

호흡기계 표준화 환자를 이용한 간호대 학생의 임상수행능력평가 타당도와 신뢰도 검증

최자윤¹ · 장금성² · 최순희² · 홍미순²

¹전남대학교 간호대학 조교수, ²전남대학교 간호대학 교수

Validity and Reliability of a Clinical Performance Examination using Standardized Patients

Choi, Ja Yun¹ · Jang, Keum Seong² · Choi, Soon Hee² · Hong, Mi Soon²

¹Assistant Professor, College of Nursing, Chonnam National University

²Professor, College of Nursing, Chonnam National University, Gwangju, Korea

Purpose: The purpose of this study was to test the validity of a modified clinical performance examination (CPX) for preclinical students in nursing. **Method:** 70 nursing students in their second semester of the junior year at C University participated in CPX. Scenarios and checklists were developed by our research team from September to October 2005. Six stations were organized. Evaluation included physical examination of a patient with lung cancer, education on usage of a metered dosage inhaler, and lobectomy postoperative care. Students were randomly assigned to a station. **Result:** There was a difference in the CPX scores according to stations. The agreement of scoring between trained faculty members and SPs was more than moderate ($r=.647$). The correlation between the CPX score and the average grade in the previous semester and between the CPX score and the average grade of a paper and pen test of the pulmonary system of adults was low ($r=.276$; $r=.048$). **Conclusion:** Traditional CPX is generally recommended, however, modified CPX is appropriate for preclinical students in the current Korean Nursing school setting if there are additional scoring systems to balance the testing level at each station.

Key words: Patient simulation, Clinical competence, Evaluation study

서 론

1. 연구의 필요성

임상실습교육은 그 대상이 인간이고 과거보다 까다로워진 임상실습환경으로 인해 교수자와 학생 모두에게 부담감을 주고 있다. 또한 임상실습의 교육영역이 지식(knowledge)영역뿐 아니라 임상수기(skill)와 태도(attitude)까지를 포함하므로 실습교육의 평가영역 역시 실무능력을 어느 정도 습득하였는지를 종합적으로 평가할 수 있는 방법의 개선이 필요하다. 임상실습

교육의 여러 문제점을 해결하고 보완하기 위해 시뮬레이션을 이용한 임상수행평가를 병행하는 방법이 대두되고 있다(Miller, 1990; Park et al., 2003; Vu et al., 1992). 이러한 접근은 병실과 유사하면서 실제로 경험할 수 있는 환자 증례를 시뮬레이션 상황으로 만들어 학생이 스트레스 없이 능숙해질 때까지 실습할 수 있게 함으로써 이론과 실제의 괴리를 줄이고, 실습현장에서의 적응력을 높여 학생의 부담을 낮추고 교수의 효율적 지도를 지원할 수 있다. 특히, 표준화 환자(Standardized Patients, SP)를 이용하여 실제 상황을 시뮬레이션하면서 임상수행능력을 평가할 수 있다면 간호대 학생의 지식, 임상수기 및 태도의

주요어 : 표준화 환자, 임상수행능력평가, 타당도

*본 논문은 2006년도 전남대학교 간호과학연구소 연구비지원에 의해 이루어졌음.

*This study was supported by the 2006 Research Fund of Chonnam Research Institute of Nursing Science.

Address reprint requests to : Choi, Ja Yun

College of Nursing, Chonnam National University, Chonnam Research Institute of Nursing Science, 5 Hak-dong, Dong-gu, Gwangju 501-746, Korea
Tel: 82-62-220-4343 Fax: 82-62-225-3307 E-mail: choijy@jnu.ac.kr

투고일 : 2007년 10월 19일 심사완료일 : 2007년 11월 27일

전반에 걸친 포괄적인 실습평가가 가능하다.

표준화 환자를 이용한 평가는 단편적인 임상수기 중심의 객관구조화 임상시험(Objective structured clinical examination, OSCE)과 좀 더 포괄적이고 임상 상황에 가깝게 구성된 임상수행능력평가(Clinical performance examination, CPX)의 두 형태로 대별된다. 국내에서는 스테이션당 허용 시간이나 문제의 구성내용을 볼 때 CPX보다는 OSCE가 임상종합평가 목적으로 활발히 이루어졌다(Park et al., 2003). OSCE는 1975년 Harden 등에 의해 처음으로 개발된 이후 의학, 간호학 등의 의료 분야에서 학생들의 임상수행능력을 평가하기 위해 개발되었으며 학생은 주어진 시간 내에 임상 상황과 유사하게 준비된 각 스테이션을 차례로 돌며 부여된 과제를 수행하고 이를 관찰하는 교수와 표준화 환자가 평가하는 방법이다(Yoo & Yoo, 2003). 최근에는 OSCE가 일련의 교육과정을 마치고 졸업 전에 임상종합평가의 목적이나 국가고시의 평가 방법으로 활용될 뿐 아니라 임상실습 경험이 없는 의학과나 간호학과 학생을 대상으로 한 임상입문 과정에서 활용되고 있다(Kim, Lee, Ahn, & Park, 2001; Yoo, 2001; Yoo & Yoo, 2003). 현실적으로 임상실습 전에 단편적인 임상수기의 습득뿐 아니라 임상 상황에서의 적절한 판단과 대처에 따른 문제해결능력 즉, 종합적인 임상수행능력을 획득할 수 있는 교실 수업과 실습실(Lab) 실습이 통합된 교육접근이 필요하다. 따라서 단순히 OSCE 차원이 아닌 임상입문평가로 활용할 수 있는 임상수행능력평가에 대한 연구가 필요하다고 생각된다.

지금까지 국내에서는 OSCE를 포함하여 임상수행능력평가 연구들이 대부분 수행능력에 대한 효과 검증에 초점을 두었으며(Yoo, 2001; Yoo & Yoo, 2003), 수행평가의 신뢰도나 타당도를 검증한 소수의 연구에서도 평가자 간의 일치도나 시험시기나 실험실 간의 평균 비교 정도만 보고되었다(Jung et al., 2005; Lee et al., 2001; Park et al., 2003). 이에 본 연구자는 수행평가를 위한 도구의 검증과 지필고사와의 준거 타당도 등 전반적인 수행평가의 타당성과 신뢰성을 평가하는 간호 연구들이 아직 미흡하다고 판단해 간호교육에 시뮬레이션과 병합된 임상수행평가 체계에 관한 기초 연구의 필요성을 느꼈다.

따라서 본 연구에서는 임상실습 경험이 없는 4년제 간호학과 2학년 학생들을 대상으로 실시한 표준화 환자를 이용한 임상수행능력평가에서 사용할 수행평가 도구 개발과 수행평가 자체의 타당도와 신뢰도를 평가하고자 한다. 수행평가 도구의 타당도는 전문가 구성 타당도로, 신뢰도는 내적일치도를 통해 확인하고, 대상자 수준과 평가 상황을 고려해 변형된 표준화 환자를 이용한 임상수행능력평가의 타당성 검증은 준거타당도로 평가하

며, 신뢰도 검증은 채점자 간 신뢰도로 평가함으로써 임상수행평가제의 타당도와 신뢰도를 평가하고자 한다.

2. 연구의 목적

본 연구의 목적은 크게 CPX를 위한 도구와 CPX의 타당도와 신뢰도 검증을 위한 것으로 구체적 목적은 다음과 같다.

첫째, 임상수행능력평가를 위한 도구의 내용 타당도와 신뢰도를 확인한다.

둘째, 임상수행능력 평가제의 타당도와 신뢰도를 확인한다. 1) 직전 학기 성적과 임상수행능력 평가 점수 간 상관관계를 확인한다. 2) 성인간호학(호흡기계, 1학점) 지필고사 점수와 임상수행능력 평가 점수 간 상관관계를 확인한다. 3) 채점자 간 임상수행능력 점수의 상관관계를 확인한다. 4) 평가실 간 난이도를 평가하기 위해 점수를 비교한다. 5) 피평가자 만족도를 확인한다.

3. 용어정의

1) 임상수행능력평가

CPX는 학생들이 한 환자의 문제에 대해 전체적으로 해결하는 종합적인 능력을 평가하는 방법이다. 즉, 임상수행능력평가를 통해 학생들은 환자의 문제를 해결하기 위해서 수행해야 하는 필요한 과제를 스스로 결정해야 하며, 이러한 과제를 정확하게 수행하고 환자와 효과적으로 상호작용하는 것을 포함한다(Lee & Kim, 1999).

본 연구의 임상수행능력평가는 표준화 환자를 이용해 자체 개발한 3개 임상 사례 중 학생 1인당 무작위 선정된 1개 사례의 과제를 해결하는 수행평가 과정을 의미한다. 본 평가는 전통적인 임상수행능력평가와는 달리 대상자들이 임상 상황에 대한 경험이 없다는 점에서 시험 1주일 전에 시험 내용을 복제한 동영상을 보고 연습을 하였다는 점, 비전문 표준화 환자로 간호사를 사용했다는 점, 그리고 6개 평가실 중에 무작위 할당 방법으로 1개 평가실만을 경험하도록 한 변형된 방식이다.

연구 방법

1. 연구 대상

연구의 대상자는 C대학교 간호대학 2005년 2학년 2학기에 성인간호학을 수강한 70명 전원이었다. 대상자들은 본 연구 적용 직전 학기인 2학년 1학기에 해부학, 생리학, 약리학, 기본간

호학 I 등 전체 총 8과목을 수강하였다. 대부분(92.9%) 여학생 이었고 평균 연령은 20.4세였다. 6개 평가실당 각각 11-12명 학생이 무작위 배정되었는데, 2학년 1학기 전체 성적의 평균 점수는 85.24 ± 5.48 점이었고 각 실별 대상자의 1학기 성적의 순위에 있어 차이는 없었다($\chi^2=8.66$, $p=.123$).

2. 연구 도구

1) 임상수행능력평가 개발팀 구성 및 사례 개발

2005년 9월 1일부터 30일까지 1개월 동안 호흡기계 관련 임상수행능력평가를 위해 개발팀을 구성하였다.

본 개발팀은 사례개발을 위한 자료 수집을 위해 첫째, 질환에 대한 이해와 전형적인 사례에 대한 개념 분석을 위해 문헌고찰을 하였고, 둘째, C대학교병원 호흡기내과 병동, 알러지 내과 교육실 및 흉부외과 병동을 각각 두 번씩 방문하여 담당 의사와 담당 간호사로부터 전형적인 환자에 대해 상담하고 사례를 추천 받았으며, 셋째, 전형적인 환자를 만나 심층 면담하였다. 전공 영역 간호학과 교수 2인과 해당 전공분야 대학원생 2인이 1차로 개발한 사례의 초안을 교육적인 목적을 위해 상황에 적절하게 수정하고 보완하였으며, 최종 개발된 사례는 폐암 환자의 건강사정, 정량식 분무요법 교육 및 폐절제술 환자 간호의 3가지 사례로 확정되었다.

2) 채점 기준표

수행평가도구 개발을 위해 2005년 10월 1일부터 31일까지 1개월 동안 본 연구의 개발팀에 의해 문헌고찰을 토대로 채점 구성문항과 문항별 세부채점기준으로 구성된 예비 채점기준표를 작성한 후 전문가에게 두 부분에 대해 각각 내용 타당도 및 신뢰도 검증을 받았다. 검증을 위한 전문가 집단은 연구팀 소속 1개 대학교병원의 3개 사례 영역별 근무 경력 5년 이상의 병동 또는 외래 간호사, 간호교육 팀장, 호흡기내과 또는 흉부외과 전문의 및 전공의 등을 포함한 20-22명을 선정하였으며, 내용 타당도는 4점 척도를 사용하여 3점 이상 100% 합의를 본 문항만을 채택하였다.

3개 사례의 채점기준표의 구성은 폐암 환자의 건강사정 영역 16문항, 정량식 분무요법 교육 영역 14문항 및 폐절제술 환자 간호 영역이 14문항으로 이루어졌고 상, 중, 하의 3점 척도를 사용하였다. 첫째, 폐암 환자의 건강사정 사례의 평가 영역은 크게 문진, 신체검진 및 총평으로 나뉘었다. 문진은 Bates가 제시한 7개 항목(Bickley, 1999)을 기준으로 하였으며 폐암 환자에 맞게 내원동기, 주증상, 과거력, 가족력, 복용약물, 식욕변화 및

체중변화를 포함하였으며 신체검진은 시진, 촉진, 타진 및 청진 순으로 정확한 방법으로 수행하며 신체검진 결과(청진음 구분)를 이해하는지를 평가하였고 마지막으로 개방형 질문, 태도 및 능숙도를 평가하는 총평을 포함하였다. 둘째, 정량식 분무요법 교육의 평가 영역은 분무요법의 목적과 중요성, 사용법, 부작용 예방법, 분무기 관리법 및 spacer 사용 시 장점에 관한 교육 내용을 포함하였고, 태도, 환자 격려 및 능숙도를 평가하는 총평을 포함하였다. 셋째, 폐절제술 환자 간호의 평가 영역은 의식상태 확인, 산소요법, 기침과 심호흡법 교육 및 중심정맥압 측정을 포함하였고 태도, 환자 격려 및 능숙도를 평가하는 총평을 포함하였다.

3) 피평가자 만족도

내용의 적절성, 난이도, 임상 유용성, 시험 기간, 시설 수준 및 다음에 참여할 의향의 6개 문항의 만족도를 본 연구팀에서 개발하였으며, 본 연구에서 신뢰도는 $\alpha=.74$ 였다.

3. 자료 수집 방법

1) 표준화 환자 및 채점자 훈련

2005년 11월 1일부터 30일까지 1개월 동안 C대학교 석사과정 대학원생 6인으로 구성된 표준화 환자 훈련을 실시하였다. 이 훈련은 1회 3시간 동안 실시하였으며, 표준화 환자의 구성은 모두 임상 경험이 3년 이상의 대학원 석사 과정생이었다. 표준화 환자 훈련은 표준화 환자가 시나리오를 암기한 후 개발팀 구성원들 앞에서 역할극을 하였고 이를 분석과 비판을 통해 여러 상황에 대비할 수 있도록 반복 훈련하였다.

이어 채점자 훈련은 앞서 언급한 표준화 환자 6인과 교수 6인으로 구성된 총 12명의 채점자를 대상으로 1시간씩 3차례 채점자 훈련을 실시하였다. 1차 채점자 훈련은 표준화 환자 1인과 교수 1인이 짝을 이루어 각 사례 별 채점기준표의 구성 문항에 대한 설명과 논의가 있었으며, 2차 훈련은 세부 채점 기준에 대한 합의가 있었고, 마지막 3차 훈련에서는 임상수행능력평가 시험에 관한 모의 훈련과 시험 당일 주의사항에 대한 설명이 이루어졌다.

2) 피평가자 오리엔테이션과 시험준비

피평가자 오리엔테이션은 평가 일주일 전인 2005년 12월 2일에 연구자 1인이 간호학과 2학년 학생을 대상으로 임상수행능력평가에 대한 설명을 하였다. 오리엔테이션 내용은 시험 일시, 시험 장소, 시험 시간, 시험 내용, 수험생 준수사항, 시험 진행 절차, 평가 방법 및 질의응답으로 구성되었다. 1시간 정도의 오

리엔테이션을 마친 후 학생들에게 본 개발팀에서 제작한 3개 사례의 동영상상을 보여주어 실습 경험이 없는 대상자들의 수행평가준비를 도왔다. 동영상 제작은 사전에 병원과 해당 병동에 승인을 얻은 후 교육 목적으로 활용된다는 점을 설명하고 참여 의사가 있는 실제 환자를 소개받아 촬영이 이루어졌다. 실제 환자분께는 시나리오 스크립트를 미리 드리고 외우게 한 후 촬영을 하였다. 3개 사례의 동영상은 1시간 정도 촬영하여 15분정도의 분량으로 편집이 이루어졌다.

3) 임상수행능력평가 시행 절차

2005년 12월 9일 총 3시간동안 실시된 임상수행능력평가에 1개 평가실당 11~12명 학생이 배정되었다. 평가 시작 1시간 전에 학생 1인당 6개의 평가실 중 한곳에만 무작위 실 배정이 이루어지도록 하여 한 학생이 1개 사례에 대한 수행평가만 받도록 했다. 무작위 실 배정에는 난수표가 사용되었고 평가 순서는 각 실로 배정된 학생들이 추첨을 하여 결정하였다. 실 배정과 평가 순서에 따라 학생 이름을 기입하여 각 실에 해당 학생의 채점표를 미리 준비해 두었다. 평가 시작 10분 전에 각 대기실에 2명의 학생을 입실 준비시켰고 15분 간격으로 대기실에서 각 평가실에 1인씩 보냈으며 평가가 끝난 후에는 나머지 대기자와의 접촉을 막기 위해 진행요원의 도움을 받았다.

신호로 호루라기를 불어 시험 시작과 끝을 알려 모든 실의 시작과 끝을 동일하게 유지하였다. 각 평가실의 시험은 13분, 추가시간 및 정리 2분의 총 15분이 소요되었다. 각 실에서는 평가자와 표준화 환자가 이미 준비된 채점 기준표에 준거하여 학생의 임상수행능력 점수를 독립적으로 매겼고 피평가자에게는 두 채점자의 평균점수가 부여되었다. 각 실에서 시험을 마치고 나오는 학생을 대상으로 훈련된 진행보조요원 6인에 의해 임상수행능력 평가에 대한 피평가자 만족도 조사를 실시하였다.

4. 자료 분석 방법

본 연구의 일반적 특성인 전 학기 성적, 2학기 호흡기계 성인 간호학 지필고사 점수, 임상수행능력평가 점수, 평가 도구의 전문가 타당도 점수 및 피평가자 만족도는 평균과 표준편차를 이용하였으며 한개 평가실당 11~12명 피평가자가 할당되어 비모수 분석방법을 활용하여 6개 평가실별 전 학기 성적과 임상수행능력평가 점수의 비교는 Kruskal-Wallis 검정을 사용하였다. 주제가 같은 2개 실의 점수 비교는 Mann-Whitney 검정을 이용하였다. 평가도구의 신뢰도는 Chronbach alpha 값을 구하였으며, 전 학기 성적과 임상수행능력평가 점수 간의 상관

관계, 성인간호학 호흡기계 지필고사 점수와 임상수행능력평가 점수 간의 상관관계 및 채점자 간 일치도는 Pearson Correlation coefficient를 사용하여 구하였다.

5. 연구 제한점

본 연구의 임상수행능력평가는 전체평가 소요시간을 고려하여 학생 1인당 1개실의 수행평가만 경험하도록 변형된 방법을 적용하였으므로 연구 결과를 일반화하는데 제한점이 있다.

연구 결과

1. 임상수행능력평가 도구의 타당도와 신뢰도

폐암 환자의 건강사정 채점기준표의 내용 타당도는 채점 구성 문항과 세부 채점기준의 평균이 각각 $3.71 \pm .41$, $3.74 \pm .41$ 이었으며 내적 일치도는 Cronbach $\alpha = .89$ 였다. 정량식 분무요법 교육 채점기준표의 내용 타당도는 채점 구성문항과 세부 채점기준의 평균이 각각 $3.78 \pm .39$, $3.81 \pm .43$ 이었으며 내적 일치도는 Cronbach $\alpha = .88$ 였다. 폐절제술 환자 간호 채점기준표의 내용 타당도는 채점 구성문항과 세부 채점기준의 평균이 각각 $3.57 \pm .42$, $3.49 \pm .42$ 이었으며 내적 일치도는 Cronbach $\alpha = .93$ 이었다(Table 1).

전문가 타당도 검사 결과 100% 합의를 이루지 못한 문항은 폐암 환자의 건강사정 영역에서 객담증상의 1개 문항, 폐절제술 환자 간호 영역에서 동공반사 확인, 운동기능 사정 및 산소공급의 3개 문항이었으며 정량식 분무요법 교육의 평가 영역은 모든 문항에서 합의가 이루어졌다. 폐암 환자의 건강사정 영역에서 객담증상에 대한 세부적인 질문이 추가되었으며 폐절제술 환자 간호 영역에서 동공반사 확인 문항이 삭제되었고 운동기능사정 문항이 간소화되었으며 산소공급량이 감량되고 산소전달체계가 단순 마스크에서 비강 캐놀라로 변경되었다.

2. 임상수행능력평가제의 타당도와 신뢰도

1) 전 학기 성적과 임상수행능력평가 점수 간의 상관성

피평가자의 2학년 1학기 8개 과목 평균성적과 임상수행능력점수의 상관관계 결과는 Table 2와 같다. 1학기 성적의 평균 점수는 85.0점이었고 임상수행능력평가의 평균 점수는 96.0점이었으며 이들 간에 상관성은 $r = .276$ ($p = .021$)으로 낮은 상관관계를 보였으며 각 실별 상관관계는 $r = .025 - .689$ 까지 분포하였다.

Table 1. Validity of Checklists for the Clinical Performance Examination

(N=70)

Subject	Category	Mean±SD	
		Items	Criteria
Physical examination	Total	3.71±.41	3.74±.41
	History taking	3.74±.38	3.75±.40
	Inspection	3.70±.47	3.70±.46
	Palpation	3.65±.46	3.72±.46
	Percussion	3.60±.50	3.72±.46
	Auscultation	3.67±.47	3.78±.43
Education of MDI use	Total	3.78±.39	3.81±.43
	Purpose of MDI	3.80±.35	3.80±.40
	Method of MDI	3.70±.42	3.80±.42
	Prevention of complication	3.78±.33	3.80±.40
	Management of MDI	3.82±.40	3.80±.40
	Advantage for spacer	3.79±.39	3.82±.45
Post operative care	Total	3.57±.42	3.49±.42
	Check M/S	3.55±.46	3.45±.41
	Oxygen therapy	3.65±.49	3.50±.51
	Education of cough and breathing	3.60±.50	3.53±.50
	Measurement of CVP	3.60±.43	3.49±.47

MDI=metered dose inhaler; M/S=mental status; CVP=central venous pressure.

Table 2. Relation between CPX Score and Average Grade at Previous Semester, and between CPX Score and Average Grade of Paper and Pens Test in Pulmonary System of Adult Nursing

(N=70)

Station	Grade		r	p	Grade		r	p
	Previous semester (Mean±SD)	CPX (Mean±SD)			Paper and pens test (Mean±SD)	CPX (Mean±SD)		
1	86.36±3.75	96.18±1.49	.139	.683	86.86±3.75	96.18±1.49	.291	.385
2	85.58±5.55	95.96±2.23	.325	.302	85.58±5.93	95.96±2.23	.002	.995
3	88.45±4.03	97.32±1.62	.055	.872	85.90±3.87	97.32±1.62	.555	.076
4	82.17±7.54	97.13±3.69	.689	.013	84.75±5.27	97.13±3.69	.545	.067
5	84.67±6.14	94.96±2.74	.208	.298	84.13±6.54	94.96±2.74	.496	.101
6	84.75±3.70	95.17±1.93	.025	.937	86.83±5.33	95.17±1.93	.006	.986
Total	85.24±5.48	96.04±3.02	.276	.021	85.49±5.28	96.04±3.02	.048	.692

CPX=clinical performance examination.

Table 3. Agreement of Scoring between Trained Faculties and Standardized Patients

(N=70)

Station	Grade		r	p
	Trained faculty (Mean±SD)	Standardized patient (Mean±SD)		
1	95.08±2.28	95.25±1.87	.688	.019
2	96.67±1.83	95.25±2.80	.854	.000
3	97.73±.91	96.91±2.34	.883	.000
4	97.25±3.28	97.00±4.57	.758	.004
5	93.92±2.97	96.00±3.30	.529	.077
6	95.25±1.87	95.08±2.28	.745	.005
Total	96.43±3.11	96.17±2.77	.647	.000

2) 호흡기계 성인간호학 지필고사 점수와 임상수행능력평가 점수 간의 상관성

피평가자의 2학기 호흡기계 성인간호학(1학점) 지필고사 점수와 임상수행능력평가 점수 간의 상관관계 결과는 Table 2와

같다. 피평가자의 지필고사 평균 점수는 85.0점이었고 임상수행능력과의 상관성은 $r=.048$ ($p=.692$)로 상관관계가 없는 것으로 나타났으며, 각 실별 상관관계 계수는 $r=.002-.555$ 로 다양하게 분포하였다.

3) 채점자 간의 임상수행능력평가 점수의 상관성(채점자 간 일치도)

각 실별 피평가자에 대한 평가는 훈련된 교수와 표준화 환자가 동시에 채점하여 두 채점자의 평균점수가 부여되었는데 전체 채점자 간 일치도는 $r=.647$ ($p<.001$)로 중정도 이상의 상관관계가 있었다. 채점자 간의 가장 높은 일치도는 3실 $r=.883$ ($p<.001$) 이었고 가장 낮은 일치도는 5실 $r=.529$ ($p=.047$)이었다(Table 3).

4) 평가실 간 점수 비교(난이도)

6개 평가실 중 1, 2실에서는 폐암 환자의 건강사정에 관한 임

Table 4. Score of Clinical Performance Examination according to Stations

(N=70)

Evaluation area	Station	n	Mean±SD	Mean rank	Z	p	χ^2	p
Physical examination	1	11	96.18±1.49	34.14	-.246	.833	14.987	.010
	2	12	95.96±2.23	32.71				
Education of MDI use	3	11	97.32±1.62	52.18	-.159	.874		
	4	12	97.13±3.69	43.46				
Postoperative care	5	12	94.96±2.74	26.25	-.029	.977		
	6	12	95.17±1.93	25.54				

MDI=metered dose inhaler.

Table 5. Student Satisfaction

(N=70)

Satisfaction	Minimum	Maximum	Mean±SD
Content appropriateness	3	5	4.48±.56
Difficulty level	2	5	4.03±.62
Clinical usefulness	3	5	4.48±.56
Testing duration	3	5	4.23±.52
Facility level	3	5	4.12±.72
Willingness to participate again	1	5	3.48±1.07
Total	2.67	5	4.14±.46

상수행능력평가가 이루어졌고 3, 4실에서는 정량식 분무요법의 교육, 5, 6실에서는 폐절제술 후 간호에 관한 임상수행능력평가가 이루어졌다. 임상수행능력평가 점수는 3실이 97.32점으로 가장 높았고 5실이 94.96점으로 가장 낮았으며 각 실별 임상수행능력평가 점수 순위에 차이가 있었다($\chi^2=14.99$; $p=.010$). 그러나 주제가 같은 1, 2실, 3, 4실 및 5, 6실의 2개 실간의 점수의 순위 차는 없었다($Z=-.249$, $p=.833$; $Z=-.159$, $p=.874$; $Z=-.029$, $p=.977$) (Table 4).

5) 만족도

피평가자 만족도는 전체 평균은 4.14점이었고, 내용의 적절성과 임상 유용성이 4.48점으로 가장 높았고 다음에도 수행능력평가에 참여할 의향은 3.48점으로 가장 낮았다(Table 5).

논 의

지금까지 대상자 수준과 평가 상황을 고려해 임상실습을 앞둔 2학년 2학기 간호대 학생을 대상으로 실제임상 상황을 시뮬레이션하여 개발한 임상수행능력평가 방법에 대한 타당도와 신뢰도를 검증하였다.

기존의 여러 연구에서 수행능력평가의 신뢰성과 타당성을 확보하기 위한 전략으로 명확한 채점 기준의 마련, 사전에 충분한 채점자 훈련 및 임상수행능력을 평가하는 총 시간을 줄이는 방법 등을 제시하고 있다(De Champlain, Margolis, King, & Klass, 1997; Tamblyn, Klass, Schanbl, & Kopelow, 1990). 먼저

명확한 채점 기준을 마련하는 것은 임상수행능력평가에 사용되는 채점표의 타당성과 신뢰성을 확보하는 것이다.

국내 수행능력평가 연구의 대부분이 CPX로 이루어지지 않아 OSCE 연구를 포함해 분석해 보면, 대부분 평가 도구의 타당도나 신뢰도를 보고하지 않았다. 간호학 연구 중 평가 도구의 타당도를 보고한 연구(Oh et al., 2006)에서는 전문가를 대상으로 5점 척도를 사용하여 문항의 기술이 구체적이고 분명했는지, 절차가 맞는지, 모든 절차를 포함했는지 등 5개 문항을 사용하여 타당도를 살폈으나 전문가 의견 수렴 후 채점표의 변경 사항에 대한 논의는 이루어지지 않았다.

본 연구에서는 평가 각 문항별로 각 수행능력을 평가하는데 적절하였는지를 전문가를 이용하여 구성타당도를 확인한 후 전문가의 100% 합의하에 문항을 채택하였고 그렇지 못한 문항에 대해서는 수정·보완이 이루어져 타당성 있는 도구로 생각된다. 본 연구에서 평가 도구의 신뢰도는 내적일치도로 확인하였는데, 알파 값이 .88에서 .93까지 분포를 보였다. 탐색적 연구 분야에서 알파 값이 .60 이상이면 충분하기 때문에, 본 연구의 평가 도구는 신뢰성있는 평가 도구로 해석할 수 있겠다.

다음으로, 사전에 충분한 채점자 훈련과 적절한 평가 시간의 배정을 통한 채점자 간 일치도를 높이는 방법을 들 수 있겠다. 표준화 환자 훈련에 필요한 최소 시간에 대한 보고(Jung et al., 2005)는 있으나 채점자 훈련에 요구되는 시간이 어느 정도인지에 대한 논의가 없어 본 연구와의 비교는 어렵다. 그러나 채점자 훈련은 채점자 간 일치도를 높이기 위해 실시되므로 채점자 간 일치 정도를 비교함으로써 수행능력평가의 신뢰성을 고려할 수 있겠다. 본 연구에서 채점자 간 일치도는 모든 실에서 중등도 이상의 상관성($r=.88-.53$)을 보였고 한 실을 제외하고는 기존 연구($r=.60-.65$)에 비해서도(Park et al., 2003) 일치도가 높았다. 그러나 국외 연구 결과는 국내 연구에 비해 훈련된 교수자와 전문 표준화 환자 간의 상관관계가 $r=.70$ 이상으로 높게 보고(Amano et al., 2004)되고 있어 채점자 간 일치도를 높일 수 있도록 더 많은 시간을 할애해서 채점자 훈련이 필요하다.

채점자 간 일치도를 높일 수 있는 또 다른 방법으로 고려할 수

있는 것은 총 평가 시간을 줄이는 방법과 전문 표준화 환자를 활용하는 방법이다. 총 평가 시간을 줄이는 방법은 재학생 수가 정해져 있기 때문에 어려움이 있다. 기존의 연구에서는 평가 시간을 줄이기 위해 2-3일에 걸쳐 평가하는 방법을 채택하였는데 평가 날짜별로 점수 차이를 보였다(Kim et al., 2001; Lee et al., 2001). 이런 방법은 평가자의 피로를 줄여 평가자 간 일치도는 높일지 모르지만 시험 날짜가 다르다 보니 늦게 보는 것이 유리하게 나타나 시험의 형평성에 문제가 제기되었다. 또한 과제가 많아지거나 채점해야 할 학생들이 너무 많아 발생할 수 있는 채점자의 피로가 채점의 일관성에 영향을 주는 주요 요인이기 때문에(Klein & Stecher, 1998) 학생에 의한 자가평가, 동료에 의한 평가 등도 수행평가의 신뢰도를 확보하기 위한 방안으로 활용될 필요가 있겠다.

본 연구에서는 무작위 배정으로 대상자가 1개 실에서만 시험을 보는 방법을 채택하였는데 이 역시 6개 실 간에 임상수행능력점수의 차이를 보여 각 실의 난이도 조율이 필요함을 시사하였다. 따라서 보다 많은 평가 인력의 지원으로 피평가자들이 6개 실 모두를 경험할 수 있는 전형적인 평가방법을 권장하며 선택적인 실 경험 시에는 표준전환 점수 방식의 채택과 같은 보다 공정한 점수체계장치가 요구된다고 생각된다.

채점자 간 일치도를 높일 수 있는 방법으로 전문 표준화 환자의 활용이 제안되고 있음을 감안할 때, 본 연구에서 대학원생을 표준화 환자로 평가자 훈련을 하였지만 전문 표준화 환자에 비해 신뢰도 확보가 더 미흡할 수 있음을 감안해야 한다고 본다. 국외의 경우 표준화 환자는 하나의 전문 직업인으로서 연합(pool)을 만들어 여러 대학이 공동 이용하는 경우가 많다(Reznick et al., 1993). 이렇게 표준화 환자가 학생들을 평가함으로써 실습을 전담하는 교수들의 시간 및 비용 절감 효과는 있겠으나 전문 표준화 환자 1인을 육성하는데 전문 트레이너를 통해 최소 15시간의 교육이 필요하다(Jung et al., 2005)는 것을 감안할 때 평가 비용적인 면의 고려가 필요하다고 생각된다. Yoo와 Yoo(2003)의 연구에서도 일반인을 표준화 환자로 선발하고 훈련하는 어려움과 비용적인 면을 고려하여 3학년 간호대 학생이 2학년 간호대 학생을 평가할 수 있도록 고안하여 학생평가에 이용한 바 있고 고학년 학생을 저학년의 임상수행능력평가에 활용하는 방안의 교육적 효과에 대한 규명이 필요함을 제시하였다. 따라서 현실적으로 비전문 표준화 환자 활용이 손쉬운 방법 이긴 하나 채점자 간 일치도를 높일 수 있는 장치를 둔 상태에서 활용을 고려할 필요가 있다고 생각된다.

본 연구에서는 임상수행능력평가가 이루어진 2학년 2학기 성 인간호화 호흡기계 지필고사 점수와 임상수행능력평가가 실시되

기 전인 2학년 1학기 8개 과목 평균 점수를 임상수행능력 점수의 준거로 삼아 준거 타당도를 확인하였는데, 각각 상관관계가 낮거나 거의 없는 것으로 나타났다. 낮은 상관성을 설명하는데 시험시기, 피시험자의 내적동기, 시험의 난이도 등의 여러 가지 변수가 있어 한 가지 요인을 단언하기는 어렵고, 제한된 수의 연구만으로 이를 주장하기는 역부족이나 간호 교육의 목표가 수행이나 실행하는 방법을 시범하거나 보여줄 수 있어야 하는데 지필고사만으로 이 목표를 달성하기에는 어려움이 있다고 생각된다. 지필고사는 지식적인 측면의 인지정도만을 평가하는데 그치므로 실무학문으로써 간호 교육의 목표 달성을 하기에 평가방법의 개선이 필요하다. 임상실습 전 학생과 임상실습 후 학생에서 시행한 OSCE 평가에 관한 연구(Park, Han, Park, & Oh, 2004)에서도 임상실습 전 학생에서는 문진, 이학적 검사, 처치술기 및 교육술기의 전체 영역에서 수행력 평가 점수와 전 과목 성적과의 상관성이 낮은 것으로 나타나 본 연구 결과와 일치하였다. 간호학이나 의학교육은 아니지만, Nam(1995)의 과학실기 평가의 타당성 연구에서 과학실기평가와 과학학력 성취도검사 간의 낮은 상관관계를 보인 예가 있었고 이러한 낮은 상관관계는 과학실기 평가의 준거 관련 타당성을 간접적으로 증명하였다고 제시하였다(Nam, 1995). 그러나 본 연구의 임상수행능력평가가 변형된 형태로 이루어져 변별력에 문제가 제기된다는 점과 학생 1인당 1개 실만을 경험한 점수이므로 준거가 되는 점수들과의 낮은 상관관계가 지필고사만의 문제로 해석하기는 어렵고 추후 임상수행능력평가 점수와 지필고사 점수의 상관성에 관한 지속적인 연구가 필요하며, 다양한 평가방식의 도입은 단일 평가방식의 단점을 보완할 수 있는 장치로 활용될 수 있다고 생각된다. 따라서 간호 교육의 평가를 지필고사에만 의존하는 것은 지양되어야 하고 수행력을 평가할 수 있는 방법의 모색이 필요하며 본 연구는 이런 점에서 교육적 의의가 있다고 생각된다. 향후, 간호학 학습목표를 수립할 때 지필평가로 목표가 달성될 수 있는 것과 수행능력평가로 목표달성이 가능한 하위 목표의 구성이 이루어져야 할 것이며, 이론 강의, 지필고사, 임상수행능력평가의 순서로 통합 교과 운영을 통해 최종적이며 궁극적인 교과 학습 목표는 임상수행능력평가를 통해 달성할 수 있도록 학습방법의 개선이 필요하다.

대상자의 임상수행능력평가 점수는 평균 95점 이상으로 높은 점수를 보였다. 일반적으로 임상 입문 전 실습이나 평가의 목적은 임상실습 전에 질환의 전형적이고 표준화된 증례를 통해 임상상황에 친숙해지면서 자신감을 얻는데 있다. 본 연구의 대상자 만족도에서 임상 유용성 문항이 가장 높은 점수를 보여 교육적 목적은 어느 정도 달성되었다고 생각된다. 문헌에서 OSCE

와 CPX를 정확하게 구별하여 사용하고 있지 않은 점을 고려하여 국내외 OSCE를 포함한 임상수행력을 평가한 연구를 분석해 본 결과 본 연구에서 사용한 주제인 문진, 신체사정, 면담기술 및 환자 교육 등의 학습에 본 연구와 같은 새로운 평가방법이 효과적인 것으로 나타났다(Mukohara et al., 2005; Park, Kim, Kim, Choi, & Kim, 1998; Roberts & Brown, 1990; Stillman et al., 1990; Suh, Lee, Kwun, & Kang, 1998; Yoo, 2001; Yoo & Yoo, 2003). 그러나 본 연구에서 피평가자가 고른 점수 분포를 보였고, 이로 인해 나타난 변별력 측면의 문제를 보완하기 위해 지필고사나 문제중심학습을 통한 튜터평가, 자가평가 및 동료평가 등의 병용이 필요하다고 생각된다.

Burch 등(2005)의 연구에서 서부 아프리카 의과대학을 3주 전에 졸업한 58명의 졸업생을 대상으로 정맥천자술, 근육주사, 골반검사, 방광카테터 삽입, 기도 삽관술, 처방전 기술 및 응급 소생술 능력을 OSCE를 이용하여 평가한 결과 58명 모두 기준 점수 미만의 점수를 얻은 것으로 나타나 졸업 전에 알아야 할 필수 실무능력에 대한 명확한 기술과 평가가 필요함을 제시하였다. 이 연구에서는 비단 연구결과가 아프리카 지역에 국한된 것이 아니며 극소수 발표된 다른 국가에서의 결과와 일치한다고 주장하였다. 국내 연구(Lee et al., 2001)에서도 OSCE에서 평가된 문제는 '임상실습기간 중 내가 배웠던 내용이다'와 '의사로서 반드시 습득해야 할 내용이다'라는 질문에 배웠다는 응답에 비해 필수 습득내용이라는 응답의 점수가 유의하게 높았다는 점도 이를 지지한 결과라고 사료된다. 이는 실무교육에서 전통적인 교육이나 평가 방법으로는 도달해야 하는 실무능력을 졸업 전에 달성하기 어렵다는 것을 극명하게 제시한 연구라고 생각된다. 졸업생들에게 요구되는 필수 실무능력 조차도 갖추지 못했다는 임상현장의 비판은 어제 오늘일이 아니다. 따라서 대학 본연의 임무를 충실히 달성한다는 의미에서도 교육 방법이나 평가 방법의 변화는 중요한 일이라고 생각된다.

결론 및 제언

본 연구는 임상실습 경험이 없는 간호대학 2학년 2학기 학생들을 대상으로 표준화 환자를 이용하여 임상 상황을 포괄적으로 제시하여 평가할 수 있는 변형된 형태의 임상수행능력평가의 타당도와 신뢰도를 검증하고자 하였다.

대상자는 C대학교 간호대학 2005년 2학기 2학기에 성인간호학을 수강한 70명 전원이었고, 6개실로 무작위 배정되었으며, 각 실의 2학년 1학기 8개 과목이 평균 성적은 차이가 없었다. 연구 도구로 임상수행능력평가 개발팀에 의해 개발된 임상

사례와 채점 기준표를 사용하였다. 임상수행능력평가 절차는 시험 시작 1시간 전에 무작위 실 배정이 이루어졌다. 모든 실의 시험은 13분, 추가시간 및 정리 2분의 총 15분이 소요되었다. 각 실에서는 평가자와 표준화 환자가 이미 준비된 채점 기준표에 준거하여 수험생의 임상수행능력 점수를 독립적으로 매겼고 수험생에게는 두 채점자의 평균점수가 부여되었다. 각 실에서 시험을 마치고 나오는 수험생을 대상으로 훈련된 진행보조요원 6인에 의해 임상수행능력평가에 대한 대상자 만족도 조사를 실시하였다. 1, 2실에서는 폐암 환자의 건강사정, 3, 4실에서는 정량식 분무요법의 교육, 5, 6실에서는 폐절제술 후 간호에 관한 임상수행능력평가가 이루어졌다.

본 연구 결과는 다음과 같다.

첫째, 임상수행능력평가를 위한 도구의 내용 타당도는 4점 척도에서 모두 3.5점 이상이었고 내적 일치도는 $\alpha=.88-.93$ 이었다.

둘째, 직전 학기 평균 성적과 임상수행능력 점수 간의 상관관계는 $r=.276$ ($p=.021$)로 낮았다.

셋째, 호흡기계 성인간호학 지필고사 점수와 임상수행능력 점수 간의 상관관계는 $r=.048$ ($p=.692$)로 상관관계가 없었다.

넷째, 채점자 간 일치도는 $r=.647$ ($p=.000$)로 중정도 이상의 상관관계가 있었다.

다섯째, 평가실 간 임상수행능력 평가 점수는 3실이 97.32점으로 가장 높았고 5실이 94.96점으로 가장 낮았으며 각 실별 점수의 순위에 차이가 있었다($\chi^2=14.99$; $p=.010$).

여섯째, 피평가자 만족도 평균은 4.14점이었다.

이상의 결과를 바탕으로 다음과 같은 제언을 하고자 한다.

첫째, 임상입문 평가로 활용할 수 있는 임상수행능력평가에 대한 지속적인 실증 연구가 필요하다.

둘째, 평가자 간 일치도를 높일 수 있는 방법의 모색이 필요하다.

셋째, 변형된 형태의 임상 입문 평가보다는 많은 평가 인력의 지원으로 대상자들이 모든 실을 경험할 수 있는 전형적인 평가방법을 권장하며 선택적인 실 경험 시에는 표준전환 점수 방식의 채택과 같은 보다 공정한 점수 체계 장치가 필요하다.

REFERENCES

- Amona, H., Sano, T., Gotoh, K., Kakuta, S., Suganuma, T., Kimura, Y., et al. (2004). Strategies for training standardized patient instructors for a competency exam. *J Dent Educ*, 68, 1104-1111.
- Bickley, L. (1999). *Bate's guide to physical examination & history taking* (7th ed.). Philadelphia: Lippincott.
- Burch, V. C., Nash, R. C., Zabow, T., Gibbs, T., Aubin, L., Jacobs,

- B., et al. (2005). A structured assessment of newly qualified medical graduates. *Med Educ*, 39, 723-731.
- De Champlain, A. F., Margolis, M. J., King, A., & Klass, D. J. (1997). Standardized patients' accuracy in recording examinees' behaviors using checklists. *Acad Med*, 72(10 Suppl 1), S85-S87.
- Jung, K. H., Choi, E. H., Bae, J. D., Na, D. Y., Son, I. K., Ko, C. W., et al. (2005). Comparison of professional standardized and amateur standardized patients in OSCE. *Korean J Med Educ*, 17(1), 97-105.
- Kim, B. S., Lee, Y. M., Ahn, D. S., & Park, J. Y. (2001). Evaluation of introduction to clinical medicine by objective structured clinical examination. *Korean J Med Educ*, 13, 289-298.
- Klein, S. P., & Stecher, B. M. (1998). Analytic versus holistic scoring of science performance tasks. *App Meas Educ*, 11, 121-137.
- Lee, Y. M., Hwang, K., Baik, S. H., Hong, M. H., Choi, Y. S., & Kim, S. H. (2001). An experience of OSCE for introducing clinical performance assessment to Korean medical licensure examination. *Korean J Med Educ*, 13(1), 47-57.
- Lee, M. S., & Kim, S. (1999). Clinical performance evaluation in nursing. *Korean J Nurs Query*, 8(1), 39-59.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Acad Med*, 65(9 Suppl), S63-S67.
- Mukohara, K., Kitamura, K., Wakabayashi, H., Abe, K., Sato, J., & Ban, N. (2005). Evaluation of a communication skills seminar for students in a Japanese medical school: a non-randomized controlled study. *BMC Med Educ*, 4, 24-29.
- Nam, M. H. (1995). A validity study of performance assessment: a comparison between science hands-on assessment, experimental report assessment, computer simulation assessment. *J Educ Eval*, 8, 115-143.
- Oh, H., Jun, S., Kim, E., Hur, M., Song, Y., Yi, G., et al. (2006). Application of OSCE in clinical nursing education: development and testing validity of checklist. *Annu Bull Bum-Suk Acad Scholarsh Found*, 10, 237-249.
- Park, H., Kim, D., Kim, D. A., Choi, H., & Kim, K. (1998). An OSCE assessment of fourth-year students as a comprehensive evaluation of medical skills before graduation. *Korean J Med Educ*, 6(1), 23-35.
- Park, H., Lee, J., Hwang, H., Lee J., Choi, Y., Kim, H., et al. (2003). The agreement of checklist recordings between faculties and standardized patients in an objective structured clinical examination (OSCE). *Korean J Med Educ*, 15, 141-150.
- Park, H., Han, J., Park, M., & Oh, J. (2004). Comparison of results from objective structured clinical examinations for medical students performed before and after clinical clerkship. *Korean J Med Educ*, 16(1), 63-71.
- Reznick, R. K., Sydney, S., Baumber, J., Cohen, R., Rothman, A., Blackmore, D., et al. (1993). Guideline for estimating the real cost of an objective structured clinical examination. *Acad Med*, 68, 513-517.
- Roberts, J., & Brown, B. (1990). Testing the OSCE: a reliable measurement of clinical nursing skills. *Can J Nurs Res*, 22(1), 51-59.
- Stillman, P. L., Regan, M. B., Swanson, D. B., Case, S., McCahan, J., Feinblatt, J., et al. (1990). An assessment of clinical skills of fourth year students at four new england medical schools. *Acad Med*, 65, 320-326.
- Suh, B. Y., Lee, D. J., Kwun, K. B., & Kang, P. S. (1998). Trial on objective structured clinical examination. *Korean J Med Educ*, 10, 363-381.
- Tamblyn, R. M., Klass, D. K., Schanbl, G. K., Kopelow, M. L. (1990). Factors associated with the accuracy of standardized patient presentation. *Acad Med*, 65(9 Suppl), S55-S56.
- Vu, N. V., Barrows, H. S., March, M. L., Verhulst, S. J., Colliver, J. A., & Travis, T. (1992). Six years of comprehensive, clinical, performance-based assessment using standardized patients at the Southern Illinois University School of Medicine. *Acad Med*, 67(1), 42-50.
- Yoo, M. (2001). The effectiveness of standardized patient managed instruction for a fundamental nursing course. *J Korean Acad Soc Nurs Educ*, 7(1), 94-109.
- Yoo, M. S., & Yoo, I. Y. (2003). Effects of OSCE method on performance of clinical skills of students in fundamentals of nursing course. *J Korean Acad Nurs*, 33, 228-235.