

전공의 평가 방법 소개

한림대학교 의과대학 가정의학교실

김 수 영

요 약

평가는 '의사결정을 위해 정보를 모으고, 합성하고 해석하는 과정'이다. 따라서 전공의 평가는 전공의가 현재 무엇을 알고 있고 어떻게 행동하고 있고 어떤 태도를 견지하고 있는지를 파악해서 앞으로 어떤 지식, 태도, 수기가 필요하며 그 전공의가 속한 전공의 프로그램은 어떤 점을 보완해야 하는지를 조사하는 것이다.

전공의 수련에는 몇 가지 원칙이 있는데 그것은 평가는 교과과정이나 프로그램 목적에 부합하여야 하고 교육목적은 해당 교육 영역을 대표해야 하며, 다양한 평가 방법을 사용해야 하고, 여러 차례 관찰하여야 하며, 여러 사람이 평가하여야 하며 평가는 미리 지정된 표준이나 기준에 따라서 이루어져야 하고 평가는 공정해야 한다는 것이다. 또한 평가도구는 타당한 자료를 얻을 수 있어야 하고 신뢰성 있는 자료를 얻을 수 있어야 하고 실현가능성이 있어야 하고 내 평가 환경에 적용할 수 있어야 하고 평가방법은 가치 있는 정보를 제공하여야 한다.

현재 이용되고 있는 평가 방법에는 360° 평가 척도(360° evaluation instrument), 의무기록에 기초한 회상 구술 시험(Chart stimulated recall oral examination, CSR), 체크리스트 평가(checklist evaluation), 실제 수행 또는 기록된 행동에 대한 전반적인 평가(Global ratings of live recorded performance), OSCE(Objective structured clinical examination), 수기, 수술, 환자 기록부(Procedure, operative, case logs), 환자 조사(patients surveys), 포트폴리오, 기록 검토(Record review), 모의시험과 모형(Simulations and models), 표준 구술 시험(STANDARDIZED ORAL EXAMINATION), 표준 환자 시험(STANDARDIZED PATIENT EXAMINATION, SP), 필기시험(WRITTEN EXAMINATION, MCQ) 등이 있다.

수련 영역은 환자진료, 의학지식, 진료 중심 학습과 향상(Learning & Improvement), 대인관계 의사소통 기술, 전문주의(Professionalism), 체계중심진료(Systems-Based Practice)로 구별할 수 있으며 각 영역과 각 영역별에 필요한 기술 그리고 이 기술을 평가하는데 가장 적합한 평가 방법이 있다.

평가는 '의사결정을 위해 정보를 모으고, 합성하고 해석하는 과정'이다. 따라서 전공의 평가는 전공의가 현재 무엇을 알고 있고 어떻게 행동하고 있고 어떤 태도를 견지하고 있는지를 파악해서 앞으로 어떤 지식, 태도, 수기가 필요하며 그 전공의가 속한 전공의 프로그램은 어떤 점을 보완해야 하는지를 조사하는 것이라고 할 수 있다. 이처럼 전공의 평가는 전공의와 프로그램 자체를 모두 평가해야 하며 전공의에 대한 평가 내용은 전공의, 지도전문의, 프로그램 과장 모두에게 지속적으로 피드백되어서 전공의는 자신의 능력을 평가하고, 지도전문의는 자신의 교육 내용에 대한 평가를 받고 프로그램 과장은 프로그램 전체 과정을 평가할 수 있어야 한다.

우리나라 일차의료의 질적 향상은 상당부분 전공의 교육의 내실화에 달려 있다고 할 수 있으며 이러한 내실화의 핵심적인 내용은 전공의 교육과정의 표준화와 함께 이에 따른 효율적인 전공의 평가라고 할 수 있다.

하지만 우리나라 가정의학에서 아직 전공의 교육과정의 표준화, 전공의 평가 모두 미진한 상태에 있으며 이에 대한 대책이 필요하다.

이 글에서는 전공의 수련 평가 원칙과 현황에 대해서 알아보고 미국에서 전공의 평가 방법으로 제시하고 있는 것들을 소개하여 우리나라 전공의 교육과 수련병원 수련 상황에 맞는 평가 방법이 무엇인지 알아보려고 한다.

전공의 수련 평가 원칙

일반적으로 모든 평가는 형성평가(formative)와 총괄평가(summative)로 구별할 수 있다. 형성평가는 매일매일 일상적으로 일어나는 행동을 피드백을 통해 교정해 나가는 과정으로 운동선수에서 코치의 역할과 비슷하다고 할 수 있다. 총괄평가는 어떤 기간이나 어떤 경험이 끝난 이후 일정수준에 도달했는지 평가하는 것이다.¹⁾ 예를 들어 가정의학회에서 실시하는 수련 중 시험(in training examination)은 형성평가이고 전문의 시험은 총괄평가이다. 수련 중 시험을 본 이후에는 자신이 부족한 부분이 피드백 되어서 교정이 가능하도록 해주지만 전문의 시험에는 이런 과정이 없기 때문이다.

평가체계가 효과적이라면 여러 가지 조건이 갖추어져야 한다. 명확한 목적(goal)과 목표(objectives)이 있어야 하고 평가자는 대상 전공의를 직접 관찰해야 한다. 또한 형성평가가 효과를 거두기 위해서는 평가자와 피평가자 사이의 신뢰관계 형성이 필수적이다.

평가를 했으면 그 결과로 학습자가 해당 상황에서 알고 있는 것은 무엇이고 믿고 있는 것은 무엇이고 할 수 있는 일이 무엇인지 추론할 수 있어야 한다.²⁾ 전공의 수련 평가를 위한 원칙은 평가 자체의 원칙과 평가 도구에 대한 원칙으로 구별할 수 있다.

1. 평가 원칙에 대한 것³⁾

1) 평가는 교과과정이나 프로그램 목적에 부합하여야 한다.

가르치는 것과 평가하는 것이 완전히 부합할 때 목표와 평가간에 일관성이 생긴다. 만일 코스를 상부 내시경 조작에 필요한 기술과 지식을 가르치는 것이라면 이 영역의 기술과 지식을 평가하여야 한다. 목표와 평가간에 일관성이 있으면 학습자는 코스 목표 영역을 넓혀서 참여할 것이며 평가되는 내용만을 학습하지는 않을 것이다.

2) 교육목적은 해당 교육 영역을 대표해야 한다.

모든 상황에서 모든 교육목적 성취 정도를 평가하는 것은 가능하지 않기 때문에 평가 대상중 추출하는 것이 필요하다. 특정 상황에서 개별 능력에 대한 대

표 행동이 무엇인지 확인해야 한다. 예를 들어 전문주의 능력(prefesionalism competence)에 대해서는 혼란 윤리적 딜레마, 환자 진료에 있어서 문화적인 문제, 환자 진료에 필요한 핵심적 태도, 특정 영역에 있어서 팀워크 등을 고려할 수 있다.

3) 다양한 평가 방법을 사용해야 한다.

능력이라는 것 자체가 다면적이고 개별 평가 방법은 이런 다면적인 능력 모두를 포착하는데에는 한계점이 있기 때문에 단일한 평가 방법이 적절하다고 보기는 어렵다. 따라서 여러 가지 평가 방법을 같이 시행하는 것이 바람직하다.

4) 여러 차례 관찰하여야 한다.

관찰 횟수를 늘리면 신뢰도, 정확도가 올라가고 시간에 따른 행동형태도 파악할 수 있다.

5) 여러 사람이 평가하여야 한다.

여러 사람이 평가하면 신뢰도, 평가 정확도가 높아지고 평가 영역도 넓어진다.

6) 평가는 미리 지정된 표준이나 기준에 따라서 이루어져야 한다.

미리 지정된 표준이란 “충분히 좋은” 혹은 “경계선상”에 대한 있는 행동이나 지식에 대해 객관적 지표가 마련되어 있다는 의미이다. 이에 대해 주관적으로 평가해서는 안 된다.

7) 평가는 공정해야 한다.

공정성은 모든 학습자가 동등하게 평가받는 것이다. 공정성은 타당도와 신뢰도가 높은 방법을 이용하면 높아지지만 여전히 결과가 능력 이외의 다른 것에 영향을 받으면 공정성은 훼손된다. 예를 들어 수련 받은 전공의와 그렇지 않은 전공의가 혼재할 때 특정 과목에 대한 시험 결과를 비교해서는 안 된다. 기초 시험(baseline test)을 제외하고 모든 피시험자는 자신이 받은 평가 자료에 대해 배울 기회도 같아야 한다. 학습자에게는 평가하는 것과 평가하지 않는 것에 대한 정보를 제공하여야 한다. 또한 평가 폼과 평가 방법에 대해서는 명확히 제시해야 한다.

2. 평가 도구(방법)에 대한 것

1) 타당한 자료를 얻을 수 있는 평가도구(방법)이어야 한다.

타당한 자료란 평가한 것에 대해 정확한 정보를 제

공하는 것이다. 타당도 추론에는 여러 가지 방법이 쓰인다. 실제 임상에서 수행 적도를 예측할 수 있도록 도움을 주는 것으로 추론할 수도 있다(예측타당도). 혹은 변화를 발견할 수 있는 능력으로 추론하기도 한다(반응성). 예들 들어 심장내과 수련을 받기 전에는 심장내과 평가에서 나쁜 점수를 받지만 순환근무를 마친 이후에는 같은 시험에서 좋은 점수를 얻을 수 있다. 혹은 다른 지표와 강력한 관계가 있다는 사실에서 추론할 수 있다(판별 타당성). 예들 들어 같은 시험을 심장내과 의사는 좋은 점수를 얻고 학생은 나쁜 점수를 얻는 것과 같은 것이다

2) 신뢰성 있는 자료를 얻을 수 있는 평가도구(방법)이어야 한다.

평가방법이 신뢰성이 있다는 것은 사용된 시기나 사용자, 측정항목이나 증례에 관계없이 일관성 있는 결과를 보인다는 의미이다. 신뢰도의 여러 측면 중 어떤 것을 중요하게 생각할지는 무엇을 평가할 것인가와 평가한 방법이 무엇인가에 따라서 달라진다. 일반적으로 신뢰도 혹은 일반화 지표는 0.8은 넘어야 한다. 관찰자간 혹은 관찰자내 신뢰도는 동일한 수행에 대해 서로 다른 평가자가 평가하는 것이다. 증례간, 문항간 신뢰도는 다른 증례, 다른 상황, 다른 문항에 대한 한 개인의 수행능력 차이의 일관성이다. 검사-재검사 신뢰도는 시간에 따른 일관성의 척도이다. 일반화 이론(generalizability theory)은 위에서 언급한 신뢰도 측정방법의 대안이 되는데 신뢰도가 떨어지는 근원을 평가하고, 전체 신뢰도를 G 계수도 제시한다.

3) 평가 도구(방법)는 실현가능성이 있어야 한다.

실현가능성은 시간, 실행에 필요한 훈련, 필요한 시설이나 기술, 피시험자 한 명에게 필요한 평가 수, 재정지출, 평가 방법 시행 정도에 따라서 달라진다.

4) 평가도구(방법)을 내 평가 환경에 적용할 수 있어야 한다(외적 타당도)

평가방법을 선택했다면 과거에 평가 상황을 고려하여야 한다. 과거 평가 상황에는 평가 목적, 피평가자와 평가자 특성, 평가가 이루어지는 환경 등이 있다. 예들 들어 시험센터에서 이루어지는 평가 도구는 클리닉이나 병동에서 쓰는 경우는 보완해서 사용해야 한다.

5) 평가방법은 가치 있는 정보를 제공하여야 한다.

가치라는 측면에서 평가방법은 교육과 학습을 촉진할 수 있는 새로운 정보를 제공하여야 한다. 예들 들어서 어 어떠한 수행을 향상되었는지에 대한 평가와 어떤 교과과정 변경이 필요한지에 대한 정보를 제공해 주어야 한다.

평가 방법⁴⁾

1. 360° 평가 척도(360° evaluation instrument)

1) 설명: 주변에 있는 여러 사람이 한 사람을 평가하는 도구이다. 대개 상사, 동료, 부하직원, 환자와 가족들이 평가하며 특정 주제(팀워크, 의사소통, 의사결정)에 대한 실행도를 설문지 혹은 실태조사 방식으로 평가한다. 대부분 평가도구는 그 행동이 얼마나 자주 일어나는 가에 대해 등급척도를 이용한다(예 5점 척도: 5점 항상...0점 없다).

2) 유용성: 평가가 총괄평가가 아닌 형성평가로 제공될 때 평가자는 더 정확하고 덜 관용적인 평가를 내린다. 이 평가 방법은 대인관계, 의사소통 기술, 직업적, 행동적인 면을 평가할 때 유용하고 때로는 환자 진료의 특정 측면이나 체계 중심 진료(systems based practice) 평가에도 유용하다.

3) 실현가능성/실행성: 대부분 이 방법을 도입하는데는 어려움이 있다. 그것은 모든 평가자에게 적절한 자동화된 방법을 이용할 수도 있다.

2. 의무기록에 기초한 회상 구술 시험(Chart stimulated recall oral examination, CSR)

1) 설명: 피평가자(전공의)의 환자 증례를 표준화된 구술 시험으로 평가한다. 시험자에게 진단검사를 실시한 이유, 임상 조건 해석, 치료 계획 등을 물어본다. 평가자는 프로토콜이나 점수 체계에 따라서 평가한다. 대부분 환자 증례 당 5-10분 정도 걸린다.

2) 유용성: 실제 환자에서 임상적 의사결정과 의학 지식 이용이나 응용을 어떻게 하는지를 평가하는 것이다. 의학지식 이해나 회상을 평가하려면 MCQ가 CSR 보다 더 낫다.

3. 체크리스트 평가(checklist evaluation)

1) 설명: 체크리스트는 복잡한 능력(competency)을 구성하는 필수적인 혹은 요구되는 요소 즉 특정 행동, 활동 등으로 이루어져 있다. 각 항목에 대해서 '네모'에 표시하거나 '예'에 표시하면 그 행동은 만족되었거나 행동이 교정되었다는 것이다.

2) 유용성: 특정 행동이나 활동으로 나눌 수 있는 능력을 평가할 때 유용하다. 환자 진료 수기(병력, 신체 검진, 수기 기술)를 평가하거나 의사 소통 기술을 평가할 때 유용하다. 진료 중심 학술 기술에 대한 자기 평가(근거중심의학)에도 유용하다. 체크리스트는 성취(performance)에 대한 피드백 방법으로 유용한데 그 이유는 어떤 과제를 성취할 때 구분된 행동에 대해서 평가하기 때문이다.

3) 실현가능성/실행성: 체크리스트를 개발하기 위해서는 평가하려는 수행 능력에 있어서 필수적인 행동이나 활동, 순서, 기준 등에 대해서 몇몇 전문가들이 합의를 이루어야 한다. 이 방법을 이용할 때는 훈련이 필요하다.

4. 실제 수행 또는 기록된 행동에 대한 전반적인 평가(Global ratings of live recorded performance)

1) 설명: 이 평가 방법이 다른 평가 방법과 다른 점은 ㉠ 평가자는 특정한 기술이나 행동을 평가하는 것이 아니라 전반적인 능력(환자 진료 기술, 의학 지식 등)을 측정하고, ㉡ 평가는 과거를 회상해서 전반적인 느낌을 기술하는 것이다. 대체로 점수 척도를 가져서 예들 들어 1=매우 좋다, 2=좋다...와 같은 형식으로 기록한다.

2) 유용성: 순회근무 마지막에 전반적인 수행 상황을 평가할 때 쓰인다. 점수체계와 함께 코멘트 형식의 내용을 받기도 한다.

3) 실현가능성/실행성: 쉽게 만들어 쓸 수 있다. 점수를 줄 때는 직접 수행을 관찰하거나 피평가자와 관계를 가져야 하는 등 시간이 걸릴 수도 있다. 평가자 훈련을 받아야 결과의 재현성이 높아진다.

5. OSCE (Objective structured clinical examination)

1) 설명: OSCE 시험을 치르는 동안 12~20개 정도 표준 환자 접촉 스테이션을 거치게 되며 각 스테이션 중간에서 피시험자는 환자기록을 완성하거나 만났던 환자에 대한 간단한 시험을 치르게 된다. 모든 피시험자는 똑같은 스케줄로 전체 station을 돌게 된다. 표준 환자가 OSCE에 이용되는 일차적 평가도 구이지만 임상 증례에 의한 자료 해석 연습, 마케팅과 임상 시나리오를 이용해서 기술을 평가하기도 한다.

2) 유용성: OSCEs는 거의 모든 미국 의과대학에서 시행되고 있고 전공의 교육 프로그램 상당수 그리고 캐나다 면허청에서는 5년 이상 시행해왔다. OSCE는 신체 진찰과 병력 취취, 환자와 가족과 의사 소통 기술, 소견을 요약하고 기록하는 능력, 감별진단하는 능력, 치료 계획을 세우는 능력, 환자 기록에 기초한 판단 능력을 평가하는 표준적인 방법이다.

3) 실현가능성/실행성: OSCEs는 특정 임상 기술과 능력을 평가하는데 매우 유용하지만 만들고 시행하기가 보통 어려운 것이 아니다. OSCEs는 한 번에 여러 명이 시험 볼 수 있어야만 비용-효과적이다. OSCEs는 대부분 외래 시설이나 CCTV가 설치된 전공의 프로그램 공간에서 실시된다. 분리된 공간이 각 station마다 필요하다. 대도시 지역이나 대학을 중심으로 자원과 전문가에 대한 콘소시엄이 필요하다.

6. 수기, 수술, 환자 기록부(Procedure, operative, case logs)

1) 설명: 진료 중 본 환자, 시행한 수술, 수기 등을 기록하는 것이다. 증례, 수술, 수기 수를 기록할 수 있다. 환자기록부는 정해진 기간 동안 본 환자 수를 기록하고 수술 기록부는 자세한 수술 기록부를 기록하기도 하고 확실한 범주에 속하는 소수의 수술이나 수기를 기록하기도 한다.

2) 유용성: 여러 종류 환자에 대한 기록부와 수술, 수기 기록부는 전공의들이 경험한 환자 범위를 결정하는데 유용하다. 주기적으로 기록부를 검토하면 전공의로 하여금 수련 필수사항이나 특정 학습목표를

충족하기 위해 어떤 환자가 수기를 해야 하는지에 대해서 알 수 있게 해준다. 전체 수련 과정에서 임상 경험을 기록한 환자 기록부는 자신이 한 경험에 대한 요약기록부가 될 수 있다. 물론 숫자가 많다고 능력 자체를 보장해 주지는 못한다.

3) **실현가능성/실행성:** 전자기록 장치를 이용하면 환자나 수기 증례를 모으고 요약하는데 도움이 된다. 이런 장치를 개발하는데는 상당한 비용이 들지만 대부분 적은 액수만 개발하고 이용할 수 있는 상업적인 프로그램도 있고 학회 등에서 무료로 제공하기도 한다. 수동으로 기록하는 것은 직원이 할 수도 있지만 전공의의 수에 따라서 혹은 요구 사항에 따라서 시간-낭비인 경우도 있다.

7. 환자 조사(patients surveys)

1) **설명:** 병원, 외래를 방문한 환자들에게 의사 진료 등에 대한 만족도를 조사하는 것이다. 환자 진료의 일반적인 부분(환자 진료 시간, 환자 진료, 능력, 예의, 공감 능력 등에 대한 전반적인 평가)에 대해 평가하는 경우가 많다. 의사가 한 설명, 청취 능력, 진찰 소견, 치료계획, 약물 부작용 등에 대해 얼마나 설명했는지와 같은 진료의 특정 측면에 대해서 조사하기도 한다. 대개 환자들에게 등급척도(매우 좋다, 좋다, 그저그렇다, 나쁘다, 매우 나쁘다)로 평가한다. 체크한 항목에 점수를 매겨서 전체 점수를 산정한다.

2) **유용성:** 환자가 한 평가 내용을 모으면 환자가 환자 진료 능력에 대한 만족(자료수집, 치료, 상담, 교육, 예방적 진료), 대인관계와 의사 소통 기술, 전문가적인 행동, 체계적인 진료(system based practice, 환자 옹호, 진료 통합) 등을 평가할 수 있다. 특정 행동에 대한 내용이 포함되어 있으면 형성평가나 수행능력향상에도 쓰일 수 있다. 총괄평가에도 쓰일 수 있지만 이런 목적으로 쓰기 위해서는 신뢰도와 타당도에 대한표준을 만족해야 한다.

3) **실현가능성/실행성:** 상업적인 회사나 여러 단체에서 환자 만족도 조사에 대한 다양한 품을 제한하고 있다 환자 인터뷰나 관심집단 토론, 설문지 등으로 새로운 조사 양식을 개발할 수도 있다. 대부분 진료 받으러 온 시간 중에 조사하지만 우편이나 전화 설문

을 하기도 한다. 환자 조사 중 닥치게 되는 어려운 점은 언어/식자 능력, 충분한 재현성을 기대할 정도의 숫자, 자료 수집, 결과 분석, 결과 통보 등에 필요한 인력 등이다. 그래서 특정 기관이나 둘 이상의 의료기관에서 같이 이루어지는 경우가 많고 각 프로그램에 보고서를 전달하기도 한다. 컴퓨터를 이용해서 더 손쉽게 조사를 진행하기도 한다.

8. 포트폴리오

1) **설명:** 전공의가 학습, 학습 성취 등에 근거를 모으는 것이다. 포트폴리오에는 문서뿐만 아니라 비디오, 오디오, 사진 등 여러 형태가 가능하다. 무엇을 배웠는가 하는 대해 숙고하는 것이 필요하다. 학습 자료뿐만 아니라 배운 것을 기록하는 것, 아직 필요한 학습내용, 그것을 어떻게 학습할 것인가에 대한 계획 등도 포함된다. 전공의 과정에서는 시행한 임상 수기 기록, 치료 방법 결정할 때 검토한 문헌 요약, 질향상 계획과 그 결과, 실제 직면한 윤리적 딜레마, 환자 결과를 추적한 컴퓨터 프로그램, 환자에게 제공한 상담 기록부 혹은 사본 등도 포트폴리오에 포함될 수 있다.

2) **유용성:** 전공의 교육의 형성평가 혹은 총괄평가 모두에 적용된다. 다른 방법으로 평가하기 힘든 능력 즉 진료 능력 향상, 환자 진료에 과학적 근거 이용하기, 전문가적 행동, 환자 옹호와 같은 것을 완성했는지에 대한 평가하는데 유용하다. 교육 기회, 아침 보고, 환자 회진, 연구 계획 등이 포트폴리오를 이용해서 전공의를 평가하는데 쓰일 수 있는 것들이다.

3) **실현가능성/실행성:** 선호하는 학습 스타일에 따라서 전공의 교육에 포트폴리오를 받아들여지가 결정된다. 일부 전공의나 지도전문의들은 이 방법을 통하지 않았으면 놓치거나 하지 않았을 행동에 집중할 수 있다고 주장하지만 다른 사람들은 얻는 것에 비해서 들이는 시간이나 노력이 너무 크다고 주장하기도 한다.

9. 기록 검토(Record review)

1) **설명:** 의무기록부서나 임상사들이 환자 의무 기록을 검토한다. 평가자는 미리 정해진 기준에 기초해한 프로토콜을 이용해서 투약, 시행한 검사, 수기,

환자 결과 등에 대한 요약지에 대해 부호화(coding)을 하며 이를 환자 진료 표준과 비교한다. 1600가지 질병에 대한 진료 표준은 Agency for HealthCare Research and Quality (<http://www.ahrq.gov/>)에서 얻을 수 있다.

2) **유용성**: 임상적 의사 결정, 환자 관리와 예방보건의서비스 계획 수립과 시행, 임상 시설과 자원 적절한 이용(적절한 병리 검사와 자문)에 대한 근거를 제시한다. 환자에 대해 기록하기 전에 다른 임상 팀과 상의하는 경우가 있기 때문에 각 개별 전공의에 대한 평가라기 보다 전체 임상 팀에 대한 평가가 되는 경우도 있다.

3) **실현가능성/실행성**: 대개 한 환자당 20~30분 정도 시간이 필요하다. 주된 제한점은 ① 의무기록이 늦어지는 경우 적절한 평가가 안되고, ② 진료 표준에 대해서 동의가 이루어져야 하고 평가하는 사람에게 맞게 코드화되어야 하며, ③ 자료의 신뢰성을 위해서 수련이 필요하다는 점 등이다.

10. 모의시험과 모형(Simulations and models)

1) **설명**: 모의시험은 실제와 비슷하지만 똑 같지는 않은 상황에서 임상적 수행을 평가하는 것이다. 모의시험의 핵심 속성은 현실과 비슷한 다양한 선택을 포괄할 수 있고, 암시를 거의 주지 않고 임상 문제에 대해 추론할 수 있도록 해주고, 피평가자가 생명을 위협하는 실수를 해도 상관없으며, 곧바로 피드백을 주어서 잘못된 것을 바로 잡을 수 있고, 다른 방법으로는 평가하기 힘든 임상 문제에 대해 평가할 수 있도록 해준다는 것이다.

모의 시험 포맷에는 필기형 문제(환자 관리 문제, patient management problems, PMPs), PMP를 컴퓨터화 한 것(clinical case simulation, CCX[®]), 역할극 상황(즉 표준 환자, standardized patients, SPs: 임상 팀 모의 시험), 해부학 모형이나 마네킹, 그리고 이들을 조합한 것 등이 있다. 마네킹은 환자 질환과 유사한 병리학적 소견을 가지고 있다.

모형은 플라스틱으로 사람 조직과 비슷하게 만들어져서 피시험자의 실제적 행동에 반응하도록 전기장치가 내장되어 있다. 가상 모의 시험 혹은 가상 환경(Virtual reality simulations or environments, VR)

은 가능한 실제 장기와 표면 이미지에 가깝게 하기 위해서 해부학 모형을 컴퓨터와 결합하는 것이며 실제 환자와 거의 같을 수도 있다. 가상 환경은 다른 방법으로는 측정하기 힘든 복잡한 임상 시험을 평가하는데 도움을 준다.

2) **유용성**: 가상환경을 이용한 모의시험은 무릎이나 다른 관절 관절경 수술, 마취가 의사가 하는 수술 중 생명을 위협하는 상황에 대한 대처, 상처 소독이나 소수술, 사람과 같은 마네킹에 실시하는 심폐소생술 같은 것을 훈련하고 평가하는 모듈이 개발되었다. 컴퓨터화된 모의시험은 다양한 임상 원칙에 대한 임상 추론, 진단과정, 치료는 전문의 시험의 한 부분으로 개발되었다.

3) **실현가능성/실행성**: 해당 영역 전문가는 실제 환자로부터 특정 기술, 능력을 평가하기 위해 임상 시나리오를 만들고 평가 기술 전문가가 이 시나리오를 표준 환자나 마네킹, 컴퓨터에 기반한 모의 시험의 형태로 바꾼다. 모의 시험은 만드는데 많이 비용이 들어가며 경제적으로 이용하기 위해서는 병리적 조건이나 임상 문제에 따라 많은 변형을 하는 것이 필요해진다.

11. 표준 구술 시험(Standardized Oral Examination)

1) **설명**: 실제 환자 증례를 이용해서 훈련된 의사가 피시험자에게 질문하여 평가하는 것이다. 환자 증례에 대한 시나리오를 주고 어떻게 관리할지를 물어 본다. 질문을 통해 임상 소견에 대한 추론 능력, 결과 해석, 치료 계획 수립 정도를 평가한다. 잘 짜여져 있으면 3~5분 정도 시간이 소요된다. 대개 서로 다른 임상 증례 18~60개에 대해서 평가하며 90분~2시간 30분 동안하고 2~4 세션으로 나누어 평가한다.

2) **유용성**: 실제 환자에 대한 임상 의사결정, 의학 지식 사용과 적용 등을 평가한다. 선다형 문제가 의학 지식 회상이나 이해를 평가하는데 좋다.

3) **실현가능성/실행성**: 대개 해당 학회가 증례를 만들고 시험전문가를 훈련한다. 표준 구술 시험을 만들고 시행하는 데는 상당한 자원과 전문가가 필요하다.

12. 표준 환자 시험(Standardized Patient Examination, SP)

1) 설명: 표준 환자(Standardized patients, SPs)는 표준화된 방법으로 의학 상황을 가장할 수 있도록 훈련 받은 사람이거나 실제 환자로 자신의 상황을 표준화된 방법으로 제시하도록 훈련받은 사람들이다. 표준환자 시험은 10~12분 환자 진료 시간 동안 서로 다른 상황을 제시하는 다수의 SP로 이루어진다. 실제 환자인 경우는 환자에 대한 신체 검진을 할 수 있다. 체크리스트나 등급 양식에 따라 관찰자나 SPs가 특정 환자 진료 업무와 예상 행동의 적절성, 정확성, 완결성 등을 평가한다. 실행기준은 미리 만든다.

여러명 SP를 이용한 시험에 부가적으로 혹은 대체해서 개별 SP들을 이용해서 특정 환자 진료 기술을 평가하기도 한다. SP는 Objective Structured Clinical Examinations (OSCE) 상황에 이용되기도 한다.

2) 유용성: SPs는 병력청취 기술, 신체 진찰 기술,

의사 소통 기술, 감별진단, 검사 이용, 치료 등을 평가할 때 쓴다. 병력 청취, 신체 진찰, 의사소통 기술에 더 신뢰성 있는 점수를 얻을 수 있다. 이 방법은 임상 기술의 총괄 평가에 가장 많이 이용된다. SP 한 명으로 특정 기술이나 지식을 평가할 수도 있다.

3) 실현가능성/실행성: 표준환자 시험을 개발한다는 것은 검사할 특정 능력 결정, 표준환자 훈련, 체크리스트나 등급 폼 개발, 기준점 등을 개발한다는 의미이다. 개발 시간은 오래 걸리지만 전공의 교육 중에 SP를 공유하여 시간을 줄일 수도 있다. 한 명의 SP는 새로운 임상 문제를 습득하는데 6~8시간이 필요하다. 만일 SP가 전공의를 평가하려면 이 보다 2배 정도 많은 시간이 필요하다. 검사에 필요한 시설은 표준환자 평가 공간과 평가 공간 사이에 전공의들이 의무기록을 할 수 있는 것이다.

13. 필기시험(Written Examination, MCQ)

1) 설명: 필기 시험 혹은 컴퓨터로 치르는 다지 선

표 1. 환자 진료에서 필요한 기술과 이에 대한 평가 방법.

	평가 방법													
	필요한 기술	의무기록검토	CSR	체크리스트	전반적 평가	SP	OSCE	모의시험과 모형화	360° 전반적 평가	포트폴리오	MCQ 시험	구술 시험	환자 증례 기록	환자 서베이
환자 진료	존중하고 보살피는 행동			3		1			2					1
	인터뷰			1		2	1							
	의사결정		1	2			2				2			
환자 진료	외래환자 관리 계획 수립, 실천	2	1	2	3			2	3					
	환자와 가족 상담과 교육			3		1	1		2					1
	신체 검진 수행			2		1	1							
	의학적 수기 수행			1	3			1	2				3	
	예방 서비스 제공					2	1			3			2	
환자 진료	팀의 일원으로 활동			3	3				1					

택형 시험(multiple-choice question, MCQ)은 단순한 사실이나 쉽게 기억할 수 있는 정보가 아닌 지식체계에 대한 이해와 의학 지식을 측정하기 위해 선택한 MCQ로 이루어져 있다. 각 문항은 지시문과 네 개 혹은 다섯 개의 보기가 있다. 피시험자는 정답에 해당하는 것을 골라서 답안지를 작성한다. 오직 한 개만이 정답이다. 지시문에는 환자 증례, 임상 소견, 도표 등이 제시될 수 있다. 그림이나 적절한 임상 정보를 별지로 주기도 한다. 수련 중 시험(in-training examination)은 대개 해당 학회별로 치르는 경우가 많으며 대개 MCQ 시험인 경우가 많다.

컴퓨터로 치르는 시험은 시험문항이 컴퓨터 모니터에 나타나고 그림이나 도표도 도표에 직접 나타난다. 이 경우 문제수가 적은 경우가 많은데 그 이유는 컴

퓨터 프로그램에서 피시험자의 능력을 바로 측정해서 문제를 선택할 수 있기 때문이다.

2) **유용성**: 의학지식과 이에 대한 이해도는 MCQ 시험으로 측정할 수 있다. 수련 중 시험 점수를 비교해서 전국 통계와 비교하면 개별 전공의들이 강한 부분과 약한 부분을 알 수 있고 이에 따라 특별한 도움을 줄 수 있다. 해당 전공의의 시험 점수 결과를 각 해마다 비교하면 전공의의 수련 경험이 향상되었는지 여부를 확인할 수 있다.

3) **실행가능성/실행성**: 해당 학회에서 심리측정 전문가의 도움을 받아서 문제를 개발한다.

표 2. 의학지식, 진료 중심 학습, 의사소통 영역에 필요한 기술과 이에 대한 평가 방법.

	평가 방법													
	필요한 기술	의무기록 검토	CSR	체크리스트	전반적 평가	SP	OSCE	모의실험과 모형화	360° 전 반적평가	포트폴리오	MCQ 시험	구술 시험	환자 증례 기록	환자 서베이
의학 지식	탐구적, 분석적 사고		1					2	3			1		
	기초과학 지식 습득 및 응용							2	3		1	1		
진료 중심 학습 과 향상	향상이 필요한 진료 내용 분석	2	2			2	2	3	3	1				2
	과학적 연구에서 근거 이용	1	1			3	2			1	1	1		
	연구와 통계 방법 적용		2	3	3						1	3		
	정보기술 이용					2	2			1	1			2
	다른 사람 학습 도와줌				2	3					1	3		
대인 관계 와 의사 소통 기술	환자와 치료적 관계 수립			3		1	1			2				1
	청취 기술			3		1	1			2				1

표 3. 직업주의, 체계에 기초한 진료 영역에 필요한 기술과 이에 대한 평가 방법

	평가 방법													
	필요한 기술	의무기록 검토	CSR	체크리스트	전반적 평가	SP	OSCE	모의실험과 모형화	360° 전반적 평가	포트폴리오	MCQ 시험	구술 시험	환자 증례 기록	환자 서베이
직업주의	존중하는, 이타적인			3			1		2					1
	윤리적으로 건전한 진료		2					2	1	3				2
	문화, 나이, 성별, 장애에 예민하게 반응		2	2			1		1	3	2			2
체계 에 기초한 진료	진료와 더 큰 체계와 반응하는 것을 이해						2			3				
	진료와 전달 체계에 대한 지식		2				3			2				
	비용-효과적인 진료	3		1					2					
	보건의료 체계 내에서 환자 옹호			3			2		1	2				1

여러 수련 영역에 따른 평가 방법⁵⁾

참 고 문 헌

논란은 있지만 수련 영역은 환자진료, 의학지식, 진료 중심 학습과 향상(Learning & Improvement), 대인관계 의사 소통 기술, 전문주의(Professionalism), 체계중심진료(Systems-Based Practice)로 구별할 수 있다. 각 영역과 각 영역별에 필요한 기술 그리고 이 기술을 평가하는데 가장 적합한 평가 방법은 표 1-3 과 같다.

1. Thomas G, Quattebaum. Technique for evaluating residents and residency programs. Pediatrics 1996; 98(6):1277-83.
2. 박훈기. 전공의평가: 형성평가. 가정의학회지. 2000. 21(11s):463-7.
3. <http://www.acgme.org/outcome/assess/keyConsider.asp>
4. <http://www.acgme.org/outcome/assess/toolbox.asp>
5. <http://www.acgme.org/Outcome/assess/ToolTable.pdf>