX sessions were compared with other areas of BE, and scores from each year were also compared.

Results: Scores from CPX sessions were significantly correlated with BE step II. Among the itemized scores of CPX sessions, clinical items including history taking and physical examination were significantly correlated with scores from other areas of BE. However, scores from global assessment from SP were strongly associated with patient-doctor relationship, history taking, and patient education.

Conclusion: Our experiences suggest that CPX utilizing SP is a useful tool to assess the clinical skills in BE. In order to produce clinically well qualified neurologists, more efforts should be made to develop cases and to improve assessment tools for CPX.

Key Words: Education, Educational measurement, Licensure, Clinical competence, Patient simulation

Received: November 9, 2010 • Revised: December 10, 2010 • Accepted: January 4, 2011

Corresponding Author: Byung-Woo Yoon

Department of Neurology, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 110-744, Korea

Tel: +82.2.2072.2875 Fax: +82.2.3672.7553 email: bwoon@snu.ac.kr

* These authors contributed equally to this work.

Korean J Med Educ 2011 Jun; 23(2): 127-135.

doi: 10.3946/kjme.2011.23.2.127.

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.

서론

의학 교육에 있어 기본적인 평가 영역은 의학 지식, 의학 술기의 사용, 의사로서 환자를 대하는 태도로 분류할 수 있다. 특히 전문의 자격시험(board licensure examination)은 양질의 의료를 추구하는 사회적 요구에 대하여, 해당 학회가 그 분야의 전문가임을 인증하기 위하여 시행하는 평가이다. 즉, 환자를 진료함에 있어 필수적인 해당 의학 및 인접 분야의 지식(knowledge), 임상 수기(skill), 태도(attitude), 각종 검사의 시행 및 판단 능력을 수험생이 갖고 있는지 전반적으로 평가하여 전문가로서의 자격 여부를 판단하고, 궁극적으로 신경과 전공의 수련 과정의 목표를 제시하는 것이다[1,2].

현재 우리나라의 전문의 자격시험은 해당 학회마다 약간의 차이는 있으나 일반적으로 서류 심사, 1차 시험의 필기고사 그리고 2차 혹은 2~3차 시험에서 슬라이드 또는 실물시험, 구술시험으로 구성되어 있다[3]. 대한신경과학회 역시 서류 심사 단계에서 수험생의 임상경험, 수련 병원에서의 교육 및 학회 참석 여부, 발표 논문을 통해 전문의 고시 수험 자격을 부여하였으며, 1차 시험에서는 기본적 지식 위주의 필기시험, 2차에서는 슬라이드 및 동영상시험, 구술시험을 통해 전문의 자격을 평가하였다. 따라서 임상 의사로서 각종 문제를 해결하는 능력을 평가할 수는 있었지만, 전문의로서 환자와 동적인 상호 관계(dynamic interaction)를 형성하고 이끌어가는 임상적인 능력은 평가할 수 없었다. 이에 대한신경과학회는 2007년도 신경과 전문의 고시부터 2차 시험에 표준화 환자(standardized patient, SP)를 활용한 진료수행시험(clinical performance examination, CPX)을 도입하여, 신경과 전문의로서 수험생들이 갖추어야 하는 병력 청취와 신체 진찰을 포함하는 정보 수집 능력과 환자-의사 관계 형성 능력을 평가하고자 하였다[1,4].

SP를 이용한 진료수행시험은 1970년대 초 영국에서 객관구조화 진료시험(objective structured clinical examination, OSCE)의 형식으로 개발되어, 미국에서는 SP를 활용하는 진료수행시험의 형태로 발전하였으며, 2004년도부터 미국 내 의과대학의 졸업 자격 인증 시험에 활용되고 있다. 국내에서도 1990년대 중반 이후 여러 의과대학에서 OSCE가 많이 이용되

고 있다가 한국 의사국가고시에서는 2010년도 필기시험과는 별도로 당락이 결정되는 의사실기시험으로 SP를 이용한 CPX가 도입되었다. 전문의 고시에서는 대한성형외과학회가 2005년도부터 OSCE를 2차 시험에 도입하였고[3], 신경과가 처음으로 실기 시험의 일환으로 전문의 고시에서 표준화 환자를 이용한 CPX 형태의 시험을 추가하게 되었다. 특히 대한신경과학회의 시도는, 국내에서 최초로 전문의 고시에 표준화 환자를 도입했다는 의의를 갖는다고 할 수 있다. 이러한 맥락에서, 본 논문은 2007년도부터 2009년도까지 신경과 전문의 시험에 도입된 CPX의 운영 경험과 시험 결과를 분석하였다. 일차적으로는 대한신경과학회에서 CPX를 도입하고 적용한 경험을 제시하고, 이차적으로는 CPX가 전체 신경과 전문의 시험 성적과 갖는 상관관계를 분석하기로 한다. 그리하여 궁극적으로 CPX의 실행 가능성 및 타당성을 확인하여, 타 전공 학회의 전문의 고시에서 표준화 환자 활용 진료 시험 도입에 도움을 주고 CPX 시험의 향후 발전 방향을 제시하고자 한다.

대상 및 방법

1. CPX 모의시험의 개발과 시행

신경과 전문의 고시에서 SP를 활용한 CPX를 도입하는 데 있어서 전공의 수련 프로그램의 지도 전문의와 수험생을 대상으로 CPX에 대한 오리엔테이션이 우선적으로 필요했다. 대한신경과학회 주관으로 지도 전문의에게 CPX 개념을 설명하고 신경과 전공의 수련 교육에 반영될 수 있도록 CPX 모의 시험을 시행하였다.

2007년 6월 신경과 CPX 시험 준비위원회(이하 준비위원회)가 처음으로 구성되었다. 대한신경과학회 고시위원장, 고시위원회 간사, 각 의과대학에서 CPX 시험 시행 경험을 갖고 있는 신경과 교수, 임상수행능력평가를 위한 컨소시엄(이하 컨소시엄)에 참여하고 있는 신경과 교수 및 의사국가고시에서 실기 시험 CPX 준비 작업에 참여 경험이 있는 교수 등이 준비위원회에 포함되었다. 준비위원회의 전체 인원은 약 6~8명으로 하였고, 되도록 신경과학의 각 세부 전공을 전공하는 교수가 빠짐없이 포함되도록 구성하였다.

CPX 시험의 응시 및 채점 과정에 대해 수험생과 채점위원들이 새로운 제도를 사전에 경험하기 위해, 준비위원회는 10월에 CPX 모의시험을 실시하기로 결정하였다. 신경과 CPX 모의시험은 당해 연도 신경과 전문의 시험에 응시 예정인 모든 신경과 4년차 전공의를 대상으로 시행되었다. 이를 위하여 2007년 7월, 각 세부 전공별로 실제 환자의 병력과 신체 진찰 소견에 바탕을 둔 사례를 작성하도록 하였다. 2007년 8월, 개발된 사례를 검토하고 수정하는 워크숍을 진행하였으며, 모의시험에서 사용할 두 개의 사례를 문제로 선택하였다. 동년 9월, 2개의 사례에 대하여 시나리오와 채점표, 채점 기준표를 최종적으로 완성하였다. 이때 시나리오의 내용과 채점 기준은 세부 전문의들의 전문가 토론으로 결정되었다. 동년 8월, 서울·경기 CPX 컨소시엄에서 활동 중인, 표준화 환자 연기 경험을 가진 배우 중에서 모의환자를 모집하였다. 선정된 후보 사례 두 개에 적절한 성별 및 연령을 가진 배우를, 각 사례당 다섯 명씩 선발하였다. 이들은 모의 CPX 시험 실시 하루 전에 소집되어 CPX 시험 준비를 시작하였다. 준비위원회 위원 중 사례별로 교수 한 명이 훈련을 담당하였고, 다섯 명의 표준화 환자에게 사례에 적합한 병력 및 신체 진찰 소견을 실제 환자와 똑같이 재현하는 연기 훈련을 지도하였다. 세 시간 동안 내용 교육 및 연기 수정으로 구성된 기본 훈련을 하였고, 이후 사례별로 서로 교체하여 두 시간 동안 교차 점검 훈련을 실시하였다. 연기 훈련 시간 중, 신경학적 이상 소견을 정확히 표현할 수 있도록 다섯 명의 표준화 환자를 일치시키는 과정이 전체 시간의 절반 정도를 차지하였다.

모의 CPX 시험의 채점위원은, 일정 기간 이상 의과대학 교육의 경험을 갖고 있으며, 매년 실시하는 대한신경과학회 고시위원회 워크숍에 참여한 신경과 교수들 중에서 자발적으로 참여 의사를 밝힌 자원자를 중심으로 구성하였다. 채점위원은 시험 시작 2시간 전에 모여, 표준화 모의환자 교육 중 촬영한 표준 동영상상을 바탕으로 사례에 관한 교육을 받았으며, 이 동영상상을 보면서 채점을 하고 상호 비교하여 채점위원 사이의 일치도를 높이는 훈련을 받았다.

사례별로 5개씩 시험 장소(station)가 배정되었으며 모의시험임을 감안하여, 각 시험 장소당 1명의 채점위원을 배치하였으며, 수험생이 해당 수련 병원의 교수와 만나지 않도록 배정하였다. CPX 모의시험은 다음과 같은 방식으로 시행되었다.

두 개의 사례를 각 다섯 개의 시험 장소에서 시행하였다. 수험생은 제비뽑기로 수험번호를 부여 받고, 두 개의 사례 중 어느 하나에 임의로 배정되었다. 수험생은 각 시험 장소의 출입문에 붙어 있는 상황지침을 통하여 환자의 간단한 인적 정보와 활력 징후(vital sign)를 파악한 후 입실하였다. 각 수험생별로 10분의 진료 시간이 주어졌으며, 이 시간 동안 수험생은 병력 청취, 진찰, 환자 상담과 교육을 완료하였다. 시험 시작 8분째에 진료 시간 마감 2분 전을 알리는 방송을 하였다. 표준화 환자 면담이 끝난 이후 5분 동안 채점위원으로부터 구술 평가를 받았다. 각 연도별 모의 CPX에서 수험생은 80명 내외이었으며, 시험 종료 후 수험생은 시험 운영에 관한 평가 설문지를 작성하였다.

채점 항목은 병력 청취, 진찰, 환자 교육, 환자-의사 관계 및 구술의 영역으로 구성되었다. 표준화 모의환자는 각 수험생의 진료 시간이 종료된 직후 대기실로 나와 수험생의 전반적인 진료에 대하여 전반적인 평가(global assessment)를 중심으로 한 평가서를 작성하였다. 모든 시험이 끝난 후 채점위원도 모의시험의 운영에 관한 설문지를 작성하였다.

2. 신경과 전문의 고시에서 CPX의 시행

신경과 CPX 준비위원회는 CPX 모의시험의 결과를 바탕으로 전문의 고시에서 CPX의 시행을 준비하였다. 2007년 10월 무렵부터 준비위원회 위원들은 각 세부 전공별로 실제 환자에 근거하여 전문의 고시 2차 시험에서 사용할 CPX 환자 사례를 개발하였으며, 12월 중에 이들 사례 중 전문의 시험에서 사용될 최종 사례를 결정하였다. 결정된 사례에 근거하여, 모의시험 때와 동일한 방식으로 서울·경기 CPX 컨소시엄 소속 표준화 모의환자 배우 중 적절한 성별과 연령을 가진 배우 10명을 선발하였다. 선발된 배우들은 시험 시행 1일 전에 소집되어, 사례별로 각 2명의 평가 위원으로부터 별도의 내용 훈련, 평가 및 교차 점검을 받았다. 전문의 시험의 보안을 유지하기 위하여, 교육자와 표준화 모의환자는 소집되면서부터 시험 준비위원회와 함께 이동하였다.

CPX의 채점위원은 대한신경과학회 고시위원회 워크숍에 참석한 경험이 있으며, 모의시험 채점위원으로 활동한 바 있는 전공의 수련 병원 소속 교수들 중 20명을 위촉하였다. 또한 지역과 병원을 안배하여 총 10개의 시험 장소 각 2명씩 배

정하였으며, 동일한 수련 병원의 수험생과 소속 교수가 만나지 않도록 배려하였다. 채점위원은 시험 시작 두 시간 전에 모여 채점 교육을 받았고, 채점위원 각자에게 배정된 표준화 모의환자의 비디오를 보며 실제로 채점 연습을 하였다.

신경과 전문의 고시 2차 시험은 2일 동안 시행되며, 첫째 날에 동영상 및 슬라이드시험을 시행하고, 둘째 날에는 CPX와 일반 구술시험을 시행하였다. 수험 장소 및 채점위원 동원 등의 제약으로 인하여, 추첨을 통해 임시 수험 번호를 배정받은 수험생을 오전과 오후 2개 조로 나누어 시행하였다. 단, CPX 시험의 보안 유지를 위하여, 오전 수험생은 오후 수험생이 집합을 완료한 이후에 해산하도록 조치하였다. 각 오전 및 오후 조 수험생은 각각 다시 2개 조로 분류하여, 구술시험-CPX 시험 및 CPX 시험-구술시험의 순서로 평가를 받았다. CPX 시험에서, 수험생은 20분 동안 환자를 대면하여 병력 청취를 하고 신체 진찰을 수행하였다. 이후 순서에 따라 수험생이 필요하다고 판단한 검사 결과를 제공하였으며, 이를 바탕으로 환자 상담 및 교육까지 완료하도록 하였다. 20분의 세션이 완료된 이후, 10분 동안 직전 진료 수행 내용을 바탕으로 채점위원 2인으로부터 구술 평가를 받았다. 채점위원의 평가 항목은 병력 청취, 신체 진찰, 환자 교육, 환자-의사 관계 및 구술의 영역으로 구성되어 있었고, 각 영역에는 해당 영역의 전반적인 평가 항목과 함께 세부 채점 항목을 기록하도록 하였다. 마지막으로 수험생의 진료 수행에 대한 전반적 평가 항목을 추가하였다. 두 명의 채점위원은 채점 과정에서 평가 기준 적용에서 논란의 여지가 있거나 기준 해석이 애매한 경우 서로 상의를 하면서 채점할 수 있었다. 두 명의 채점위원이 각

각 평가한 세부 항목 점수는, 통계적으로 유의한 상관관계를 보였다. 표준화 모의환자는 첫 20분의 대면 진료 세션이 종료된 후 별도의 대기실에서 수험생의 태도 및 진료 능력에 대하여 전반적 평가를 하였다. 시험이 종료된 이후, 수험생 및 채점위원을 대상으로 표준화 환자 시험 시행에 관한 설문 조사를 시행하였다.

연도별 신경과 전문의 고시 총점의 분포는 2007년 154.49 ± 10.13점, 2008년 153.24 ± 9.50점, 2009년 156.31 ± 12.37점으로 자세한 내용은 Table 1에서와 같다.

3. 분석

CPX 시험이 시행된 2007년, 2008년 및 2009년도 신경과 전문의 시험의 원자료를 이용하여 분석을 시행하였다. 신경과적 지식 위주의 1차 시험, CPX 시험 점수를 제외한 나머지 2차 시험, CPX 시험 및 CPX 모의시험의 점수 사이의 상관관계를 분석하였다. 이때, 연도별로 만점에 차이가 있음을 고려하여, 만점에 대한 백분율의 평균 및 표준 오차를 서로 비교하였다. 연도별 점수의 차이는 일원 분산분석(analysis of variance)을 시행하였고, 두 개 연도 간의 사후 검정은 Bonferroni 보정을 적용한 짝지음 분석(pairwise analysis)을 시행하였다. 또한 CPX를 구성하는 4~5개의 세부 항목 점수가 전문의 고시의 다른 평가 영역과 갖는 상관관계를 분석하였고, 연도별 CPX의 각 세부 항목 원점수와 1차, CPX 시험 점수를 제외한 나머지 2차 시험 및 CPX 모의시험 총 원점수 사이의 상관관계도 분석하였다. 2008년도 수험생에 대한 CPX 표준화 환자의 평가는 A~E의 5단계 전반적 평가와 기

Table 1. Baseline Characteristics of Board Licensure Examination of Korean Neurological Association in Year 2007 to 2009

	2007	2008	2009	p-value
Workshop				
No. of attendees	69	73	76	
Score	65.00 ± 16.57	60.57 ± 11.21	57.32 ± 13.20	<0.01
Board examination				
No. of attendees	76	79	88	
First step	82.75 ± 4.40	83.98 ± 4.99	81.99 ± 6.64	0.08
Second step	70.54 ± 10.96	68.79 ± 6.39	73.83 ± 7.27	<0.01
CPX	9.82 ± 1.98	8.78 ± 2.05	11.06 ± 2.34	<0.01

Scores of second step including movie, slide, oral test, and clinical performance examination (CPX). Scores of workshop was re-calculated on the basis of 100 points. p-values were calculated by analysis of variance tests.

술적 평가(descriptive evaluation)으로 구성되어 있었는데, 이 중 전반적 평가를 A=5, B=4, C=3, D=2, E=1로 정량화하여 다른 시험 점수와의 상관관계를 분석하였다. 2007년도 및 2008년도 CPX 시험에서 두 채점위원 사이의 일치도는 스피어만 상관 계수(Spearman correlation coefficient)를 이용하여 분석하였다. 그리고 채점위원과 수험생의 설문 조사 응답을 기술하였다. 각 연도별 고시 영역의 내적 일치도는 Cronbach's alpha로 제시하였다. 본 분석에서 상관관계는 피어슨 상관계수(Pearson correlation coefficient)를 이용하였고, 유의 수준은 $p < 0.05$ 로 설정하였다. 통계 분석은 SPSS version 17.0.1 (SPSS Inc., Chicago, USA)를 사용하였다.

수는 유의하게 증가하였다($p < 0.01$).

2. CPX와 기타 시험 영역 사이의 상관관계

CPX와 전문의 고시의 다른 영역 사이의 상관관계를 분석한 결과(Table 2), 3개 연도에 걸쳐 CPX 시험은 동영상, 슬라이드 및 구술시험으로 구성된 2차 시험의 점수와 유의한 양의 상관관계를 보인다는 사실이 확인되었다. 2007년도와 2009년도에는 1차 시험과도 유의한 상관관계를 보였다. CPX 시험 첫 해인 2007년도의 경우, CPX는 모의시험과 유의한 양의 상관관계를 보였으나, 2008년도 및 2009년도의 경우 상관관계는 통계적으로 유의하지 않았다.

결과

1. 신경과 전문의 시험의 응시 현황

2007년도부터 3개 년도에 걸쳐 전문의 고시 및 모의시험의 응시 인원과 점수 분포는 Table 1에 제시되어 있다. CPX 시험 첫 해인 2007년도와 비교할 때, 2009년도 고시의 모의시험 점수는 유의하게 감소되어 있으며($p < 0.01$), CPX 시험의 점

Table 2. Correlations between Clinical Performance Examination (CPX) Score and Other Areas of Board Licensure Examination

	2007	2008	2009
First step	0.279*	0.154	0.248*
Second step (excluding CPX score)	0.408**	0.395**	0.339**
Workshop	0.302*	0.092	0.109

p-values were calculated using pearson's correlation coefficient. * $p < 0.05$, ** $p < 0.01$.

Table 3. Correlations between Clinical Performance Examination (CPX) Items Scores and Other Areas of Board Licensure Examination

Year	Items	Item score	First step	Second step	Workshop
2007	History taking	3.36±0.73	0.302**	0.369**	0.184
	Physical examination	3.23±0.82	0.152	0.288*	0.276*
	Explanation	3.51±0.64	0.256*	0.310**	0.236*
	Summarization	3.17±0.81	0.257*	0.345**	0.314**
	Oral presentation	3.24±0.99	0.225	0.382**	0.210
2008	History taking	9.56±2.40	0.015	0.193	0.121
	Physical examination	7.75±2.45	0.202	0.444**	0.094
	Patient education	9.78±2.07	0.196	0.335**	0.158
	Patient-doctor relationship	9.64±2.01	0.046	0.164	0.068
	Oral presentation	8.24±2.79	0.180	0.435**	-0.031
2009	History taking	3.88±0.72	0.224*	0.183	0.216
	Physical examination	3.55±0.68	0.254*	0.313**	0.216
	Patient education	4.02±0.68	0.198	0.172	0.283*
	Oral presentation	3.74±0.77	0.291**	0.322**	0.172

Item scores from year 2007 and 2009 were based on 5-point scale, while item scores from year 2008 were on 15-point scale. Item score presented as mean±standard errors. Scores of second step excluding CPX score.

* $p < 0.05$, ** $p < 0.01$.

CPX의 각 평가 항목에 따라, 전문의 고시의 기타 평가 영역과 갖는 상관관계는 서로 상이하였다(Table 3). 2007년도 고시의 경우, 1차, CPX를 제외한 2차 및 CPX 시험과 공히 통계적으로 유의한 양의 상관관계를 보이는 세부 항목은 병력 청취, 설명 및 요약이었다. 2008년도 고시의 경우, 각 세부항목은 1차 시험 점수와 유의한 상관관계를 보이지 않았다. 그러나 신체 진찰, 환자 교육 및 구술 항목은 CPX를 제외한 2차 및 CPX 시험과 유의한 양의 상관관계를 보였다. 2009년도 고시의 경우 병력 청취, 신체 진찰 및 구술 항목이 세 가지 고시 세부 영역과 유의한 양의 상관관계를 보이고 있었다. 삼 개년도의 경험을 종합할 때, 주로 병력 청취, 신체 진찰 등 환자로부터 필요한 정보를 획득하는 과정을 평가하는 항목이 전문의 고시의 다른 평가 영역과 유의한 상관관계를 보이고 있었다. 그러나 환자 교육 및 의사 환자 관계 등 전문적인 지식 이외에 환자와의 역동적인 상호 관계를 형성하는 과정을 평가하는 항목은 전문의 고시의 다른 평가 영역과 뚜렷한 상관관계를 보이지 않음을 알 수 있었다.

연도별 각 고시 영역의 내적 일치도는 2007년 0.87, 2008년 0.89, 2009년 0.81로 측정되었다.

3. 표준화 환자의 평가

2008년도 고시에서, 수험생에 대한 SP의 전반적 평가는 1차($r=0.145$), CPX를 제외한 2차($r=0.082$) 및 모의시험($r=0.182$)과 유의한 상관관계를 보이지 않았으며, CPX 시험 점수와 유의한 상관관계를 나타내었다($r=0.385$, $p<0.01$). SP의 평가는 CPX의 모든 세부 항목과 유의한 양의 상관관계를 보이고 있었다. 세부 항목 중 의사 환자 관계($r=0.444$, $p<0.01$), 병력 청취($r=0.369$, $p<0.01$) 및 환자 교육($r=0.313$, $p<0.01$)과 강한 상관관계를 형성하였다. 그러나 구술($r=0.296$, $p=0.01$) 및 진찰($r=0.278$, $p=0.02$) 항목은 비교적 약한 상관관계를 보이고 있었다.

4. CPX 시험에 대한 수험생 및 채점위원의 반응

CPX 시험에 대한 수험생 및 채점위원의 평가는 대체로 긍정적이었다. 수험생들 중 CPX가 필요하다고 생각하는 응답자는 38.4%이었고, 이는 CPX가 필요하지 않다고 생각하는 응답자 13.7%에 비해 월등히 높은 응답이었다(보통 47.9%).

또한 CPX의 시험 시간과 시행 방법의 적절성을 묻는 질문에 대하여, 38.4%의 수험생이 만족한다는 반응을 보였고, 이는 미흡하다는 응답자 4.1%에 비하여 높은 수준이었다(보통 57.5%). 24명의 채점위원 중 22명(91.6%)의 채점위원이 CPX 시험이 필요하다고 응답하였으며, 필요하지 않다고 응답한 위원은 0명이었다. CPX 시험에 대한 긍정적인 평가는 2008년 전문의 고시의 채점위원도 동일하였다. 총 19명의 채점위원이 질문에 응답하였으며, 이 중 94.7%가 필요성에 공감하였으며 78.9%가 시험의 변별력을 긍정적으로 평가하였다. 시험 시행에 관한 정성적인 지적으로는, 수험생의 전반적인 진단 능력 및 문제 해결 능력을 평가하기 위하여, 현재 10분으로 설정되어 있는 구술 평가 시간을 연장하고 다양한 질문 항목을 개발할 필요가 있다는 의견이 많았다.

고찰

SP를 이용한 CPX 시험은 이미 미국 대다수의 의과대학에서 학생 교육 및 평가에 널리 사용되고 있다. 국내에서도 1994년 OSCE의 형식으로 서울대학교 의과대학에서 처음 도입한 이래, 2008년까지 전국적으로 4개의 의과대학 컨소시엄 체제로 표준화 환자 활용 진료시험이 운영되고 있다[5]. CPX는 실제로 임상에서 접할 수 있는 환경 속에서, 수험생이 의사로서 환자와 역동적인 관계를 형성하면서 환자의 감정적인 부분을 헤아리고 나아가 환자를 교육하는 능력을 평가할 수 있다는 장점을 갖고 있다. 특히 전문의 시험에 도입된 CPX의 경우, 신경과 전문의로서 보다 심화된 지식과 전문가적 식견을 가지고 이를 환자에게 적용하고 또 환자에게서 신뢰를 얻을 수 있는 수험생의 능력을 평가하는 데 필수적인 방법이라고 할 수 있다.

CPX는 많은 비용과 많은 준비 시간을 필요로 하며, 익숙하지 않은 시험 및 평가 방식에 대한 수험생과 채점위원의 거부감을 해소하는 데 상당한 노력이 요구된다. 대한신경과학회 및 고시위원회에서도 이와 같은 점을 인식하고, 매년 모의시험을 준비하여 수험생 및 채점위원이 CPX에 익숙해질 수 있도록 배려하였다. 그 결과, CPX 모의시험 성적은 2007년에 비하여 2009년도에서 다소 저하되었지만, 본 시험의 경우에는 오히려 시간이 경과하면서 CPX 시험 점수가 더 향상되는

양상이 관찰되었다. 이를 통하여 대한신경과학회 전문의 고시에서 CPX 시험의 운영이 점차 안정되어 가고 있음을 짐작할 수 있다. 또한 누적된 CPX 시험 시행을 통하여, 임상적인 실기 능력의 중요성이 다시 한 번 강조되면서 전공의 수련 및 교육 과정에 임상적 능력에 대한 고려가 이전보다 증가되었다고 생각할 수 있다. 하지만 매년 동일한 시험 문제가 출제된 것이 아니었기 때문에, 이를 결과 해석에 고려해야 할 것이다. 지필 고사인 1차 시험 성적과 CPX 성적의 상관성은, 연도별로 약간 차이는 있지만 서로 상관성이 없거나 피어슨 상관계수가 0.25, 혹은 0.28로서 약한 양의 상관관계를 보이며 이는 6~8%의 설명력을 의미한다. 따라서 1차 시험과 CPX는 서로 다른 영역을 평가하고 있다고 해석할 수 있다. 2차 시험의 다른 평가인 동영상 및 슬라이드시험, 일반 구술시험 성적의 총점은 CPX 점수와 0.34~0.41의 피어슨 상관계수를 보여주었으며 이는 11~17%의 설명력을 의미한다. 따라서 2차 시험의 다른 영역 시험과 CPX는 중등도의 상관성을 보이지만 역시 슬라이드/동영상시험, 구술시험과는 서로 다른 영역을 평가하고 있다고 해석할 수 있다. CPX 본 시험과 모의시험 성적 간에는 2007년도에는 서로 약한 상관성을 보이다가 2008년도 이후로는 통계적으로 유의한 상관성이 없었다. 이는 몇 가지로 해석이 가능한데 수험생들이 모의시험에서 보여준 수행능력이 본 시험에서는 시험의 준비라는 과정을 거치면서 변했다는 것을 시사하며 이는 모의시험에는 같은 시험 중례가 3년 동안 반복적으로 출제되었다는 점에서 수험생이 미리 준비할 수 있어 정확한 수험생에 대한 평가를 반영하지 못했을 가능성이 있다. 또 다른 해석으로는 이처럼 모의시험 CPX 성적과 본 시험 CPX 성적이 서로 상관성이 없는 이유가 10월 이후 수험생의 수련과정에서의 수행능력 향상을 반영하고 있다고도 해석할 수 있지만 시기적으로 이미 실제적인 수련은 끝난 시점이므로 시험 준비 효과가 더 크게 작용한 것으로 해석해야 할 것이다. 모의 CPX의 평가 목표는 원래 수험생과 채점위원으로 하여금 시험 운영에 대하여 친숙해지도록 하는 데 있으므로 이러한 소기의 목적은 달성한 것으로 판단된다.

Table 3에서 CPX 영역별 점수 중 특히 신체진찰 점수가 상대적으로 낮다. 이는 신경과 전문의로서 요구되는 신체진찰수기 교육이 수련과정에서 상대적으로 불충분하거나 채점위원간의 표준진찰수기에 대한 이견으로 점수를 낮게 받았다고 해

석할 수도 있다. CPX 세부 영역 점수 중 병력 청취, 환자 교육 관련 점수는 1차 시험 성적과 상관성이 없거나 혹은 약한 상관성만을 보여주고 있다. 반면 신체 진찰영역 CPX 점수는 1차 시험성적과 2007~2008년도에는 상관성이 없었고 2009년도에는 약한 상관성을 보였다. 이는 신경과 전문의 고시 CPX에서 병력 정보수집 기능은 사전 지식 체계가 약간 도움이 되었고 신체진찰 능력은 사전지식체계가 별 도움이 되지 않았다고 해석할 수 있다. CPX 세부영역과 CPX를 제외한 다른 2차 시험 분야와의 상관성을 평가하였을 때 신체진찰과 CPX에 포함된 구술 영역의 점수가 중등도의 상관성을 보였다. 이는 2차 시험이 CPX가 아니더라도 필기가 아닌 실기 시험이고 통합, 분석, 비교, 판단의 기능을 신경과 전문의로서의 기준을 적용하여 평가하였으므로 신체진찰이나 구술 평가의 영역이 2차 시험의 다른 영역과도 비교적 높은 상관성을 보였다고 해석할 수 있다. 환자 교육 및 의사-환자관계 세부 영역은 1차 시험, 2차 시험(CPX 제외)성적과 약한 상관관계를 보이거나 유의한 상관성이 없었던 점은 몇 가지 해석이 가능할 것이다, 하나는 실제로 지식과 수기 능력이 태도와 면담 기술과 직접적인 연관성이 없었다고 해석할 수 있고, 다른 하나는 이들 영역이 수험생이 환자와 역동적인 상호 관계를 형성하며 환자에게 적절히 개입하는 과정을 평가하는 항목으로써 다른 시험 방법으로 평가할 수 없는 영역을 CPX를 통하여 평가하였다고 해석할 수 있을 것이다, CPX가 의학 지식을 넘어서 한 사람의 의사로서 환자를 전인적으로 돌보고 환자의 감정을 고려하며 환자에게 개입하는 과정을 평가하는 도구라면 이를 신경과 전문의 시험에 도입한 것이 타당하다는 것을 보여주는 결과라고 하겠다.

CPX 시험 점수가 전문의 고시의 다른 영역인 1차 및 2차 시험의 점수와 갖는 상관성이, 경험이 축적되면서 점차 증가하고 있다. 물론 CPX의 시행은 기존의 지식 및 결과 위주의 평가에서, 수행 및 과정 위주의 평가로 이행하는 데 방점을 찍고 있다. 그러나 환자에 대한 적절한 접근은 충분한 지식 및 이를 적용할 수 있는 능력에서 비롯된다는 점을 고려할 때, 궁극적으로 CPX에서 획득한 점수는 전체적인 시험 점수와 어느 정도 비례할 것으로 생각된다. 또한 CPX 시험의 시행 및 평가 방법에 대한 수험생의 만족도나 적정성의 평가는 13.7%와 4.1%만이 만족하지 않거나 적절하지 않다고 대답하여 만족도나 적정성은 높다고 판단된다. 하지만 중도의 입장에 있

는 수험생들이 상대적으로 많다는 점을 의미 있게 받아 들여야 한다. 또한 90% 이상의 채점위원들이 CPX 시험에 대하여 그 필요성을 인정하였다는 점은 주목할 만하다. 이는 지필 위주의 수동적인 기존 시험 방식에 비하여 실제 임상에서의 진단 및 환자-의사 관계 형성 능력을 평가하는, 새로운 흐름이 요청되었음을 시사한다고 하겠다. 그리고 CPX 시험은 그러한 변화의 흐름에 부합하는 방식임을 반증한다고 할 수 있다. 다만 여러 채점위원들이 다면적이고 심층적인 수험생 평가가 필요함을 지적하고 있으며, 이는 향후 신경과 고시위원회에서 해결해야 할 숙제라고 하겠다. 또한 채점위원들이 신경과 전문의 및 수련 지도 교수 일반의 여론을 반영하지 못할 가능성을 고려할 필요가 있다.

지난 3년간의 신경과 CPX 시행 경험을 살펴볼 때, 다음과 같은 지점에 관심을 가져야 한다. 첫째, CPX는 1차 필기시험과는 약한 상관성을 보이고 CPX를 제외한 다른 영역의 2차 시험 성적과는 중등도의 상관성을 보이는 바 SP를 활용한 진료 시험은 2차 시험 중 독립적인 한 평가 영역으로 인정받아야 한다. 둘째, 신경과 전문의 고시에서 2008년도를 제외하고는 환자-의사관계의 평가를 적극적으로 활용하지 않았는데 이는 의사소통능력 보다는 오히려 전문가적인 정보수집과 통합 및 판단 능력을 평가하고자 하는 목표가 더 중요하다고 판단하였을 수 있다. 하지만 향후 CPX 문항의 확대를 전제로 하여 환자-의사관계를 포함한 면담술 및 의사소통능력도 실기시험에서 적극적으로 평가되어야 할 것이다. CPX의 세부항목 가운데 문진, 병력 청취 및 진찰 등 진단에 필요한 정보를 환자로부터 획득하는 과정을 평가하는 항목이, 주로 지식 등을 평가하는 전문의 고시의 다른 평가 영역과 높은 상관관계를 보이고 있었다. 이에 비하여 환자 교육 및 의사-환자 관계 등 수험생이 신경과 전문의로서 환자와 역동적인 상호 관계를 형성하며 환자에게 적절히 개입하는 과정을 평가하는 항목은 전문의 고시의 다른 평가 영역과 비교적 낮은 상관관계를 보이고 있었다. 이는 CPX가 의학 지식을 넘어서 수험생이 한 사람의 의사로서 환자를 전인적으로 돌보고 환자의 감정을 고려하며 환자에게 개입하는 과정을 평가하는 측면에서도 타당성을 확보하고 있음을 보여주는 결과라고 하겠다. 셋째, CPX의 시행에 있어 준비 시간 및 비용의 부담 이외에, 비정형의 새로운 시험 방식과 과정 및 평가 방식으로 인하여 CPX가 갖는 타당성에 대한

2/3 정도의 수험생이 갖는 의문 또한 CPX를 준비하는 주체가 극복하여야 하는 문제일 것이다. 하지만 신경과 고시에서 시행된 CPX는 모의시험을 포함한 충분한 사전 준비와 노력으로 이러한 어려움을 순조롭게 극복해나가고 있으며, CPX가 신경과 고시의 중요한 평가 항목으로 자리매김하게 되었다고 할 수 있겠다. 넷째, 수험생 평가의 측면에서 CPX의 변별력은 채점위원 사이에서 공감대가 형성되고 있으나, 현재의 CPX 시행 방식으로는 다소 단면적이고 획일적인 평가에 그칠 우려가 있다는 지적이 제기되었다. 이에 따라 추가적인 시험 사례 개발, 시험 시간 연장 등 향후 다면적이고 심층적인 수험생 평가가 가능한 CPX 개발이 요청된다고 하겠다. 다섯째, 신경과 고시에서의 CPX 시행 경험에서 수험생에 대한 SP의 평가가 기존의 평가 방식과는 사뭇 다른 결과행태를 보였다는 점에 주목할 필요가 있다. 평가자로서 SP가 갖는 의의, 타당도 및 신뢰도에 대해서 국내외에서 많은 연구가 이루어진 바 있다 [6,7,8,9,10,11]. 일반적으로 훈련을 잘 받은 SP의 경우 타당성 있는 채점을 할 수 있다는 결과도 발표된 바 있다. 그러나 신경과 고시에서는 몇 가지 현실적인 한계를 고려하여 SP를 평가자로 도입하지는 않고, 향후의 연구를 위한 기초 자료로 환자 만족도에 한하여 SP의 평가 자료를 수집하였다. 그런데 SP가 평가하는 수험생의 점수는, 지식수준 및 지식의 적용을 평가하는 1차 및 2차 시험과 상당히 낮은 상관성을 보이고 있었다. 동시에 모의환자의 평가는 CPX의 세부 항목 중 의사-환자 관계 및 병력 청취 등 직접적으로 의사가 환자와 관계를 형성하는 항목과 높은 상관성을 보였다. 이러한 표준화 모의환자의 평가 결과는 앞서의 다른 논문에서는 SP가 수험생과의 상호 작용에서 특정한 행동을 하였거나 느낌을 주었는지를 체크하도록 하는 채점에 비해 수험생에 대한 전반적 평가를 요구하였다는 점에서 주로 차이점이 발생하였다고 판단되나 앞으로 표준화 환자를 평가자로 활용할 때 어떤 방식으로 또는 어떤 영역의 평가를 맡길지, 어느 정도 훈련을 시켜야 평가의 일관성을 유지시킬 수 있는지에 대한 좀 더 신중한 고려가 있어야 할 것이다. 또한 다양한 표준화 환자 사례를 개발하고 심층적인 질문 문항을 마련하여, 수험생의 진단 및 문제 해결 능력을 자세히 평가할 수 있도록 개량할 필요가 있음이 지적되었다.

본 연구는 첫째, 대조군이 없고, 둘째, 응시생의 숫자가 적어 다변수 분석이 불가능하며, 셋째, 연도별 응시생의 수험 능

력 및 출제 문제의 변이 여부를 반영하지 못했다는 한계를 갖고 있다고 하겠다.

신경과 전문의 고시에서 3년 동안 시행한 CPX를 검토해 본 결과는 기존의 필기 및 슬라이드/동영상시험과 같이 자료의 해석과 신경과적 전문 지식을 평가하는 영역항목을 보완하면서, 보다 전인적인 능력을 함양하는 신경과 전문의를 배출하는 도구로 활용할 수 있을 것이다. 또한 표준화 환자 시험은 수험생 및 평가자 모두에게 긍정적인 평가를 받는 평가 도구로 자리잡고 있음을 확인할 수 있었다. 특히 향후 CPX 문항이 확대되어 2차 실기 시험에서 비중이 높은 평가 도구로 자리 잡으면, 신경과 전공의 교육 및 수련 과정 역시 전문의로서 환자와 역동적인 관계를 형성하는 능력을 배양하는 과정을 증가시키는 방향으로 변화할 것으로 기대된다. 그러한 의미에서 기존의 신경과 전공의 교육 및 전문의 고시에서의 평가가 다분히 전문적인 지식을 함양하고 지식의 축적 정도를 가려내는 쪽으로만 기능하였다는 반성의 한 대안이 전문의 고시에서의 CPX의 도입이라 할 수 있겠다. 하지만 3년간의 고시 시행 경험만으로 CPX가 신경과 전공의 교육 및 전문의 배출 제도에 현격한 변화를 가져왔다고 보기는 어려울 것이다. 향후 보다 장기간의 경험이 누적되고 신경과 이외의 전문 과목에서 시행된 CPX 경험이 공유된다면, CPX의 실질적인 의의가 더욱 명확히 밝혀지리라 기대한다.

Acknowledgements: None.

Funding: None.

Conflicts of interest: None.

REFERENCES

1. Lee YM, Hwang K, Baik SH, Hong MH, Choi YS, Kim SH. An experience of OSCE for introducing clinical performance assessment to Korean Medical Licensure Examination. *Korean J Med Educ* 2001; 13: 47-57.
2. Board Examination Committee of Korean Neurological Association. Manual of work: Board Examination Committee of Korean Neurological Association. Seoul, Korea: Korean Neurological Association; 2007.
3. Kim SW, Lim SY, Mun GH, Shin JS. Introduction of OSCE (objective structured clinical examination) to board licensure examination of plastic surgeons. *Korean J Med Educ* 2006; 18: 153-160.
4. Solomon DJ, Ferenchick G. Sources of measurement error in an ECG examination: implications for performance-based assessments. *Adv Health Sci Educ Theory Pract* 2004; 9: 283-290.
5. Han JJ, Park H, Kwon I, Ryu KH, Eo E, Kim N, Jung J, Kim KH, Lee SN. The comparison of clinical performance examination scores according to the different testing time: six medical schools in Seoul. *Gyeonggi CPX Consortium* 2005. *Korean J Med Educ* 2007; 19: 31-38.
6. Norman GR, Neufeld VR, Walsh A, Woodward CA, McConvey GA. Measuring physicians' performances by using simulated patients. *J Med Educ* 1985; 60: 925-934.
7. Rethans JJ, van Boven CP. Simulated patients in general practice: a different look at the consultation. *Br Med J (Clin Res Ed)* 1987; 294: 809-812.
8. Kim S, Park S, Hur Y, Lee SJ. The appropriateness of using standardized patients' (SPs) assessment scores in clinical performance examination (CPX). *Korean J Med Educ* 2005; 17: 163-172.
9. Park WB, Lee SA, Kim EA, Kim YS, Kim SW, Shin JS, Lee YS. Correlation of CPX Scores with the scores of the clinical clerkship assessments and written examinations. *Korean J Med Educ* 2005; 17: 297-303.
10. Kwon I, Kim N, Lee SN, Eo E, Park H, Lee DH, Park MH, Oh J, Han JJ, Huh J, Ryu KH. Comparison of the evaluation results of faculty with those of standardized patients in a clinical performance examination experience. *Korean J Med Educ* 2005; 17: 173-184.
11. Shin J, Lee S, Park H. Standardized patients' accuracy in recording checklist items during clinical performance examinations. *Korean J Med Educ* 2005; 17: 197-203.