

보건의료 전문직 교육에서의 평가

한재진^{1,2}

¹이화여자대학교 의과대학 명예교수

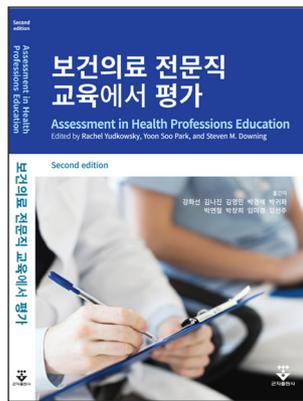
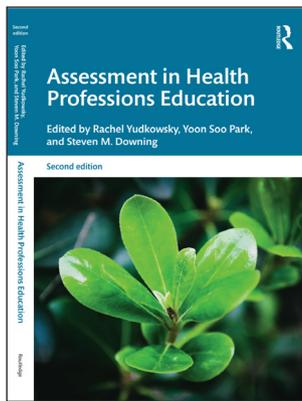
²한국의학교육평가원 의사전문역량인증단장

Assessment in Health Professions Education

Jae Jin Han^{1,2}

¹Professor Emeritus, Ewha Womans University School of Medicine, Seoul, Korea

²Director of ABMPC, Korean Institute of Medical Education & Evaluation, Seoul, Korea



원저: Rachel Yudkowsky, Yoon Soo Park, Steven M. Downing.
Assessment in health professions education. 2nd ed. New York (NY): Routledge; 2019. 321 p.

저서: 보건의료 전문직 교육에서 평가

저자: Rachel Yudkowsky, Yoon Soo Park, Steven M. Downing

역자: 강화선, 김나진, 김영민, 박경혜, 박귀화, 박연철, 박장희, 임미경, 임선주

출판사: 군자출판사(주)

출판연도: 2023년

쪽수: 329쪽

우리는 교육에서 ‘측정과 평가’(평가) 분야의 중요성이 더욱 강조 되는, 즉 “From AOL (assessment of learning) to AAL (assessment affecting learning)”의 시대에 살고 있다. 특히 보건의료 영역의 확대와 전문성의 심화 그리고 사회적 관심이 높아지는 경향은 한국 뿐 아니라 세계적인 추세이다. 따라서 보건의료 전문직 교육 (health professions education, HPE) 영역에서의 평가는 그 중요성에 걸맞게 지속적인 연구와 함께 새로운 성과들이 계속 나오고 있으며 향후 AI(artificial intelligence)가 주도하는 미래 보건의료와 의학환경에서는 더욱 중요해질 것으로 보인다.

이러한 상황에서 HPE에서의 평가와 관련된 거의 모든 주제를 망라하여 체계적으로 다루면서도 진진된 이론을 소개하고 현장의 적용사례를 적절히 들어 이해도를 높인 이 역서는, 모든 HPE 관련자가 필독할 교과서로서 손색이 없다. 이는 HPE 전문가뿐 아니라 평가 분야의 연구를 시작하고자 하는 학생, 교육 실무 담당자 및 교육정책 관련자에게도 큰 도움이 될 것으로 보인다.

이 책은 다음 세 가지 특징을 가지고 있다.

첫 번째 특징은, 평가이론의 수학-통계학적 배경을 중시하는 기존 교과서의 전형을 탈피하여, HPE와 평가현장에서 많이 다루어지는 주제를 이론에 치우치지 않고 실용적인 면을 강조하며 비중 있게 설명하고, 특히 주제 관련 최신 연구동향과 결과들도 자세히 소개하고 있다. Part 1에서는 평가의 기본적인 주제를 다루고, part 2에서는 실제 보건의료 교육현장에서 많이 활용되고 있는 여러 평가방법들을 소개하고 있으며, part 3에서는 새롭게 나타난 주제들을 다루고 있다. Part 1의 제1장은 이 책의 서론 장으로 평가의 전반적인 맥락과 책의 개요를 요약하였으며, 2장과 3장에서는 각각 ‘타당도’와 ‘신뢰도’를 주제로 하여 핵심 이론을 간략하게 기술하면서 최신 논거들을 소개하고 있다. 특히 타당도에서는 평가의 목적과 질 평가와의 연계성을 강조하고 대규모 표준화 평가체계를 관리하는 틀로 ‘Kane틀’과 ‘Messick틀’을 소개하면서 각 틀에서의 추론 혹은 근거 범주를 활용한 평가 해석과 타당도 위협요소에 대처하는 예를 상세히 설명한다. 평가 담론의 핵심이라 할 수 있는 신뢰도는 심리측정학적 추론의 이해를 위해

기본적인 고전검사이론을 설명하면서 3가지 신뢰도지수인 (1) 검사-재검사 신뢰도, (2) 반분검사 신뢰도, (3) 내적 일관성 신뢰도 (Cronbach α)와 세 가지 평정(채점)자 간 신뢰도계수인 (1) 일치도 계수, (2) Kappa 계수, (3) 급내상관계수(intraclass correlation)의 내용을 기본 공식, 도표와 함께 간단한 예시를 들어 설명하고 있는데, 독자는 직접 계산을 해보면서 이해도를 높일 수 있다. 또한 신뢰도를 높이는 세 가지 방법으로 (1) 문항 등 평가 구성 요소의 추가, (2) 문항 분석을 통한 수정, 제거 및 대체, (3) 합성점수 생성 등을 제안하고 있다. 타당도와 신뢰도 간의 관련성은 “신뢰도는 타당도를 위한 필요조건이지만 충분조건은 아니며, 평가점수는 일관성이 있으나(신뢰도는 높으나) 정확하지 않을(타당도는 낮을) 수 있다”라고 정리하였다. 4장에서는 근래에 많이 활용되고 있는 ‘일반화가능도 이론’의 배경을 설명하면서, 앞의 3장에서 다루었던 고전검사이론에서의 ‘신뢰도’와 비교하여, 진점수를 구성하는 오차점수의 다양한 성분을 추정하고 각각의 분산성분을 분석하는 주요 개념과 함께 공식을 제시하였으며, G-study와 D-study의 예제를 통한 활용사례를 소개하였다. 이어서 5장에서는 기존의 평가 관련 교과서에서 일반적으로 우선하여 다루어 왔던, 검사 혹은 점수의 통계치 관련 내용으로 점수 유형과 문항 분석에서 나오는 용어들의 정의와 핵심 개념을 설명하고 Kuder-Richardson 21, 측정의 표준오차(standard error of the measurement) 및 스피어만-브라운 예측공식을 예제와 더불어 제시하였다. Part 1의 마지막 장은 준거 설정 주제인데, 8단계의 준거 설정 절차를 제안하고 Angoff, Ebel, Hoftee, 경계선 집단, 대비 집단, body of work 및 환자안전 등 7가지 준거 설정방법을 기본 개념, 실행절차, 예시 및 도표 등으로 설명하였다. 이러한 준거 설정의 모든 절차와 방법은 보건의료 전문직 교육과 고부담 시험에서의 수용 가능한 합격점수 결정과 판정에 있어서 타당성 논증과 방어를 담보하고 시험의 목적에 적합해야 함을 강조하였다.

이 책의 두 번째 특징은, 현재 많이 활용되고 있는 평가의 방법(시험)들을 각 장별로 하나씩 다루면서 역사나 배경, 정의, 활용 현황, 형태나 형식, 시험 운영, 준거 설정 및 질 관리 등의 내용과 사례 제시, 각 시험방법과 관련된 최신 연구결과 등을 기술하고 있다. Part 2의 7장 지필시험 주제에서는 평가대상이 습득한 지식 뿐 아니라 기본원리의 이해능력과 문제 해결을 위한 추론능력까지 평가할 수 있는 다양한 지필시험의 형태와 컴퓨터 기반을 포함하여 채점 관련 논점들을 기술하고 있다. 8장의 주제인 구술시험은 오랜 역사와 Miller의 역량 피라미드 상층부인 ‘does’와 ‘shows how’를 평가할 수 있는 장점에도 불구하고 타당도에 있어서의 논란이 계속 있어 왔는데, 이와 관련된 제한점과 이러한 제한점을 극복할 방안과 질 보장 전략을 제시하고 있다. 9장에서는 수행평가 주제로 우리나라 의사 국시에도 포함된 ‘객관구조화진

료시험(objective structured clinical examination)’의 준비와 실행, 채점, 준거 설정 및 타당도 등을 설명하고 있다. 이어서 우리나라에서도 활용도가 점차로 증가하고 있는 현장바탕평가와 위임가능진료활동(entrustable professional activities) (10장), 평가포트폴리오(12장), 시뮬레이션교육에서의 평가(part 3의 14장)에 관하여 기술하고 있으며 우리나라에서는 아직 생소한 내러티브 평가에 대해서도 part 2의 11장에서 소개하고 있다. 위의 내용들에서 특히 유용한 부분은 개별 장에서 해당 평가방법(시험)을 실행할 때 발생할 수 있는 ‘타당도의 위협요소’를 분석하고 그 해결방안을 정리한 표들이다.

이 책의 세 번째 특징은 보건의료 평가 분야에서 새롭게 떠오르고 있는 주제들을 다루고 있는데 part 3의 13장에서는 임상 상황에서 주요한 특징(key features, KF)을 파악하고 개별적이고 도전적인 결정들을 평가하는 ‘KF 시험’을 소개하고, 15장에서는 우리나라 심층면접 시험에서도 사용되기 시작한 다중미니면접(multiple mini-interview)으로 대표되는 상황판단검사(situational judgement test)의 주제를 인사심리학의 성과와 의학교육에서의 임상수행평가의 성과가 만나게 된 역사적 배경과 함께 전문직업성 평가의 필요성 증대에 대응하기 위한 객관화된 평가도구의 설계와 예측타당도를 포함한 평가의 타당도 연구성과 등을 예제(15장)를 통해 설명하였다. 그 외에 평가문화의 측면에서, 전통적인 평가의 문제점들을 극복하고자 나타난 ‘계획적 평가(programmatic assessment)’(16장), 형성평가와 총괄평가의 포괄적 설계와 실행을 통한 ‘평가가 학습에 미치는 영향’(17장), 교육시간보다 입증된 수행을 바탕으로 교육의 성과를 판단하고자 하는 교육적 접근 문화인 ‘완전학습(mastery learning)’(18장), 그리고 19장에서는 ‘문항반응이론(item response theory)’을 소개하면서 표본에 따른 평가점수와 문항 특성을 동일 척도로 연결할 수 있는 장점으로 인하여 고전검사이론(classical test theory)의 강력한 대안의 역할을 하면서 ‘컴퓨터능력적응시험(computer adaptive test)이나 검사동등화를 위한 바탕이 될 수 있음을 설명하고 있으며, 마지막으로 HPE 분야의 다양한 평가사업에서 통계학자와의 협력문화(20장)의 중요성을 강조하였다.

책의 소개를 마무리하면서 특별히 언급하고자 하는 한가지는 이 책의 저자 중 한 분이며, 이 책에서 ‘한국 독자들을 위한 서문’을 쓴 박운수 교수의 약력이다. 박 교수는 현재 일리노이 대학교 시카고(University of Illinois at Chicago) 의과대학의 의학교육학교실 주임교수로 재직 중이면서 Academic Medicine, Medical Education 등 저명한 의학교육 저널의 편집진으로 활동하고 있으며, 특히 미국 의학교육계의 측정평가 전문가와 연구자들이 모두 모이는 ‘Research in Medical Education’이라는 Association of American Medical Colleges 산하 그룹의 의장을 한국인 혈통으로는 최초로

역임한 바 있다(2018-2021).

끝으로, 가톨릭대학교 의과대학 김영민 교수를 비롯하여 국내 의학교육계에서 왕성한 활동을 하고 있는 총 아홉 분이 이 책의 번역을 위해 기울인 세심한 작업과 한국 의학교육평가의 발전을 위한 사명감과 열정이 느껴지는 노고에 독자로서 심심한 감사를 드린다.

Jae Jin Han

ORCID

<https://orcid.org/0000-0002-6499-7642>