

Correlation of CPX Scores with the Scores on Written Multiple-Choice Examinations on the Certifying Examination for Family Medicine in 2009 to 2011

Jung Jin Cho¹, Ji Yong Kim², Hoon Ki Park³ and In Hong Hwang⁴

¹Department of Family Medicine, Hallym University Sacred Heart Hospital, College of Medicine, Hallym University, Anyang, ²Kangbuk Samsung Hospital Seocho Medical Clinic, Seoul, ³Department of Family Medicine, Hanyang University Hospital, Seoul, and ⁴Department of Family Medicine, Hallym University Kangdong Sacred Heart Hospital, College of Medicine, Hallym University, Seoul, Korea

3개년도 가정의학과 전문의 자격시험에서 임상진료시험 결과 및 필기시험 성적 간의 상관관계

¹한림대학교성심병원 가정의학과, ²강북삼성병원 서초클리닉, ³한양대학교병원 가정의학과, ⁴한림대학교강동성심병원 가정의학과

조정진¹, 김지용², 박훈기³, 황인홍⁴

Purpose: Recently, the clinical practice examination (CPX) using standardized patients has been introduced into several specialty certifying examinations in Korea. The purpose of this paper was to determine the correlation of a resident's performance on the CPX with the comprehensive written multiple-choice question (MCQ) examination on the certifying examination for family medicine.

Methods: The subjects of this study were 1,023 residents who completed the 1st and 2nd certifying examination for family medicine between 2009 and 2011. We determined the correlation between the total scores and 4 domain scores (history taking, physical examination, patient education, and patient-physician interaction) on the CPX with the MCQ scores of the 1st written test and 2nd slide examination and the correlation between the total CPX score and scores on the CPX domains.

Results: The correlation between CPX score with each MCQ examination (0.21~0.45 with 1st written MCQ, 0.15~0.33 with 2nd slide MCQ) was lower than that between each MCQ examination (0.46~0.59). The CPX score on patient education did not correlate with the 1st written and 2nd MCQ scores. The CPX scores on history taking and physical examination correlated slightly with the 1st written MCQ scores. The global ratings of preceptor examiners had the highest correlation ($r=0.68\sim0.82$) with the total CPX scores.

Conclusion: Considering the mild correlation of CPX scores with each MCQ examination, the CPX is more likely to measure other qualities, such as critical thinking and communication skills.

Key Words: Certification, Clinical competence, Educational measurement, Family medicine

Received: July 27, 2011 • Revised: September 28, 2011 • Accepted: October 4, 2011

Corresponding Author: Jung Jin Cho

Department of Family Medicine, Hallym University Sacred Heart Hospital, College of Medicine, Hallym University, 896 Pyeongchon-dong, Dongan-gu, Anyang 431-070, Korea
Tel: +82.31.380.1783 Fax: +82.31.380.1782 email: threej@hallym.or.kr

A part of this paper was presented at the 47th Annual Spring Congress of the Korean Academy of Family Medicine in March 2009.

Korean J Med Educ 2011 Dec; 23(4): 315-322.

doi: 10.3946/kjme.2011.23.4.315.

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

전문의 자격시험은 전문의로서의 진료능력이 있는가 총괄적으로 평가하는 시험이며 타당도와 신뢰도가 높아야 한다. 진료능력은 다양한 정의가 있지만 미국의 전공의 수련교육인증위원회(Accreditation Council for Graduate Medical Education, ACGME)는 환자 진료, 의학적 지식, 진료에 기초한 학습과 향상, 대인 및 의사소통기술, 계통적 진료 등 5가지 영역을 전문의가 되기 위한 진료능력으로 규정하고 있다 [1]. 이러한 진료능력 평가는 의학교육의 주요 영역이기도 하다. Crossley et al. [2]은 교육 목적에 따라 인지영역(cognition), 기술(Skill)과 태도(attitude)로 분류하였으며 의학교육평가도 마땅히 이 3가지 영역을 포괄하여야 한다. Miller [3]는 인지영역과 행동영역(behaviors; performance, action)으로 의학교육 평가단계를 구분하였다. 실제로 전문의 자격시험은 의학교육 평가단계를 반영하여 1차 시험은 필기시험, 2차 시험은 실기시험이라고 법으로 규정되어 있다. 하지만 전문의 2차 시험에 행동영역을 평가하는 실제적인 실기시험이 시행되는 전문학회는 많지 않다. 대부분의 전문학회는 1차에서는 필기시험을 운영하고 2차에서는 슬라이드를 이용한 필기시험이나 구술면접 등을 활용하고 있다. 1차 객관식 필기시험은 인지영역의 평가이며, 2차 시험에서 흔히 사용되는 슬라이드 시험도 그림이나 사진을 이용한 변형된 인지영역 평가일 뿐 인지기능의 상위 개념인 행동영역을 평가하지 못한다는 한계를 가진다.

의학교육평가에서 행동영역을 평가하는 대표적인 평가지표로는 객관구조화진료시험(objective structured clinical examination, OSCE)이나 임상진료시험(clinical practice examination, CPX) 등이 있다. CPX는 수험생이 표준화 환자(standardized patient, SP)를 대상으로 실제 환자처럼 병력을 청취하고, 신체검사를 수행하며, 임상검사를 계획하고, 진단을 내리며 치료 계획을 세우고, 상담을 수행하는 과정을 평가하는 시험이다. OSCE에도 SP를 사용할 수 있지만 OSCE는 특정 술기의 수행능력을 단편적으로 확인하는 반면, CPX는 특정 환자를 대상으로 포괄적인 임상수행능력을 평가하는 평가방법이다. 캐나다에서는 1992년부터 OSCE가 의사

국가시험에 도입이 되었고[4], 미국에서는 1998년도부터 Educational Commission for Foreign Medical Graduates (ECFMG)에 도입되었고, 2005년부터 임상술기에 대한 내용이 United States Medical Licensing Examination (USMLE)에서도 시행되고 있다[5]. 일본에서도 1994년 의과대학 2학년 말에 선택적으로 OSCE를 실시하고 있다[6]. 국내에서도 2010년부터 의사국가시험에 도입되었다.

국내 전문의 자격시험에서도 2차 시험이 고유의 실기시험 기능을 하기 위해 행동영역을 평가하려는 변화가 나타나고 있다. 2006년 대한성형외과학회 전문의 자격시험에 OSCE가 도입되었으며[7], 대한신경과학회 전문의 고시에서는 2007년 신경학적 진찰을 중심 내용으로 하는 CPX 한 사례를 2차 실기시험에서 운영하고 있다[8]. 대한가정의학회에서는 2009년부터 가정의학과 전문의로서 최소한 갖춰야 할 진료능력과 태도를 보유하고 있는지 평가하기 위해 전문의 고시에 SP를 이용한 CPX를 도입하였다.

최선답형(multiple choice questions, MCQ) 필기시험 성적과 실기시험인 CPX 성적 간의 상관성 분석은 CPX의 도입이 의학교육평가 이론을 얼마나 제대로 반영하고 있는지 살펴볼 수 있으며 아울러 전문의 고시에서 CPX 도입 타당성을 분석하는 의미를 가질 수 있다. 기존 연구를 살펴보면 캐나다 치과대학생을 대상으로 한 결과[9], 국내에서 의과대학생에서 필기시험과 실기시험 성적 간의 상관성 분석을 한 연구결과[10,11] 등이 있다. 전문의 자격시험 결과를 활용한 연구는 캐나다에서 치과의사 국가시험[12] 결과를 분석한 연구와 국내 신경과 전문의 자격시험 결과를 분석한 연구[8]로 매우 적은 편이다.

이에 국내 대한가정의학회 전문의 자격시험에서 실시된 MCQ 필기시험 성적과 CPX 성적 간의 상관관계를 살펴보고, CPX 세부 평가항목별로 MCQ 필기시험 성적과의 연관성을 알아보고자 한다. 이러한 분석을 통하여 전문의 자격시험에서 CPX의 타당성을 확인하여 향후 타 전문학회 전문의 자격시험에서 활용하는 데 도움이 되고자 한다.

대상 및 방법

1. 연구 대상

2009년부터 2011년까지 가정의학과 전문의 자격시험 1차 필기시험, 2차 슬라이드 시험 및 2차 CPX 시험 모두 응시한 1,013명의 성적을 대상으로 하였다. 1차 시험에 불합격하여 2차 시험에 응시하지 못하였거나, 1차 시험 면제, 2차 시험 중 일부 불응시한 경우는 제외하고 분석하였다.

2. MCQ 필기시험의 진행

2009년 1월, 2010년 1월, 2011년 1월에 MCQ 250문항으로 1차 필기시험이 실시되었다. 시험 시간은 4시간이 배정되었다. 시험 성적은 각 문항당 0.4점으로 총점을 구하였고 100점 만점으로 환산하여 산출하였으며 이하 '필기MCQ'로 표기하였다.

각 연도별 1차 필기시험 합격자를 상대로 2009년 1월, 2010년 1월, 2011년 1월 슬라이드 그림이나 사진을 이용한 MCQ 50문항으로 2차 시험이 실시되었다. 문항에 필요한 그림이나 사진이 포함된 슬라이드를 조명을 끈 상태에서 30초간 비추다가 조명을 밝힌 상태에서 30초간 슬라이드를 보여 주어 각 문항당 1분씩 배정하여 시험 시간은 총 50분이 배정되었다. 시험 성적은 각 문항당 2점으로 총점을 구하였으며 100점 만점으로 환산하여 산출하였으며 '슬라이드MCQ'로 표기하였다.

3. CPX의 진행

1차 필기시험 합격자를 대상으로 2009년 1월, 2010년 1월, 2011년 1월 SP를 활용한 2사례로 구성된 2차 CPX가 실시되었다. 수험생을 오전/오후 조로 배분하고, 총 4개의 사례를 개발하여 오전/오후 문제를 다르게 운영하였으며, 각 수험생이 두 시험 방을 연속적으로 순환하며 시험을 치렀다. SP 대면 시간은 한 환자당 10분씩 총 20분이 배정되었으며, 상황 지침을 읽는 시간은 1분을 배정하였다.

각 케이스당 SP는 2009년, 2010년에는 10명씩을 복제하였고, 2011년에는 12명씩을 복제하였다. 한 명의 SP는 16명의

Table 1. Distribution of Domains in the Clinical Practice Examination on the Certifying Examination for Family Medicine in 2009 to 2011

	HX	PE	IS	PPI	GR of SP	GR of Dr
2009	39.5	17.3	13.4	22.3	3.7	3.7
2010	37.3	6.3	24.6	23.9	4.0	4.0
2011	42.5	20.8	10.0	20.0	3.3	3.3

Values are presented as percent.

HX: History taking, PE: Physical examination, IS: Information sharing, PPI: Patient and physician interaction, GR: Global rating, SP: Standardized patient, Dr: Preceptor doctor.

수험생의 진료를 받았다. 수험생은 오전 오후에 면허번호가 비슷하게 배정하고, 채점위원은 각 시험 방마다 수련병원 지도 전문의 1인이 담당하였으며, 수험생의 출신학교, 수련병원의 이해관계가 채점에 영향을 주지 않도록 배정하였다.

위촉된 채점위원을 대상으로 환자-의사관계의 채점 일치도 향상을 위해 모의 사례에 대한 SP와 훈련교수 간의 시험상황을 연출한 동영상을 활용하여 모의 채점을 통한 일치도 향상 훈련을 실시하였다. 채점표 한 항목당 1점으로 환산하여 총점을 구하였다. 수험생의 성적은 100점 만점으로 환산하여 산출하였으며 이하 'CPX총점'이라 표기하였다. 세부 평가영역은 병력 청취, 신체 진찰, 환자 교육(정보 공유), 환자-의사관계, 지도 전문의의 전반적 만족도, 환자 만족도로 구성되었다. 연도별 진료수행시험의 하위 평가영역별 구성 분포는 Table 1과 같다.

병력 청취와 환자 교육의 평가영역은 개발된 이분법의 체크리스트에 따라 '했다=1점', '안했다=0점'으로 채점하였다. 신체검사항목은 '제대로 했다=1점', '제대로 못했다=0.5점', '하지 않았다=0점'으로 채점하였다. 수험생 의사의 진료에 대한 만족도는 지도 전문의와 SP가 4점 리커트(Likert) 척도에 따라 채점하였다. CPX 하위 평가영역당 점수는 100점 만점으로 환산하여 산출하여 분석에 활용하였다.

4. 통계분석

시험연도별 성별 분포는 카이제곱검정, 연령 분포는 ANOVA 검정을 적용하였으며, 성적 간의 관련성은 피어슨 상관계수(r)를 이용하여 각 변수 간의 연관성을 측정하였다.

상관계수가 0~0.5 이하이면 약한 양적 선형관계, 0.5 이상이면 강한 양적 선형관계로 간주하였다[13]. $p < 0.05$ 를 의미 있는 수준으로 설정하였다.

결과

1. 연구 대상의 기본 특성

연구 대상은 2009년 308명, 2010년 338명, 2011년 367명으로 총 1,013명이었다. 남자는 631명으로 62.3%, 여자는 382명으로 37.7%였다. 성별 분포는 시험 연도별로 차이가 있었다($p < 0.05$). 연령은 평균 34.8 ± 3.3 세였으며 시험 연도별로 차이가 있었다($p < 0.01$) (Table 2).

2. CPX 총점과 1차 필기 성적 및 2차 슬라이드 성적과의 상관관계

1차 필기MCQ 총점과 2차 슬라이드MCQ 총점 간의 상관계수는 0.46~0.59로 대체적으로 강한 양적 상관관계를 보였다. 필기MCQ와 CPX 총점과의 상관관계는 0.21~0.45로 약한 상관관계를 보였다. CPX 총점과 슬라이드MCQ 간의 상관관계는 0.15~0.33으로 전반적으로 약한 상관관계를 나타냈다 (Tables 3, 4, 5).

3. CPX 세부 평가영역별 성적과 1차 필기 성적 및 2차 슬라이드 성적과의 상관관계

1차 필기MCQ 총점과 CPX 세부 평가영역 간의 상관계수를 살펴보면 병력 청취가 0.07~0.26, 신체 진찰 0.06~0.24, 환자 교육 0.04~0.22, 환자-의사관계 0.16~0.36로 약한 상관

Table 2. Baseline Characteristics of Attendees of the Certifying Examination for Family Medicine in 2009 to 2011

Characteristics		2009 (n=308)	2010 (n=338)	2011 (n=367)	Total (n=1,013)
Sex	Male	211 (68.5)	197 (58.3)	223 (60.8)	631 (62.3)
	Female	97 (31.5)	14 (41.7)	144 (39.2)	382 (37.7)
Age		34.0 ± 3.2	33.9 ± 2.9	36.3 ± 3.3	34.8 ± 3.3

Values are presented as number (%) or mean \pm standard deviation (SD).

Table 3. Correlation between Written MCQ Scores, Slide MCQ Scores, Total CPX Scores, and the Categorical Scores of the CPX on the Certifying Examination for Family Medicine in 2009

	W-MCQ	S-MCQ	T-CPX	Hx	PE	IS	PPI	GR of DR	GR of SP
W-MCQ	1								
S-MCQ	0.57**	1							
T-CPX	0.32**	0.16**	1						
HX	0.17**	0.07	0.73**	1					
PE	0.26**	0.13*	0.70**	0.34**	1				
IS	0.11	0.06	0.43**	-0.00	0.28**	1			
PPI	0.24**	0.11	0.76**	0.39**	0.96**	0.31**	1		
GR of DR	0.19**	0.08	0.68**	0.48**	0.62**	0.20**	0.71**	1	
GR of SP	0.15**	0.06	0.52**	0.21**	0.50**	0.16**	0.54**	0.45**	1

W-MCQ: Score of 1st written multiple choice question examination, S-MCQ: Score of 2nd slide multiple choice question examination, CPX: Clinical practice examination, T-CPX: Total score of CPX, HX: History taking, PE: Physical examination IS: Information sharing, PPI: Patient and physician interaction, GR: Global rating, SP: Standardized patient, Dr: Preceptor doctor.

* $p < 0.5$ by Pearson's correlation, ** $p < 0.01$ by Pearson's correlation.

Table 4. Correlation between Written MCQ Scores, Slide MCQ Scores, Total CPX Scores, and the Categorical Scores of the CPX on the Certifying Examination for Family Medicine in 2010

	W-MCQ	S-MCQ	T-CPX	Hx	PE	IS	PPI	GR of DR	GR of SP
W-MCQ	1								
S-MCQ	0.46**	1							
T-CPX	0.21**	0.15**	1						
HX	0.07	0.07	0.66**	1					
PE	0.18**	0.06	0.31**	0.15**	1				
IS	0.08	0.04	0.52**	0.13*	-0.02	1			
PPI	0.16**	0.15**	0.71**	0.42**	0.19**	0.27**	1		
GR of DR	0.13*	0.18**	0.77**	0.42**	0.23**	0.38**	0.79**	1	
GR of SP	0.16**	0.20**	0.47**	0.30**	0.19**	0.14*	0.48**	0.47**	1

W-MCQ: Score of 1st written multiple choice question examination, S-MCQ: Score of 2nd slide multiple choice question examination, CPX: Clinical practice examination, T-CPX: Total score of CPX, HX: History taking, PE: Physical examination, IS: Information sharing, PPI: Patient and physician interaction, GR: Global rating, SP: Standardized patient, Dr: Preceptor doctor.

*p<0.5 by Pearson's correlation, **p<0.01 by Pearson's correlation.

Table 5. Correlation between Written MCQ Scores, Slide MCQ Scores, Total CPX Scores, and Categorical Scores of the CPX on the Certifying Examination for Family Medicine in 2011

	W-MCQ	S-MCQ	T-CPX	Hx	PE	IS	PPI	GR of DR	GR of SP
W-MCQ	1								
S-MCQ	0.59**	1							
T-CPX	0.45**	0.33**	1						
HX	0.26*	0.26**	0.77**	1					
PE	0.24*	0.21*	0.62*	0.44**	1				
IS	0.22*	0.08	0.39**	0.21**	0.26**	1			
PPI	0.36**	0.29*	0.79**	0.50**	0.81**	0.34**	1		
GR of DR	0.40*	0.29**	0.82**	0.55**	0.68**	0.39**	0.88**	1	
GR of SP	0.30**	0.19**	0.51*	0.29**	0.49**	0.26**	0.56**	0.57**	1

W-MCQ: Score of 1st written multiple choice question examination, S-MCQ: Score of 2nd slide multiple choice question examination, CPX: Clinical practice examination, T-CPX: Total score of CPX, HX: History taking, PE: Physical examination, IS: Information sharing, PPI: Patient and physician interaction, GR: Global rating, SP: Standardized patient, Dr: Preceptor doctor.

*p<0.5 by Pearson's correlation, **p<0.01 by Pearson's correlation.

계수를 보였으며 그 중에서도 환자 교육과 병력 청취의 상관관계가 가장 낮았다.

2차 슬라이드MCQ와 CPX 세부 평가영역 간의 상관계수는 병력 청취가 0.07~0.45, 신체 진찰 0.18~0.40, 환자 교육 0.08~0.22, 환자-의사관계 0.11~0.29로 약한 상관계수를 보였으며 그 중에서도 환자 교육과 환자-의사관계의 상관관계가 낮았다(Tables 3, 4, 5).

4. CPX 총점과 세부 평가영역별 성적 간의 상관관계

CPX총점과 지도 전문의의 총괄평가 점수 간의 상관계수는 0.68~0.82로 강한 양적 상관관계를 SP의 만족도 간의 상관계수는 0.47~0.52로 양적 상관관계를 보였다. CPX총점과 병력 청취는 0.66~0.77, 환자-의사관계는 0.71~0.79로 강한 상관관계를 보였으며, 신체 진찰은 0.31~0.70, 환자 교육은 0.39~0.52로 연도별로 다양한 상관관계를 보였다(Tables 3, 4, 5).

고찰

본 연구는 3개 연도의 가정의학과 전문의 자격시험 성적을 이용하여 최선답형 필기시험과 실기시험 성적 간의 상관관계를 분석하고자 하였다.

인지영역 평가시험인 1차 필기MCQ와 2차 슬라이드MCQ 간의 상관계수가 가장 높았다. 이는 1차 필기시험과 2차 슬라이드시험이 형식이 다르더라도 같은 인지영역 평가라는 점에 비추어 예상되는 결과이다.

상관분석에서 일반적으로 상관계수가 0.5 미만이면 약한 양적 선형관계, 0.5 이상이면 강한 양적 선형관계로 간주한다 [13]. 본 연구에서 행동영역 평가인 CXP 총점과 필기시험 간의 상관계수를 보면 1차 필기 MCQ나 2차 슬라이드MCQ와는 약한 상관계수를 보였다. 기존 국내 의과대학생에서 CPX 총점과 필기시험 성적 간의 상관계수인 0.13~0.28, 내과 슬라이드시험과의 상관계수 0.19, 전체 학교 성적 0.28~0.32 [10], 신경과 전문의 고시의 필기시험과의 상관계수 0.15~0.28 [8]과 유사한 수준이다. 이는 행동영역 평가가 사전 지식과 문제 해결능력에 영향을 받을 수 있는 가능성이 있으므로 CPX 점수와 필기시험 성적이 어느 정도의 상관성은 보인다고 우선적으로 해석할 수 있다. 하지만 한편으로는 두 가지 형식의 시험 성적의 연관성의 강도가 높지 않은 것은 평가영역이 서로 달랐기 때문으로 해석되어 가정의학과 전문의 고시에서 인지영역의 시험에 추가적으로 행동평가영역을 포함시켜야 한다는 데 합리적인 근거를 제시한다고 해석할 수 있겠다. 즉, CPX는 기존 전문의 자격시험의 최선답형 필기시험과 비교하여 고유한 평가영역을 가지고 있음을 확인할 수 있다. 행동영역의 CPX 성적도 인지영역과 상대적으로 약하지만 상관관계가 있음을 보여주는 이유는, 의학지식이 CPX의 환자대면진료에서 도구적 지식으로 활용되고 있기 때문이다. 즉, 진료능력이라는 고위 능력은 사전지식체계를 기초로 하고 있기 때문에 보여진다. 필기시험 성적과 CPX 시험성적 간의 상관계수는 시험 연도별로 다르게 나타났는데, 2011년에 비해 2009년, 2010년에 보다 약한 상관관계를 보여주었다. 시험 연도별 차이는 신경과 전문의 고시[8]와도 유사한 현상인데, 이는 CPX 문항에 따라 인지영역의 반영도가 달라질 가능성을 제기하고

있다.

CPX에서 주요하게 평가되는 병력 청취, 신체 진찰, 환자 교육(정보 공유) 및 환자-의사관계 4가지 세부 평가영역별 성적과 1차 필기MCQ 성적 간의 상관계수를 살펴보면 시험 연도에 따라 다르게 나타나 2011년에는 약한 상관계수를 보였으며 2010년도에는 연관성이 없다고도 나타났다. 필기MCQ 성적과 CPX 평가영역간의 상관계수를 살펴보면 병력 청취, 신체 진찰, 환자 교육, 환자-의사관계 등 4가지 영역 모두 약한 상관계수를 보였으며 그 중에서도 환자 교육과 병력 청취 간의 상관관계가 가장 낮았다.

CPX 하위 4가지 평가항목 중 어떠한 영역이 인지능력 평가와 다른 영역인가 하는 것은 연구마다 결과가 상이하다. 전공의를 대상으로 한 외국의 연구에서는 신체 진찰이 환자 면담이나 감별 진단에 비해 상관계수가 낮았으며[14], 국내 3학년 의과대학생을 대상으로 한 연구에서도 신체 진찰은 학생 성적과 무관하였다[3]. 하지만 국내 의과대학생을 대상으로 한 연구에서는 신체 진찰과의 상관계수가 가장 높고, 환자 교육, 병력 청취 순이었으며 임상 예절만 상관성이 없다고 나타났다[7]. 신경과 전문의 자격시험 결과를 분석한 연구에서는 시험 연도별로 환자 교육에서 상관성이 없거나, 신체 진찰과의 상관성이 없다고 다르게 나타났다[8].

본 연구는 국내 의과대학생과 유사하게 신체 진찰이 다른 평가항목에 비해 상대적으로 인지영역 평가인 필기시험 성적과 상관계수가 높고 환자 교육과 병력 청취가 상관계수가 낮거나 없다고 나타났다. 신체 진찰과의 상관계수가 시험 연도에 따라 많이 달라지는 것에 비추어 CPX 시나리오 및 채점표 개발에 따라 신체 진찰이 평가하는 영역이 달라진다고 볼 수 있다. 신체진찰에 술기라는 행동영역을 대표적으로 평가하는 영역임에 비추어 볼 때 이러한 결과가 어떠한 요인에 의해 나타나는지 추가 연구가 필요한 영역이다.

2차 슬라이드MCQ 총점과 CPX 평가영역 간의 상관계수를 살펴보면 연도별로 다르지만 약한 상관계수를 보였으며, 그 중에서도 환자 교육과 환자-의사관계의 상관관계가 낮았다.

2차 슬라이드MCQ와 CPX 평가영역 간의 상관계수는 1차 필기MCQ 성적과 CPX 평가영역 간의 상관계수와 조금 다른 양상을 보이는데, 이는 국내 의과대학생을 대상으로 한 연구 결과에도 나타났던 현상이다. 전문의 시험의 경우 동일한 슬

라이드 시험과 분석한 결과가 없어 비교할 수 없었다. 슬라이드MCQ와 CPX 평가영역 간의 상관계수는 시험 연도별로 수치가 차이가 많이 난다. 이에 대한 요인은 다양한 요인이 작용하리라 보이지만, 한 가지 요인은 슬라이드시험 문항수가 50개로 적어 같은 인지영역에 대한 평가일지라도 250문항인 필기시험에 비해 신뢰도가 낮을 수밖에 없기 때문으로 보인다.

CPX 평가항목 중 환자 교육이 1차 필기MCQ, 2차 슬라이드MCQ 모두와 가장 낮은 상관계수를 보이는 것은 두 사례의 CPX에서 환자 교육의 채점표 항목이 적었던 것이 한 가지 이유가 될 수 있다. 국내 의과대학생을 대상으로 한 연구 결과와도 유사한 결과로 이전 연구에서 지적한 대로 환자 교육을 위해서는 단순한 지식의 축적 이외에 다른 요소가 필요함을 반증한다고 볼 수 있다[3]. 즉, 필기시험을 보완할 수 있는 평가영역임에 비추어 환자 교육은 향후 CPX 시험에서 강화할 필요성이 있는 영역으로 보인다.

CPX 총점과 각 평가항목별 성적 간의 상관관계를 살펴보면 시험 연도와 관계없이 일정하게 상관계수가 높은 평가영역은 병력 청취, 환자-의사관계로 나타났다. 이는 평가영역별 구성 비율이 일정하기 때문이기도 하지만, 한편으로 반영 비율이 일정하면 시험 연도에 따라 수험생이나 개발 문항 내용에 영향을 적게 받는다는 것을 의미한다. 따라서 CPX 시험성적은 신뢰도 변동이 적어 전문의시험에서 CPX 시험 적용의 유용성을 보여주는 것이다.

CPX 총점과 신체 진찰 간의 상관계수는 시험 연도에 따라 약한 상관관계이거나 강한 상관관계로 나타나고 있다. 이는 CPX 문항 중 신체 진찰 평가영역의 비중이 문항에 따라 시험 연도별로 6.3%부터 20.8%로 달라서 CPX 총점에 기여하는 정도가 달라지기 때문이라고 해석할 수 있다.

CPX 총점과 지도 전문의의 총괄평가 점수와의 상관계수(0.68~0.82)는 SP의 총괄평가와의 상관계수(0.47~0.52)보다 높았으며 문항수가 1개로 낮은 비율(3.3~4.0%)임에도 불구하고 CPX의 모든 평가영역에 비해 가장 높은 상관계수를 보여주었다. 이는 지도 전문의나 SP의 총괄점수가 전체 점수와의 상관성이 높다는 기존의 연구 결과[15]와 유사한 결과이다. 가정의학과 전문의 자격시험에서도 지도 전문의 총괄점수의 타당성이 유지된다고 보여지는 결과이다. 이는 CPX를 이용한 전문의 고시에서 합격선 설정방법으로 지도전문의의

총괄점수의 유용성을 뒷받침할 수 있는 소견이기도 하며 경제집단합격선 결정방법의 타당성에 대한 근거가 될 수 있다 [16].

본 연구의 제한점은 각 시험 연도별 CPX 사례가 두 사례로 적어 CPX 점수의 신뢰도가 상대적으로 낮을 수 있다는 점과 각 영역별 채점항목의 수가 일부 영역에서 충분치 않았다는 점이다. 이러한 이유로 시험 연도별 상관계수의 분포의 폭이 넓게 나타나게 된 것으로 보인다.

본 연구는 전문의시험에서 필기시험과 CPX 실기시험 간의 상관성을 연구한 기초자료인데 국내외에서 의사전문의 자격 시험 성적 자료로 분석한 연구가 매우 적어, 이 연구 결과를 확장시킨다면 CPX 평가영역과 필기시험 평가영역 개발에 있어서 상호 도움을 줄 수 있으리라 보인다. 그러한 목적을 위해서는 인지영역평가인 필기시험도 영역을 구분하여 점수를 산출하고 CPX의 세부 평가항목과 비교 분석하는 연구가 필요하다. 하지만 현재 전문의 시험에서 총점 이외에 개별 문항반응 점수를 얻을 수 있는 원자료가 공개되지 않아서 이러한 분석을 할 수 없었다. 향후 의학교육 평가 목적에 한하여 필요한 원자료가 공개된다면 각 평가에서 중복되는 영역을 감소시키고 강화할 영역을 개발하는 자료로 활용될 수 있으리라 본다.

Acknowledgements: None.

Funding: None.

Conflicts of interest: None.

REFERENCES

1. Leach DC. Evaluation of competency: an ACGME perspective. Accreditation Council for Graduate Medical Education. Am J Phys Med Rehabil 2000; 79: 487-489.
2. Crossley J, Humphris G, Jolly B. Assessing health professionals. Med Educ 2002; 36: 800-804.
3. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990; 65(9 Suppl): S63-S67.
4. Reznick R, Smee S, Rothman A, Chalmers A, Swanson

- D, Dufresne L, Lacombe G, Baumber J, Poldre P, Levasseur L. An objective structured clinical examination for the licentiate: report of the pilot project of the Medical Council of Canada. *Acad Med* 1992; 67: 487-494.
5. Papadakis MA. The Step 2 clinical-skills examination. *N Engl J Med* 2004; 350: 1703-1705.
6. Onishi H, Yoshida I. Rapid change in Japanese medical education. *Med Teach* 2004; 26: 403-408.
7. 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.
8. Kim BJ, Sung JJ, Park HK, Seo DW, Chung CS, Yoon BW. Clinical performance examination utilizing standardized patients in board examination: based on the board examination of Korean Neurological Association for three years. *Korean J Med Educ* 2011; 23: 127-135.
9. Dennehy PC, Susarla SM, Karimbux NY. Relationship between dental students' performance on standardized multiple-choice examinations and OSCEs. *J Dent Educ* 2008; 72: 585-592.
10. 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.
11. Kim JJ, Lee KJ, Choi KY, Lee DW. Analysis of the evaluation for clinical performance examination using standardized patients in one medical school. *Korean J Med Educ* 2004; 16: 51-61.
12. Gerrow JD, Murphy HJ, Boyd MA, Scott DA. Concurrent validity of written and OSCE components of the Canadian dental certification examinations. *J Dent Educ* 2003; 67: 896-901.
13. Glaser AN. High-yield biostatistics. Philadelphia, USA: Lippincott Williams & Wilkins; 2001.
14. Stillman P, Swanson D, Regan MB, Philbin MM, Nelson V, Ebert T, Ley B, Parrino T, Shorey J, Stillman A, Alpert E, Caslowitz J, Clive D, Florek J, Hamolsky M, Hatem C, Kizirian J, Kopelman R, Levenson D, Levinson G, McCue J, Pohl H, Schiffman F, Schwartz J, Thane M, Wolf M. Assessment of clinical skills of residents utilizing standardized patients. A follow-up study and recommendations for application. *Ann Intern Med* 1991; 114: 393-401.
15. Solomon DJ, Szauter K, Rosebraugh CJ, Callaway MR. Global ratings of student performance in a standardized patient examination: is the whole more than the sum of the parts? *Adv Health Sci Educ Theory Pract* 2000; 5: 131-140.
16. Wilkinson TJ, Newble DI, Frampton CM. Standard setting in an objective structured clinical examination: use of global ratings of borderline performance to determine the passing score. *Med Educ* 2001; 35: 1043-1049.