

# 의사면허 실기시험 제도의 성과와 과제

김종훈

인하대학교 의학전문대학원 의학교육실

## The Effects and Challenges of Clinical Skills Assessment in the Korean Medical License Examination

Jong Hoon Kim

Office of Medical Education, Inha University School of Medicine, Incheon, Korea

Clinical skills assessment was recently introduced to the Korean Medical License Examination to test medical school graduates' competencies in clinical skills. Various measures, including research and rehearsals, had been undertaken to prepare for the assessment for several years before the clinical skills assessment was first implemented. The assessment has been repeated annually for about 3,500 examinees over the course of 50 days per year for the past 4 years. The introduction of the assessment had significant effects on education in Korea's medical schools. Many schools have established clinical skills labs and the teaching of clinical skills has also been strengthened. The residents who have taken the clinical skills exam now express more confidence in caring for patients. To improve the quality of the assessment, it should be performed on a year-round schedule and a pilot test and various forms of the items should be introduced.

### Corresponding author

Jong Hoon Kim  
Office of Medical Education, Inha  
University School of Medicine,  
366 Seohaee-daero, Jung-gu,  
Incheon 400-103, Korea  
Tel: +82-32-890-0918  
Fax: +82-32-888-7224  
E-mail: minasun01@inha.ac.kr

Received: September 27, 2013

Revised: October 22, 2013

Accepted: October 22, 2013

**Keywords:** Clinical competence, Assessment, Licensure, Medical

## 서론

우리나라 의사면허 국가시험에 실기시험이 도입된 지도 4년이 지나고, 2013년 올해로 5년째를 맞이하고 있다. 그동안 많은 사람들이 노력한 결과 실기시험은 큰 문제없이 안정적으로 시행되고 있으며 의학교육에도 많은 영향을 미치고 있다. 이 시점에서 그동안의 시행과정과 그 성과를 살펴보고 앞으로 개선방향을 정리할 필요성이 제기되었다. 이에 본 논문에서는 기존의 연구보고서와 논문 등을 바탕으로 실기시험 도입의 취지와 그 준비과정, 그동안의 실제 시험의 시행 경과, 실기시험 도입에 따른 영향과 성과 등을 간략하게 기술하고, 저자 나름의 실기시험의 개선과 발전을 위한 방안을 제시해 보고자 한다.

### 의사면허 국가시험에 실기시험 도입의 취지

의사면허 시험에 실기시험을 도입할 필요가 있다는 주장은 실제

로 실기시험을 시작하기 오래전부터 제기되어 왔다. 저자가 판단하기에 이러한 필요성이 대두된 근본적인 원인은 1970년대부터 미주에서 제기되기 시작한 의학교육 패러다임의 변화에서 찾아야 한다고 생각한다(Caraccio et al., 2002). 20세기 초 의학교육은 Flexner의 연구보고에 따라 혁명적인 변화를 가져온다. 이는 기존의 비체계적인 도제식 교육과정을 체계적인 구조와 과정에 기반을 둔 교육과정으로 변화시키는 계기가 되었다. 그런데 그로부터 몇 십 년이 흐른 1970년대부터 새로운 의학교육의 흐름이 태동하기 시작했다. 이는 대중과 사회로부터 의사 스스로가 사회가 원하는 의사로서 적격(competent)한지 책임(accountability and responsibility) 있게 담보하라는 요구가 커졌기 때문이다. 그래서 의사의 적격성(competency)을 정의하고 적격성을 구성하는 속성(attributes)을 파악하며 각 속성마다 일정 기간(의학대학 재학기간 혹은 전공의 수련기간) 안에 달성해야 할 수준을 설정하여 이를 달성하기 위한 교육과정을 만들어야 할 필요성이 제기되었다. 더하여 이러한 교육과정을 마친 결과 실제로 그 수준에 도달했는지를 각 속성별로 검증하

기 위한 평가(measurement of outcomes) 전략을 모색하게 되었다. 의학교육의 적격성을 구성하는 속성은 흔히 Miller가 제안한 pyramid로(Figure 1) 설명되고 있다(Miller, 1990). 그러나 1990년대 초까지만 해도 이러한 속성들을 제대로 평가할 수 있는 전략이 마련되지 않아 평가가 필기시험을 통한 ‘know’와 ‘know how’ 영역의 속성에 국한되어 왔다. 그런데 미국과 캐나다에서 연구와 임상경험을 통하여 몇 의사들의 경우 지식은 풍부하나 환자로부터 의학적인 정보를 획득하는 능력이나 환자와의 의사소통능력에 문제가 있음이 밝혀졌다(Ramsey et al., 1998; Stillman et al., 1986, 1990; Suchman et al., 1997). 그리고 이러한 능력의 부재가 의료과오나 실수에 관련되어 있음이 밝혀졌다(Beckman et al., 1995; Moore et al., 2000; Vincent et al., 1994).

이러한 상황을 개선하기 위해 Figure 1에서도 나타난 바와 같이 4가지 영역의 속성 중에서 더 높은 직업적 신뢰성(professional authenticity)을 담보하기 위해서는 pyramid의 상위 속성인 ‘shows how’와 ‘does’ 영역의 속성을 평가할 필요가 제기되었다. 직업적 신뢰성이란 어떤 평가의 내용과 형식이 평가를 받는 사람의 일(work)과 관련된 능력과 얼마나 밀접한 관계를 가지고 있는가를 나타내는 척도를 말한다. 그런 의미에서 ‘knows how’의 영역보다는 ‘shows how’의 영역이 평가를 받는 사람의 직업적 신뢰성에 더 밀접한 관련이 있다고 말할 수 있다. 여기서 ‘shows how’의 영역은 실제와 유사한 상황에서 적절하게 행동할 수 있는 개인의 능력을 보여주는 것을 말하는데 이러한 능력을 측정하기 위해 가장 많이 사용되는 방법이 표준화 환자나 기계적 simulator를 이용한 실기평가로 1990년대부터 objective structured clinical examination (OSCE)나 임상수행평가(clinical performance examination, CPX) 같은 방법이 널리 전파되면서 이 ‘shows how’의 영역에 속하는 속성들이 실제로 평가의 대상이 되기 시작했다. ‘Does’는 습관적인 실행에 있어 실제 수행(actual performance in habitual practice)을 의미하는데 이 영역이 ‘shows how’에 비하여 직업적 신뢰도가 더 높다고 할 수 있으나 ‘does’ 영역은 실제로 업무를 수행하면서 현장에

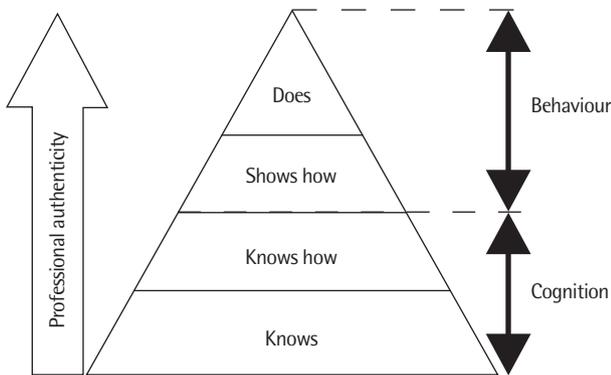


Figure 1. Miller pyramid. From Miller, G. E. (1990). *Acad Med*, 65 (9 Suppl), S63-S67.

서 받는 평가이므로 의사면허가 없는 학생을 대상으로 일정한 체계를 갖추어 정확한 평가를 하는 것은 매우 어려운 일이다.

우리나라에도 이러한 의학교육 패러다임의 변화가 알려지면서 ‘shows how’ 영역 평가의 필요성이 대두되고 OSCE, CPX와 같은 평가방법이 실행할 수 있게 됨에 따라 1990년대 후반부터 각 학교 별로 혹은 여러 학교가 컨소시엄을 형성하여 실기시험을 시행하게 되었다. 이에 따라 국가적 차원에서도 기본의학과정을 수료한 사람이 실제 상황에서도 의사로서 적절한 수준의 임상적 능력을 발휘할 수 있는지 검증함으로써 의사의 직업적 신뢰성을 담보하고자 의사면허시험에 실기시험을 도입하게 된 것이다.

### 실기시험 도입을 위한 준비과정

국가면허시험으로 실기시험을 시행하기 위해서는 학교나 컨소시엄 차원의 준비보다 더 많은 노력과 투자가 필요했다. 그 기술적인 내용은 한국보건교육인국가시험원(이하 국시험원)에서 발간한 백서에 자세하게 서술되어 있다(Kim KS, 2011). 본 논문에서는 준비 과정을 몇 개의 범주로 나누어 각 과정의 필요성과 그 의미를 간략하게 기술하였다.

#### 1. 연구사업

1999년부터 실기시험을 시행하기 위한 준비과정으로 여러 연구가 수행되었다. 먼저 의사국가시험에서 임상수기 수행능력을 측정하기 위한 방안을 모색하는 연구를 시작으로 의사시험을 당시의 필기시험 일단계 평가에서 다단계 평가로 이행하면서 여기에 필기 이외의 단계로 실기시험을 시행하는 방안의 타당성과 구체적인 실행방법에 대한 연구가 시행되어 다단계 시험을 시행할 경우 실기시험이 우선적으로 시행되어야 한다는 결론이 내려졌으며 시험을 시행하기 위해 필요한 준비사항과 다양한 시험의 시행방법들이 제시되었다. 이후에 모든 진료문항에 공통적으로 사용될 환자의사관계를 평가하기 위한 항목개발연구와 실기시험의 합격기준 설정방법을 고안하기 위한 연구가 시행되었고 그 결과물이 현재 실기시험에 적용되고 있다. 그 외에도 실기시험을 실제로 실행하기 위한 구체적인 사항에 대한 연구가 진행되었다.

#### 2. 시험문항 선정

2006년 당시 한국의과대학장협회에서 발간한 학습목표집과 실행방안연구를 바탕으로 1차 의사 수준의 환자진료능력을 평가할 수 있는 항목을 추출하고 환자가 호소하는 임상표현으로 분류하여 진료문항(초기에는 CPX문항이라고 명명함)의 제목(예를 들어, 복통, 발열 등)을 결정하였으며, 1차 의사 수준에서 시행할 수 있어야 하는 단순임상술기와 진찰방법을 추출하여 수기문항(초기에는 OSCE문항이라고 명명함)의 제목(예를 들어, 기관 내 삽관, 복부진

찰 등)을 결정하였다. 1차로 선정된 문항들을 대상으로 빈도와 위험도 같은 임상적인 중요도, 시험에서의 실현 가능성, 의과대학의 교육상황 등을 감안하여 최종적으로 실제 시험에 사용될 문항을 결정하였다.

### 3. 시험문항 개발

진료문항은 일차적으로 각 임상표현의 환자를 주로 진료할 것으로 예상되는 전문가에게 환자 연기를 위한 시나리오와 수험생의 진료행위를 평가하기 위한 채점표를 의뢰하여 초고를 받은 후에 문항심의회위원회의 집중 심의작업을 통하여 문항의 완성도를 높였으며 문항들 간의 구조와 내용에 있어 통일성을 기하였다. 수기문항도 각 문항의 전문가에게 수험생이 특정 술기나 진찰을 완수하기 위해 시행해야 하는 세부 항목으로 구성된 채점표를 의뢰하여 초고를 받은 후에 문항심의회위원회의 심의를 통하여 1차 의사가 수행할 수 있는 수준으로 완성하였다.

### 4. 평가체계 개발

일정 기일 내에 많은 수의 수험생을 평가하면서도 진료와 수기로 여러 영역의 문항을 고루 평가하면서 일정 수준 이상의 평가 신뢰도를 담보하기 위하여 최대한 많은 문항을 사용하기 위한 평가체계를 개발하였다. 그 결과 진료문항과 수기문항을 각 6개의 범주로 나누고, 각 범주에서 1개 문항을 무작위로 추출하여 총 12개 문항으로 이루어진 문항 set를 상당수 만든 후에 진료와 수기문항에서 같은 내용을 평가하는 문항이 있는지 검토(예를 들어, 진료 영역의 복통문항과 수기 영역의 복부진찰문항이 한 set에 있는 경우)하여 평가내용이 겹치는 문항이 없도록 조정하였다.

### 5. 예행연습

실제 평가를 시행하기에 앞서 준비상태를 점검하기 위하여 3차례 예행연습을 시행하였다. 첫 예행연습은 소수의 학생을 대상으로 개발된 평가체계가 실제 시험에 적용하기에 무리가 없는지를 집중 점검하여 문제가 없음을 확인하였다. 두 번째, 세 번째 예행연습은 점차 많은 학생을 대상으로 실제 실기시험이 시행될 국시원 실기시험센터에서 시행되어 센터의 준비사항을 점검하고 교수 채점위원과 표준화 환자의 채점을 비교 검토하여 진료문항은 표준화 환자가, 수기문항은 교수 채점위원이 채점표를 완성하기로 결정되었다. 또한 예행연습을 통하여 수집한 평가결과자료를 이용하여 연구사업을 통하여 제안된 합격선 결정방법을 실제로 적용하여 실제 시험에 사용이 가능함을 검증하였다.

이 외에 실기시험의 시행을 홍보하기 위해 의학교육학회에서 실기시험 시행에 대한 토론회를 하였고 각 대학에는 국시원에서 정식으로 공문을 발송하였으며, 예행연습을 위해 각 대학에 참가학생을 모집하는 과정과, 그리고 준비과정에서 한국의과대학·의전원

협회와 긴밀하게 의사소통을 하였다. 시험의 시행시기는 법률상 4학년 학생에게 졸업예정자의 자격이 주어지는 졸업예정일 전년도 9월부터 시행하기로 결정하였다.

## 실기시험의 시행 경과

2009년 9월에 처음으로 약 50일간에 걸쳐 3,500여 명의 수험생을 대상으로 실기시험이 시행되었으며 작년까지 4년간 시행되었다. 실기시험의 시행하기 위해서 새로운 문항의 개발, 시험에 사용될 문항의 선정과 수정, 표준화 환자 훈련과 수기문항을 위한 기구 준비, 시험의 시행, 합격선 설정, 시험에 사용될 문항의 정리 등의 작업이 매년 시행되고 있다. 새로운 문항의 개발과정은 위에서 ‘시험문항 개발’과 거의 동일하므로 여기서는 다른 작업에 대하여 간략하게 기술하였다.

### 1. 문항 선정과 수정

진료문항의 경우 기출문항과 새로 개발된 문항을 대상으로 당해 연도에 사용될 문항을 6개의 범주에서 골고루 선정한다. 선정된 문항을 문항 심의회원들에게 전문성과 경험을 기준으로 분배하여 1차로 점검을 한 후에 시연(1차 점검본으로 표준화 환자를 훈련시켜 심사위원 중 한 명이 수험생 역할을 하고 다른 심의회원이 채점자 역할을 하며 문항을 점검하는 과정)을 통하여 문항의 문제점을 발견하고 2차로 다시 수정을 한 후에 심의회원 전원이 참석하는 토론을 거쳐 실제 시험에 사용될 최종문항을 완성한다. 시연 후에 수정해야 하는 부분이 많거나 중요한 수정이 있다고 판단되면 다시 시연을 시행하기도 한다. 이때 새로 개발된 문항은 기존문항과의 비교 검토를 통하여 채점표의 내용이나 난이도 등의 측면에서 기존문항과 차이가 나는 부분이 있는지 점검한다.

수기문항의 경우 기존 시험에 사용되었던 문항은 다음에 기술할 ‘시험 후 문항정리’ 작업에서 수정을 마치고 다음 해에 다시 문항심의를 거쳐 재사용 여부가 결정되는데, 필요할 경우 시연을 하여 문제점이 있는지 다시 점검한다. 새로 선정된 문항은 진료문항과 같은 경로로 1차 점검을 마치고 시연을 통하여 문항의 문제점을 점검한 후에 2차로 수정을 한다. 이후 기존과 신규 수기문항 모두 심의회원 전원이 참석하는 토론을 통하여 최종문항을 확정한다. 앞서 ‘평가체계 개발’에서 기술한 것처럼 이들 진료와 수기문항을 조합하여 각 수험생이 수행할 시험 set를 완성한다.

### 2. 표준화 환자 훈련

위에서 기술한 문항심의회작업을 거쳐 당해 연도에 사용될 진료문항이 확정되면 이에 따라 표준화 환자의 모집이 이루어진다. 모집된 표준화 환자들은 국시원의 표준화 환자 교육자와 대학교수로 이루어진 교육위원들로부터 표준화 환자가 되기 위한 훈련을 받는다. 훈

련은 배당받은 진료문항의 환자로서 행동하기, 채점표 완성하기로 이루어져 있다. 몇 차례 훈련을 마친 후에 교육위원으로부터 합격 판정(문항의 환자를 올바르게 표현하고 채점표를 정확하게 완성)을 받은 표준화 환자를 대상으로 실제 시험을 시행하기 전에 여러 교육위원이 최종 예행연습을 시행하여 여기서 최종 합격을 받은 표준화 환자가 실제 시험에 투입된다. 훈련과정이나 최종 예행연습에서 불합격 판정을 받은 표준화 환자는 재훈련을 실시하여 다시 판정을 받는다. 여기서 합격 판정을 받으면 시험에 투입되지만 불합격 판정을 받으면 표준화 환자를 교체하여 다시 훈련을 시행한다.

### 3. 시험의 시행

대개 9월 초부터 11월 말까지 50여 일간 하루에 최대 3회(한 회에 24명씩, 2개 center에서 각 12명씩) 72명의 학생이 시험을 보게 된다. 수험생은 진료문항과 수기문항 중에서 각기 6개가 추출된 모두 12개 문항으로 이루어진 시험 set를 진료문항과 수기문항을 번갈아가며 수행하게 된다. 진료문항에서 표준화 환자를 진료한 수험생은 바로 이어 사이시험(대개 진료한 표준화 환자에서 감별해야 할 질환과 필요한 진단이나 치료계획을 기술하는 시험)을 완성하고 이어 수기문항을 수행하고 수기문항을 마치면 바로 다음 진료문항을 수행하는 형태로 시험이 이루어진다. 진료문항은 표준화 환자가 짝을 이루어 연기와 채점표 완성을 담당하게 되는데 한 표준화 환자가 환자 연기를 하며 수험생의 진료를 받을 때, 동시에 한 표준화 환자는 시험장 외부에서 그 모습을 관찰하며 실시간으로 채점표를 완성하게 된다. 수기문항의 경우에는 각 대학에서 추천된 교수채점위원이 시험방 내에서 학생의 수기 수행을 관찰하며 실시간으로 채점표를 완성하는데, 이 과정에서 채점위원의 문항에 대한 개선 의견도 취합하고 있다.

매 시험일 각 시험 center마다 1명의 책임채점위원이 전체적인 시험의 진행을 감시하며 돌발사항이나 채점 시 의문사항 등에 대하여 대처하고 있다. 책임채점위원은 사이시험의 채점도 담당하고 있다.

### 4. 합격선 설정

모든 수험생이 시험을 마치고 모든 자료가 정리되면 매년 시험에 사용된 모든 문항에 대하여 각 문항마다 그 문항의 합격점수를 결정하게 되며 이를 토대로 각 수험생의 합격 여부를 결정하게 된다. 상세한 합격점수 결정방법은 보건복지부장관의 고시(제2009-106호)로 공표되어 있다. 각 문항의 합격선을 결정할 후에는 각 시험 set에 따른 점수와 합격률 등을 분석하여 특정 시험 set에 따른 편차가 있는지를 재차 검증한다.

### 5. 시험 후 문항정리

시험에 사용된 문항에 대하여 시험결과 채점표 항목에 대한 학생의 반응도와 표준화 환자나 채점위원의 의견 등을 분석하여 시

나리오나 채점표가 표준화 환자의 연거나 평가의 신뢰도를 향상시키는 방향으로 수정 보완이 가능한지를 논의하여 문항정리를 하고 문제은행에 입고한다.

### 6. 문항 풀 관리

수기문항의 경우 새로운 시험항목이 추가될 때마다 새로운 문항을 만들고 있으며 몇 개의 특수한 문항의 경우 동일항목에 여러 형태의 문항을 추가하는 방법으로 문항을 관리하고 있다. 진료문항의 경우 새로운 시험항목과 기존문항에 대하여 일정 수의 신규문항을 만들고 있으며 차후 충분한 수의 문항이 확보되면 오래된 문항의 도태를 고려하고 있다.

## 실기시험 도입에 따른 영향과 성과

지난 2011년에 저자도 참여한 설문조사와 전문가 토론을 통하여 실기시험의 성과를 분석하는 연구가 시행되어 그 결과가 보고서와 논문으로 발표된 바 있다(Park, 2012). 여기서는 그 내용을 간략하게 기술하였다.

설문조사결과 실기시험의 도입은 각 학교의 교육에 상당한 영향을 미친 것으로 나타났다. 임상실기를 실습하기 위한 공간을 갖춘 대학이 크게 늘어났으며 임상실습 준비과정, 임상실기평가 등을 교과과정에 도입한 대학도 많아졌다. 그리고 임상실기 시험을 시행하기 위한 각 지역별 컨소시엄에 거의 모든 대학이 가입하면서 컨소시엄이 활성화되는 계기가 되었다. 또한 졸업자격을 검증하는 평가에 실기시험이 포함되게 되었다. 임상실습에 있어서도 학생이 환자를 접할 수 있는 기회가 증대되었으며 교수들이 실습에 참여하는 시간도 늘었다고 답하였다. 교수와 간호사, 환자를 대상으로 실기시험을 도입한 후에 배출된 수련의나 전공의에 대한 평가를 묻은 결과 기본 임상실기, 환자에 대한 태도, 자신감, 의사소통의 측면에서 전에 비해 향상이 되었다는 응답이 있었다. 반면에 의료과오나 의무기록 작성능력에는 의사실기시험의 도입이 큰 영향을 주지 않은 것으로 나타났다. 실기시험에 합격하여 의사가 된 수련의나 전공의들은 실기시험을 위한 준비과정이 자신의 일차의료 능력 배양, 환자와의 대화능력 향상, 전반적인 임상능력 배양에 도움이 되었다고 스스로를 평가하였다.

전문가 집중 토의모임에서는 의사실기시험의 도입이 의학교육의 긍정적인 변화를 가져온 측면이 많이 있지만 아직까지 임상실습의 내실화 등 변화의 정도와 방향에 대해서는 더 많은 발전과 정립이 필요하다는 것이 공통적인 의견이었다. 또한 새로이 수련과정을 시작하는 의사가 과거에 비해 임상 수행에 대한 자신감이 향상되었고 실제 수행능력도 좋아졌다는 의견이 지배적이었다. 하지만 이러한 능력을 더욱 개발하고 발전시켜야 할 교육체계의 미비로 향상되었던 능력이 시간이 지남에 따라 오히려 퇴보하는 문제가 지적되었

다. 이를 개선하기 위해 졸업 후 교육을 더욱 체계화시킬 필요가 있으며 이러한 교육의 성과를 평가하는 체계도 마련될 필요가 있다는 의견이 있었다.

미국의 경우 2004년도에 의사면허시험에 실기시험을 도입한 바 있는데 Hauer et al. (2006)은 미국 내 25개 의과대학의 학생교육 지도자와 반구조화된 대담을 하여 이를 분석한 결과 면허 실기시험의 도입이 각 학교가 임상능력 평가와 교육과정의 변화에 대한 토론을 할 수 있는 기회와 이를 실행할 수 있는 동기를 제공하는 역할을 했다고 결론지었다. 또한 몇 년 후에 발표된 설문조사연구에서 많은 대학에서 실기시험이 도입됨에 따라서 표준화 환자의 이용, 각종 수기교육, 임상실습, 교수개발 프로그램이 증가하거나 새로 도입되었다고 답하였다(Gilliland et al., 2008). 대만에서는 의사면허시험에 실기시험을 도입한다는 발표가 교육병원의 OSCE 시행에 미친 영향을 설문조사한 연구에서 교육병원의 OSCE가 확대되었으며 OSCE를 시행하기 위한 시설도 확충되었다고 보고하였다(Lin et al., 2013). 양국의 예에서도 우리나라와 유사하게 실기시험의 도입이 교육기관의 실기교육을 강화하는 효과가 있었음을 보여주고 있다.

### 실기시험의 개선과 발전을 위한 방안

앞서 언급했듯이 지난 4년간 우리나라 의사면허시험에서 실기시험이 시행되어 왔다. 모든 평가가 그 목적인 바에 100% 충실하게 시행되는 것은 불가능한 일이나 제도의 발전을 위해서는 그 제도 시행의 근본 취지를 다시 한 번 생각하며 그 취지를 달성하기 위하여 어떤 가치가 중요하고 그 가치를 실현하기 위한 대책을 강구하는 것이 필요하다고 생각된다. 물론 이러한 과정에서 실제로 그 대책의 실현이 가능한지를 면밀히 따져서 추진의 우선순위를 결정하여야 할 것이다.

우리나라가 의사국가고시에 실기시험을 시행하는 목적은 개인에 따라 생각하는 바에 차이가 있을 수 있지만 지난 실기시험 평가 목표 개발연구과정에서 결론을 내린 내용은 다음과 같다. ‘의사실기시험은 의과대학/의학전문대학원 졸업생 혹은 졸업예정자가 일차의료에 필요한 기본적인 지식, 술기, 태도를 적용하여 환자중심으로 안전하고 효과적인 진료를 수행하는지 평가하는 데 있다(Park et al., 2012).’ 그러므로 수험생이 그렇게 수행하는지를 정확하게 판별하는 것이 무엇보다도 중요하다고 본다(평가의 정확성). 실기시험의 특성상 수험생이 어떤 시기와 센터에서 평가를 받느냐에 따라 평가에 사용되는 내용이 상당 부분 다를 수 있으나 그럼에도 불구하고 모든 학생이 평등한 기준으로 평가를 받을 수 있어야 한다(평가의 공정성). 현재 국가시험을 개선하기 위해 어떤 제도나 시설이 필요한가에 대한 논의가 여러 연구와 회의상에서 이루어지고 있다. 여기서는 각각의 과제를 정확성과 공정성의 개선 측면에서 그 필요성을 기술하고 실제로 시행하기 위해서는 무엇이 필요한지

저자의 의견을 기술하였다.

먼저 제도의 개선으로 실기시험의 시행을 상시화하는 방안이 논의될 필요가 있다. 현재 시행되는 의사국가시험은 필기시험이든 실기시험이든 졸업예정자(최종학년 학생으로 6개월 이내에 졸업이 가능한 자)나 졸업생에게 1년에 단 1차례의 응시 기회만을 부여하고 있다. 그리고 이 1회의 시험에서 불합격의 판정을 받으면 정식 졸업 후 교육과정(여기서는 수련과정을 말함)에 들어가지 못하고 그 다음 시험까지 1년을 기다려야 하는 교육의 공백기간이 생긴다. 그런데 미국은 수험생이 첫 응시에 불합격하더라도 1년 이내의 기간에 다시 응시할 기회가 두 차례 더 부여되고 있으며(United States Medical Licensing Examination, 2012) 캐나다의 경우에는 졸업 후 1년 이상의 수련을 받은 후에 응시할 수 있으며 1년에 응시 기회는 봄, 가을로 2회만 부여되지만 이 시험에 불합격하더라도 수련이 중지되는 것이 아니므로 수련을 마치고 독자진료를 하는 시점까지 수련의 중단 없이 여러 차례의 응시 기회가 부여된다(Medical Council of Canada, 2013). 우리나라도 단 1회가 아니라 수험생이 첫 응시에 불합격하더라도 졸업하기 전에(이미 졸업생이라면 다시 1년을 더 기다리기 전에) 다시 응시 기회를 주기 위해 미국과 같은 상시시험 제도의 도입이 연구된 바 있다(Han et al., 2010). 상시시험의 취지는 학생이 자신의 의사로서의 실력 이외에 다른 조건에 의하여 이익이나 불이익을 받지 않도록 하겠다는 것이라고 말할 수 있다. 예를 들어 원래 실력은 부족하지 않으나 시험 당일의 수험생의 신체적, 정신적 상황으로 자기 실력을 발휘하지 못하는 경우, 당일의 시험문항이 수험생의 실력 중에서 제일 취약한 분야에서만 출제되는 경우, 당일의 시험문항이 다른 날에 비해 난이도가 높은 경우 등은 수험생이 불이익을 받을 수 있다. 반대로 수험생의 실력이 부족한데 당일의 시험이 다른 날에 비해서 난이도가 낮으면 운이 좋아서 시험에 합격하는 상황이 벌어질 수 있다. 이와 같이 서로 다른 문항을 사용하여 수험생을 평가해야 하는 경우 1회의 시험만으로는 학생의 실력이 정확하게 합격할 수준을 상회하는지 모든 학생에게 공정한 평가가 이루어지는지 보장하기 어렵다. 이러한 문제를 완전히 해결하는 것은 불가능하지만 상시시험을 도입한다면 ‘운’보다는 ‘실력’이 평가에 더 큰 영향을 미칠 수 있도록 시험제도를 수정할 수 있는 길이 있다고 생각한다. 예를 들어 현행 합격선 결정방법에 의하면 각 문항의 합격선을 결정할 때 위원들이 제시한 합격선을 평균하여 그 값에서 표준오차를 빼도록 규정하고 있다(Ministry Health and Welfare, 2009). 이는 합격선 설정과정에서의 오류가 수험생에게 미칠 부정적 영향을 우려한 결과라고 할 수 있다. 캐나다의 경우에는 오히려 평균에서 측정오차를 더하여 합격선을 설정하고 있다(Smee & Blackmore, 2001). 이는 기준을 엄격하게 하여 수험생이 ‘운’이 좋아서 시험에 합격할 가능성을 줄이기 위한 것이다. 단 수험생이 엄격한 기준에 의해 ‘실력’이 있는데 ‘운’이 없어서 불합격한 경우라도 정식 교육의 단절 없이 다시 시험에 응시할 수 있으

므로 정식교육이 단절되어야 하는 우리나라의 상황에 비해서는 큰 문제가 아니다. 그러므로 우리나라에서도 상시시험을 도입하고 합격선을 현재보다 엄격하게 관리한다면 ‘운’이 좋아서 합격하는 경우를 줄이면서도, 학생이 ‘운’이 없어서 불합격하여도 정식교육의 단절이라는 엄청난 불이익을 줄일 수 있어 도입을 검토할 필요가 있다고 생각한다. 다만 이 제도를 시행하기 위해서는 준비해야 할 사항이 많이 있다. 우선 법적인 조치가 취해져야 하는데 현행 의료법에 의하면 졸업예정 6개월 전부터 국가시험에 응시할 수 있도록 되어 있는데 상시시험이므로 최소한 이 기간을 1년 이상으로 늘릴 필요가 있다(Korea Ministry of Government Legislation, 2011). 다음으로 이런 상시시험제도를 감당할 수 있는 국시원의 시설과 특히 인력의 보강이 필요하다. 또한 시험센터의 운영, 합격선 결정방법 등 세부제도의 정비도 필요하다. 상시시험제도가 도입되면 학생 1인당 시험시간을 현재보다 더 연장(시험문항 수 확대와 문항당 소요시간 연장, 모의시험[pilot test] 문항 도입)하고 시험의 형태도 더 평가에 적합한 형태(진료수기 결합형 문항, 진료 후 구두시험 등)로 개선하는 데 필요한 여건을 제공할 수 있을 것이다.

저자가 상시시험의 도입이 필요하다고 생각하는 이유가 하나 더 있다. 앞에서 언급했지만 현행 제도하에서도 먼저 시험을 본 수험생들이 시험을 앞둔 수험생들에게 시험에 대한 정보를 제공하고 있을 것으로 예상되고 있다. 이런 상황에 대하여 법적으로 도덕적으로 문제가 제기된 적이 있고(Baek & An, 2011) 앞으로 제기될 소지도 다분하다. 수험생이 시험에 합격하기 위해 불법적이고 비도덕적인 수단을 동원하는 것에 결코 찬성할 수는 없지만 수험생의 입장에서 생각해 볼 때 저자는 그들이 그렇게까지 필사적으로 정보를 공유하는 연유에 다음과 같은 측면이 있다고 생각한다. 앞서 언급했듯이 미국이나 캐나다의 수험생은 한 번의 시험에서 불합격하더라도 경력의 공백 기간 없이 다시 응시가 가능하지만 우리나라에서는 한 번의 시험 실패가 수험생에게 1년의 공백기를 가져온다. 다시 말해 시험을 앞둔 우리나라의 수험생이 느끼는 합격에 대한 절박함이 미국이나 캐나다 수험생보다 매우 클 것은 자명할 것이다. 특히 시험의 실패가 병역 처분과 관계가 있다면 그 절박함의 강도는 훨씬 강할 것이다. 이런 상황이라면 아무리 높은 도덕적 관념을 가지고 있다고 해도 합격의 가능성을 높일 수 있다고 생각하는 수단이 있으면 그 수단이 비도덕적이라고 하여도 은밀히 시행할 수 있다면 수험생이 이를 이용하고 싶은 유혹을 떨치기 어려울 것이다. 물론 상시시험을 시행한다고 해서 정보의 공유가 근절될 것이라고 보기는 어렵다. 아마 미국이나 캐나다 같은 외국에서도 시험에 대한 정보의 공유가 없지는 않을 것이다. 일례로 미국의 의사국가시험을 관장하는 the national board of medical examiners (NBME)의 소식지에 학생을 대상으로 한 ‘Sharing isn’t always caring’이라는 글이 실려 있는데 그 내용은 의사로서의 직업정신을 생각하여 정보공유를 자제하고 공유를 발견하면 신고해 달라는 내용이다(Na-

tional Board of Medical Examiners, 2010). 이런 호소가 얼마나 수험생의 공감을 얻었는지 알 길은 없지만 우리나라에서도 시험을 상시화하여 단번에 합격해야만 경력과 교육의 단절을 피할 수 있다는 학생의 부담을 덜어준다면 저자의 생각으로는 학생이 비도덕적 수단에 의존하고자 하는 절박함도 줄어들 것이고 무엇보다도 시험을 주관하는 입장에서 더욱 떳떳하게 수험생에게 의사로서의 도덕성과 직업정신을 요구할 수 있는 명분도 세울 수 있을 것으로 보인다.

다음으로 진료수기 결합형 문항과 같은 새로운 형식의 문항을 도입할 필요가 있다. 이러한 형식의 문항이 필요한 이유는 정확성의 측면, 즉 실제 의사가 진료현장에서 발휘해야 할 능력을 보다 현실과 가까운 환경을 만들어서 평가하기 위해서이다. 현행 수기문항의 경우 많은 문항에서 실제 환자에게 어떤 수기를 시행할 때 수기를 시행하기 전에, 수기를 시행하는 중에 그리고 수기를 끝낸 후에 반드시 동반될 수밖에 없는 환자와 의사 사이의 의사소통에 대한 평가가 충분히 이루어지지 않고 있다. 이는 시험의 운영상 수기문항에 5분이라는 짧은 시간이 배정되어 있어 그 시간에 학생의 술기 테크닉과 환자와의 의사소통을 모두 평가하기가 어렵기 때문이다. 물론 환자와의 의사소통은 대개 진료문항을 통하여 평가되고 있으므로 현 시험에서도 평가가 이루어지지 않고 있는 것은 아니지만 환자에게 보다 침습적인 수술을 해야 하는 술기를 시행함에 있어 의사소통은 진료 시 의사소통과는 다른 측면이 있으므로 별도로 평가할 필요가 있다. 그러므로 앞으로 시험운영체계의 개선을 모색할 때 현재의 수기문항은 가능한 수기 자체의 수행과 그 수기를 시행받는 모의환자와 의사소통을 결합하는 형식의 평가가 가능해지는 방향으로 논의를 할 필요가 있다. 진료수기 결합형 이외에도 진료문항과 환자상태보고나 구두시험 결합형 문항의 도입도 고려할 수 있다. 또 진료문항에 있어 표준화 환자의 특성상 표현 가능한 신체진찰 소견이 제한적일 수밖에 없는데 표준화 환자로써 표현 불가능한 소견(예를 들어, 심 잡음이나 비정상 폐음, 직장수지검사나 골반내진 검사의 이상소견 등)도 현재 개발되고 있는 기술을 이용한다면 보다 풍부하게 문항을 만들 수 있다(Cardionics, 2012; Verma et al., 2011).

현행 시험제도에서는 새로 개발된 문항을 실제 시험에 사용하기 전에 모의시험을 시행할 기회가 제한적이다. 물론 문항심의회와 표준화 환자 훈련, 최종 예행연습을 통하여 시험에 사용하는 데 문제가 없는지 철저히 검증하고 있지만 보다 자세하고 정확하게 문항을 분석하기 위해 실제 수험생을 대상으로 문항을 점검하는 모의시험을 시행할 방법이 없다. 미국이나 캐나다의 경우 새로 출제되는 문항을 평가에 사용되는 문항과 섞어서 실제 국가시험에서 수험생이 이 새 문항에 대하여 어떤 반응을 보이는지 자료를 축적하여 실제 시험에 사용이 가능한지 여부를 마지막으로 점검하고 있다(Han et al., 2012). 이는 새로운 문항이 정확하게 수험생의 능력을 측정할 수 있는지 판단하기 위해 필요한 과정이라고 생각된다. 그러므로 우

리나라에서도 장차 시험의 운영방식에 변화를 주고자 할 때 새로운 형식의 문항도입과 수험생을 상대로 새로운 문항에 대한 모의시험이 가능할 수 있도록 논의를 할 필요가 있다.

마지막으로 채점체계의 개선을 고려할 필요가 있다. 현 시험운영 방식에 의하면 수험생이 표준화 환자를 진료하고 나서 사이시험에 감별진단과 앞으로의 진단이나 치료계획을 5분 내에 기술하도록 되어 있다. USMLE step 2 clinical skills (CS)의 경우 표준화 환자를 진료한 후 10분 내에 중요한 병력과 신체진찰소견, 감별진단과 앞으로의 계획을 기술하도록 되어 있다. 더하여 감별진단에는 반드시 수험생이 이런 감별진단을 해야 한다고 판단한 병력과 신체진찰소견상의 근거를 3개 이내로 기술하도록 되어 있다. 우리나라에서 한 대학의 학생을 대상으로 사이시험에 병력과 신체진찰소견을 기술하도록 하고 그 점수와 병력, 신체진찰 점수 간의 관련성을 조사한 연구결과 관련성이 높지 않은 것으로 나타났다(Kim JH, 2010, 2011). 이는 학생이 병력을 통하여 환자에게서 어떤 질문을 했는지, 그리고 어떤 신체진찰을 했는지와 병력이나 신체진찰을 통하여 얻은 정보의 중요성에 대하여 어떻게 판단하는지가 큰 관련성이 없다고 해석할 수 있다. 그러므로 학생의 임상적 능력을 보다 정확하게 측정하려면 사이시험에 병력과 신체진찰을 주요 소견을 기술하도록 하는 것이 필요하다고 본다. 또한 실기시험에서는 학생의 정보획득 능력을 측정하기 위해 학생이 표준화 환자에게 특정 정보를 얻기 위한 질문을 했는지(혹은 표준화 환자에게서 어떤 정보를 들었는지) 여부를 파악하기 위해 checklist 형태의 채점표를 사용하고 있다. 그런데 작년부터 USMLE step 2 CS에서는 이 병력 checklist를 폐지하고 병력청취에 관한 부분은 사이시험을 통해서만 평가하고 있다. 이런 변화를 시행한 이유를 정확하게 알 수는 없으나 지난 2011년 NBME의 assessment scientist인 Ann King이 강의에서 그 이유의 하나로 표준화 환자가 다양한 수험생의 다양한 질문에 답을 할 수 있는 여지를 주어 보다 실제 환자에 가까운 역할을 수행할 수 있도록 하기 위해서라고 발표하였다(Lewis, 2011). 우리나라에서도 표준화 환자가 병력채점표를 정확하게 완성하기 위해 수험생이 특정 질문을 했는지에 과도하게 집중하게 되면 표준화 환자의 태도가 실제 환자와 괴리되는 현상이 벌어질 수 있다. 물론 이러한 현행 실기시험의 채점체계가 평가의 정확성과 공정성을 크게 저해한다고 말할 수는 없지만 보다 시험의 완성도를 높이기 위해서는 채점표와 사이시험체계의 변화를 고려해야 하며 특히 시험운영체계를 개선하여 모의시험이 가능해진다면 반드시 그 시행 가능성과 효과를 검증해 볼 필요가 있다.

## 결 론

이상에서 현행 우리나라의 의사국가시험 실기시험의 앞으로의 개선 과제를 열거하였다. 결국 현행 제도의 문제를 개선하기 위해

서는 국가시험 자체를 통하여 개선책(새로운 형태의 문항도입, 채점체계의 개선 등)의 효용을 검증하는 체계를 도입할 필요가 있다고 판단된다. 그리고 이런 체계를 도입하기 위해서는 현행 시험운영체계의 개선(실기시험 상시화, 실기시험 중에 모의시험 도입 등)이 필수적이다. 이를 통하여 보다 정확하고 공정한 수험생의 실기능력 평가가 이루어질 수 있을 것이라고 기대된다.

## REFERENCES

- Baek, M. K., & An, S. (2011). *Yebi euisa byungdeon yangsim 'euisa kukgashim' jojicjuc yuchul*. Retrieved from <http://www.seoul.co.kr/news/news-View.php?id=20110401010014>
- Beckman, H. B., Markakis, K. M., Suchman, A. L., & Frankel, R. M. (1994). The doctor-patient relationship and malpractice: lessons from plaintiff depositions. *Arch Intern Med*, 154(12), 1365-1370.
- Cardionics. (2012). *SimScope (the hybrid simulator)*. Retrieved from <http://www.cardionics.com/simscope.html>
- Carraccio, C., Wolfsthal, S. D., Englander, R., Ferentz, K., & Martin, C. (2002). Shifting paradigms: from Flexner to competencies. *Acad Med*, 77(5), 361-367.
- Gilliland, W. R., La Rochelle, J., Hawkins, R., Dillon, G. F., Mechaber, A. J., & Dyrbye, L., & Durning, S. J. (2008). Changes in clinical skills education resulting from the introduction of the USMLE step 2 clinical skills (CS) examination. *Med Teach*, 30(3), 325-327.
- Han, J. J., Kim, J. H., Park, H. K., & Im, H. J. (2012). *Euisa silkisihumeseoui sunhengsihum (pretest) silhengbangan yeongu*. Seoul: National Health Personnel Licensing Examination Board.
- Han, J. J., Park, H. K., & Kim, J. H. (2010). *Euisa silki sangsihumjedo doip tadangsung yeongu*. Seoul: National Health Personnel Licensing Examination Board.
- Hauer, K. E., Teherani, A., Kerr, K. M., O'sullivan, P. S., & Irby, D. M. (2006). Impact of the United States Medical Licensing Examination Step 2 Clinical Skills exam on medical school clinical skills assessment. *Acad Med*, 81(10 Suppl), S13-S16.
- Kim, J. H. (2010). Correlations of information gathering scores between checklists and interstation works in a clinical performance examination. *Korean Med Educ Rev*, 12(2), 19-26.
- Kim, J. H. (2011). *Post-encounter note is necessary to evaluate information gathering ability in clinical performance examination*. Association for Medical Education in Europe Annual Conference Abstract Book, 336.
- Kim, K. S. (2011). *National Health Personnel Licensing Examination Board beksio*. Seoul: Hyeonmoon.
- Korea Ministry of Government Legislation. (2011). *Euiryobub*. Retrieved from <http://www.law.go.kr/lsInfoP.do?lsiSeq=106850#0000>
- Lewis, K. (2011). *Assessing Students with SPs: what's ahead?* Retrieved from <http://aspnews.org/2011/09/26/assessing-students-with-sps-what%E2%80%99s-ahead/>
- Lin, C. W., Tsai, T. C., Sun, C. K., Chen, D. F., & Liu, K. M. (2013). Power of the policy: how the announcement of high-stakes clinical examination altered OSCE implementation at institutional level. *BMC Med Educ*, 13, 8.
- Medical Council of Canada. (2013). *Application information*. Retrieved from <http://mcc.ca/examinations/mccq-c-part-ii/application-information/>
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Acad Med*, 65(9 Suppl), S63-S67.
- Ministry Health and Welfare. (2009). *Euisa kukgashimeui silkisihum hab-*

- kyeokja gyeoljungbangbub*. Retrieved from [http://www.mw.go.kr/front\\_new/jb/sjb0402vw.jsp?PAR\\_MENU\\_ID=03&MENU\\_ID=030402&page=62&CONT\\_SEQ=214991](http://www.mw.go.kr/front_new/jb/sjb0402vw.jsp?PAR_MENU_ID=03&MENU_ID=030402&page=62&CONT_SEQ=214991)
- Moore, P. J., Adler, N. E., & Robertson, P. A. (2000). Medical malpractice: the effect of doctor-patient relations on medical patient perceptions and malpractice intentions. *West J Med*, 173(4), 244-250.
- National Board of Medical Examiners. (2010). *Sharing isn't always caring*. Retrieved from <http://www.nbme.org/PDF/Publications/Examiner-2010-Fall-Winter.pdf>
- Park, H. K. (2012). The impact of introducing the Korean Medical Licensing Examination clinical skills assessment on medical education. *J Korean Med Assoc*, 55(2), 116-123.
- Park, H. K., Han, J. J., Kim, J. H., & Park, W. B. (2012). *Euisa silkisihum pyeonggamokpyo gaebalyeongu*. Seoul: National Health Personnel Licensing Examination Board.
- Ramsey, P. G., Curtis, J. R., Paauw, D. S., Carline, J. D., & Wenrich, M. D. (1998). History-taking and preventive medicine skills among primary care physicians: an assessment using standardized patients. *Am J Med*, 104(2), 152-158.
- Smee, S. M., & Blackmore, D. E. (2001). Setting standards for an objective structured clinical examination: the borderline group method gains ground on Angoff. *Med Educ*, 35(11), 1009-1010.
- Stillman, P. L., Regan, M. B., Swanson, D. B., Case, S., McCahan, J., Feinblatt, J., & Nelson, D. V. (1990). An assessment of the clinical skills of fourth-year students at four New England medical schools. *Acad Med*, 65(5), 320-326.
- Stillman, P. L., Swanson, D. B., Smee, S., Stillman, A. E., Ebert, T. H., Emmel, V. S., & Hatem, C. (1986). Assessing clinical skills of residents with standardized patients. *Ann Intern Med*, 105(5), 762-771.
- Suchman, A. L., Markakis, K., Beckman, H. B., & Frankel, R. (1997). A model of empathic communication in the medical interview. *JAMA*, 277(8), 678-682.
- United States Medical Licensing Examination. (2012). *2013 Bulletin of information: eligibility*. Retrieved from <http://www.usmle.org/pdfs/bulletin/2013bulletin.pdf>
- Verma, A., Bhatt, H., Booton, P., & Kneebone, R. (2011). The Ventriroscope® as an innovative tool for assessing clinical examination skills: appraisal of a novel method of simulating auscultatory findings. *Med Teach*, 33(7), e388-e396.
- Vincent, C., Young, M., & Phillips, A. (1994). Why do people sue doctors?: a study of patients and relatives taking legal action. *Lancet*, 343(8913), 1609-1613.