

표준화 환자를 활용한 의학교육

Medical Education Using Standardized Patients

박훈기

한양대학교 의학교육학교실

Hoon-Ki Park, M.D., Ph.D.

Department of Medical Education, Hanyang University
College of Medicine, Seoul, Korea

책임저자 주소: 133-791, 서울시 성동구 왕십리로 222

한양대학교 의과대학 의학교육학교실

Tel: 02-2290-8740, Fax: 02-2281-7279

E-mail: hoonkp@hanyang.ac.kr

투고일: 2011년 12월 20일, 심사일: 2011년 12월 23일, 게재확정일: 2012년 1월 30일

Abstract

This article is intended to review up to date knowledge about standardized patients (SPs) and SP-based objective structured clinical examinations (OSCE) in medical education. SPs have become almost indispensable in the education and training of health care professionals. The utilization of SPs provides an invaluable contribution to the creation of a safe, yet realistic, learner-centered environment. While SPs provide a realistic experience of patient interaction that can be evaluated by a third party observer, they can also be trained to assess the student's performance and provide feedback to the student with consistency. Standardization of the performance, recording student behaviors by well-developed checklists, and/or receiving feedback from the SPs are important parameters of the educational experience and provide a meaningful evaluation of student performance. In assessment, SPs are used most commonly in the context of formal examinations, often referred to as

OSCEs. An OSCE should be prepared step by step according to principles of appropriate logistics. The validity of the OSCE assessment is increased by matching educational content with the balanced blue print matrix for comprehensive performance examination. The thorough training of all physician/SP raters involved during the whole observation time in which students interact with SPs is critical to the reliable OSCE. SP-based OSCE has its own limitations, so it is to be judged as one among many evaluation methods for competency and/or performance.

Key Words: Patient Simulation; Medical Education; Undergraduate; Competency-Based Education

서론

표준화 환자는 1963년 미국 남가주 대학의 신경과 의사이며 의학교육학자인 Barrows의 아이디어로 탄생하였다 [1]. 당시에는 표준화 환자의 개념이 없이 다만 실제 환자와 비슷한 연기를 하는 교육 모듈을 개발한다는 의미가 강했으나 그 후 연기의 표준화라는 개념이 추가되면서 표준화 환자라는 용어가 생겨났다. 임상 능력(competency)을 측정하는 한 가지 방법이 바로 객관구조화진료시험(objective structured clinical examination, OSCE)이다. Harden 등이 1975년 처음으로 지식, 태도, 수기 등을 동시에 측정할 수 있는 시험의 형태로 소개한 이후 OSCE는 의학교육에서 평가와 교육 방법으로서 발전해왔다[2]. 표준화 환자의 진료만으로 구성된 OSCE를 임상진료시험/임상수행평가(clinical practice examination or clinical performance examination, CPX)라고 칭한다. 표준화 환자는 이제 의학

교육뿐 아니라 간호학, 치의학, 물리치료[3], 영양학[4], 약학[5] 등 의료인 및 관련직업인 교육에서 다양하게 활용되고 있다. OSCE는 평가뿐 아니라 피드백을 제공하여 면담, 환자 상담 등의 교육을 위해서도 활용되고 있다[6]. 우리나라에서는 의사국가시험에서 실기시험이 도입되면서 표준화 환자를 활용한 교육과 평가가 확대되고 있다.

본론

1. 표준화 환자의 개념

표준화 환자는 실제 환자와 똑같이 훈련된 일반 사람을 의미하며 진료를 하는 사람이 달라져도 일관성 있는 연기와 임무를 수행할 수 있어야 한다. 모의 환자(simulated patient)와 표준화 환자(standardized patient)는 서로 구별이 되어 사용되어야 한다. 모의 환자는 실제 환자의 증상이나 징후를 표현하는 것(simulation)에 초점을 맞추고 있는 용어이다. 표준화 환자는 시뮬레이션 과정 자체가 일관성(consistency)이 있다는 것을 강조하는 용어이다. 따라서 표준화 환자는 대면하는 학생들 간에 차이가 없이 일관성 있는 표현을 해야 하고 표준화 환자들 간의 연기 차이도 없어야 한다[7]. 하지만 모의 환자는 이런 대상의 차이에 따라 조금의 연기 차이는 허용된다고 할 수 있다. 표준화 환자는 모의 환자라고 할 수 있지만 모의 환자가 곧 표준화 환자라고 간주할 수는 없다. 예를 들면 우리나라에서 신체 진찰이나 의사소통술 교육에서 환자의 이야기만 따오고 실제로 연기 표준화나 피드백 표준화가 되지 않은 환자는 표준화 환자라고 부르기보다는 모의 환자라고 명명되어야 한다[8]. 지역마다 용어의 차이는 있어서 유럽이나 아시아 쪽에서는 표준화 환자라는 용어 보다는 모의 환자라는 용어를 더 흔하게 사용하고, 미국 쪽에서는 모의 환자나 진정한 의미의 표준화의 의미 구분 없이 표준화 환자라고 용어가 통용되는 경향이 있다[9]. 표준화 환자는 일반적으로 본래 자신은 질병을 갖고 있지 않지만 특정한 환자 사례를 일관성 있게 연기하도록 훈련된 사람을 말한다. 반면 환자로서 소견을 직접 갖고 있으면서 교육에 투입된 사람은 실제 환자라고 분류하는 것이 바람직하다. 표준화 환자의 의미에서 강조되는 표준화의 영역에는 연기, 평가, 피드

백의 세 가지가 있고 교육 목적에 따라 이 모든 기능의 일관성을 요구될 수도 있으나 경우에 따라서는 연기와 다른 어느 하나를 합쳐서 표준화 기준으로 요구되기도 한다. 환자 강사(patient instructor)는 환자의 역할을 하면서 학생들에게 피드백을 제공하는 사람을 일컫는 말인데 반드시 모의 환자일 필요는 없고 어떤 경우에는 의사나 교수가 그 역할을 담당하기도 한다. 이 밖에도 부인과 진찰을 가르칠 수 있는 Gynecological Teaching Associate (GTA) [10], 남성의 생식기진찰을 교육 할 수 있는 Male Urological Teaching Associate (MUTA) [11] 등의 특정 용어가 있다.

2. 표준화 환자의 장점 및 단점

표준화 환자는 실제 환자와 비교하여 여러 가지 장점을 갖고 있다(Table 1). 표준화 환자는 실제 환자와 달리 교육적인 목적에 따라 시기에 구애 받지 않고 이용이 가능하다. 표준화 환자는 실제 환자가 한꺼번에는 나타내기가 힘든 다양한 임상 상황을 만들어 훈련이 가능하게 한다. 표준화 환자를 개발하면 복잡한 환자의 사례를 질병의 진행 단계나 진료의 스펙트럼에 따라 조정이 가능하며 초진과 재진을 묶어서 사례로 연결하는 것도 가능하다. 표준화 환자는 실제 환자로서는 접근하기 어려운 교육적인 상황을 부담 없이 연출할 수 있다. 예를 들면, 말기 환자에게 나쁜 소식을 전하는 사례를 만들어 학생들에게 안전하게 정보 공유에 대한 교육을 특별한 상황에서 시행할 수 있다. 표준화 환자는 학습자의 수준에 따라 사례의 수준을 조절하여 학습자 중심의 안전한 교육 모듈을 개발 할 수 있게 한다[12]. 표준화 환자를 활용하는 수업에서는 원한다면 얼마든지 반복

Table 1. Educational Advantages of Standardized Patients Over Real Patients

Available at any time
Available in non-clinical settings
A standardized patient problem
Provide clinical training for the beginning learners
Eliminate the risks of using an actual patients
Provide opportunity for work with emergent conditions
Offer practice with difficult patients or sensitive situations
Provide direct feedback about interpersonal skills
Offer a flexible educational tool
Modulate pressures of time and urgency
Time can be telescoped

학습을 할 수 있다. 실제 환자와 달리 표준화 환자는 특정한 피드백을 제공할 수 있다. 의사에게 자문 기술을 교육시키는데 강의보다 표준화 환자교육이 더 효과적이라고 보고하고 있다. 표준화 환자를 활용한 교육은 의료인 뿐 아니라 학생들에게 수용성이 높고 아울러 역할극에 비하여 훨씬 효과가 높은 것으로 평가 받고 있다[13].

표준화 환자를 교육에서 활용할 때 가장 문제가 되는 점은 비용의 문제이다. 표준화 환자를 훈련시키고 질 관리를 하기 위해서는 표준화 환자 교육자가 필요하고 표준화 환자를 활용하는 교육에서는 표준화 환자에게 적절한 보상을 하는 것이 원칙이다. 우리나라의 경우 표준화 환자로 구성된 8 사례의 종합 시험을 운영하는데 소요되는 비용은 학생 1인당 10만 원 정도로 추산되었다. 한 명의 표준화 환자를 대상으로 여러 명의 학생들이 진료 실습을 하고 이들로부터 피드백을 받는 형태의 표준화 환자 교육 프로그램은 비용이 적게 든다. 하지만 평가와 피드백을 동시에 일대일로 진행하는 표준화 환자 활용 교육에서는 임상진료시험에서의와 비슷한 규모의 비용이 들어간다. 표준화 환자의 두 번째 단점은 실제 환자가 아닌 훈련된 환자라는 데서 비롯된다. 물론 훈련이 제대로 된 표준화 환자라면 진짜 환자와 구별을 할 수 없어야 한다. 한 연구에 의하면 대상 의사 중 0-18% 정도만 표준화 환자를 실제 환자라고 인식하였다고 한다[14]. 연구 대상 의사에게 동의서를 받고 난 후, 실제 환자처럼 표준화 환자가 투입되는 기간이 길어질수록, 의료보험카드처럼 실제 환자가 지참하는 것들을 추가시키면 시킬수록 의사는 표준화 환자를 실제 환자로 받아들이는 경향이 강했다[14].

3. 표준화 환자의 임상적 활용

1) 교육 목적(teaching)의 표준화 환자 활용

표준화 환자는 면담 기술 훈련(면담 시작하기, 정보 얻기/병력 청취, 정보 공유/환자 교육, 면담 종결하기, 의사소통능력)을 위해 사용될 수 있다. 좀 더 복잡한 면담의 기술 향상을 위해서 표준화 환자를 활용할 수 있다. 예를 들면 의료 과오에 대한 환자 상담[15], 성 관련 병력 청취나 에이즈 환자 상담[16], 가정 폭력 상담[17] 등의 면담 기법 교육에서 표준화 환자를 활용하고 있다. 이 때 표준화 환자는 실제 진

료에서 만날 수 있는 환자의 병력과 신체 진찰 소견을 연기하고 학생은 이 표준화 환자를 대상으로 실제 환자를 만났을 때와 똑 같이 현실감 있게 의사의 역할을 수행한다. 수행 능력과 보완점 및 장점에 대한 피드백은 직접 관찰자로 참여하는 교수나 전문 강사가 시행할 수도 있고, 표준화 환자가 이러한 피드백을 서면 혹은 직접 제공할 수 있다. 단순한 병력을 청취하는 능력을 훈련시키는 교육은 저학년에서 이루어지고 나쁜 소식 전하기처럼 난이도가 높은 면담술 교육은 고학년에서 이루어질 수 있다. 이러한 표준화 환자를 활용한 면담이나 상담 교육은 알코올 중독 환자처럼 학생들의 면담 자체를 거부할 수 있는 상황도 교육이 가능하게 한다[18]. 실제 환자와 비교하여 표준화 환자를 활용한 교육은 학생들의 의사소통술 훈련이나 병력 청취 능력, 신체진찰, 환자 관리 능력 향상에는 차이가 없거나 혹은 좀 더 나은 것으로 보고하고 있다[19]. 상황의 현실감 제공 혹은 난이도 높은 수준의 연기는 실제 환자가 더 잘 할 수 있다. 하지만 안전한 진료 교육 환경, 피드백 제공에서는 표준화 환자가 실제 환자보다 더 나은 능력을 발휘하였다.

표준화 환자는 일반적인 기본 신체진찰 능력 향상 교육에도 활용할 수 있다. 정상인이라도 정확한 신체진찰 절차나 요령을 교육하는데 있어서 환자 역할을 해주면 그만큼 교육의 효과는 더 커진다. 신체 진찰 소견의 시뮬레이션에는 한계가 있긴 하지만 창의성을 가미한다면 많은 소견을 실제 환자와 같이 표준화 환자가 표현할 수 있다[20]. 대개 신경 소견은 시뮬레이션이 쉬운 반면, 신체의 변형을 가져오게 하는 황달을 보여주는 공막 소견이나 혹은 심잡음은 쉽게 시뮬레이션할 수 없다. 기흉 환자의 청진 소견은 비교적 쉽게 시뮬레이션이 가능하지만 타진 소견을 동시에 시뮬레이션하기는 힘들다. 신체 소견의 시뮬레이션이 실제 환자와 아주 똑 같아지는 것이 표준화 환자 훈련의 목표이지만 교육의 목적 상 부분적인 표현만 가능하더라도 정상인을 대상으로 하는 신체 진찰에 비하여 학습자가 더욱 현실감을 가질 수 있다는 점을 중요시해야 한다. 표준화 환자와의 대면진료상황과 술기 교육을 병합한 교육 모델이 개발 되고 있다[21]. 표준화 환자와 모형의 통합 모듈(integrated procedural performance instrument, IPPI)은 모형만을 이용한 교육보다 현실감을 더 높일 뿐 아니라 실제 상황에서 일어날 수 있는 중요한 절차를 교육할 수 있다는 점에서 중요한 의미를 갖고 있다. 예를 들면, 표준화 환

자와 대화를 하면서 모형을 통해 정맥 주사를 시행하게 하는 교육 모듈이다[22]. 시청각 가상현실을 이용한 위내시경 교육도 그 한 예이다[23]. 만성질환을 갖고 있는 표준화 환자는 초진뿐 아니라 재진 진료 교육에도 활용될 수 있다. 특히 학생의사가 표준화 환자를 처음 진료한 후 피드백을 받고, 두 번째 표준화 환자 진료에서 진료 능력이 향상된 것을 확인하는 방법은 표준화 환자나 학생 모두 의미 있는 교육이라고 평가받고 있다[24]. 표준화 환자의 피드백 역시 처음 진료와 비교하여 두 번째 진료에서 그 내용이 좀 더 구체적이었다[25]. 하지만 이러한 표준화 환자를 개발하고 운영하는 과정은 단순한 한 사례를 개발하는 것보다는 더 복잡하고 좀 더 많은 훈련이 필요하다.

2) 평가(assessment)에서의 표준화 환자 활용

의과대학의 공통 교육목표는 졸업생들이 의학교육을 마친 후 일차의료를 수행하는 데 필요한 자질을 갖추게 하는데 있다. 최근 의학교육의 추세는 임상 능력(competency) 바탕 교과과정과 평가 방법을 강조하고 있다[26]. 객관구조화진료시험(OSCE)은 1975년 Harden 등이 처음으로 지식, 태도, 수기 등을 동시에 측정할 수 있는 시험의 형태로 소개한 이후 의학교육에서 평가와 교육 방법으로서 발전해왔다[27]. Liaison Committee on Medical Education (LCME)의 2002-2003 미국 의대 교과과정 실태조사에 의하면 전체 126개 의대 중 77%가 임상의학입문과정 중 병력청취와 신체진찰 교육에서, 55%는 3, 4학년 최종임상종합시험의 형태로 OSCE/SP (Standardized patient)를 한 번 이상 시행하고 있다[28]. 필수임상실습교육 과정 중에 OSCE/SP의 활용 정도는 교과목별로 보면 가정의학, 내과, 산부인과, 소아과, 정신과, 외과, 신경과의 순으로 높았다. 임상종합시험에서 스테이션의 수는 5개부터 15개 이상까지 다양하였고 측정 영역은 병력청취, 신체진찰, 정보통합, 의사소통능력 등의 네 영역에서 모두 90% 이상으로 일정하게 높은 사용빈도를 보였다. 우리나라에서는 서울의대에서 OSCE를 처음 실시한 이래 새로운 임상수행평가방법으로 많은 의과대학이 OSCE를 형성평가 혹은 종합사정으로 시행하고 있다. 특히 2009년도 의사국가고시에서 실기시험이 도입되면서 OSCE 형태의 시험은 의과대학의 중요한 교육과정으로 자리 매김을 하였다. 캐나다에서는 2년간의 예비시험을 거쳐 1994년

도부터 OSCE가 의사국가고시에서 이차실기시험으로 시행되고 있다[29].

OSCE는 임상능력평가시험의 한 형태로 수험생이 여러 개의 5-20분 단위 스테이션을 돌면서 각 스테이션에서 요구하는 특정한 수기 혹은 일련의 수기들을 수행하고, 1-2명의 평가자가 미리 정해진 체크리스트에 따라 객관적으로 채점을 하는 실기시험이다[27,30]. 고전적인 OSCE의 유형과 조금 다른 여러 가지 변형이 생겨났다. 예를 들면 스테이션 당 시간을 더 늘리거나 모의환자가 직접 채점을 하기도 한다. 선다형 문제나 지필고사를 OSCE에 삽입하는 것도 또 다른 변형이다[31]. OSCE는 전통적인 임상 평가방법인 구술고사나 실제 환자를 통한 평가자 관찰시험과는 달리 채점 기준으로 구체적 체크리스트가 있기 때문에 평가자 별 차이를 없애고 객관적으로 평가할 수 있다는 장점이 있다. 채점표를 만들 때 예상 답은 시험 시행 전에 미리 준비해야 하고 이는 가장 최소 단위의 하위기능(sub-skills)인 채점항목(items)으로 나뉘어야 한다. 각각의 채점항목은 중요도에 따라 가중치를 둘 수도 있다. 채점 항목은 그 속성을 측정 영역(domain)과 기능(skills)으로 분류하여 행렬표(matrix)를 만들어 볼 수 있다. 예를 들면 응급질환의 의사소통기술을 측정하는 어떤 한 문항을 행렬표에 넣어 보면 어느 분야를 측정하고자 하는지 다른 문항과 더불어 전체 시험 내용의 균형을 점검해 볼 수 있다(Table 2).

OSCE를 임상적으로 어떻게 운영하는가는 학생 수, 지원의 정도, 교육과정의 특성에 따라 달라진다[32]. OSCE를 임상실습교육에서 평가 목적으로 운영하고자 할 때 단계별로 유의할 사항이 있다[33]. OSCE는 미리 측정하고자 했던 문제를 출제해야 한다. 임상실습 교과과정 혹은 과목별 학습 목표를 기준으로 중요한 것을 골라 시험 문항으로 선정해야 한다. 졸업 전 임상종합시험의 일환으로 시행하는 OSCE는 일차의료를 수행하는데 필수적으로 요구되는 임상능력을 평가해야 한다. 어느 한 교과목에 국한하지 않고 여러 과목을 종합하는 임상술기와 임무수행능력을 측정해야 한다.

OSCE는 학습 성취도의 수준을 구분해 주어야 한다. 시험의 타당도를 유지하려면 교육과 평가가 학습자 수준과 일치해야 한다[34]. 예를 들어 임상종합시험에서는 병력청취, 진찰능력, 자료해석, 문제해결능력, 응급질환처치능력, 면담 및 의사소통기술 등 영역별로 의과대학 3-4학년에서 교육했던 능력 군이 OSCE 내용에 골고루 포함되어야 한다.

Table 2. A Sample of Blueprint Matrix During a Comprehensive SP-Based Examination Composed of 16 Cases

Clinical Area	Acute/Emergent	Chronic	Ill-defined	Well Care	Behavior	Grave Prognosis
History (%Hx)	Chest Pain (50 Hx, 25 Px, 25 PPI)	Alcoholism (50 Hx, 25 Px, 25 PPI)	Hypothyroid (75 Hx, 25 Px) Headache (75 Hx, 25 Px)	High Risk Sexual Behavior (50 Hx 50 PPI)	Depression (75 Hx 25 IS) Eating Disorder (100 Hx)	
Physical (%Px)	Abdominal Pain (75 Px, 25 Hx) Dyspnea (75 Px, 25 Hx)	Claudication (50 Px 25 Hx 25 IS)		Breast Screening (50 Px, 25 IS, 25 Hx)		
Information Sharing (%IS)		Non-Compliance (50 IS, 25 Hx, 25 PPI)		Smoking Cessation (75 IS 25 PPI)		End-stage COPD (50 IS 50 PPI)
Physician/Patient Interaction (%PPI)		Hostile With Pain (75 PPI, 25 IS)				Giving bad news to a family member (75 PPI 25 IS)

SP, standardized patient; Hx, history taking; Px, physical examination; IS, information sharing; PPI, patient-physician interaction. Numbers in parentheses are percent of checklist items in a case.

이 밖에도 OSCE는 학생의 의사로서 태도와 윤리적 자세도 측정할 수 있어야 한다. 스테이션의 난이도는 수험생의 수준에 따라 적절하게 조정되어야 한다.

시험 출제표(blueprint matrix)는 평가 영역(content)을 제시하며 시험의 내용과 포괄성을 한눈에 알아 볼 수 있게 한다. OSCE 목적이 형성평가인가 총합평가인가에 따라 시험의 구성이 달라진다. 형성 평가인 경우 각 스테이션 별로 구조화된 평가표가 있어야 하고 이에 따라 평가자가 학생 능력을 평가한다. 각 스테이션의 학생 임무가 끝나면 평가 교수는 학생의 평가표를 검토해보고 이상적인 수준은 어느 정도여야 하는지를 학생에게 조언 해준다. 전체 점수는 교육자에게 다음 OSCE에서 어느 영역에 치중할 것 인지와 어떤 스테이션을 또 출제하고 어떤 스테이션은 새로 만들어야 하는지를 결정하는데 기초 자료를 제공한다. 반면 총합 평가에서는 즉각적인 피드백 보다는 학생의 수행능력의 당락판정과 프로그램의 질에 대한 평가가 주목적이 된다.

몇 명의 학생이 OSCE를 치르는가에 따라 스테이션 수, 문제 수, 문제 유형, 일정, 복제 수, 시험 장소가 결정된다. 125명의 학생을 대상으로 시행하는 OSCE의 경우 세군 데로 시험장을 복제하고 시간차를 두고 2부로 나누어 운영하려면 각 고사장 별로 21개 스테이션을 마련해야 하며 이 중 3-4개의 휴식 스테이션을 두면 적절한 시험문제 수는 15-17개 정도가 된다. 일반적으로 OSCE의 효율성이나 타당성을 적절히 유지하려면 15개 정도(범위: 10-20)의 스테이션은 되어야 한다[35]. 스테이션 수가 10개 미만이면 시험의 주

제를 충분히 고를 수가 없고 20개가 넘으면 운영이 힘들고 평가교수의 수가 많아져 동원이 어려워진다. 필요에 따라 중간에 휴식 스테이션(rest station)을 둘 수 있으며, 두 스테이션을 연결하여 학생들이 시간을 두 배로 사용하게 해주는 연결스테이션(link station)을 만들 수도 있다. 시험장을 두 개 세트 이상으로 복제하는 경우에는 표준화 환자의 훈련뿐 아니라 평가자의 훈련도 동시에 충분하게 이루어져야 신뢰도를 유지할 수 있다[36].

공간, 운영요원, 평가자, 시간기록원, 학생 수 등이 OSCE 일정을 결정한다. 스테이션 별 시간은 3-20분 범위 안에서 합리적인 선택을 해야 한다. 일반적인 교육과정에서 시험이나 심리적 측정의 신뢰도는 0.8을 넘어야 한다. 표준화 환자를 이용한 시험의 신뢰도는 0.41-0.85정도의 범위에 있다고 보고하고 있다. 전체 시험시간이 표준화 환자를 이용한 시험의 신뢰도와 관련이 있는데 3-4시간이 적절한 시간이라고 제시한 연구가 있다[37]. 하지만 아직까지 OSCE 형식의 시험에서 정확하게 어느 정도의 스테이션 수와 스테이션 당 시간이 적절한지 밝혀진 바는 없다.

OSCE의 시험 시간에는 스테이션 간 이동시간, 휴식 시간, 수험생의 등록/오리엔테이션 시간 등이 함께 고려되어야 한다. 스테이션을 이동하는 시간은 10-30초 정도를 할당하는 것이 좋고 이동시간이 특히 길어지는 구간은 중간에 휴식 스테이션을 두는 것이 바람직하다.

학생과 평가 교수에게 각각 오리엔테이션을 실시해야 한다. 학생 개인 별로 시험장의 지도를 나누어주고 스테이션

별로 주어진 시간, 메모 기록 가능여부, 스테이션 별 점수, 수험 요령 등을 간단명료하게 설명해 주어야 한다. 학생이 필히 지참해야 할 물품과 가지고 오면 안 되는 것을 미리 알려 주고 명찰을 배부한다. 스테이션의 끝을 알리는 종과 예령의 규칙을 알려 준다. 이동 시간을 30초 정도 주고 시작과 끝의 종만을 울려주는 운영방법도 있다. 학생들이 모의환자를 실제 환자로 대할 것을 강조해 준다. 만약 피드백이 예정되어 있으면 언제 어떻게 제공되는지를 알려 준다.

OSCE는 학생뿐 아니라 교육자에게 형성평가로서 이용될 수 있다. 학생 각자에게 관찰자인 교수가 직접 각 방마다 피드백을 줄 수 있다. 즉각적인 피드백은 다음에 이어지는 스테이션에서 학생능력을 향상시킬 뿐 아니라 수험생의 학습경험의 질을 높이는 효과가 있다. 그룹 별 수행능력에 대한 피드백은 교육프로그램에서 부족한 점을 찾아 낼 수 있다. OSCE에서 피드백은 스테이션을 마친 후 즉각적으로 시행하는 것이 가장 효과적이라고 알려져 있다. 이상적인 피드백 과정은 전문가의 진료 시범을 녹화한 모델 비디오를 보고 스스로 판단하게 하거나 아니면 평가자가 직접 평가표를 근거로 장점과 발전방향을 동시에 조언해 줄 수 있다. OSCE 도중 각 스테이션 별로 짧게 피드백을 해주는 경우에는 스테이션 종료 30초-1분 전에 해당 학생에게 평가자가 직접 관찰한 사항을 근거로 중요한 점 한 두 가지를 피드백해준다.

OSCE의 내용 타당도를 높이려면 다음 세 가지 단계로 접근해야 한다[30]. 제 1단계는 수험생이 제대로 수행할 수 있어야 하는 문제(problems) 혹은 상황(conditions)을 찾아 내는 일이다. 전문가 집단의 의견을 종합하거나 좀 더 체계적인 연구를 통해서 이 작업을 수행할 수 있다. 한 명의 출제자보다는 여러 명의 전문가가 평가항목을 선정하는 것이 중요한 문항 선정에 더 바람직하다[38]. 두 번째 단계는 앞의 문제 혹은 임상 상황에 따라 수험생이 꼭 할 수 있어야 하는 임무(tasks)를 결정하는 단계이다. 예를 들어 가슴 통증이라는 문제를 놓고 협심증 환자로부터 병력청취, 심전도를 찍고 해석하기, 심폐소생술 시행하기, 협심증 약 복용이나 식이요법에 대한 환자교육하기 등이 임무에 해당한다. 출제계획표를 작성해 보는 것이 제 3단계 작업이다. 이 과정은 시험에 포함시킬 문항(items)의 표본을 정의하는 단계이다. 한 축에는 고유능력(generic competencies, 예 병력청취, 의사소통기술, 신체진찰, 검사, 처치), 다른 한 축에

는 수험생이 능력을 보여 주어야 하는 문제 혹은 상황을 담은 이차원 행렬표를 작성한다. 영국에서 외국 졸업자를 대상으로 시행한 의사자격시험에서 세로축에는 병력청취, 신체진찰, 의사소통술, 진료수기, 응급진료의 5개 영역이 있고 가로축에는 장기시스템 별로 15개 분야로 나누어 좀 더 세분화된 시험 문항의 청사진을 만들어 사용했다[39].

우리나라 의과대학에서 OSCE를 도입하고 의사국가시험에서 OSCE를 시행하는 것 자체만으로 의학 교육과정에 지대한 변화를 가져왔다. 그렇다면 이미 출제된 문제를 그대로 다음 번에 출제하는 경우 과연 원래 의도했던 대로 능력군을 변함없이 측정할 수 있는가가 의문시된다. OSCE의 스테이션이 무슨 과에서 출제되었느냐 혹은 어떤 과목의 교수가 평가자인가에 따라 똑같은 표준환자와 면담 과정 중 선택하는 행동반응 유형이 달라질 수 있다. 반면 OSCE에서 보여준 반응 행동이 다음에 임상에서 일정한 행동으로 계속 유지될 것인지 확인할 수가 없다. 따라서 OSCE의 타당도 연구에서는 기존의 정량적인 연구를 넘어선 질적 연구가 추가적으로 더 필요하다.

OSCE처럼 임상수행능력을 평가하는 시험은 증례와 문제 수를 증가시켜야 하고 적어도 4-8시간 정도로 시험시간이 길어져야 OSCE가 신뢰도를 유지할 수 있다고 한다[40]. 표준화 환자가 채점을 할 때 채점자간 일치도는 90%까지 보고된 바 있으며 일반적으로는 80%이상의 수준을 보인다고 한다. 채점자의 훈련부족은 신뢰도를 저하시키는 중요한 요인으로 지적 받고 있다[41]. 일반화(generalization) 이론에 입각한 연구결과를 보면 평가자간 일치도보다 더 중요한 것이 바로 증례 특이성(case specificity)이다[30]. 이러한 증례 특이성이란 현실적 문제를 줄이기 위해 몇 가지 대안이 제시되었다. OSCE와 다른 형식의 시험을 접목하는 것이 한 가지 방법이다[42]. 예를 들어 90분 분량 OSCE는 신뢰도가 0.6으로 낮았지만 여기에 90분짜리 필기시험을 추가하면 신뢰도가 0.8로 올라갔다.

다른 한 돌파구는 평가 척도를 혼합하는 방법이다[30]. OSCE 평가표를 만들다 보면 체크리스트가 신뢰성 있게 자세한 항목으로 만들어지긴 하지만 중간에 측정이 곤란하여 중요한 항목은 고의적으로 빠져 실제로 수행능력을 평가하는데 있어서는 질이 떨어지는 평가표로 전락할 수 있다. 체크리스트를 이용한 평가에 비하여 총괄채점(global rating)의 우수성이 최근 각광을 받고 있다[43]. 일반적으로 기술적

인 요소가 많이 가미된 수기 스테이션은 체크리스트 척도가 적당하고 의사소통기술이나 진단 해결형 스테이션의 경우에는 총괄채점이 적절하다고 한다. 체크리스트 점수는 스테이션 사이에 그다지 큰 연관성을 보이지 않지만 총괄채점은 스테이션간에 서로 중등도의 높은 상관성을 보였다 [44]. 특정한 수기를 평가할 때는 체크리스트를 적용하고 환자교육이나 의사의 임무 등 전반적인 접근방법에 관해서는 총괄채점척도를 사용하여 이 두 가지 척도를 적절하게 섞어서 사용하는 것이 OSCE의 신뢰도를 높이는 방법이다.

일반적으로 합격기준 설정에는 상대 평가인 분포군 거방법(norm-referenced)과 절대 평가인 준거기준방법(criterion-referenced)의 두 가지가 있다[45]. 수행능력평가에서는 절대평가방법이 더 바람직하다. 이는 경계선 집단 방법(borderline group method)과 Angoff 방법의 두 가지로 다시 나뉜다[46]. Angoff 방법에서는 증례 개발자 혹은 일정 수의 전문가가 경계선상에 있는 수험생의 점수를 미리 예상해 보게 하고 이를 기준으로 합격기준을 정하는 방법이다. 경계선 집단 방법은 대규모 OSCE나 국가시험 등에서뿐 아니라 소규모 OSCE에서도 널리 이용하고 있는 방법이다. 평가자는 체크리스트 채점과 상관없이 해당 학생의 수행능력을 전반적 평가척도(global scale), 예를 들면 통과/경계선/낙제의 기준으로 평가한다. 경계선의 점수를 받은 수험생의 점수의 평균값을 그 스테이션의 통과 기준으로 잡고 전체 OSCE의 당락평정은 모든 스테이션의 경계선에 걸린 학생들의 총 평균값을 기준으로 결정한다. 이 방법은 Angoff 방법에 비하여 비교적 쉽고 시간이 덜 들며 직접관찰을 통하여 기준을 설정했다는 측면에서 신뢰성이 더 높은 것으로 알려져 있다. 변형경계군 방법(modified borderline-group method)에 의한 기준점수는 증례개발자가 제시한 기준점수(case-author method)에 비하여 더 높게 나왔으며 전자의 경우 더 많은 학생이 탈락하는 것으로 나타났다[47].

OSCE는 전체 스테이션이 현실적인 임상 상황을 얼마나 재현할 수 있느냐에 있어서 세 가지의 제한점을 갖고 있다 [32]. 첫째, 각 스테이션별로 정해진 시간만 사용해야 하므로 의사-환자 접촉을 임의로 분리하여 임상환자면담을 고립된 단면으로 만들어 수험생 수행능력을 평가한다는 한계를 갖고 있다. 따라서 형성평가로서는 취약한 점이 있지만 스테이션 수를 늘리면 신뢰도가 올라가고 총괄평가로 이용

하기에는 오히려 유리해진다. 둘째, OSCE는 임무 단위의 문항으로 구성된 평가표를 사용하는데 이는 의사환자간의 접촉이 행동의 나열로 설명될 수 있다는 것을 전제로 하고 있다. 따라서 평가표는 수험생의 완벽한 행동을 요구하지만 수험생의 수준이 올라가면 이런 완벽성에 입각한 기준은 중요성이 상대적으로 감소한다. 마지막으로 시뮬레이션에 한계가 있기 때문에 실제로 시험에 포함시킬 수 있는 환자 문제가 제한적이라는 점이다. 이 또한 수험생의 경험 및 교육 수준이 올라가면 더욱 제한점으로 부각된다. OSCE가 전통적인 임상교육평가방법의 제한점을 수용하는 대안으로 발전해 온 것은 사실이지만 임상실기 성취도를 완벽하게 평가할 수 있는 황금률은 아니고 오히려 다른 임상실습 시험과 서로 보완적으로 활용될 때 그 효과가 더 크다고 할 수 있다[48].

3) 피드백 제공을 위한 표준화 환자의 활용

표준화 환자는 학생의 수행능력을 평가하기도 하지만 피드백을 제공하여 학생의 행동이 변할 수 있게 하는데 또 중요한 의의를 가진다[49]. 피드백은 표준화 환자 진료가 끝난 후 서면으로 혹은 표준화 환자가 직접 학생의 행동에 대해 장단점을 알려 줄 수 있다. OSCE는 원칙대로 제대로 시행하면 학생의 수행 능력 정도와 실습교육의 질을 알 수 있게 해준다[50]. OSCE 결과에 대한 피드백은 스테이션이 끝난 직후 중요한 점 한 두 가지를 평가자가 즉각적으로 학생에게 피드백을 제공해 주거나 나중에 성취도를 서면보고를 통해서 할 수 있다. 스테이션 수가 적고 OSCE를 교육적인 목적으로 활용할 때는 표준 모델을 담은 비디오를 통해 스테이션이 끝난 후 학생에게 바로 피드백을 줄 수 있다. OSCE 성적 자체는 그 문제에 해당하는 실습교육과정의 교육성으로 바로 피드백 될 수 있으며 해당 교과목에서는 학생들의 성취도를 보고 실습교육의 질을 개선할 수 있다.

결론

표준화 환자는 의학뿐 아니라 의료인 양성 교육에서 넓게 활용되고 있으며 아울러 다른 직종의 교육이나 평가에서도 점점 그 유용성이 증가하고 있다. 표준화 환자는 임상

수행능력을 평가할 수 있을 뿐 아니라 학생들에게 구체적이고 유용한 피드백을 제공할 수 있다. 표준화 환자의 역할에 따라 연기, 채점, 피드백 제공 능력의 표준화는 표준화 환자를 활용한 교육에서 매우 중요하다. 표준화 환자를 평가에서 활용하는 가장 중요한 예가 OSCE이다. OSCE를 성공적으로 계획하고 운영하려면 단계별로 지침에 따라 적절하게 준비를 해야 하고 원칙에 따라 시행을 해야 한다. OSCE의 타당도는 임상교육의 학습목표를 바탕으로 하여 내용 타당도를 유지하는 것이 중요하다. OSCE의 신뢰도 유지에는 평가자의 훈련이 중요하고 시험 시간과 시험의 구성은 신뢰도를 결정하는 또 다른 요인이다. OSCE는 그 자체의 약점을 충분히 고려하고 다른 형태의 시험과 병행하여 포괄적인 평가목적으로 사용하는 것이 바람직하다.

References

1. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. *AAMC, Acad Med* 1993;68:443-51.
2. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *Br Med J* 1975;1:447-51.
3. Lane C, Rollnick S. The use of simulated patients and role-play in communication skills training: a review of the literature to August 2005. *Patient Educ Couns* 2007;67:13-20.
4. Beshgetoor D, Wade D. Use of actors as simulated patients in nutritional counseling. *J Nutr Educ Behav* 2007;39:101-2.
5. Watson MC, Skelton JR, Bond CM, Croft P, Wiskin CM, Grimshaw JM, et al. Simulated patients in the community pharmacy setting. Using simulated patients to measure practice in the community pharmacy setting. *Pharm World Sci* 2004;26:32-7.
6. Brazeau C, Boyd L, Crosson J. Changing an existing OSCE to a teaching tool: the making of a teaching OSCE. *Acad Med* 2002;77:932.
7. Norman GR, Tugwell P, Feightner JW. A comparison of resident performance on real and simulated patients. *J Med Educ* 1982;57:708-15.
8. Adamo G. Simulated and standardized patients in OSCEs: achievements and challenges 1992-2003. *Med Teach* 2003;25:262-70.
9. Cleland JA, Abe K, Rethans JJ. The use of simulated patients in medical education: AMEE Guide No 42. *Med Teach* 2009;31:477-86.
10. Robertson K, Hegarty K, O'Connor V, Gunn J. Women teaching women's health: issues in the establishment of a clinical teaching associate program for the well woman check. *Women Health* 2003;37:49-65.
11. Fairbank C. Men's health: it is imperative to teach scrotal and rectal examination. *Clin Teach* 2011;8:101-4.
12. Ker JS, Dowie A, Dowell J, Dewar G, Dent JA, Ramsay J, et al. Twelve tips for developing and maintaining a simulated patient bank. *Med Teach* 2005;27:4-9.
13. Bokken L, Rethans JJ, van Heurn L, Duviervier R, Scherpbier A, van der Vleuten C. Students' views on the use of real patients and simulated patients in undergraduate medical education. *Acad Med* 2009;84:958-63.
14. Rethans JJ, Gorter S, Bokken L, Morrison L. Unannounced standardised patients in real practice: a systematic literature review. *Med Educ* 2007;41:537-49.
15. Halbach JL, Sullivan LL. Teaching medical students about medical errors and patient safety: evaluation of a required curriculum. *Acad Med* 2005;80:600-6.
16. Haist SA, Griffith IC, Hoellein AR, Talente G, Montgomery T, Wilson JF. Improving students' sexual history inquiry and HIV counseling with an interactive workshop using standardized patients. *J Gen Intern Med* 2004;19:549-53.
17. Haist SA, Wilson JF, Pursley HG, Jessup ML, Gibson JS, Kwolek DG, et al. Domestic violence: increasing knowledge and improving skills with a four-hour workshop using standardized patients. *Acad Med* 2003;78:S24-6.
18. Eagles JM, Calder SA, Nicoll KS, Walker LG. A comparison of real patients, simulated patients and videotaped interview in teaching medical students about alcohol misuse. *Med Teach* 2001;23:490-3.
19. Bokken L, Rethans JJ, Jobsis Q, Duviervier R, Scherpbier

- A, van der Vleuten C. Instructiveness of real patients and simulated patients in undergraduate medical education: a randomized experiment. *Acad Med* 2010;85:148-54.
20. Barrows HS. Training standardized patients to have physical findings. Illinois: Southern Illinois University School of Medicine; 1999:1-32.
 21. Kneebone R, Nestel D, Yadollahi F, Brown R, Nolan C, Durack J, et al. Assessing procedural skills in context: Exploring the feasibility of an Integrated Procedural Performance Instrument (IPPI). *Med Educ* 2006;40:1105-14.
 22. Kneebone R, Kidd J, Nestel D, Asvall S, Paraskeva P, Darzi A. An innovative model for teaching and learning clinical procedures. *Med Educ* 2002;36:628-34.
 23. Kneebone R. Evaluating clinical simulations for learning procedural skills: a theory-based approach. *Acad Med* 2005;80:549-53.
 24. Linssen T, Bokken L, Rethans JJ. Return visits by simulated patients. *Med Educ* 2008;42:536.
 25. Bokken L, Linssen T, Scherpbier A, van der Vleuten C, Rethans JJ. The longitudinal simulated patient program: evaluations by teachers and students and feasibility. *Med Teach* 2009;31:613-20.
 26. Small PA, Jr., Stevens CB, Duerson MC. Issues in medical education: basic problems and potential solutions. *Acad Med* 1993;68:S89-98.
 27. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979;13:41-54.
 28. Barzansky B, Etzel SI. Educational programs in US medical schools, 2002-2003. *JAMA* 2003;290:1190-6.
 29. Reznick RK, Blackmore D, Cohen R, Baumber J, Rothman A, Smee S, et al. An objective structured clinical examination for the licentiate of the Medical Council of Canada: from research to reality. *Acad Med* 1993;68:S4-6.
 30. Newble D. Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ* 2004;38:199-203.
 31. Williams RG, McLaughlin MA, Eulenberg B, Hurm M, Nendaz MR. The patient findings questionnaire: one solution to an important standardized patient examination problem. *Acad Med* 1999;74:1118-24.
 32. Smee S. Skill based assessment. *BMJ* 2003;326:703-6.
 33. Tervo RC, Dimitrievich E, Trujillo AL, Whittle K, Redinius P, Wellman L. The Objective Structured Clinical Examination (OSCE) in the clinical clerkship: an overview. *S D J Med* 1997;50:153-6.
 34. Wilkes M, Bligh J. Evaluating educational interventions. *BMJ* 1999;318:1269-72.
 35. Selby C, Osman L, Davis M, Lee M. Set up and run an objective structured clinical exam. *BMJ* 1995;310:1187-90.
 36. Park H, Lee J, Kim S, Kim K, Park H. The effect of using two duplicated examination sites to simulate the same cases on the OSCE reliability. *Korean J Med Educ* 1999; 11:37-52.
 37. Colliver JA, Williams RG. Technical issues: test application. *AAMC. Acad Med* 1993;68:454-60.
 38. Valentino J, Donnelly MB, Sloan DA, Schwartz RW, Haydon RC, 3rd. The reliability of six faculty members in identifying important OSCE items. *Acad Med* 1998;73: 204-5.
 39. Tombleson P, Fox RA, Dacre JA. Defining the content for the objective structured clinical examination component of the professional and linguistic assessments board examination: development of a blueprint. *Med Educ* 2000; 34:566-72.
 40. Vu NV, Barrows HS, Marcy ML, Verhulst SJ, Colliver JA, Travis T. Six years of comprehensive, clinical, performance-based assessment using standardized patients at the Southern Illinois University School of Medicine. *Acad Med* 1992;67:42-50.
 41. Park H, Lee J, Hwang H, Lee J, Choi Y, Kim H, et al. The agreement of checklist recordings between faculties and standardized patients in an Objective Structured Clinical Examination (OSCE). *Korean J Med Educ* 2003;15:143-52.
 42. Wass V, McGibbon D, Van der Vleuten C. Composite undergraduate clinical examinations: how should the components be combined to maximize reliability? *Med Educ* 2001;35:326-30.
 43. Hodges B. Validity and the OSCE. *Med Teach* 2003;25: 250-4.

44. Regehr G, Freeman R, Hodges B, Russell L. Assessing the generalizability of OSCE measures across content domains. *Acad Med* 1999;74:1320-2.
45. Cusimano MD, Rothman A, Keystone J. Defining standards of competent performance on an OSCE. *Acad Med* 1998;73:S112-3.
46. Kaufman DM, Mann KV, Muijtjens AM, van der Vleuten CP. A comparison of standard-setting procedures for an OSCE in undergraduate medical education. *Acad Med* 2000;75:267-71.
47. Humphrey-Murto S, MacFadyen JC. Standard setting: a comparison of case-author and modified borderline-group methods in a small-scale OSCE. *Acad Med* 2002;77:729-32.
48. Carraccio C, Englander R. The objective structured clinical examination: a step in the direction of competency-based evaluation. *Arch Pediatr Adolesc Med* 2000;154:736-41.
49. Blake KD, Gusella J, Greaven S, Wakefield S. The risks and benefits of being a young female adolescent standardised patient. *Med Educ* 2006;40:26-35.
50. Sloan DA, Donnelly MB, Schwartz RW, Strodel WE. The Objective Structured Clinical Examination. The new gold standard for evaluating postgraduate clinical performance. *Ann Surg* 1995;222:735-42.