

성과중심교육에서 학생평가

임선주

부산대학교 의학전문대학원 의학교육실

Assessment in Outcome-Based Education

Sun Ju Im

Department of Medical Education, Pusan National University School of Medicine, Yangsan, Korea

Assessment plays a vital role in outcome-based education (OBE). This article describes the characteristics of assessment and appropriate assessment instruments for measuring learning outcomes in OBE. Assessment in OBE needs to be formative, continuous, and frequent. Miller's pyramid is useful for selecting the appropriate assessment instruments to reflect a specific outcome; different methods can be applied to evaluate one outcome. Outcomes and competency mean that student must 'do'; therefore, performance tests are emphasized. Qualitative methods as well as quantitative methods are used to evaluate the outcomes of areas such as professionalism or ethics. An absolute criterion-based standard is usually applied to decide whether students pass or fail, but the decision should be based on gathering value judgments and reaching consensus. Active participation of faculty members and students in assessment is crucial.

Corresponding author

Sunju Im
Department of Medical Education,
Pusan National University School of
Medicine, 49 Busandaehak-ro,
Yangsan 626-815, Korea
Tel: +82-51-510-8021
Fax: +82-51-510-8125
E-mail: sunjuim11@hanmail.net

Received: January 31, 2013

Revised: February 18, 2013

Accepted: February 19, 2013

Keywords: Outcome assessment, Competency-based education

서론

성과중심교육에서 '성과(outcome)'는 교육과정을 이수한 학생들이 달성해야 하는 역량(competency)을 의미하고, 교수는 학생들에게 현재의 수준에 대해 지속적으로 피드백을 제공함으로써 모든 학생들이 성과에 도달할 수 있도록 해야 한다(Shumway et al., 2003). 이 과정에서 평가는 학생과 교수에게 현재의 수준에 대한 정보를 제공하는 핵심적인 역할을 하며, 적절한 평가 없이 성과중심교육은 성공하기 어렵다. 지금까지 의학교육현장에서 평가는, 교육결과인 성과에 대하여 고려한 것이 아니라, 교수가 중요하다고 생각하는 지식위주의 학습목표에 대하여 지필시험으로 학생을 평가하고 있다. 학생에게 현재 수준에 대하여 알려주고 피드백을 제공하는 의도보다는, 진급과 졸업 여부를 결정하는 종합평가로서의 목적이 강하다(Carraccio et al., 2002). 이러한 현실에서 교육의 가치를 학생들의 역량 향상에 두고 있는 성과중심교육의 도입은 평가방법과 체계에 대한 변화를 요구한다.

성과중심교육을 설계하는 과정은, 우선 학습성과를 선정한다음, 교육과정을 설계하기 이전에 평가기준과 방법을 먼저 설계하는

것을 권장한다. 이것은 학습목표설정, 교육과정설계, 평가방법설계의 순서로 일어나는 전통적인 설계방법과 비교하여, 후향적으로 설계하는 방식을 취한다(Dent & Harden, 2009). 평가는 성과와 직접적인 연관이 있으며 평가설계는 성과중심교육 설계의 초기에 일어난다. 평가를 설계하기 위해서는 왜 평가해야 하고, 무엇을 평가하며, 어떻게 평가할 것인지 고려해야 하며, 평가도구의 신뢰도, 타당도, 실현가능성을 생각해 볼 필요가 있다. '무엇'을 평가하느냐에 해당하는 것이 '성과 또는 역량'으로 볼 수 있으며, 성과의 특성을 잘 측정할 수 방법을 선택하는 것이 중요하겠다.

일반적으로 평가는, 평가의 목적(피드백을 주기 위한 형성평가인가, 서열을 결정하기 위한 총괄평가인가?), 평가방법의 선택(지필시험으로 판단할 것인가, 수행평가로 판단할 것인가?), 평가도구의 개발(수행평가를 시행한다면 체크리스트, 루브리, 평정척도 중 어떤 도구를 어떻게 개발할 것인가?) 및 기준설정(합격/불합격을 어떻게 결정할 것인가?)의 사항을 고려한다.

본 연구에서는 위와 같은 일반적인 평가에서의 고려사항에 대하여 문헌고찰을 통해 성과중심교육에서 평가의 특성을 알아보고자 하였다. 또한, 성과중심교육을 조기에 도입한 대학에서 시행한 평

가방법과 이들 대학에서 제시한 모형을 통하여 성과를 판단하는 적절한 평가방법과 도구를 살펴보고자 하였다. 마지막으로, 일반적인 기준설정방법과 임상수행평가에서 기준설정방법에 대해 고찰하였다.

성과중심교육에서 평가의 특성

평가의 목적, 방법 및 기준설정과 관련하여 성과중심교육에서 평가의 특성은 다음과 같다.

첫째, 성과중심교육에서 평가의 목적은 학생들에게 단계별로 향상 정도를 알려주기 위한 것이다. 평가는 목적에 따라, 학생의 학업 성취도를 평가하여 서열을 정하기 위한 총괄평가(summative assessment)와 학생에게 학업성취도에 대한 정보를 제공함으로써 장 점과 취약점에 대한 피드백을 주기 위한 형성평가(formative assessment)로 나눌 수 있다. 성과중심교육에서는 학생이 졸업할 때 갖추어야 할 역량을 정의한 다음(exit outcome), 이 목표에 도달하기 위한 단계별 수준을 정하게 된다. 이것은 학년이 올라감에 따라 교육내용의 폭이 넓어지고 깊이가 깊어지도록 교육과정을 설계하는 나선형 교육과정과 일치한다(Harden, 1999). 이러한 교육과정에서 학생이 최종 졸업성과를 달성할 수 있도록 단계별로 향상 정도를 파악하는 것이 성과중심교육에서 평가의 목적이라 할 수 있다(Harden, 2007; Holmboe et al., 2010). 평가의 목적이 학생에게 향상 정도를 알려주기 위한 것이라면, 평가는 총괄적이기보다 형성평가의 성격을 띤다. Carraccio et al. (2002)은 성과중심교육에서 총괄평가보다 형성평가가 중요함을 강조하였고, Burch et al. (2006)은 형성평가를 통해 임상실습에서 학습을 향상시켰다고 보고하였다. 그 밖에 성과중심교육에서 평가의 목적은 교육과정에 대한 피드백을 제공함으로써 교육과정의 개선을 위해 시행하며, 우수한 교육 프로그램의 제공은 대학의 사회적 책무성과도 연관되며, 이것은 보편적인 평가의 목적과도 일치한다.

둘째, 성과중심교육에서 평가는 지속적으로 자주 일어난다(Holmboe et al., 2010). 학생들의 역량을 향상시키기 위해 피드백을 제공하기 위해서 평가의 빈도가 증가한다. 이러한 예로 수업에서는 퀴즈나 이해도에 대한 점검을 시행할 수 있으며 임상실습현장에서 수행에 대한 피드백을 지속적으로 시행할 수 있다. 평가는 교수와 학습이 일어나는 수업, 실습 중에 빈번히 일어나게 되므로 평가-수업 간 경계가 종종 불분명하게 된다.

셋째, 역량의 특성에 따라 합당한 방법으로 성과를 평가한다. 평가방법은 지필시험(written examination), 임상수행평가(clinical or practical assessments), 관찰(observation), 포트폴리오와 다른 기록평가(portfolio, other records of performance), 동료평가와 자기평가(peer report and self-report)의 다섯 가지로 분류할 수 있다(Shumway et al., 2003). 이 중 성과에 명시된 역량을 잘 나타낼 수

있는 평가방법을 선택하여야 하며, 각 평가방법의 타당도, 신뢰도, 적용효과, 비용 등을 고려하여야 한다.

넷째, 같은 역량에 대해서 다양한 평가방법으로 성과를 평가한다. Miller (1990)은 어떤 단일한 평가방법으로 의사의 전문적인 수행과 같은 복합적 능력을 판단할 수 없다고 하였다. 예를 들어 의학 지식 역량은 보편적으로 지필시험으로 평가하나, 포트폴리오, 관찰, 수행평가를 통해서도 판단할 수 있다(Shumway et al., 2003).

다섯째, 성과는 궁극적으로는 실제 상황에서 행하는 것을 의미하므로 수행평가가 강조된다(Gruppen et al., 2012; Holmboe et al., 2010). 저학년 시기의 성과는 지식적인 내용을 습득하는 것일 수 있지만, 임상실습을 수행하는 3, 4학년 학생들의 평가는 실제 상황에서 수행하는 것을 목표로 해야 한다. 따라서 임상실습 상황에서는 실제 수행을 평가하도록 권장하며, 불가능한 경우 실제 상황과 비슷하게 조성하여 수행평가를 시행해야 한다.

여섯째, 성과를 판단하기 위해서 양적 평가와 더불어 질적 평가를 시행한다(Holmboe et al., 2010; Shumway et al., 2003). 의료윤리와 전문직업성처럼 양적 평가가 어려운 성과항목들은 질적 평가가 합당하다(Davis et al., 2001; Smith et al., 2007). 질적 평가는 신뢰도를 확보하는 일이 중요하며, 삼각기법(triangulation), 지속적이고 잦은 평가, 평가자 훈련을 통해 객관성을 확보할 수 있다(Shumway et al., 2003). 삼각기법은 여러 평가자가 다양한 각도에서 평가하는 것을 의미하며, 한 번의 시점에서 평가한 것은 높은 신뢰도를 확보할 수 없으므로 지속적으로 자주 평가해야 하며, 평가자 간의 편차를 줄이기 위해서는 평가자 훈련을 시행할 수 있다.

일곱째, 성과중심교육에서 평가기준은 일반적으로 절대적 기준(standards, criterion)을 사용한다(Gruppen et al., 2012; Holmboe et al., 2010). 성과중심교육에서 합격/불합격을 결정하기 위해 평가기준을 설정하는 일은 한 가지 뚜렷한 방법은 없지만, 대체로 다른 학생들의 수행수준과 관련이 없는 기준을 사용한다. 이것은 모든 학생은 학습 후에 명시된 성과수준에 도달해야 하는 것을 의미한다.

여덟째, 성과중심교육의 평가에 교수와 학생이 적극적으로 참여한다(Holmboe et al., 2010). 성과중심교육에서 다양한 평가가 지속적으로 이루어지게 되므로 교수의 이해와 참여가 필수적이다. 학생 또한 자신이 부족한 역량을 스스로 파악하고 보완하기 위해, 달성해야 하는 성과기준뿐만 아니라 그것을 판단하는 기준과 평가에도 깊이 관여해야 한다.

현재의 교육과정은 구조 또는 과정을 강조한 교육과정으로서(structured and process-based education), 성과중심교육에서 평가의 특성과 비교하여 Table 1에 제시하였다(Carraccio et al., 2002; Gruppen et al., 2012; Holmboe et al., 2010). 과정중심교육의 현재 교육과정에서 평가는 종합평가의 의도로서 과정 종료시점에 간헐적으로 시행하는 경우가 많다. 또, 학습성과는 고려하지 않고 지필 시험방법으로 지식습득에 대하여 평가하는 경우가 대부분이다. 질

Table 1. A comparison of the assessments in structure- and process-based versus outcome-based education

Assessments	Structure- and process-based	Outcome-based
Purpose	Summative	Formative
Frequency	Intermittent	Continuous, frequent
Methods	No consideration for outcomes	Match assessment methods with specific outcome
	Single measure, usually written test	Multiple measures for a competency
Measurement component	Knowledge acquisition	Knowledge application, performance in real-world setting
Property	Usually quantitative	Qualitative and quantitative
Standards	Relative, norm-referenced	Absolute, criterion-based
Engagement	Teacher	Student and teacher

적 평가는 드물고 학생들의 상대적인 결과에 따라 판정하는 경우가 많으며 학생은 평가기준이나 과정을 잘 알지 못한 채 통보받은 경우가 많다. 이렇게 현재 시행 중인 평가의 특성과 차이를 보임으로 새로운 성과중심교육의 도입을 위해서는 평가체계의 개선이 필요하다.

성과중심교육에서 평가방법의 선택

유럽의학교육학회(Association for Medical Education in Europe)에서 발간한 성과평가에 대한 지침을 중심으로, 평가방법의 종류와 성과에 합당한 평가방법을 살펴보고자 한다(Shumway et al., 2003).

1. 평가방법의 분류

이 지침서에서는 평가방법을 5가지로 분류하여 장단점과 적용 효과를 제시하였다. 지필시험(written assessments)은 지식의 회상 또는 적용하는 성과를 평가하기 위하여 사용하며 단일정답형 또는 확장결합형 선택형 시험문항(multiple-choice of questions, MCQ)을 보편적으로 사용하고 있다. 지필시험은 신뢰성이 높은 장점이 있지만, 깊은 이해보다는 표면적 지식을 평가하는 경향이 있고 실제 수행하는 것과 차이가 있다는 단점이 있다. 임상수행평가(clinical/practical assessments)는 임상수기 역량을 평가하는 신뢰성이 높은 시험이다. 표준화 환자 또는 시뮬레이션을 사용한 시험은 실제 어려운 임상 상황을 표현할 수 있을 뿐만 아니라 즉각적인 피드백을 제공할 수 있는 장점이 있지만, 학생이 시험을 준비할 때 임상수기의 단편을 수행하도록 훈련받아 임상수기 전체의 흐름 연결을 이해하기에는 부족하며 표준화 환자 훈련 등 많은 비용이 수반되는 단점이 있다. 관찰(observation)은 문제바탕학습이나 임상실습을 수행하는 동안 시행할 수 있는 방법이며 의사소통기술이나 태도 등의 역량을 평가하기에 합당하다. 체크리스트(checklist) 또는 평점척도(rating scale)의 방법으로 평가하며 학생들에게 피드백을 주기 위한 효과적인 방법이지만, 교수 간 차이로 인해 신뢰도를 확보하기 어려운 단점이 있다. 포트폴리오(portfolio)는 수행한 내용과

수행에 대한 성찰을 포함하는 것으로서 비판적 사고와 자기평가와 같은 성과를 판단하는 유용한 방법일 수 있으나(Davis et al., 2001), 신뢰도를 확보하기 위해 삼각기법 등의 방법이 필요하다. 로그북(logbooks)은 포트폴리오와 유사하나 단순히 학생들의 경험을 기록한 문서라 평가로서 한계가 많으며 학생이 정확하게 기록을 한 것인지 명확하지 않은 단점이 있다. 동료와 자기평가(peer and self-assessment)는 교수평가와 함께 사용할 수 있으며 태도와 의사소통기술을 평가하기에 유용하나 신뢰도를 확보하기 위해서는 훈련이 필요하며 신뢰를 기본으로 한 교육환경의 조성이 전제되어야 한다. 평가방법을 선택할 때 측정하고자 하는 성과의 특성에 합당한 방법을 선택해야 하며 각 평가방법의 장단점, 교육에 적용할 때 고려사항 등을 살펴보아야 할 것이다.

2. 성과에 따른 평가방법

Shumway et al. (2003)은 유럽의학교육학회 지침서를 통하여 Miller의 평가 피라미드와 성과를 연결하여 평가방법을 선택하는 모형을 제시하고자 하였다.

Miller (1990)는 학생을 평가하는 목적에 따라 아는 것(knows), 어떻게 해야 하는지를 아는 것(knows how), 어떻게 해야 하는지를 보여주는 것(shows how), 실제 하는 것(does)을 구분하여 평가해야 한다고 하였다. Miller가 제시한 평가체계 피라미드와 평가방법을 연결할 수 있는데, MCQ와 같은 지필시험은 knows와 knows how와, objective structured clinical examination (OSCE)와 같은 임상수행시험은 shows how, 실제 상황에서 행하는 것을 관찰하는 방법이나 포트폴리오 평가는 does에 해당한다(Figure 1).

Miller의 평가 피라미드와 던디 의과대학(University of Dundee School of Medicine)의 12가지 성과를 연결하면 Figure 2와 같다. 의학지식(medical sciences), 진단 및 처치(investigation and management)는 knows와 knows how에, 임상수기(clinical skills)와 의사소통(communication)은 shows how, 윤리(ethics)와 의사의 역할(role of doctor)은 does에 해당한다.

Table 2는 성과의 특성을 잘 측정할 수 있는 평가방법을 제시하였다. 병력청취와 신체진찰 등을 포함하는 임상수기 역량은 OSCE

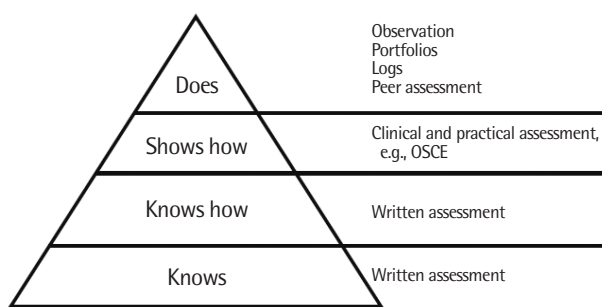


Figure 1. The learning assessment pyramid. From Shumway, J. M., Harden, R. M., & Association for Medical Education in Europe. (2003). *Med Teach*, 25(6), 569-584, with permission from Informa Healthcare. OSCE, objective structured clinical examination.

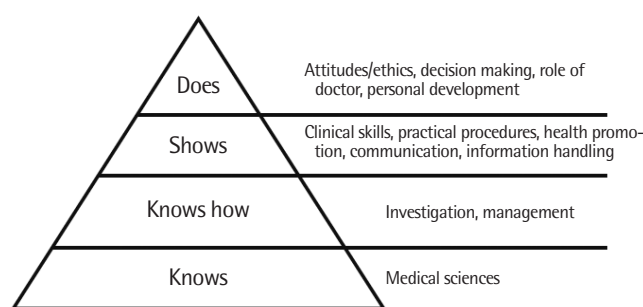


Figure 2. The 12 learning outcomes matched for assessment purposes against the most appropriate level of the Miller Pyramid. From Shumway, J. M., Harden, R. M., & Association for Medical Education in Europe. (2003). *Med Teach*, 25(6), 569-584, with permission from Informa Healthcare.

Table 2. Recommended assessment methods for the 12 learning outcomes

Learning outcome	Assessment methods
What the doctor is able to do	
Clinical skills	OSCE, observation, logbooks, written examination
Practical procedures	OSCE, portfolios and logbooks, observation, written examination
Patient investigation	Written examination, OSCE, observation, portfolio
Patient management	Written examination, OSCE, observation, portfolios
Health promotion and disease prevention	OSCE, portfolios, observation, written assessment
Communication	OSCE, observation, peer/self assessment, portfolio
Information management skills	Portfolio, OSCE, observation, written examination
How doctors approach their practice	
Principles of social, basic and clinical sciences	Written examination, portfolios, observation, OSCE
Attitudes, ethics and legal responsibilities	Observation, portfolio, peer/self assessment, OSCE, written examination
Decision making, clinical reasoning and judgement	Portfolio, observation, written examination, OSCE, peer/self assessment
Doctors as professionals	
Role as a professional	Observation, peer/self assessment, portfolio, OSCE, written examination
Personal development	Portfolio, observation, peer/self assessment, OSCE, written examination

From Shumway, J. M., Harden, R. M., & Association for Medical Education in Europe. (2003). *Med Teach*, 25(6), 569-584, with permission from Informa Healthcare. OSCE, objective structured clinical examination.

와 같은 수행평가로 판단할 수 있을 뿐만 아니라 임상실습을 수행할 때 관찰방법이나 로그북을 이용하여 평가할 수 있다. 의학 지식 역량은 지필시험으로 평가할 수 있지만, 포트폴리오, 관찰 및 수행 평가로도 판단이 가능하며, 태도와 윤리성은 실제 임상 상황에서 판단하는 것이 가장 바람직하다.

3. 임상실습에서의 평가

성공중심교육에서는 실제 임상 상황에서 행하는 것을 강조하므로 임상실습에서 평가는 더욱 중요하며 가치가 있다. 의사소통기술이나 임상수기 역량은 임상수행평가를 통해서 가능하며 우리나라는 2009년부터 의사실기시험이 도입되면서 임상실습평가로서 수행시험이 정착되고 있다. 그러나 표준화 환자를 훈련하여 시행하는 임상수행평가는 임상수기의 단면을 평가하는 것으로서 진료 전체의 흐름을 판단하기에 한계가 있으며 실제 수행과는 차이가 있다

(Shumway et al., 2003).

Davis & Harden (2003)은 실제 임상 상황에서의 수행을 평가하는 것을 강조하며, 직무중심평가(workplace-based assessment)와 같은 새로운 평가방법을 제안하였다. 직무중심평가는 매일의 실습 현장에서 형성평가의 일환으로 지속적인 평가가 일어나며 피드백을 통해 학생들의 수행 향상을 도모하는 것이다. Burch et al. (2006)은 임상실습 중에 장기간의 직무중심평가를 시행한 결과, 학생들은 자신의 역량수준을 알 수 있을 뿐만 아니라 학습동기유발, 실습활동에 적극적으로 참여하는 등의 긍정적 효과가 있었고, 교수 또한 실습 중 형성평가가 학생들의 학습촉진에 효과적이었다고 하였다. 피드백은 학생의 요구와 일치하여야 하며, 중요한 사안에 대해 구체적으로 제시해야 효과적인데, 임상교수들의 평가에 대한 이해를 높이기 위해서 교수개발프로그램을 시행할 필요가 있다(Norcini & Burch, 2007).

직무중심평가의 평가방법으로는 수행의 직접관찰, 증례에 대한 토의 및 동료, 간호사 및 환자에 의한 다각도의 360° 평가 등이 있다 (Miller & Archer, 2010).

포트폴리오는 전문직업성, 평생학습능력과 같이 기존의 일반적인 평가방법으로 측정하기 어려운 역량을 평가할 수 있다. 포트폴리오 평가는 활동내용, 성취에 대한 근거, 학습경험에 대한 성찰을 포함하게 된다. Friedman Ben David et al. (2001)은 성공적인 포트폴리오 평가를 위해 절차를 제시하였는데, 1) 포트폴리오 평가의 목적을 분명히 하고, 2) 평가하는 역량을 결정하며, 3) 근거자료를 정하고, 4) 채점표를 개발하며, 5) 관련 교수를 훈련하며, 6) 평가절차에 대하여 계획하며, 7) 학생에게 알려주며, 8) 평가지침을 개발하고, 9) 신뢰도와 타당도 근거를 확보하며, 10) 개선절차를 마련한다.

4. 강의와 실습에서의 실제 적용

브라운 의대(Brown Medical School)에서 9가지 역량에 대해 평가한 방법을 살펴보고자 한다(Smith et al., 2003). 첫째, 의사소통 기술은 소집단수업에서 의사소통을 직접 관찰하거나 비디오로 녹화하여 평가할 수 있는데, 실제 해부학수업에서 해부실습을 하면서 발표, 질문과 응답을 통하여 의사소통기술을 평가하였다. 또, 선택과정에서 의사소통기술에 대하여 서술식으로 작성한 동료평가는 효과적이었다고 보고하였다. 임상실습을 수행하는 동안에는 직접 관찰 또는 OSCE시험을 통해 의사소통기술 평가가 이루어졌다. 둘째, 임상수기 역량은 2학년 임상의학입문수업에서 피드백을 위하여 표준화 환자를 활용하였으며, 임상실습에서는 직접관찰, OSCE 시험으로 평가하였다. 셋째, 기초의학지식의 적용 역량은 문제바탕 학습(problem-based learning, PBL) 수업 또는 임상 상황으로 주어진 MCQ, modified essay question (MEQ) 문항으로 평가하였다. 넷째, 진단·치료·예방 역량은 PBL수업에서 평가가 가능하며, 임상실습에서는 지필 또는 구술시험, OSCE시험을 통해 평가하였다. 다섯째, 평생학습 역량에 대한 평가는 PBL에서 학습과제 준비를 평가하거나, 임상실습 논문지도교수가 장기간 학생들을 관찰하면서 시행하였다. 여섯째, 자기인식·자기관리·개인적 성장 역량 또한 멘토교수가 오랫동안 학생과 교류하면서 평가하였다. 일곱째, 건강관리의 사회적 맥락은 보고서, 구술시험, 전시물 및 성찰수업시간에 교수 직접관찰 등을 통해 평가하였으며, 지역사회 관련 프로그램이나 프로젝트, 일반인 대상의 발표과제로 평가가 가능하였다. 여덟째, 도덕적 사유와 임상윤리는 윤리적 이슈를 분석하는 지필과제, OSCE에서 윤리적 문제를 접근하는 방법으로 평가하였으며 멘토교수가 평가하기도 하였다. 아홉째, 문제해결 역량은 병원조직에서 복잡한 업무를 처리해야 할 때 우선순위를 정하고 효과적으로 일을 처리하는 능력으로, 임상 상황 또는 지필시험분석, 실습에서 실제 복잡한 일이 주어졌을 때 처리방법 및 학생인턴수행 시 업무관리능력을 평가하였다.

5. 기준설정(Setting Standard)

성과에 도달하였는지를 판단하는 것은 수집한 평가결과들을 취합하여 기준점수를 설정하는 복잡하고 어려운 일이다. 이러한 기준 설정은 기준의 유형과 기준 설정방법을 고려하여야 한다.

기준의 유형은 상대적(relative/norm-referenced) 기준과 절대적(absolute/criterion-referenced) 기준으로 분류할 수 있다. 상대적 기준은 피험자의 시험결과에 따라 특정 수나 퍼센트를 선발할 때 적절하며 운영하기가 쉬운 장점이 있으나 역량의 성취도를 판단하기는 어렵다(Norcini, 2003; Ricketts, 2009).

기준설정방법에는, 통과시킬 학생의 일정 비율을 정하는 고정비율방법(fixed percentage method, relative method), 무작위로 표본을 추출한 학생집단에 대해 결정을 내린 후 합격과 불합격의 점수를 그래프로 나타내어 비교하는 비교집단방법(contrasting groups method, examinee-centered absolute method), 시험문항의 내용을 고려하여 경계그룹(borderline group) 학생들을 판단하는 엔고프(Angoff)와 에벨(Abel)방법(test-centered absolute method), 상대적&절대적 평가방법을 절충한 호프스티(Hofstee)방법 등이 있다(Norcini, 2003).

이와 같은 기준설정방법은 지필시험에 적용하기 위해 개발되었지만, 임상수행평가에서 변형하여 사용할 수 있다(Boulet et al., 2003). 엔고프와 에벨방법을 적용하면 각 시험방에 사용된 증례의 채점표를 검토하여 최소수행기준을 정한 다음 각 시험방의 점수를 합하여 기준점수를 설정한다. 비교집단방법에서도 유사하게 각 증례마다 합격과 불합격을 판단할 수 있는 점수를 정의하고, 모든 시험방의 점수를 합하여 기준점수를 정한다.

성과중심교육에서 평가는 상대적 서열을 결정하기 위한 것이 아니라 역량에 도달하였는지를 판단하는 것으로서 절대적 기준을 적용하는 것이 합당해 보인다. 그러나 George et al. (2006)에 의하면 두 가지 기준유형과 설정방법에 따라 판단결과가 차이를 보이며, 각 방법은 장단점이 있다. 사전에 결정된 기준점수(cut-off scores)로 판단하는 준거참조법(절대적 평가)은 실패율에서 변동을 보이며, 상대적 평가는 기준점수 자체의 변동을 보이는 단점이 있으므로 병합하여 판단하는 것이 유리하다는 의견도 있다(Cohen-Schotanus & van der Vleuten, 2010).

기준 설정은 시험의 목적, 학생의 특성, 역량의 특성을 고려하여 다양한 견해를 통합하여 합리적인 판단을 하는 것으로서 평가도구가 신뢰도를 확보했는지, 합격률이 역량성취도와 일치하는지 속고 해야 하며, 향후 수행과도 비교하여야 한다(Norcini, 2003).

결론

성과중심교육을 실천하기 위해 평가를 체계적으로 시행하는 것은 핵심적인 일이다. 현재 의학교육현장에서 평가는 성과중심교육

에서 평가와 차이를 보이며, 다음과 같은 성과중심의 평가의 특성을 고려하여 평가체계를 개선할 필요가 있다.

성과중심교육에서 평가는 학생들에게 성과성취수준에 대하여 피드백을 주기 위한 형성평가의 목적이 강하며 지속적으로 이루어져야 한다. 성과를 평가하기 위해서는 적절한 평가방법을 선택하고 같은 성과에 대해서 다양한 방법으로 평가를 시행한다. 성과는 학생들이 실제 상황에서 수행하는 것을 목표로 하므로 수행평가가 특히 중요하며, 양적 평가뿐만 아니라 질적 평가를 시행할 필요가 있다. 학생의 성과도달 여부의 판단은 대체로 절대적 기준을 사용하는 것이 원칙이지만 다양한 자료의 분석과 의견을 종합하여 판단하는 것이 중요하다.

성과중심교육에서 평가의 이러한 특성은 교수의 적극적인 참여를 필요로 한다. 지식뿐만 아니라 윤리성, 평생학습능력 등 다양한 성과에 대하여 평가가 필요하며, 특히 전문직업성 등의 성과를 판단하기 위해서는 새로운 평가방법의 개발이 요구된다. 또 다양한 평가방법의 장단점을 알고 적절히 적용하여 합리적으로 판단할 수 있어야 하며, 효과적으로 피드백을 제시하는 방법에 대해서도 잘 알고 있어야 한다. 성과중심교육에서 학생은 중심적인 역할을 하는데, 성과의 정의와 판단기준을 명확히 이해하고 반드시 달성해야 하는 목표라는 점을 인지할 필요가 있으며, 기준을 충족하지 못한 역량을 파악하여 개선하도록 스스로 노력해야 할 것이다. 이를 위해서 성과중심교육과정과 평가는 교수와 학생이 함께 설계하여 공유의식을 높이고, 지원체계와 시스템을 갖추어 수월성을 확보하도록 노력해야 할 것이다.

ACKNOWLEDGMENT

이 논문은 부산대학교 자유과제 학술연구비(2년)에 의하여 연구되었다.

REFERENCES

Boulet, J. R., De Champlain, A. F., & McKinley, D. W. (2003). Setting defensible performance standards on OSCEs and standardized patient examinations. *Med Teach, 25*(3), 245-249.
 Burch, V. C., Seggie, J. L., & Gary, N. E. (2006). Formative assessment promotes

learning in undergraduate clinical clerkships. *S Afr Med J, 96*(5), 430-433.
 Carraccio, C., Wolfsthal, S. D., Englander, R., Ferentz, K., & Martin, C. (2002). Shifting paradigms: from Flexner to competencies. *Acad Med, 77*(5), 361-367.
 Cohen-Schotanus, J., & van der Vleuten, C. P. (2010). A standard setting method with the best performing students as point of reference: practical and affordable. *Med Teach, 32*(2), 154-160.
 Davis, M. H., Friedman Ben-David, M., Harden, R. M., Howie, P., Ker, J., McGhee, C., ... Snadden, D. (2001). Portfolio assessment in medical students' final examinations. *Med Teach, 23*(4), 357-366.
 Davis, M. H., & Harden, R. M. (2003). Competency-based assessment: making it a reality. *Med Teach, 25*(6), 565-568.
 Dent, J. A., & Harden, R. M. (2009). *A practical guide for medical teachers* (3rd ed.). Philadelphia: Churchill Livingstone.
 Friedman Ben David, M., Davis, M. H., Harden, R. M., Howie, P. W., Ker, J., & Pippard, M. J. (2001). AMEE medical education guide no. 24: portfolios as a method of student assessment. *Med Teach, 23*(6), 535-551.
 George, S., Haque, M. S., & Oyebo, F. (2006). Standard setting: comparison of two methods. *BMC Med Educ, 6*, 46.
 Gruppen, L. D., Mangrulkar, R. S., & Kolars, J. C. (2012). The promise of competency-based education in the health professions for improving global health. *Hum Resour Health, 10*(1), 43.
 Harden, R. M. (1999). What is a spiral curriculum? *Med Teach, 21*(2), 141-143.
 Harden, R. M. (2007). Learning outcomes as a tool to assess progression. *Med Teach, 29*(7), 678-682.
 Holmboe, E. S., Sherbino, J., Long, D. M., Swing, S. R., & Frank, J. R. (2010). The role of assessment in competency-based medical education. *Med Teach, 32*(8), 676-682.
 Miller, A., & Archer, J. (2010). Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ, 341*, c5064.
 Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Acad Med, 65*(9 Suppl), S63-S67.
 Norcini, J., & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE guide no. 31. *Med Teach, 29*(9), 855-871.
 Norcini, J. J. (2003). Setting standards on educational tests. *Med Educ, 37*(5), 464-469.
 Ricketts, C. (2009). A plea for the proper use of criterion-referenced tests in medical assessment. *Med Educ, 43*(12), 1141-1146.
 Shumway, J. M., Harden, R. M., & Association for Medical Education in Europe. (2003). AMEE guide no. 25: the assessment of learning outcomes for the competent and reflective physician. *Med Teach, 25*(6), 569-584.
 Smith, S. R., Dollase, R. H., & Boss, J. A. (2003). Assessing students' performances in a competency-based curriculum. *Acad Med, 78*(1), 97-107.
 Smith, S. R., Goldman, R. E., Dollase, R. H., & Taylor, J. S. (2007). Assessing medical students for non-traditional competencies. *Med Teach, 29*(7), 711-716.