

Comprehension of Patient-Physician Interaction through Analysis of Relationships between Domains in Clinical Performance Examination

Jae Jin Han¹, Myoung Jin Lee¹ and Hyun-jung Im²

¹Department of Medical Education, Ewha Womans University School of Medicine, and

²Korean Educational Development Institute, Seoul, Korea

임상수행시험에서 ‘환자의사관계’ 영역의 특성 – 평가영역 간 관계분석을 통하여

¹이화여자대학교 의학전문대학원 의학교육실, ²한국교육개발원

한재진¹, 이명진¹, 임현정²

Purpose: This research investigated the psychometric properties of the patient-physician interaction (PPI) domain in the clinical performance examination (CPX). This research aimed to understand the PPI domain in the CPX through a psychometric and relationship analysis between the domains.

Methods: Data were drawn from 1,302 examinees on a set of 6 common CPX cases and 1,066 on a 'bad news delivery' case. All cases included 7 PPI items, among which we calculated internal consistency reliability. Correlations were made between PPI and the other domains. Analyses using the structural equation model (SEM) were conducted to assess the relationships between latent factors and controlled measurement errors. To calculate the disparity between colleges, we performed a multi-level analysis. Also, we conducted t-tests to investigate the consistency of the PPI and information sharing (IS) scores in the 'bad news delivery' case.

Results: Correlation between the mean PPI score and the total CPX score was high (0.707). The correlations between PPI and other domains were; 0.904 for the patient's overall satisfaction, 0.41 for history taking, and 0.327 for patient education. In SEM, these correlations between latent variables increased. The proportion of level-2 (between-school component) variance in PPI was 4.1%. For 'bad news delivery', the group that checked 'yes' on the IS items had higher PPI scores (mostly $p < 0.01$).

Conclusion: PPI is an influential domain of the CPX and is highly related with the patient's overall satisfaction, clinical courtesy, and history taking. Disparities between schools in PPI are relatively small, such that the PPI could be due to individual factors rather than the school.

Key Words: Clinical competence, Undergraduate medical education, Educational measurement

Received: April 6, 2010 • Revised: May 10, 2010 • Accepted: May 17, 2010

Corresponding Author: Jae Jin Han

Department of Medical Education, Ewha Womans University School of Medicine, 911-1 Mok 6-dong, Yangcheon-gu, Seoul 158-710, Korea

Tel: +82.2650.5629 Fax: +82.2650.5836 email: jjhan@ewha.ac.kr

*This paper was presented at the poster session of the 5th Asia Pacific Medical Education Conference, 2008.

Korean J Med Educ 2010 Sep; 22(3): 177-184.

doi: 10.3946/kjme.2010.22.3.177.

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

임상수행능력을 평가하기 위한 대표적인 평가도구인 객관 구조화진료시험(objective structured clinical examination) 혹은 임상수행시험(clinical performance examination, CPX)은 평가 도구로서 적합한 신뢰도와 타당도를 갖추기 위하여 실제 진료 현장에서 행해지는 환자 진료 형태, 즉 의사와 환자의 대화, 신체진찰을 포함한 정보수집, 환자교육, 환자관리 등과 같이 보다 복잡한 임상 상황을 표준화하여 시험 형태로 만들게 된다[1,2]. 하지만 진료 모습을 분절화하여 개발하게 되는 CPX 평가 도구는 그 속성상 사용하는 증례의 내용, 증례 수, 평가 문항, 표준화 환자, 평가자 및 평가 방법 등의 변수들이 합적선 결정을 포함한 표준화 평가 도구의 신뢰도와 타당도의 적정화에 매우 심대한 영향을 미치게 된다[3]. 대부분의 CPX 증례는 실제 임상 수행의 핵심 요소인 ‘환자로부터의 정보 수집’, 이를 통한 ‘임상추론’ 및 ‘환자-의사 관계(patient-physician interaction, PPI)’의 세 가지 공통된 구성 요소를 갖고 있다. 하지만 이러한 핵심 구성 요소는 평가의 신뢰도를 높이기 위하여 다시 하위 구조로 분절되는데 이러한 하위 구조 혹은 세부 영역을 어떻게 구성하고 평가 도구화하는 것이 적절한지, 특히 PPI 영역의 문항 구성은 어떤 것이 가장 이상적인지는 논란이 있다[4]. 따라서 CPX 내의 분절화된 하위 구조와 각각의 세부 영역 간의 관계를 분석해 보는 것은 타당도 검증의 측면에서 의미가 있다고 할 수 있다. 그동안 PPI 영역에 관련된 국내의 선행 연구로는 조기임상교육과정으로 시행한 ‘의료대학수업’의 성적과 추후 시행한 CPX에서의 PPI 점수 및 PPI 항목과의 연관성이나[5], CPX의 정보 공유가 PPI 점수에 미치는 영향 등이 분석되었다[6,7].

본 연구에서는 서울-경기 컨소시엄의 CPX를 구성하고 있는 하위 영역 중에서 PPI 영역과 타 영역(병력청취, 신체진찰, 환자교육, 임상예절, 정보 나누기 등)과의 다각적인 관계 분석을 통해 사례와 문항에 따른 PPI 점수 분포의 차이, PPI 점수와 총점 및 타 영역 점수의 상관관계, PPI 점수의 학교별 차이 등을 파악하고자 하였다. 이를 통하여 검사 도구의 본질이라고 할 수 있는 타당도 검증을 위한 일부 증거를 수집할 수 있을 것이며, 이와 같은 분석과 수집된 증거는 향후 보다

통합적으로 검사 도구의 타당도를 검증하기 위한 기초 자료로 활용 될 수 있다. 또한 사례와 문항에 따른 PPI 점수의 차이 검증, 유사 영역 점수와의 일치 정도, 임상수행능력에 학교의 특성이 기여하는 정도를 분석함으로써 타 영역에 비해 PPI 영역이 가지는 고유한 특성을 이해하고자 하였다.

대상 및 방법

1. 연구 대상

서울/경기 CPX 컨소시엄에서 시행한 1,640명의 의과대학 의학과 4학년생을 대상으로 하였다. 컨소시엄에 참여한 의과대학은 18개교로 각 대학별 학생 수는 38~173명으로 다양하였으며, 각 대학의 실정에 맞게 5~10개 사례에 대한 평가를 실시하였다. 시행한 사례의 수와 종류가 다를 경우에도 개인의 CPX 점수는 100점 만점인 사례 점수의 평균으로 계산되므로 어떤 사례를 시행했는가가 CPX 점수에 영향을 미친다. CPX에는 일정 부분 사례 특이성이 존재하는 것으로 보고되었으며[8], 이러한 사례 특이성이 각 사례의 평균을 이용하여 영역 점수를 산출하는 경우에 오차원(source of error)으로 작용할 수 있다. 그러므로 본 연구에서는 각 대학에서 대부분 공통적으로 시행한 6개 사례를 모두 치른 학생만을 연구대상으로 하였으며 추출된 연구대상의 수는 15개 대학에 소속된 의과대학생 1,302명이었다. 다만, ‘나쁜 소식 전하기(사례 7)’에서 정보나누기 영역 점수와 PPI와의 일치 여부를 분석할 때에는 이 사례에 참여한 모든 학생을 대상으로 분석하였으며, 12개 대학 1,066명의 의과대학생이 이에 해당하였다.

2. 측정 도구

CPX 사례는 전반적 평가(GR), 병력청취(HT), 신체진찰(PE), 임상 의사 예절(CC), 환자-의사 관계(PPI), 환자교육(EDU)과 기타 정신과적 평가(Psy), 정보 나누기(IS)로 구성되어 있다. 이 중 PPI 영역은 0점부터 5점까지의 6단계 척도 7문항으로 모든 사례에 동일한 형태로 포함되어 있으며, 각 문항은 ‘관계 형성’, ‘청취 능력’, ‘환자 격려’, ‘분위기 조성’, ‘예의’, ‘정보 전달’, ‘전문가적 태도’를 측정한다.

3. 자료 분석

1) 본 연구에서는 ‘인사하기’ 영역을 총점에서 제외하였으며, ‘신체진찰’ 영역 점수와 ‘임상예절’ 영역이 구조상 연동이 되어 있으므로 ‘임상예절’ 점수가 ‘해당 없음’으로 채점된 경우에는 0점으로 수정하였다. 이와 같이 자료를 수정하고 공통된 6개 사례의 평균으로 CPX의 각 영역 점수 및 총점을 재계산하였다.

2) PPI 문항의 Cronbach α 계수를 산출하여 내적일관성 신뢰도를 검증하고 PPI 점수 및 사례별 총점의 기술 통계를 산출하여, 각 사례별 PPI 점수와 사례 총점의 분포가 어떠한 경향을 보이고 있는가를 파악하였다.

3) PPI 영역과 타 영역 간 상관계수를 산출하고, 구조방정식모형(structural equation modeling, SEM)을 이용하여 측정오차를 고려한 잠재요인 간 관계를 분석하였다. 설정된 모형의 적합도 지수로는 NNFI, CFI와 RMSEA를 이용하며 관련 문헌의 기준에 따라 설정된 모형의 적합성 여부를 판단하였다[9]. 보통 NNFI는 값이 높아질수록 모형의 적합도가 좋은 것을 나타내며 대체로 0.9 이상이면 좋은 적합도라고 볼 수 있고, RMSEA가 0.05 이하이면 좋은 적합도(close fit), RMSEA가 0.08 이하이면 괜찮은 적합도(reasonable fit), RMSEA가 0.10 이하이면 보통 적합도(mediocre fit), RMSEA가 0.10 이상이면 나쁜 적합도(unacceptable fit)로 판단한다[10].

4) PPI 점수 및 타 영역 점수가 학교에 따라 다르게 나타나는 것을 파악하기 위하여 학교수준에 학생수준이 포함되는 2수준 다층분석(multilevel analysis)을 통해서 학교수준 분산을 산출하고, 전체 분산에서 학교차 분산이 차지하는 비율을 계산하였다. 학교(j)에 다니는 학생(i)의 점수(Y_{ij})는 다음 식과 같이 나타낼 수 있다.

[1수준: 학생수준]

$$Y_{ij} = \beta_{0j} + r_{ij}, r_{ij} \sim N(0, \sigma^2)$$

[2수준: 학교수준]

$$\beta_{0j} = \gamma_{00} + \mu_{0j}, \mu_{0j} \sim N(0, \tau_{00})$$

위의 식에서 전체 학생의 만족도 분산을 고려해 보면, 총분산은 해당 학교에 소속됨으로써 나타나는 학교의 고유한 분산(τ_{00})과 학생 개인의 특성에서 비롯된 개인차 분산(σ^2)

으로 구성됨을 알 수 있다. 전체 만족도의 분산 중에서 학교수준 분산이 차지하는 비율($\tau_{00}/(\sigma^2 + \tau_{00})$)을 산출함으로써 학생의 만족도에 있어서 학교 간 차이로 인하여 발생하는 분산의 크기를 파악할 수 있다.

5) 나쁜 소식 전하기에 포함되어 있는 정보나누기 문항에 대한 응답 결과와 PPI 점수의 일치 여부를 파악하기 위하여 t검정을 이용하여 정답 여부에 따른 PPI 점수의 평균을 비교하였다.

결과

1. PPI 점수의 분포

본 연구를 위해 선택된 6개 사례에서 100점 만점으로 환산된 PPI 점수와 사례 총점의 기술 통계를 산출하여 비교하였다(Table 1, Fig. 1).

PPI 점수는 1번 사례에서 가장 높게 나타났고 2번 사례에서 가장 낮게 나타났으며 약 5점의 차이를 보였다. 사례 총점은 5번 사례에서 가장 높았고 2번 사례에서 가장 낮게 나타났으며, 이 둘 사이의 차이는 14점이었다. PPI 영역에서나 병력 청취, 신체 진찰 등 다른 임상수행능력 영역에서 사례 2가 가장 어려웠다고 볼 수 있다. 그러나 PPI 점수는 사례별로 큰 차이를 나타내지 않는데 반하여 사례 총점은 사례에 따라서 점수 차이가 큰 편으로 나타났다. 그러므로 PPI 이외에 임상수행능력을 구성하는 나머지 요인이 사례에 따라서 편차가 크다는 것을 알 수 있다. PPI 점수와 사례 총점 모두 표준편

Table 1. Descriptive Statistics: PPI Score by Case Total Score (n=1,302)

Case	PPI score		Total score		
	Mean	SD	Mean	SD	Cronbach' α
1	64.34	11.78	56.99	12.86	0.904
2	59.35	12.72	47.95	11.41	0.897
3	60.69	9.32	50.54	12.17	0.889
4	59.88	13.29	55.04	10.25	0.913
5	63.27	9.56	62.03	11.02	0.867
6	60.92	9.88	60.68	9.68	0.881

PPI: Patient-physician interaction, SD: Standard deviation.

Fig. 1. Patient-Physician Interaction (PPI) Scores by Case

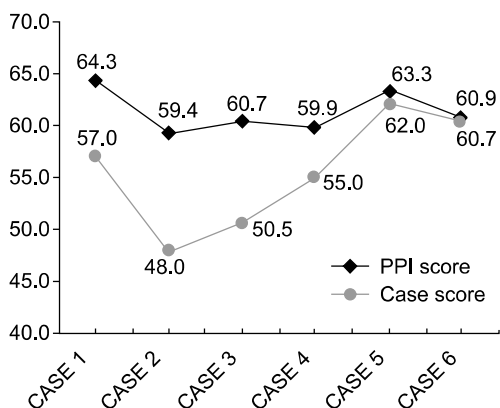


Table 2. Correlation Coefficients between Case Score, PPI Score and CPX Final Score

	Correlation	Mean correlation ^{a)}
Case 1-PPI 1	0.637 ^{b)}	0.595
Case 2-PPI 2	0.672 ^{b)}	0.624
Case 3-PPI 3	0.640 ^{b)}	0.644
Case 4-PPI 4	0.683 ^{b)}	0.705
Case 5-PPI 5	0.636 ^{b)}	0.665
Case 6-PPI 6	0.619 ^{b)}	0.595
CPX-PPI	0.707 ^{b)}	

PPI: Patient-physician interaction, CPX: Clinical performance examination.

^{a)}Mean correlation between standardized patients, ^{b)}p<0.01.

차는 유사하게 나타났으나 4번 사례에서의 PPI 점수 표준편차가 비교적 크게 나타나서 해당 사례에서 학생의 PPI 점수에서 개인차가 상대적으로 컸다는 것을 알 수 있다.

2. PPI 점수와 사례 총점과의 상관관계

PPI 점수와 사례 총점과의 관계를 살펴보기 위하여 상관계수를 산출하였다. 일반적으로 사례마다 4~5명의 표준화 환자가 투입되었는데, 이 차이로 인하여 상관관계가 다르게 나타날 수 있으므로 표준화 환자별 상관계수를 산출하여 평균값을 같이 제시하였다(Table 2).

전체적으로 PPI 점수와 사례 총점과의 상관계수는 0.6에서 0.7 사이에 분포하였고 표준화 환자별 상관계수의 평균값도 유사한 분포를 보였다. 표준화 환자의 구성에 따라서 채점자 간 상관계수의 편차가 큰 사례도 있었으나 평균적으로 유사

Table 3. Correlations between PPI and Other Domains of CPX

Domain of CPX	GR	HT	PE	EDU	CC	PPI
HT	0.445 ^{a)}					
PE	0.374 ^{a)}	0.465 ^{a)}				
EDU	0.317 ^{a)}	0.147 ^{a)}	0.138 ^{a)}			
CC	0.353 ^{a)}	0.256 ^{a)}	0.507 ^{a)}	0.106 ^{a)}		
PPI	0.904 ^{a)}	0.410 ^{a)}	0.332 ^{a)}	0.327 ^{a)}	0.375 ^{a)}	
CPX score	0.715 ^{a)}	0.816 ^{a)}	0.725 ^{a)}	0.342 ^{a)}	0.587 ^{a)}	0.707 ^{a)}

CPX: Clinical performance examination, GR: Global rating, HT: History taking, PE: Physical exam, EDU: Patient education, CC: Clinical courtesy, PPI: Patient-physician interaction.

^{a)}p<0.01.

한 관계를 나타냈다. 전체 PPI 점수 평균과 CPX 총점과의 상관관은 0.707로 나타났다. PPI와 사례 총점, CPX 총점과의 상관계수는 모두 유의수준 0.01에서 통계적으로 유의하였으므로, PPI가 임상수행능력과 상관이 높은 주요 영역임을 알 수 있다. 한편 PPI 점수와 사례 총점의 상관계수가 각 사례에 관계없이 어느 정도 일정 크기를 유지하고 있는 것으로 보아 사례 특이성에도 불구하고 안정적인 경향을 나타낸다고 볼 수 있다. 각 사례별 PPI의 신뢰도는 0.867~0.904로 매우 높은 것으로 나타나서 측정 문항의 내적 일관성이 높음을 알 수 있다.

3. PPI 점수와 타 영역 점수와의 상관관계

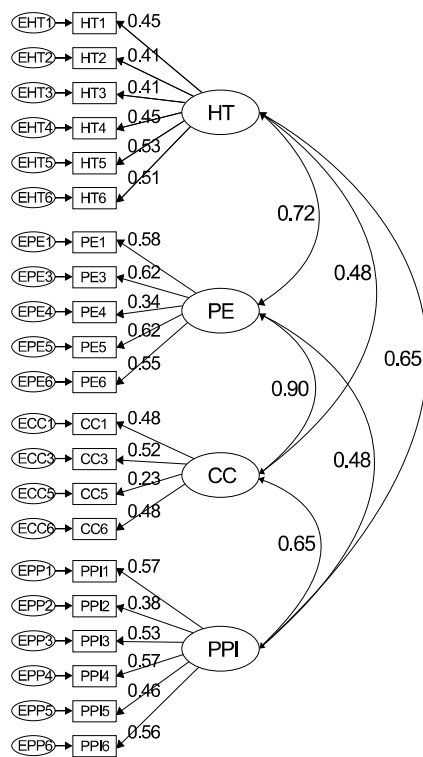
PPI 영역이 임상수행능력의 다른 영역과 어떤 관계가 있는가를 파악하기 위하여 각 영역의 6개 사례 평균 점수 간 상관계수를 산출한 결과는 Table 3과 같다. PPI와 타 영역 간의 상관계수는 대략 0.3에서 0.4 사이에 분포하였고, 그 중 병력 청취와의 상관이 0.410으로 가장 높았으며, 환자교육과의 상관이 가장 낮았다(r=0.327). PPI 영역이 어느 특정 영역과 상관이 크게 높지 않기 때문에 임상수행능력 평가에 있어서 고유한 기능을 수행하고 있는 영역임을 알 수 있다.

CPX의 각 사례에는 평가지 제일 처음에 학생 의사에 대한 표준화 환자의 전반적인 만족도를 묻는 문항이 있는데, 이 문항과 PPI의 상관계수가 0.904로 매우 높게 나타났다. 이는 통계적으로 유의할 뿐만 아니라 전반적인 만족도를 묻는 1개 문항이 PPI 점수 변화의 80% 이상을 설명해 줄 수 있다는 의미로 환자의 전반적인 만족도가 PPI에 큰 영향을 미친다는 것을 알 수 있다.

CPX 총점과 하위 영역 간의 상관계수를 살펴보면, 대체적으로 배점이 높은 병력청취, 신체진찰 영역이 총점과의 상관성이 높게 나타났다. PPI와 CPX 총점과의 상관은 0.707로 높게 나타났다. 또한 PPI와 매우 높은 상관을 보인 전반적인 만족도와 총점의 상관도 0.715로 높았다. 반면 임상예절은 보통 수준의 상관을 보였고($r=0.587$), 환자교육 영역이 총점과 상관성이 낮은 것으로 나타났다.

측정오차를 통제된 상태에서 잠재변수 간 관계를 분석함으로써 보다 정확한 요인 간 관계를 파악하기 위하여 구조방정식모형을 사용한 분석을 수행하였다. 자료를 잘 설명해 줄 수 있는 최적의 모형을 찾기 위해서 상관계수 분석에서 사용한 5개 영역을 투입한 5요인 모형과 다른 영역과의 상관성이 낮은 뿐더러 각 사례별 점수 간 안정적인 측정 결과를 보여주지 못하고 있는 환자교육 영역을 제외한 4요인 모형을 대안적으로 설정하였다. 두 모형의 적합도를 비교한 결과, 보다 적합하다

Fig. 2. Correlations between Latent Variables of Clinical Performance Examination (CPX)



Goodness of fit index: $\chi^2 = 1447.59$, $df = 183$, ($p = 0.000$). NNFI = 0.714, CFI = 0.750, RMSEA = 0.073. HT: History taking, PE: Physical exam, CC: Clinical courtesy, PPI: Patient-physician interaction.

고 평가되는 4요인 모형을 Fig. 2에 제시하였다. 선택된 모형의 적합도는 RMSEA=0.073으로 적절한 수준의 적합도를 보였다[9,10].

Table 3에 제시된 상관계수와 비교해 볼 때, 잠재변수 간 상관은 전체적으로 기존 영역 점수 간 상관보다는 높게 나타났다. PPI 영역은 병력청취와 임상예절과 0.65, 신체진찰과 0.48의 상관을 나타내어 중간 수준의 상관을 보였다. 그러므로 병력청취와 임상예절 점수가 높을수록 PPI 점수가 높은 경향이 있다고 볼 수 있다. PPI를 제외한 다른 영역 간 상관도 높게 나타났으며, 특히 신체진찰과 임상예절은 0.90의 높은 상관을 보였다.

4. PPI 점수의 학교차 분석

PPI 점수의 학교 간 편차를 알아보기 위하여 다층분석을 이용해 학교 간 분산 비율을 산출하였다(Table 4). 본 연구의 자료는 학생이 학교에 포함된 2수준 다층자료로 볼 수 있으므로 다층분석을 통하여 전체 분산을 분석하면, 학교차에 기인한 학교 간 분산과 해당 학교에 소속된 학생들의 개인차에 기인한 학교 내 분산으로 구분하여 산출할 수 있다.

분산성분 분석 결과, PPI 점수의 학교 간 분산은 4.1%로 타 영역에 비하여 학교차가 적은 것으로 나타났다. 학교차가 큰 영역으로는 신체진찰(33.4%), 병력청취(32.6%)를 들 수 있으며, 이로 인하여 전체 CPX 점수 분산의 31.7%가 학교차에 기인하는 것으로 나타났다. 이는 학생이 받은 CPX 점수의 30% 이상이 학생이 소속된 학교에 의해 결정된다는 것을 의미하는 것이므로 학교의 영향이 매우 크게 작용했다고 볼 수 있다.

Table 4. Variance Component Ratio Using Multilevel Analysis

	Variance component (%)	
	Between-school	Within-school
HT	32.31 (32.6)	66.95 (67.4)
PE	68.25 (33.4)	136.14 (66.6)
EDU	14.82 (6.6)	210.48 (93.4)
CC	38.25 (17.6)	178.71 (82.4)
PPI	1.73 (4.1)	40.58 (95.9)
CPX final score	16.60 (31.7)	35.79 (68.3)

HT: History taking, PE: Physical exam, EDU: Patient education, CC: Clinical courtesy, PPI: Patient-physician interaction, CPX: Clinical performance examination.

Table 5. Comparisons of PPI Scores with Results of Performance on Checklist Items in the 'Bad-News-Delivery' Case

Items	Results of performance	PPI score			t-statistics
		n	Mean	SD	
Relationship 1	Yes	551	69.3	11.0	12.65 ^{a)}
	No	515	61.7	8.3	
Relationship 2	Yes	998	66.4	10.2	9.35 ^{a)}
	No	68	54.5	8.9	
Relationship 3	Yes	892	67.3	10.0	12.59 ^{a)}
	No	174	57.1	8.4	
Explaining 1	Yes	831	66.3	10.8	3.89 ^{a)}
	No	235	63.3	8.8	
Explaining 2	Yes	756	67.2	10.9	8.10 ^{a)}
	No	310	61.7	8.1	
Explaining 3	Yes	1,044	65.9	10.2	6.10 ^{a)}
	No	22	52.3	13.6	
Listening	Yes	1,054	65.7	10.5	2.26 ^{b)}
	No	12	58.8	9.3	

PPI: Patient-physician interaction, SD: Standard deviation.

^{a)}p<0.01, ^{b)}p<0.05.

PPI 영역은 다른 영역보다 학교에 따른 점수 차이가 크지 않고 학생 개인의 차이가 대부분임을 알 수 있다.

5. '나쁜 소식 전하기'에서의 PPI

'나쁜 소식 전하기'에 해당하는 사례 7에는 정보 나누기 영역을 측정하는 14문항이 '예/아니오'의 2점 척도로 포함되어 있다. 이 중 PPI의 문항과 내용상 직·간접적으로 일치하는 7개 문항을 선정하여 해당 문항에 대한 응답 여부에 따라 PPI 점수에 차이가 있는가를 검정하였다(Table 5).

'나쁜 소식 전하기' 사례의 정보나누기 문항 중에서 PPI 문항과 유사한 내용을 묻는 것으로 선택된 7문항은 공감이나 충분한 설명 등과 같이 분위기 조성, 정보전달, 청취능력과 관련된 수행을 측정하는 문항이다. 해당 문항에 '예'라고 응답한 집단과 그렇지 못한 집단의 PPI 점수를 비교해 본 결과 모두 유의수준 0.05에서 통계적으로 유의한 차이를 보였으며, '예'라고 응답한 집단의 PPI 평균이 최소 3점에서 10점 이상 높게 나타났다. 그러므로 해당 문항에 대해 적절한 반응을 보인 학생이 PPI에서도 높은 점수를 받았음을 알 수 있다.

고찰

임상수행능력을 평가하는 도구로서의 CPX는 평가의 타당도와 신뢰도를 적절하게 유지하기 위하여 증례 내용 및 수, 표준화 환자의 연기, 채점자, 시험 시간 등의 여러 요소들을 객관화하고 표준화하려는 노력이 요구된다[2]. 또한 증례의 세부 평가 문항의 채점 방법을 개별 행위의 단순한 수행 여부에 대하여 '예/아니오'로 평가하는 '체크리스트' 채점 도구와 개별 행위의 수행 정도에 대하여 '매우 잘함/.../매우 못함'의 기준 중에서 선택하는 '전반적' 채점 도구의 특성 역시 평가의 타당도와 신뢰도에 영향을 미친다. 병력 청취나 신체진찰 영역에서의 채점은 '체크리스트' 채점 도구가 보다 적절하다는 공감감이 있어 왔다. 한편, 개별 행위나 내용 외에 의사와 환자와의 소통, 공감, 대화, 배려, 신뢰 등의 관계를 평가하기 위한 PPI 영역의 채점 도구로서는 '전반적' 채점 도구가 일반적으로 사용되고 있으나 그 적절함에 대한 논란은 계속 있어 왔으며 채점 도구의 타당도와 신뢰도를 보다 높이기 위한 노력들이 요구되고 있다[11]. 또한 CPX 문항의 분절화된 영역과 문항들은 서로가 다른 독립된 평가 목적으로 구성되어 있기는 하지만 임상수행이라는 전체적인 틀에서 볼 때 서로 연관이 있을 수밖에 없다. 특히 환자와의 관계에 연관된 심리평가측면의 다양성은 PPI 영역의 평가를 어렵게 하면서도 임상 수행의 전체적인 능력과 매우 밀접한 관련이 있을 수밖에 없다[12]. CPX가 의학교육과정의 보편적인 평가 방법이 된 점을 고려하면 각 영역과 PPI 영역의 관계를 다양한 방법을 통해 다각적으로 분석해 보려는 시도는 의미가 있을 것이다. 특히 고부담시험의 합격선 결정에 CPX 각 영역별 점수의 적용을 어떻게 하는 것이 적절한가를 결정하기 위한 기본 정보로서 PPI 영역에 대한 분석이 더욱 중요하다고 하겠다[13].

본 연구에서 PPI 점수와 표준화 환자의 전반적 만족도는 매우 높은 상관관계를 나타냈다. 이는 PPI가 환자의 만족도에 중대한 영향을 미친다는 것을 의미한다. PPI에서 측정하고 있는 환자와의 상호작용에서 발생하는 정서적인 요인들이 환자로 하여금 의사의 진료에 만족하도록 하는 중요한 요인임을 다시 한 번 확인할 수 있었다[14]. PPI 점수와 타 영역과의 관계를 분석한 결과, '신체 진찰' 영역에 비해 '병력 청취'

영역과 더 관련이 높았다. 또한 PPI 점수와 CPX 총점 간 상관은 0.7로 높게 나타났다. 그러므로 PPI가 임상수행능력에 미치는 영향이 크다는 것을 알 수 있다(Table 3).

PPI는 모든 사례에 동일한 형태로 포함되어 있는 공통 문항이라고 볼 수 있으므로 PPI 영역은 사례 특이성 및 학교 간 편차가 상대적으로 적은 영역이어야 하며, 다양한 사례에 대해 여러 학교가 응시한 경우에도 PPI 영역에 대한 비교는 어느 정도 안정적으로 이루어질 수 있다. CPX의 타 영역과 비교해 보았을 때, PPI 영역의 학교차 분산이 상대적으로 작게 나타났다. 그러므로 다른 임상수행 영역의 능력에 비해서 PPI 영역에서 측정하는 능력은 학교마다 비슷한 수준이며, PPI 점수의 차이는 개인차에 기인한 부분이 크다는 것을 알 수 있다. PPI 점수에 대한 학교 특색이 없다는 것은 PPI 교육에 대한 각 학교 차원의 노력이 부족하다는 의미일 수도 있으므로 PPI 교육이 학교 차원에서 이루어지는가의 여부 및 내용에 대한 점검이 필요할 것이다. 일반적으로 CPX 점수는 학교 차이보다는 각각의 개인차에 기인한다고 할 수 있다. 본 연구의 개인 및 학교 요인에 대한 다층분석 결과를 볼 때, 개인차가 큰 부분을 차지하고는 있지만 학교 간 분산 비율이 약 30%였다(Table 4). 이는 미국 United States Medical Licensing Examination Step 2 clinical knowledge 점수를 이용하여 다층분석을 수행한 연구[14]에서 학교 차 분산이 약 6%였다는 것과 비교했을 때, 상당히 큰 수치라고 할 수 있으며 본 연구의 대상이 되는 서울/경기 CPX 컨소시엄에 속한 18개의 과대학 학생들의 임상수행능력에 학교 간 격차가 어느 정도 존재한다는 것을 의미한다.

증례 특성상, 환자와의 관계가 대부분의 측정요소인 '나쁜 소식 전하기'의 경우에 채점 방법이 다른 항목들을 비교한 결과 대부분 의미 있게 일치하였다(Table 5). 비슷한 PPI 속성을 가진 유사 문항의 2점 척도 '체크리스트'와 6점 척도 '전반적 채점 도구'의 채점 결과는 모두 학생 평가에 유용할 수 있으므로 두 가지 방법 모두를 이용하여 채점 문항수를 늘리고 문항 목적에 맞추어 연관된 세부 내용을 측정함으로써 문항 신뢰도를 높일 수도 있고 채점자 신뢰도를 점검할 수도 있을 것이다.

요인분석은 관찰변수 간의 상관관계를 소수의 잠재변수에 의해 설명하는 분석 방법으로 임상수행능력평가의 분석에 적

용된 예로는 고부담시험에서 사용된 각각의 증례에 대하여 요인분석을 시행하여 타당한 합격선 결정을 위해 총점에서 제외할 증례를 선택한 연구가 있었으며[15], 전문가적 행동의 개발을 위한 교육 과정에서 자기 성찰이나 자기 통찰 등의 요소들의 기여도를 관찰하여 그 평가도구의 타당도를 분석한 연구도 있었다[16].

본 연구에서도 구조방정식모형을 이용하여 측정오차를 고려한 잠재변수 간 관계를 파악하였으며 기존의 상관관계수 분석과 잠재변수 분석을 모두 수행하여 그 구인 간 관계를 수치화하여 비교하였다(Fig. 2). 즉, 직접적으로 측정할 수 없는 다양한 구인 간 관계를 측정하기 위하여 관찰점수를 그대로 사용하지 않고 측정의 오차를 제거한 진점수를 고려하였다는 점이 보다 실제에 가까운 구인 간 관계 분석이라고 볼 수 있다. 임상수행평가에 관련된 기존의 연구들이 채점의 객관성과 일관성을 나타내는 신뢰도에 많은 관심이 집중된 데 비하여, 평가의 핵심적인 기준으로서 개발된 검사 도구의 타당도 검증을 위한 노력은 부족한 실정이다. 타당도는 검사 점수로부터 유추되는 해석과 추론의 적합성을 의미하는 가장 근본적이고 본질적인 특성이다. 따라서 본 연구에서와 같이 검사 점수의 해석과 활용의 타당성에 대한 종합적인 증거들을 수집하는 것은 개발된 검사 도구의 질을 평가하기 위한 기본적인 절차로서의 의의가 있다. 향후 연구과제로는 사용 증례의 특성별로 PPI 영역의 특성이 어떻게 다른지를 보기 위한 분석이나 이를 통하여 증례의 범주나 구조화 표준을 달리하는 시도를 할 수 있을 것으로 사료된다.

CONFLICT OF INTEREST

None.

REFERENCES

1. Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical examination. *Med Educ* 1988; 22: 325-334.

2. Howley LD. Performance assessment in medical education: where we've been and where we're going. *Eval Health Prof* 2004; 27: 285-303.
3. Cusimano MD. Standard setting in medical education. *Acad Med* 1996; 71(10 Suppl): S112-S120.
4. Boon H, Stewart M. Patient-physician communication assessment instruments: 1986 to 1996 in review. *Patient Educ Couns* 1998; 35: 161-176.
5. Lee YM, Kim BS. Association between student performance in a medical communication skills course and patient-physician interaction scores on a clinical performance examination. *Korean J Med Educ* 2008; 20: 313-320.
6. 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.
7. Kim J, Lee K, Yoo D, Yang E. The effects of information sharing between students on results of clinical performance examination. *Korean J Med Educ* 2006; 18: 239-247.
8. Kim J, Lee K, Yoo D, Yang E. Effects of case type and standardized patient gender on student performance in clinical performance examination. *Korean J Med Educ* 2007; 19: 23-30.
9. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999; 6: 1-55.
10. Hong S. The criteria for selecting appropriate fit indices in structural equation modeling and their rationales. *Korean J Clin Psychol* 2000; 19: 161-177.
11. Cohen DS, Colliver JA, Marcy MS, Fried ED, Swartz MH. Psychometric properties of a standardized-patient checklist and rating-scale form used to assess interpersonal and communication skills. *Acad Med* 1996; 71(1 Suppl): S87-S89.
12. Colliver JA, Swartz MH, Robbs RS, Cohen DS. Relationship between clinical competence and interpersonal and communication skills in standardized patient assessment. *Acad Med* 1999; 74: 271-274.
13. Whelan GP, Boulet JR, McKinley DW, Norcini JJ, van Zanten M, Hambleton RK, Burdick WP, Peitzman SJ. Scoring standardized patient examinations: lessons learned from the development and administration of ECFMG clinical skills assessment (CSA). *Med Teach* 2005; 27: 200-206.
14. Cuddy MM, Swanson DB, Dillon GF, Holtman MC, Clauser BE. A multilevel analysis of the relationships between selected examinee characteristics and United States Medical Licensing Examination Step 2 clinical knowledge performance: revisiting old findings and asking new questions. *Acad Med* 2006; 81(10 Suppl): S103-S107.
15. Chesser AM, Laing MR, Miedzybrodzka ZH, Brittenden J, Heys SD. Factor analysis can be a useful standard setting tool in a high stakes OSCE assessment. *Med Educ* 2004; 38: 825-831.
16. Roberts C, Stark P. Readiness for self-directed change in professional behaviours: factorial validation of the self-reflection and insight scale. *Med Educ* 2008; 42: 1054-1063.