

성과중심교육 측면에서 우리나라 의과대학 학생평가의 현실과 과제

박장희

연세대학교 의과대학 의학교육학과

Current and Future Challenges of Student Assessment in Medical Education from an Outcome-based Education Perspective

Jang Hee Park

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

Most medical colleges in Korea have been shifting from traditional education to outcome-based education, which is the general trend in medical education. The purpose of this study was to make some suggestions in light of the reality and challenges of student assessment in medical education from the perspective of outcome-based education. First, those who are responsible for student assessment should be diversified to include faculty, residents, students, and evaluation committee members. They need separate roles in educational evaluation, so evaluation competencies are required for them. Second, various methods for evaluation and score interpretation can be used for effective evaluation. We can adopt diagnostic, formative, and summative evaluation functionally, and the norm-referenced, criterion-referenced, growth-referenced, and ability-referenced evaluation based on criteria for score interpretation. Finally, various evaluation domains and test forms can be administered together in the common lectures in the medical school. We can test not only knowledge but also skills and attitudes, with diverse test forms such as supply and performance types.

Corresponding author

Jang Hee Park
Department of Medical Education,
Yonsei University College of Medicine,
50 Yonsei-ro, Seodaemun-gu, Seoul
120-752, Korea
Tel: +82-2-2228-2507
Fax: +82-2-364-5450
E-mail: jhpark2@yuhs.ac,
janepark2@hanmail.net

Received: October 4, 2013

Revised: October 21, 2013

Accepted: October 22, 2013

Keywords: Medical education, Student evaluation, Outcome-based education

서론

평가(evaluation)란 우리의 일상생활에서 벌어지는 각종 선택과 관련된다. 출근은 어떻게 해야 할지, 점심은 무엇을 먹을지, 어떤 물건을 사야 할지 등 일상생활에서 벌어지는 각종 선택의 상황에서 나름의 기준을 세우고 판단(결정)하는 것이다. 최종 판단에 있어 관련 자료를 수집하고, 적절한 기준을 세우는 것이 최선의 선택을 할 수 있는 방법이다. 교육평가(educational evaluation)란 교육이 벌어지는 장면에서 일어나는 선택의 상황에서 각종 정보수집과 판단을 내리는 것이다. Sung (2009)은 교육평가란 교육과 관련된 모든 것의 양, 정도, 질, 가치, 장점 등을 체계적으로 측정하여 판단하는 주관적 행위로서 교육목적에 대한 가치를 판단하는 행위라고 보다 광범위하게 정의하고 있다. 이러한 맥락에서 의학교육에서 일

어나는 학생성취와 관련된 정보수집과 판단에 관한 것을 의학교육 평가라고 말할 수 있다. 평가의 대상은 인적 대상으로 학생, 교수, 학부모, 교육행정이 등이 있으며, 물적 대상으로 교수개발프로그램 등의 소프트웨어와 교육환경과 예산의 하드웨어 그리고 평가의 평가로 분류할 수 있다(Kim, 2009; Sung, 2009). 본 연구에서는 학생 대상의 평가에 한정하여 기술하겠다.

현재 의학교육은 학생의 역량 설정과 이에 대한 구체적인 성과(outcome)중심으로 환경이 변화하고 있다. 이는 가르치는 사람 위주의 교육에서 배운 사람이 실제로 습득한 성과를 규정하고 이를 중심으로 교육과정과 평가를 실시하는 것이다(Harden, 2007). 성과중심교육은 지식 위주의 교육에서 직업전문성 등을 포함한 전인적 차원의 교육프로그램을 만들고, 학생들이 도달하고자 하는 수준을 학습성과로서 구체적으로 설정하고, 학습성과 항목을 평가

항목으로 활용하여 학습자들의 달성 정도를 평가하여 확인하는 것까지 포함한다. 단적으로 말하면 평가할 내용과 수준을 시기별로 정해 놓고 이를 주기적으로 학생이 도달했는지 평가를 통해 학습성과를 확인하겠다는 것이다. 이 과정에서 많은 의과대학 교수들은 ‘형성평가’, ‘포트폴리오평가’, ‘합격선 설정’, ‘행동용어적 목표 기술’, ‘도달수준 설정’ 등 생소한 평가 관련 용어들과 접하게 된다. 성과비탕교육은 1985년 미국 의과대학을 대상으로 하는 인증평가의 기준으로 채택되었고, 우리나라도 2012년 ‘post 2주기 의과대학 평가인증기준’으로 발표되면서 교육평가 관련 지식은 단순한 학문적 경향이 아니라 대학의 생존을 위해 반드시 알아야 하는 분야가 되었다(Ahn & Yang, 2013; Han, 2013). 많은 대학들이 이러한 의학 교육의 흐름에 발맞추기 위해 성과중심교육체제로 변화하면서 가장 난제로 여기는 것이 학생평가와 관련된 부분이다. 이에 본 논문에서는 본문에서 의과대학의 학생평가를 성과비탕교육과 연계하여 그 문제점과 과제를 중심으로 몇 가지 대안을 제시하였다.

성과중심교육에서 바람직한 학생평가의 10가지 제안

본문에서는 성과중심교육 측면에서 효과적인 학생평가방법을 현실과 과제를 근거로 제안하였다. 제안 항목은 크게 평가인력, 평가기능, 참조준거, 평가내용, 평가방법으로 구분하여 기술하였다.

1. 평가인력

〈제안 1〉 교육과정 운영자들의 평가전문성이 강화되어야 한다.

의과대학은 성과중심교육체제로의 변화를 요구받고 있다. 교육자들이 나름의 철학을 가지고 학생들을 교육하고 평가하는 체제에서, 학생들의 학습성과(결과)를 측정 가능한 수준에서 미리 설정하고 교육 후 이에 대한 검증을 받아야 하는 체제로 변화하고 있다. 그러므로 평가의 결과가 교육의 질을 평가하는 중요한 기준이 된다(Ahn & Yang, 2013; Im, 2013; McAleer, 2005). 이러한 체제로 변화하기 위해서는 대학은 시기나 과정별로 측정이 가능한 수준과 방법을 사전에 고려하여 학습목표를 설정해야 한다. 교수들은 측정 가능한 학습목표(성과)로 기존 강의를 변화시키는 것뿐만 아니라, 새롭게 요구되는 형성평가, 학습시기별 성과, 새로운 교육방법의 도입과 평가 등 기존에 익숙하지 않은 용어들로 인해 지적인 피로감을 호소하고 있다.

의과대학교육의 특징 중 하나가 대부분의 과정이 오톨버스강의로 진행되어 있고, 교수들은 전체 수업 안에 부분적으로 참여하는 것이다. 학문의 융합 차원에서는 바람직하지만 전체 교육과정의 철학이나 원리에 대한 공유가 없이는 교수 개인의 교육철학이나 능력에 의해 자칫 교육의 내용이나 질이 달라질 수 있다.

성과중심교육체제로의 변화에서는 성과를 확인하는 것이 관건이다. 학습성과를 확인하는 평가방법에 대한 다양한 접근과, 시기

별, 과정별 성과를 주기적으로 평가 및 관리하기 위해서는 이를 의과대학 중앙에서 관리하는 위원회나 의학교육 주체들이 평가에 대한 이해가 필요하다.

대학의 평가 관련 위원회들은 매우 다르게 운영되고 있고, 위원마다 새로운 평가방법에 대한 이해 정도에 차이가 있으며, 대부분 평가에 관한 총괄적이면서 조직 차원의 비전이 없다(Elizondo-Montemayor, 2004). 우리나라도 이와 크게 다르지 않을 것이다. 그러므로 의과대학 전체 교육과정을 계획하고, 주기적으로 학습성과를 검증하고 필요한 교육적 조치나 판단을 내리는 관련 주체들의 교육평가 전문성이 강화되어야 한다. 개별 교수들은 능력에 상관없이 전체 과정의 부분으로 참여하기 때문에 전체적인 가이드와 관리해주는 전문성 있는 주체가 더욱 필요하다.

〈제안 2〉 평가에 다양한 인력을 활용해야 한다(위원회평가, 전공의 평가, 동료평가, 자기평가).

진료와 연구를 우선으로 생각하는 대부분의 의과대학 교수들에게 다양한 교육방법과 평가방법을 이용하라고 강조하는 것은 상당한 스트레스로 작용할 수 있다. 더군다나 평가기능에 따라 진단, 형성, 종합평가까지 평가의 시기가 다양화하는 것도 그렇다. 강의 이외의 방법으로 수업을 진행하고 이를 반영한 평가를 하려면 수업 및 평가를 도와주는 추가 인력이 필요하다.

평가적 측면만 보면 평가인력을 다양하게 활용할 필요가 있다(Amin & Khoo, 2006). 먼저 학생의 자기평가(self-assessment)를 늘려야 한다. 이는 진단평가 시 스스로 수업 관련 학습성과에 비추어 자신의 능력을 평가할 수 있으며, 그 결과를 교수는 수업에 활용할 수 있다. 다양한 학습활동과 소그룹활동에서도 활용할 수 있다. 더불어 학생들이 서로를 평가하는 동료평가를 늘릴 필요가 있다. 소그룹활동 등에서 학생들은 서로에 대해 가장 잘 파악하고 있으므로 이를 평가에 반영할 수 있다. 자기평가와 동료평가에 구체적인 평가기준을 제시한다면 평가과정에서 학생들의 자기반성 및 성찰의 기회로 삼을 수 있는 의미 있는 평가가 될 수 있다. 또한 전공의나 대학원생을 교육자로 합류시킬 수 있다. 이들에게 교육내용과 평가방법 등을 사전교육하고 교육이나 평가에 임하도록 한다면 교수는 평가부담을 줄일 수 있을 뿐 아니라 전공의나 대학원생들은 평가내용과 방법 등을 익힐 수 있는 기회가 되며, 교육자의 역할을 수행하므로 교육에 대한 흥미와 자신감도 높일 수 있을 것이다. 외국의 경우는 전공의에게 교육자로서 역할수행을 위한 교육 및 훈련을 체계적으로 시행하고 있다(Association of American Medical Colleges, 2000; Mann et al., 2007; Morrison et al., 2005). 아울러 교수들 간에도 평가에 관한 업무분담을 해야 할 것이다. 과정과 시기에 따라 주기적인 평가를 시행한다면 과정별 학습성과(course outcomes) 담당자와 시기별 학습성과(phase outcomes) 담당자와도 적절한 업무분담 및 협력이 필요하다. 이때 과정별 학습성과는 과정

별(인체의 정상구조, 성장과 노화 등)로 구분되는 학습성과를, 시기별 학습성과는 학년별 또는 기초의학, 임상의학, 임상실습 등 시기별 학습성과를 의미한다(Korean Institute of Medical Education and Evaluation, 2012).

2. 평가기능

〈제안 3〉 평가기능에 따른 평가가 제대로 이뤄져야 한다(진단평가, 형성평가, 총합평가).

현대 의학교육평가의 방향은 직무수행의 실제 세계(real world)에 필요한 역량을 종합적으로 평가하여 피평가자의 능력을 타당하고 신뢰 있게 평가하자는 것이다. 이는 평가목적에 따라 지속적으로, 여러 차례에, 다양한 형태로, 다양한 사람이 평가하는 것을 추구한다는 의미이다. 지속적으로 여러 차례 평가하는 차원에서 고려해야 하는 것이 평가의 기능이다. 평가는 기능에 따라 진단평가, 형성평가, 총합평가로 분류된다. 이는 의사가 환자를 진료하고 치료하는 과정과 유사하다. 진단평가는 환자를 치료하기 이전에 환자의 상태를 다양한 방법으로 진단하듯이 수업에 들어가기 전에 가르쳐야 하는 학생들의 학습상태를 다양한 방법으로 확인하는 것이다. 형성평가는 치료의 과정에서 환자에서 투여한 의료적 조치들이 효과를 발휘하는지를 환자에게 질문이나 각종 검사를 통해 확인하고 필요한 경우 다양한 처방을 내리는 것과 유사하다. 교육자가 가르치는 과정에서 피교육자의 학습상황을 점검하여 피드백해 주고 개별적으로 필요한 조치를 내려 주기도 한다. 총합평가는 환자의 치료를 중단하기 이전에 최종 검사를 통해 완치 등의 판단을 내리는 것과 유사하다. 학습과정을 모두 마친 학생들에 대해 최종 성적을 통해 교육적 판단을 내리는 것이다.

1) 진단평가

진단평가(diagnostic evaluation)는 의학교육에서 잘 언급되고 있지 않다. 학습자에게 필요한 진단목적은 지식적으로는 해당 수업과 관련된 선행지식의 정도를 파악하는 것이고, 수업에 대한 흥미도 등을 파악하는 것이다. 진단평가의 방법은 형식적인 방법과 비형식적 방법이 있다. 퀴즈 등 간단한 시험을 통해 형식적인 방법으로 점검할 수 있다. 하지만 옴니버스형태의 블록(block)강의에서 몇 시간만을 할당받은 교수가 진단평가하기란 쉽지 않다. 이를 위해서 교수들은 동일 강의의 작년 성취 정도와 강의평가 결과를 다시 확인하거나 현 강의 이전 과정에서의 학생들 성취 정도나 강의평가 등의 결과를 비형식적으로 점검할 필요가 있다. 아울러 교육과정 담당자가 수업의뢰를 할 때 전체 교육과정과 사전학습내용과 각 수업에서 기대하는 학습성과를 자세하게 제시한다면 수업을 계획하는데 도움을 줄 것이다. 이러한 것들이 어렵다면 수업을 하면서 중요한 사전지식에 관해 어느 정도 알고 있는지 학생에게 간단한 질문들을 통해 확인하거나 학습성과 자기점검표를 작성하여 학생 스

로 점검토록 하는 방법도 있다.

2) 형성평가

형성평가(formative evaluation)는 1967년 Michael Scriven에 의해 소개된 개념으로 교육과정 중간에 시행하는 평가이다. 평가결과를 통해 학생에게는 성취 정도 등에 대해 가르치는 사람에게 교육내용 및 방법 등 개선에 필요한 정보를 제공한다(Sung, 2009). 형성평가는 학생에게 학습성취 등에 대해 피드백(feedback)을 주고 교육프로그램을 개선하는 것을 주목적으로 하는 평가이다. 일반적으로 평가결과를 성적에 반영하지 않는다. 하지만 형성평가를 일선에서 반영하기란 쉽지 않다. 성적에 민감한 의과대학 학생들이 성적에 반영되지도 않는 시험을 제대로 볼 것인가, 평가방법은 어떻게 해야 하는가, 피드백은 개별교수가 일일이 학생들을 대상으로 주어야 하는가, 어떤 방법으로 평가할 것인가, 블록에서 일부 강의만 담당하는 교수들이 형성평가까지 해야 하는가 등의 문제 등이 도출될 수 있다.

형성평가를 교육과정에 정착하기 위한 방안을 생각할 수 있다. 의사가 환자의 치료과정에서 환자에게 질문을 통해 확인하는 경우와 검사를 통해 확인하는 방법이 있을 것이다. 이와 같이 수업의 과정에서 강의만 하는 것이 아니라 수업과 관련된 질문들을 수시로 다양한 학생들에게 함으로 확인할 수도 있고, 전체 과정에서 일정 기간이 경과한 이후에 해당 시점까지 배운 내용을 간단하게 평가할 수도 있다. 평가할 경우 컴퓨터기반시험체제가 갖춰진 경우에는 기출문제를 이용한 시험을 통해 간단하게 성취 정도를 확인하거나 간단한 문항형태를 활용하면 부담이 적을 것이다. 간단한 문항의 예가 진위형(일명 OX문제) 문항(absolutely true-false type)이다. 진위형 문항은 추측도(확률)가 50%로 높아서 총합평가와 같은 고부담시험에서는 잘 활용되지 않는다. 하지만 문항 출제가 용이하고 많은 내용을 빠른 시간에 평가할 수 있고 문항 수가 많을수록 신뢰도도 상승한다는 장점이 있어, 성적반영률이 낮고 학생의 최저 수행 정도만 확인하는 차원에서는 적절하다고 볼 수 있다. 추측률을 낮추기 위해서는 틀릴 경우(X일 경우) 틀린 부분을 수정하라고 할 수도 있다.

형성평가에 대한 피드백은 개별면담이 가장 좋겠으나, 각 수업 차원에서는 결과와 피드백을 간단하게 서면이나 인터넷으로 공지하는 것도 방법일 수 있고, 개별면담을 하되 대상을 성적하위집단, 무작위추출집단, 또는 두 방법을 혼합할 수 있을 것이다. 아울러 형성평가는 피드백이 주목적이긴 하지만 성적에 민감한 의과대학의 특성을 반영하여 학습자의 학습동기 차원에서 최소한의 성적을 반영하거나, 형성평가의 일부 내용을 총합평가에서 형태를 달리 하여 출제할 수 있을 것이다. 수업시간이 상대적으로 짧은 과정은 비형식적 형성평가만을 활용할 수도 있다.

3) 총합평가

총합평가(summative evaluation)는 수업이 종료된 이후에 학습 성과 도달 여부를 종합적으로 판정하는 평가이다(Sung, 2009) (Summative evaluation은 ‘총괄평가’라는 용어로 흔히 사용되기도 하지만 학생성취 관련 자료를 모두 모아서 최종 판단한다는 의미를 살리기 위해 ‘총합평가’라는 용어를 사용하겠다). 학생들의 최종 성적 산출을 위한 평가로 의과대학에서 가장 많이 사용되는 것이다. 총합평가는 배운 내용을 모두 포함하는 평가이다. 앞에서 언급했지만 형성평가의 일부 내용을 총합평가에 반영할 수 있을 것이다. 왜냐하면 형성평가는 각 과정에서 최소한의 중요한 내용들로 문항을 구성했기 때문이다. 물론 총합평가에 포함시킬 때는 동일 내용을 문항수준이나 평가형태를 달리하여 출제할 수 있으며, 이를 통해 형성평가에 대한 동기도 함께 유발할 수 있다.

3. 참조준거

학생이 시험을 통해 90점이라는 점수를 받았으면 성취한 점수가 학생이 학습을 통해서 성취한 성과와 관련해 어떤 의미가 있는 것인지에 대해 해석과 판단이 필요하며, 이를 위해서 기준이 필요하다. 가장 흔한 평가가 다른 학생들 성적에 비교해 해석하는 규준참조평가(norm-referenced evaluation)로서 상대평가로 불린다. 이에 비하여 합격기준에 근거해 합격 여부를 판정하는 법이 절대평가로 불리는 준거참조평가(criterion-referenced evaluation)이다. 이외에도 학생의 수업 시작시점의 성적에 비해(초기치) 얼마나 어느 방향으로 성적 변화가 있었는가를 통해 해석하는 성장참조평가(growth-referenced evaluation)가 있고, 학생의 능력(예, 입학성적, 지난 학기 평점 등)에 근거해 성취 정도를 해석하는 능력참조평가(ability-referenced evaluation)가 있다. 이 중에서 성과중심교육과 관련하여 준거참조평가와 성장참조평가를 논의하겠다.

〈제안 4〉 준거참조평가로 학생의 학습성취 도달 여부를 평가해야 한다.

준거참조평가(criterion-referenced evaluation)는 학생의 학습을 통해서 성취해야 할 교육목표의 기준을 준거점수(예, 합격점수 등)로서 미리 설정하고 이를 기준으로 학생점수를 해석하는 것이다. 만일 합격 여부만 결정할 때는 준거가 한 개 필요하고, 상·중·하로 구분하기 위해서는 준거가 두 개 필요하다. 성과중심교육에서는 다양한 평가를 통해 학습성과를 파악하고 피드백을 주며, 다음 시기로 나아가야 하는 등의 결정을 해야 한다. 이때 의사결정에 해당하는 점수를 타당하고 신뢰 있게 설정해야 한다.

준거참조평가에서는 준거설정이 중요하다. 준거를 설정하는 방법은 평가의 목적과 형태에 따라 다양하며 같은 집단에 대한 준거 설정이라도 사용하는 방법에 따라 그 결과가 다르게 나올 수 있다. 준거설정방법으로는 규준적인 준거설정, 피험자집단의 특성을 고

려한 준거설정, 검사도구의 내용분석을 근거로 한 준거설정 그리고 방법들을 병행한 준거설정방법들이 있다. 다양한 준거설정방법 중에서 평가의 목적과 현실에 맞는 것을 선택하되 병행하여 쓰는 방법으로 신뢰도를 높일 수 있다.

각 대학이 성과중심교육을 시행하면서 가장 먼저 직면하게 되는 것이 준거참조평가를 의과대학에 도입해야 하는 것이다. 의과대학이 기존의 규준참조평가체제에서 준거참조평가체제로 변화하기 위해서는 여러 가지 어려움이 따르리라 예상된다. 현재 의사국가시험의 실기시험은 준거참조평가를 시행하고 있으며, 필기시험에서도 준거참조평가를 시행하기 위해 연구가 진행되고 있다고 밝혔다(National Health Personnel Licensing Examination Board, 2013). 교육과학기술부도 준거참조평가인 ‘성취평가제’를 2013년에는 현재 중학교 1, 2학년부, 2016년에는 모든 중고등학교에 실시할 것을 공표하였다(Korea Institute for Curriculum and Evaluation, 2013).

〈제안 5〉 학습성장검사로 성장참조평가해야 한다.

성장참조평가(growth-referenced evaluation)는 교육과정을 통하여 학생의 능력이 얼마나 성장하였는가에 관심을 두는 평가이다(Sung, 2009). 성장참조평가는 학습의 초반 상황(초기치)에 비해 교육이 진행되면서 학생 성취 정도가 얼마나 변화했는지(변화율)를 통해 학생을 평가하는 것이다. 개인적인 차원에서는 학생상담 등에 많이 활용되고, 국가적인 차원에서는 학습자와 소속 학교의 성취 등 변화와 그 원인을 파악하는 데 활용된다. 성장참조평가를 하기 위해서는 동일 대상에 대한 여러 번의 시험결과를 바탕으로 의사결정 등이 이뤄진다. 초·중·고등학생 대상으로 학생의 성장을 장기적으로 추적하는 기관별 평가로는 한국교육과정평가원의 ‘국가수준학업성취도평가(Korea Institute for Curriculum and Evaluation, 2013),’ 한국교육개발원의 ‘한국교육중단연구(Korean Educational Development Institute, 2013),’ 한국청소년개발원의 ‘한국아동·청소년패널조사(National Youth Policy Institute, 2013)’ 등이 있다.

의학교육에서는 학습성과와 관련하여 progress test가 소개되고 있다(Friedman, 2005). Progress test는 직역하면 진도검사, 진척검사, 진보검사 등으로 번역할 수 있다. 하지만 저자는 ‘학습성장검사(progress test)’로 번역을 제안한다. 이는 성장참조평가와 연계할 수 있으며, 경험중심교육과정에서 교육의 목적은 학생경험의 성장과 진보라고 정의하고 있고, 교육의 목적은 학생의 바람직한 변화라는 일반적인 정의와도 맥을 같이하기 때문이다.

졸업성과를 의과대학의 최종 목표로 설정했다면 졸업성과에 관련해 학생의 초기에는 능력이 어느 정도였는지, 과정을 거치며 시기별 평가를 통해 어느 정도 목표치에 다다르고 있는지 확인할 필요가 있다. 이때 달성 여부 확인과 더불어 그 변화 정도를 파악해야

한다. 변화가 많거나 없거나 정제될 때 각 시기에는 어떤 특성이 있는지, 이러한 변화는 초기 능력과 어떤 연관이 있는지 등을 분석을 통해 확인할 수 있기 때문이다. 분석결과는 교육과정 전반에 다양한 정보를 제공해 주고 개인 차원에 피드백을 제공할 수 있다. 이를 위해서는 측정학적인 설계와 장기간의 자료수집이 필요하다.

4. 평가내용

〈제안 6〉 대학의 모든 평가가 학생의 졸업역량 달성도를 평가할 때 내용타당도가 향상된다.

타당도(Validity)란 한 검사 혹은 평가도구가 ‘측정하려고 의도하는 것’을 어느 정도로 충실히 측정하고 있는가로 간단하게 검사도구 목적의 적합성(適合性)이다(Sung, 2009). 내용타당도란 검사내용이 검사도구의 목적에 적합한 정도를 말하며, 여기에는 교과타당도(curriculum validity)와 교수타당도(instructional validity)로 구분된다. 교과타당도는 검사가 교육과정에 있는 내용을 얼마나 잘 포함하고 있으나의 문제이고, 교수타당도는 교수· 학습 중에 가르치고 배운 내용이 얼마나 포함되었느냐를 말한다(Sung, 2009). 즉 수업한 내용이 평가에 얼마나 반영되었는지를 분석하는 것이다. 만일 수업은 교수가 하고 시험문제는 문제집에 있는 문제를 그대로 인용하여 출제하거나 전공의 등에게 맡겨서 출제하여 가르친 내용과 평가한 내용이 서로 맞지 않는다면 낮은 타당도를 보일 것이다.

학습성과 측면에서 보면 각 대학이 졸업역량과 각 시기 및 단계별 성과를 설정하였다면 이것이 반드시 각 수업에 반영되어 평가로

이뤄져야 한다. Figure 1과 같이 대학의 모든 구성원이 공동의 목표 달성을 위해 실제 교육에서 노력한다면 보다 높은 내용타당도를 이룰 것이다. 즉 타당도란 과녁의 중심과 같이 검사의 내용이 목적에 얼마나 부합되었는가가기 때문이다(Sung, 2009).

〈제안 7〉 기출문제의 반복 출제는 학생의 반응과정에 기초한 타당도 근거를 낮춘다.

타당도 중 반응과정에 기초한 근거(evidence based on response process)란 문항 출제자의 의도에 맞게 학생들이 반응했는가를 질적으로 분석하는 것이다(Sung, 2009). 예를 들어 출제자가 문제해결형 문항이라고 출제하였으나 학생들은 기출문제로 이미 답을 알았다면 암기형 문항수준으로 떨어져 타당도가 낮은 문항으로 간주된다. 의과대학은 정보공유문화가 체계적으로 구축되어 있어 문항을 출제하는 교수들에게는 상당한 부담이 된다. 실제 좋은 문제임에도 불구하고 문항분석을 하면 분별도가 상당히 낮은 문항으로 처리되는 경우들이 많은데 이는 대부분 기출문제였다. 꼭 알아야 하는 문항이라 중복 출제가 불가피하다면 문항의 수준이나 형태를 바꾸어서 출제하는 것이 타당도를 높일 수 있다.

5. 평가방법

〈제안 8〉 평가기능과 참조준거에 맞게 이(삼)원분류표를 작성해야 한다.

앞에서 평가기능에 따라서는 진단, 형성, 총합평가가 있으며, 참조준거에 따라서는 기준, 준거, 성장, 능력참조평가를 제시하였다. 이 두 가지를 연관하기 위한 비유로 운동경기 중 허들을 예로 들 수 있다. 운동선수가 출발하기 전 출발선에 제대로 섰는지, 자세는 적절한지, 부정출발은 없는지를 확인하는데 이는 준거참조평가에 따라 진단평가를 하는 것과 유사하다. 출발신호와 함께 선수는 달리면서 허들을 차레로 넘어서 달려간다. 이때 각 허들을 제대로 뛰어 넘는지 평가하는 것은 형성평가를 준거참조평가하는 것과 유사하며, 도착지점을 제대로 빠르게 통과한 사람 순으로 메달을 부여하는 것은 총합평가를 기준참조평가하는 것과 유사하다. 각각의 평가는 그 목적에 따른 기준에 맞게 구성되고 해석된다. 현재 지식 영

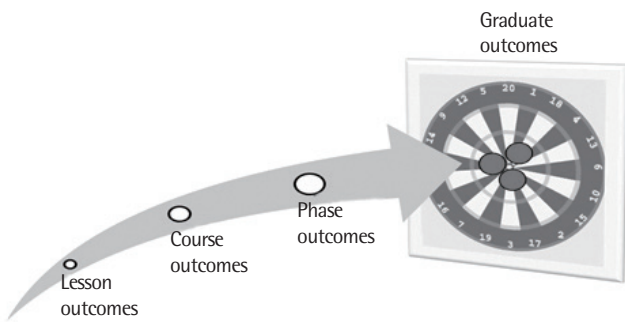


Figure 1. The elevating the content validity in outcome-based education.

Table 1. The comparison of diagnostic, formative & summative evaluation

Evaluation	Diagnostic evaluation	Formative evaluation	Summative evaluation
Time	Before instruction	During instruction	After instruction
Purpose	To provide an adequate program	To improve the program/feedback	To make decisions about the program/accountability
Method	Formal, informal	Frequent/formal, informal	Formal
By whom	Instructor, content expert	Instructor	Content expert, evaluation expert
Standard	Criterion	Criterion	Norm or criterion
Content level of test item	Comply with criterion level	Comply with criterion level	Norm-referenced: various difficulty levels Criterion-referenced: comply with criterion level

From Sung, T. J. (2009). Educational evaluation. Seoul: Hakjisa.

역의 평가에 있어서 우리나라 의과대학은 주로 총합평가를 기준점 조평가하는 경향이 많다.

평가를 위해서는 각 평가기능과 참조준거에 맞는 출제계획을 세워야 한다. 많은 대학이 대학에서 출제계획을 세울 때는 주로 각 수업별(교수별) 출제문항 수와 배점을 고려한다. 시험범위는 배운 내용 전체이고, 학생들에게는 특히 지식 영역에서 평가와 관련된 별다른 정보를 제공하지 않는 것이 일반적이다.

평가를 위해서는 2(3)가지 차원에서 계획하는 이(삼)원분류표를 작성할 필요가 있다. 3가지 차원이란 평가내용범위(시험범위), 각 문항의 내용수준, 평가문항의 맥락(외래, 입원, 응급 등)이다. 맥락을 제외한 이원분류표가 주로 교육현장에서 활용되고 있다. 출제계획에 대한 사항은 수업 전에 작성되어야 수업 시 평가내용을 고려하면서 수업을 진행할 수 있다.

평가기능에 따른 진단, 형성, 총합평가와 참조준거를 고려하여 구성한 것이 Table 1이다. 평가내용(시험범위)에서 학교에서 합의한 각 수준의 학습성도가 반영되어야 한다. 이를 통해 졸업역량과 맥을 같이하여 평가도구의 내용타당도를 올릴 수 있다. 진단평가는 주로 수업 진행을 위한 학생의 능력(주로 출발점행동)인 최소수행능력 확인을 위해 준거참조평가에 의해 이뤄지는데, 형식적으로 평가할 때는 수업 전까지 내용 중 최소능력수준에 해당하는 문항수준의 내용을 평가한다. 형성평가는 각 수업의 과정 도중에 이루어지는 것으로 상대적으로 범위가 좁으며, 총합평가는 배운 내용을 모두 포함하므로 범위가 가장 넓다.

문항수준은 진단평가와 형성평가가 주로 준거참조평가를 하기 때문에 수업 전의 최소능력을 기준으로, 형성평가는 수업과정에서의 최소능력을 기준으로 준거에 맞는 수준의 문항들로 구성한다. 총합평가를 기준점조평가할 때는 다양한 지식수준의 분별도 높은 문항으로 학생들을 능력별로 구분할 수 있도록 구성한다.

〈제안 9〉 문항형태별 문항제작원리를 준수해서 문항을 제작해야 한다.

각 대학이 교수대상으로 문항제작법을 워크숍을 통해 교육시키고 있지만, 실제 의과대학에서 이뤄지는 문제들을 검토해 보면 기본적인 문항제작원리에 맞지 않은 문항들을 여전히 발견하게 된다. 출제자는 피험자가 문항 관련 능력이 있을 때 충분히 풀 수 있도록 문제를 구성해야 한다. 피험자가 문항 관련 능력 이외에 실수 등의 이유로 문제를 틀렸다면 이것은 점수의 오차(error)를 일으켜서 신뢰도에 영향을 준다.

의과대학에서 가장 많이 쓰이는 문항이 선택형이고 선택형 문항에서 오차를 줄이기 위한 몇 가지 문항제작원리를 생각해 볼 수 있다. 신뢰도란 측정결과의 일관성 정도이고 일관적인 결과를 얻기 위해서는 측정의 오차를 줄여야 한다. 문항구성 시 학생의 능력 이외의 것이 결과에 작용하지 않게 해야 한다. 즉 실력발휘를 제대로

할 수 있는 문항을 구성하기 위한 몇 가지를 정리하였다. 첫째, 문항은 짧고 명확하게 제시되어야 한다. 선택형 문항은 진술문과 답가지로 구성되는데, 진술문에서 임상상황을 반영한다고 장황하게 기술하는 경우가 있다. 임상상황에 여러 가지 단서를 제공하는 것은 좋지만, 일반시험 상황에서 너무 장황한 설명은 제한된 시간에 시험을 치러야 하는 학생들에게는 상당한 피로와 혼란을 줄 수 있다. 아울러 답가지들도 동일한 문장이나 단어의 반복을 피하고, 비슷한 문장구조로 구성해야 한다. 둘째, 실수로 틀릴 수 있는 것들에 대해서는 주의를 기울이도록 표시해 주어야 한다. 학생들이 실수하는 것들로는 부정문, 답의 개수, 정답 선택조건(최선정답이나 옳은 답이나) 등이 있다. 실수할 수 있는 부분에 대해서는 밑줄이나 굵은 글씨로 표시하여 주의를 요하도록 해야 한다. 셋째, 피험자들이 답가지만으로 정답을 유추할 수 있도록 하지 말아야 한다. 즉 답가지들이 서로 유사한 문장구조, 문항길이, 동일 영역의 내용들, 정답 번호가 치우치지 않도록 하는 등을 고려하여 피험자들에게 정답의 단서를 제공하지 않도록 해야 한다. 넷째, 틀린 답도 어느 정도 매력적이어야 한다. 정확한 답을 모르는 피험자들이 정답과 유사하게 간주하는 틀린 답들이 있어야 그 능력을 충분히 구분할 수 있기 때문이다. 이러한 문항들이 많을 경우 학생들의 능력을 구분하는 분별도가 상승한다.

〈제안 10〉 일반수업에서도 전체 영역을 반영하는 다양한 형태의 문항으로 평가해야 한다.

평가는 평가 영역에 따라 인지적 영역(cognitive domain), 심동적 영역(psychomotor domain), 정의적 영역(affective domain)으로 구분된다. 인지적 영역은 지식을, 심동적 영역은 신체와 관련된 기능의 숙달과 발달로, 정의적 영역은 인간의 마음과 관련된 특성으로 태도, 인성, 가치관, 도덕성 등을 포함한다. 의과대학의 졸업역량은 대부분 이 세 영역을 모두 포함하고 있다. 의과대학생의 역량과 관련하여 미국의과대학협의회인 Association of American Medical Colleges (1998)의 의과대학 공통의 학습목표는 knowledge, skillful, altruism, dutiful이다. 미국인증평가기관인 Accreditation Council for Graduate Medical Education (2006)은 medical knowledge, patient care, interpersonal & communication skills, system based practice, professionalism을, 영국인증기관인 General Medical Council (2009)은 Tomorrow's doctor에서 the doctor as a scholar and a scientist, a practitioner, a professional로 설정하였다. 우리나라 의과대학들도 다양한 영역을 포함한 졸업역량을 설정하였다.

대부분 의과대학이 지식전달 위주의 강의로 수업을 진행하고 있으며 선택형 위주의 문항형태로 학생을 평가하고 있다. 강의는 많은 사람들을 대상으로 많은 지식을 효율적으로 전달하는 데 유용한 교육방법이지만 학생들의 자발적인 사고과정을 이끄는 데 한계

가 있다. 선택형은 문항 출제는 용이하지 않지만 많은 내용을 평가할 수 있고, 신뢰도가 높기 때문에 고부담시험(high-stakes test)에서 주로 사용되고 있다. 하지만 학습자들이 주어진 선택형 답가지 안에서 사고하게 되고, 추측이나 시험요령 등으로 평가결과가 달라질 수 있는 한계가 있다.

지식수준을 평가함에 있어 선택형 문항(selection type) 이외에 서답형(supply type or essay type)이나 이를 포함한 수행형(performance type) 평가가 가능하다. 서답형 문항은 쉽게 주관식이라고 할 수 있는데, 수준에 따라 괄호형(close form), 완성형(completion form), 단답형(short-answer form), 논술형(essay type)이 있고, 논술의 채점용이성을 높이기 위해 논술의 내용을 제한하는 제한된 논술형이 있다. 수행형으로는 구술시험, 찬반토론에 대한 평가, 소그룹 토의, 과제(연구) 수행 및 발표 등을 통해 학생의 지식수준을 보다 심층적으로 파악하고, 태도, 자기주도력, 발표력 등을 함께 평가할 수 있다(Sung, 2009).

의과대학은 대부분 의학직업전문성(medical professionalism)을 졸업역량으로 포함하고 있는데, 직업전문성을 구분하여 별도의 교육을 시킬 수도 있지만, 기존의 지식 위주 평가방법을 논술형, 구술시험, 토의와 발표 등으로 전환한다면 지식을 보다 깊이 있게 평가할 수 있을 뿐 아니라 직업전문성과 관련된 의사소통능력, 리더십, 발표력, 협동심 등을 함께 평가할 수 있다(Huxham et al., 2012).

서답형이나 수행형 평가는 많은 주제를 다룰 수 없으므로 평가하는 내용이 수업의 과정을 대표할 만한 내용타당성이 확보되어야 하고 여러 상황을 고려한 자세한 채점기준표가 함께 개발되어야 한다. 채점자의 신뢰도 확보도 중요한 문제이므로 채점자 훈련을 통

해 신뢰도를 높이거나 성적반영률이 낮은 형성평가 등에 활용할 수 있다. 타당도가 높은 문항형태이므로 신뢰도 확보가 어렵고 시간노력이 필요하지만, 신뢰도가 높은 방법들(선택형 문항 등)을 함께 신뢰도를 보완한다면 의사로서의 종합적 능력 개발 및 평가에 효과가 클 것이다. 지금까지 본론에서 논의한 내용을 Table 2와 같이 요약할 수 있다.

결론

본 논문에서는 우리나라 의과대학에서 학생평가의 현황과 문제점을 학습성과와 연관하여 평가인력, 평가기능과 참조준거, 평가내용과 방법으로 나누어 논의하였다. 이들 논의를 근거로 한 종합적인 결론은 다음과 같다.

첫째, 성과중심교육에는 학생에 대한 여러 평가가 주기적으로 이뤄지므로 많은 평가인력이 필요하고 이들에 대해 평가전문성이 요구된다. 아울러 평가인력들 간에 효율적인 역할분담도 필수적이다. 성과중심교육에서 학습성과를 중심으로 한 교육이란 학생들이 도달해야 할 성과(outcomes)를 미리 정하고 이에 도달했는지 여부를 확인하여 교육적 판단을 내리는 것이다. Post 2주기 의과대학 평가인증기준이 발표되면서 각 의과대학은 성과중심교육의 교육과정 운영에 많은 관심을 기울이고 있다. 각 수준에서 학습성과에 도달했는지 여부를 지속적으로 평가하고 이를 기준으로 학생에게 피드백하고 교육프로그램 등을 운영 및 개선하기 위해서는 체계적이고 지속적인 평가가 필요하다. 의과대학 교수 중 의학교육 관련 지식은 대체로 부족한 상태인데, 평가전문성까지 익히게 하기는 쉽지

Table 2. Ten suggestions for effective student assessment in outcome-based education

Category	Suggestion	Present	Future challenge
Personnel	1) Strengthen the evaluation expertise of course facilitator	Unfamiliarity with outcome-based education	Establish an evaluation committee and strengthen their expertise on educational evaluation
	2) Utilizing a diverse workforce for evaluation	Professor-led evaluation and residents' assistance	Student assessment by the professors, students themselves, residents, and committee
Function	3) Evaluation based on function	Most summative evaluation	Adopting evaluation based on function, such as diagnosis, formative, and summative evaluation
Reference	4) Score interpretation with various references	Most norm-referenced evaluation	Criterion-referenced and growth-referenced evaluation
Content	5) Improve the content validity	Individual lessons without considering the graduate outcomes of the college	All lessons considering the graduate outcomes of the college
	6) Improve the response process validity	Recycle past test questions without revising them	If necessary, recycle past test question with revision
	7) Well-rounded education and evaluation	Most knowledge-intensive instruction and evaluation	An inquiry into well-rounded assessment including items about knowledge, skills, and attitudes
Method	8) Develop a three-way test specification table	Mostly one-way test specification tables	Develop three-way test specification tables with the range, level, and context of test items
	9) Fully meet the guidelines for developing test items	Do not fully meet the guidelines for developing test items	Develop the test items considering the validity and reliability
	10) Use various test item forms	Most multiple choice question about knowledge	Recommend the supply and performance test forms on professionalism, etc.

않을 것이다. 요구되는 평가인력 중 전공의에게 기초적인 교육자로서의 자질을 습득도록 하여 의학교육에 참여토록 할 수도 있다. 외국에서는 교육자로서 역할을 수행하는 전공의에게 resident as learner & teacher, resident as teacher, clinical teacher 등으로 명명하고 많은 교육훈련들이 이뤄지고 있다. 현재 우리나라도 ‘배우며 가르치는 전공의(resident as learner & teacher)’에 대한 연구가 진행되고 있다(Korean Institute of Medical Education and Evaluation, 2012). 학생들도 자기평가와 동료평가 등을 적극 활용하여, 태도 등 지식 이외의 학습 영역이 골고루 자극되고 평가될 수 있도록 해야 한다. 학생의 전인교육이 일상 교육현장에서 이뤄지도록 교육과정 운영자, 담당 교수, 학생, 전공의 등이 함께 역할을 분담하고 노력하는 체제로 구성할 수 있다.

둘째, 효과적인 학생평가를 위해 평가의 목적과 기능에 따라 다양한 평가방법과 점수해석방법을 활용해야 한다. 평가기능에 따라서는 진단평가, 형성평가, 총합평가가 있으며, 점수해석을 위한 참조기준은 규준참조평가, 준거참조평가, 성장참조평가, 능력참조평가가 있다. 이 중에서 성과중심교육에서는 학생의 성취 도달 여부를 기준으로 하는 준거참조평가와 학생의 능력발달 및 성장을 기준으로 하는 성장참조평가에 관심을 가질 필요가 있다. 준거참조평가는 준거설정방법이, 성장참조평가는 장기적인 평가 설계와 동등화(equating) 등의 통계적인 고려가 있어야 한다.

셋째, 일반수업에서도 학생의 다양한 영역을 다양한 문항형태로 평가해야 한다. 대부분 지식전달을 목적으로 하는 교육은 강의와 선택형 문항 위주의 평가로 이뤄졌다. 하지만 리더십, 직업전문성, 자기주도학습 등의 졸업역량을 이끌어 내려면 이와 같은 능력을 발휘하도록 수업과 평가에 반영되어야 한다. 지식습득도 자기주도학습이나 소그룹활동으로 이뤄지도록 하고 평가도 지필검사 이외에 서답형 문항, 수행형 문항으로 구성한다면 보다 높은 수준의 지식 평가와 더불어 태도와 관련된 각종 역량들도 함께 평가할 수 있을 것이다.

REFERENCES

Accreditation Council for Graduate Medical Education. (2013, October 1). *ACGME Common Program Requirement*. Retrieved from http://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/CPRs_07012016.pdf

Ahn, J. H., & Yang, E. (2013). An outcome-based approach in medical cur-

riculum development. *Korean Med Educ Rev*, 15(1), 9-18.

Amin, Z., & Khoo, H. E. (2006). *Basics in medical education*. Singapore: World Scientific Publishing.

Association of American Medical Colleges. (1998). *Learning objectives for medical student education: guideline for medical schools*. Washington (DC): Association of American Medical Colleges.

Association of American Medical Colleges. (2000). *Graduate medical education core curriculum: core curriculum working group report*. Washington (DC): Association of American Medical Colleges.

Elizondo-Montemayor, L. L. (2004). How we assess students using an holistic standardized assessment system. *Med Teach*, 26(5), 400-402.

Friedman, B. M. (2005). Principles of assessment. In J. A. Dent, & R. M. Harden (Eds.). *A practical guide for medical teachers* (2nd ed., pp. 282-292). Philadelphia: Churchill Livingstone.

General Medical Council. (2009). *Tomorrow's doctor*. Retrieved from http://www.gmc-uk.org/TomorrowsDoctors_2009.pdf_39260971.pdf

Han, J. J. (2013). The development of outcome-based curriculum in medical schools outside Korea. *Korean Med Educ Rev*, 15(1), 19-24.

Harden, R. M. (2007). Outcome-based education: the future is today. *Med Teach*, 29(7), 625-629.

Huxham, M., Campbell, F., & Westwood, J. (2012). Oral versus written assessments: a test of student performance and attitudes. *Assess Eval High Educ*, 37(1), 125-136.

Im, S. J. (2013). Assessment in outcome-based education. *Korean Med Educ Rev*, 15(1), 25-30.

Kim, S. W. (2009). *Educational evaluation*. Seoul: Hakjisa.

Korean Educational Development Institute. (2013, October 15). *Korean educational longitudinal study*. Retrieved from <http://eng.kedi.re.kr/khome/eng/archive/report/viewReports.do>

Korea Institute for Curriculum and Evaluation. (2013, October 10). *Criterion-referenced evaluation*. Retrieved from <http://assess.kice.re.kr/leaflet/sub1.html>

Korea Institute for Curriculum and Evaluation. (2013, October 15). *National level test*. Retrieved from http://naea.kice.re.kr/open_01a.jsp

Korean Institute of Medical Education and Evaluation. (2012). *The guideline of outcome-based system in medical educational curriculum*. Seoul: Korean Institute of Medical Education and Evaluation.

Mann, K. V., Sutton, E., & Frank, B. (2007). Twelve tips for preparing residents as teachers. *Med Teach*, 29(4), 301-306.

McAleer, S. (2005). Choosing assessment instruments. In J. A. Dent, & R. M. Harden (Eds.). *A practical guide for medical teachers* (2nd ed., pp. 302-310). Philadelphia: Churchill Livingstone.

Morrison, E. H., Shapiro, J. F., & Harthill, M. (2005). Resident doctors' understanding of their roles as clinical teachers. *Med Educ*, 39(2), 137-144.

National Health Personnel Licensing Examination Board. (2013). *Competency Based Assessment for Health Professions: the 21st celebration symposium*. Seoul: National Health Personnel Licensing Examination Board.

National Youth Policy Institute. (2013, October 1). *Korean youth panel*. Retrieved from <http://www.nypi.re.kr/eng/activities.np>

Sung, T. J. (2009). *Educational evaluation*. Seoul: Hakjisa.