

일개 국내 의과대학 교육프로그램 질 향상을 위한 평가체계 틀 개발에 관한 연구

강유림^{1,2}, 김혜원^{1,2}, 최준용^{1,3}¹연세대학교 의과대학 연세동곡의학교육원 교육평가센터²연세대학교 의과대학 의학교육학교실³연세대학교 의과대학 내과학교실

Development of a Program Evaluation Framework for Improving the Quality of Undergraduate Medical Education

Yulim Kang^{1,2}, Hae Won Kim^{1,2}, Jun Yong Choi^{1,3}¹Center for Educational Evaluation, Yonsei Dongguk Medical Education Institute, Yonsei University College of Medicine, Seoul, Korea²Department of Medical Education, Yonsei University College of Medicine, Seoul, Korea³Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Korea

The development and implementation of a systematic program evaluation framework is critical for improving the quality of the undergraduate medical education. At Yonsei University College of Medicine, we established the following five evaluation domains that encompass various aspects of educational experiences: (1) preclinical curriculum; (2) clinical curriculum; (3) educational environment, resources, and systems; (4) performance of students and graduates and program outcomes; and (5) implementation and outcomes of the curriculum. Specific evaluation indicators were designed within these domains and validated through the Delphi technique, which integrated expert opinions. In total, 98 indicators were identified across five domains. These indicators will function as a comprehensive tool for assessing medical education programs. The proposed evaluation framework addresses both short- and long-term educational changes, facilitating systematic monitoring, continuous quality improvement of curricula, and better outcomes for students. As this framework is grounded in the unique context of the institution, it is appropriate for a comprehensive evaluation of interactions at various educational stages. Furthermore, it may serve as a strategic foundation for identifying areas that require improvement, ensuring that the curriculum aligns with current medical education standards and practices. The framework's structured approach and continuing evaluation processes may make it possible to obtain essential data for ongoing development, potentially contributing to a robust system for quality improvement in medical education. The findings of this study are expected to serve as a valuable reference for developing similar evaluation frameworks in other medical schools.

Keywords: Delphi technique; Program evaluation; Quality improvement; Undergraduate medical education

Received: October 15, 2024 Revised: December 18, 2024

Accepted: January 6, 2025

Corresponding author: Hae Won KimDepartment of Medical Education, Yonsei University College of Medicine,
50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
Tel: +82-2-2228-2506 Fax: +82-2-364-5450 E-mail: peacemaker1010@yuhs.ac**Co-corresponding author:** Jun Yong ChoiDepartment of Internal Medicine, Yonsei University College of Medicine, 50-1
Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
Tel: +82-2-2228-1974 Fax: +82-2-393-6884 E-mail: seran@yuhs.ac

서론

의과대학은 미래의 의료인이 사회의 건강수요를 충족시킬 수 있는 의학지식, 임상술기, 전문직업성 등 총체적인 역량을 갖추도록 교육해야 하는 책무성이 있다[1]. 따라서 의학교육은 사회적 필요성에 부응하고 교육의 효과와 효율을 증대하기 위해 지속적인 질적 제고가 필수적인 영역이며, 이러한 맥락에서 교육프로그램 평가의 중요성이 강조되고 있다[2,3].

교육프로그램 평가는 교육 관련 의사 결정을 위해 유용한 정보를 획득, 제공하는 의도적인 개입으로[4], 교육프로그램의 실효

성을 평가하고 교육활동 전반의 종합적인 질 관리방안을 모색하는 데 도움이 된다. 평가를 통해 교육프로그램의 목표 성취 여부나 교육의 각 단계에 내재된 가치를 확인할 수 있으며[5,6], 더 나아가서는 교육에 영향을 주는 프로그램 내·외부 요인을 식별할 수 있기 때문에 이해관계자의 요구를 파악하고 프로그램 발전에 효과적인 결정변수를 정하는 데 효과적이다[7]. 교육프로그램의 복잡한 성과를 측정, 분석, 평가하기 위해서는 평가요소들을 총체적으로 포괄하는 체계적인 평가체계의 확립이 선행되어야 하고, 이를 바탕으로 프로그램의 발전방안을 탐색할 수 있다[8]. 실제로 미국 및 캐나다의 의학교육프로그램 인증기준은 의과대학이 단기 또는 장기 교육목표를 계획하고 측정 가능한 성과를 달성해 나감으로써 지속적인 질 향상 활동을 하도록 요구하고 있으며[9], 국내 의학교육 평가인증기준에서도 의과대학 단위에서 자체적인 평가체계를 수립하고 주기적인 교육평가를 통해 교육 개선활동을 하도록 장려하고 있다[10].

보건의료분야에서 교육프로그램 평가의 초기 연구는 체크리스트를 기반으로 한 투입 대비 산출 중심의 총괄평가가 주를 이뤘으며, 교육의 최종적인 성과를 확인하려는 경향이 두드러졌다[11]. 그러나 이는 교육의 과정에 대한 깊이 있는 이해와 분석이 결여된 블랙박스모형적 접근에 해당한다. 즉 교육의 효과성을 평가하는 데 있어 표면적인 결과만 고려하게 되며, 학습환경, 자원, 교육효과성 간의 상호작용에 대한 근거를 제공하지 못한다. 또한 다양한 이해관계자 사이의 역동성을 충분히 반영하지 못하기 때문에 평가의 타당성과 유용성 측면에서 문제가 제기되었다[12,13].

지난 40여 년간 보건의료분야에서는 교육프로그램의 성과를 네 단계의 위계로 구분하는 Kirkpatrick 평가모형이 널리 활용되었다[11]. 이는 프로그램 평가모형에 기반한 체계적인 평가를 시도한 것이라 볼 수 있지만 여전히 교육적 개입으로 인한 최종적인 결과만을 제한적으로 평가한다는 한계가 있다[7,14,15]. 교육프로그램을 결과 중심으로 평가하는 것은 교육의 여러 요소들이 유기적으로 작용하는 복잡한 현상을 제대로 설명할 수 없으므로 진정한 개선을 이끌어내기 어렵다. 최근에는 교육 전반에 대한 종합적인 평가를 시도한 선행연구들이 수행되고 있으나[15-21], 이론에 근거한 평가체계가 부재한 비정형적인 평가를 활용하고 있어[22,23], 교육프로그램의 목표 달성 여부 및 프로그램의 장·단점을 효과적으로 도출하는 데 제한점이 있다[24]. 아울러, 일부 연구는 교육과정에서 이루어지는 단편적인 교육활동의 적절성을 확인하는 데 그치고 있어 의과대학 교육프로그램의 총체적 평가체제로 제시하기에는 한계가 있다[2,25-27].

다양한 교육활동을 통해 발생하는 역동성을 질 향상에 반영하기 위해서는 교육의 실행단계와 구성요소를 통합적으로 고려한 평가체계가 필요하다[8]. 특히 교육의 결과가 의미 있는 변화로

이어지기까지는 상당한 시간이 소요되므로 의과대학 자체의 견고한 개념적 평가체계를 구축하여 주기적인 교육프로그램 평가가 시행되어야 할 것이다. 또한 교육의 효과성을 극대화하기 위해서는 대학 고유의 맥락을 반영하여 교육의 여러 요인 간 유기적 연관성 및 정합성을 제고한 평가체계를 구축해야 한다.

최근 프로그램 평가의 동향은 과정과 결과 측정을 동시에 고려할 수 있는 다양한 접근방식을 채택하는 것으로 이는 프로그램의 최종적인 결과뿐만 아니라, 프로그램 실행 중 발생할 수 있는 다양한 상황적인 요소 및 참여자의 반응을 반영함으로써 프로그램의 점진적 개선을 위한 실질적인 자료를 제공할 수 있다[28]. 이에 따라 본 연구는 대학의 교육환경과 교육과정의 맥락적 특징을 반영한 통합적인 접근을 통해 의과대학 교육프로그램 운영의 전반적인 과정을 포괄할 수 있는 체계적이고 구조화된 평가체계를 구축하는 것을 목표로 하였다.

연구대상 및 방법

1. 연구개요

연세대학교 의과대학 교육프로그램 평가체계 구축은 두 단계로 실행되었다. 첫째, 교육프로그램 평가의 핵심 구성요소를 확인하고 연세대학교 의과대학의 맥락과 평가인증기준에 부합하는 평가지표를 구성하기 위해 현행 교육과정 및 교육 관련 문서, 평가인증기준, 선행연구를 검토하고 분석하였다. 이후 연구진의 논의 및 교육평가센터 교원들의 협의를 바탕으로 교육과정과 교육의 활동 전체를 포괄하여 종합적인 평가를 실행할 수 있는 교육프로그램 평가체계 틀의 초안을 구성하였다. 둘째, 델파이 기법을 적용하여 평가체계 구성요소의 타당성을 검증하고, 최종적인 평가지표 확정을 위한 합의를 도출하였다. 연구절차와 각 단계를 통해 도출된 결과는 Figure 1에 요약하여 제시하였다.

2. 교육프로그램 평가체계 틀 초안 개발

1) 연세대학교 의과대학 교육 맥락 및 미충족 과제 검토

의과대학 교육프로그램의 효과적인 평가와 개선을 위해서는 현재 교육프로그램에 내재된 다양한 물리적, 사회적, 정책적 환경요인들을 면밀히 분석할 필요가 있다. 이러한 맥락적 요인 및 요인 간 상호작용은 교육프로그램의 실행과정과 결과에 중요한 영향을 미친다[29]. 또한 의도한 교육성과와 실제 교육 결과 간의 차이를 파악하고 그 원인을 분석하여 미충족 과제 및 개선이 필요한 부분을 명확히 할 필요가 있다. 본 연구에서는 이를 위해 의과대학 교육과정 개발 및 평가, 교육정책 등과 관련하여 발간된 교내 보고서, 규정, 회의자료와 같은 문서를 포괄적으로 검토하

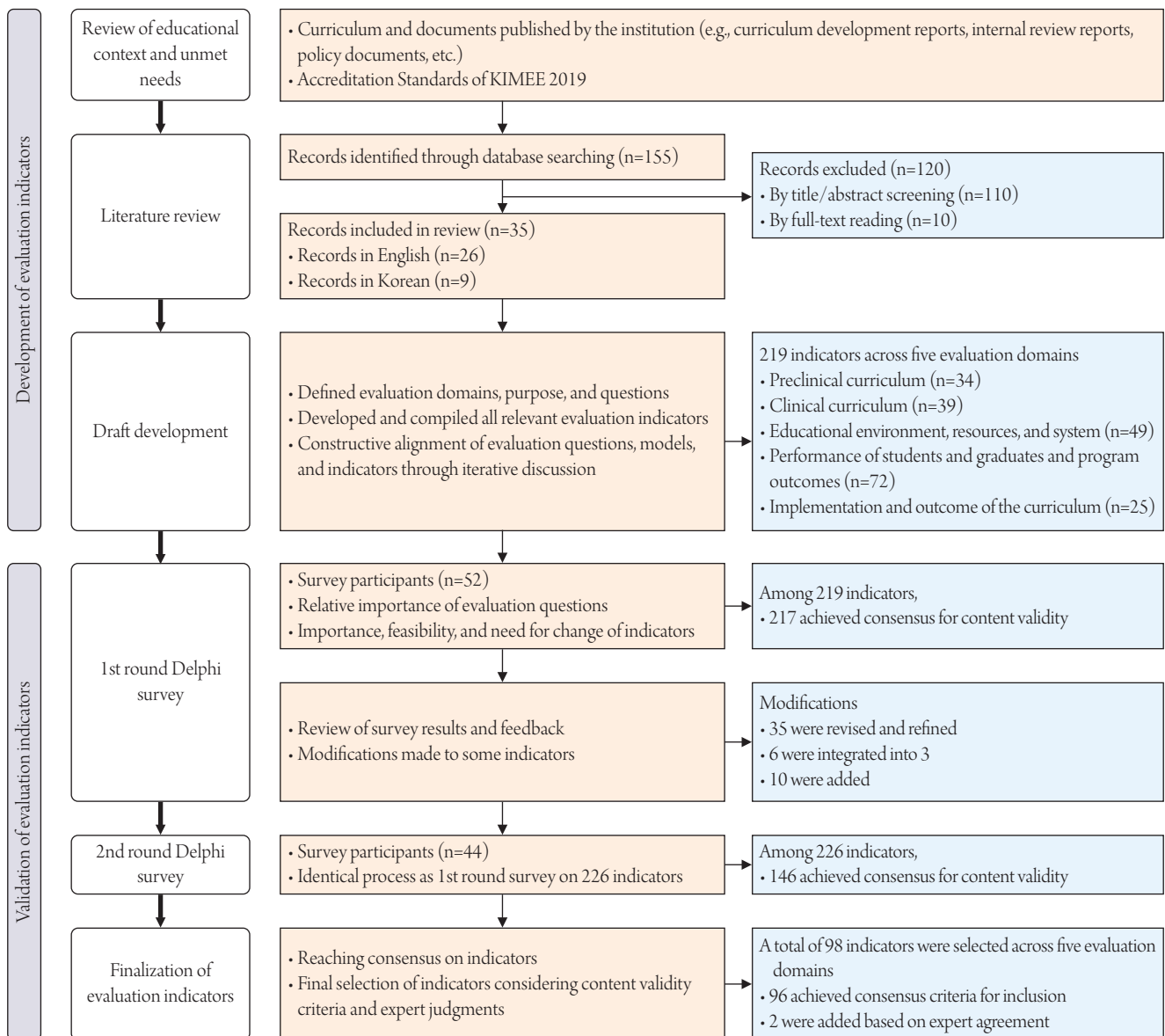


Figure 1. Overview and results of the evaluation framework development process.

였고, 국내 의학교육 평가인증기준의 세부 항목과 이전 평가인증 결과를 분석하였다.

2) 문헌고찰

의과대학 교육프로그램 평가의 방향성을 설계하고 교육활동 전개에 요구되는 핵심적인 교육성과의 개념을 정의하기 위해 국내외 교육프로그램 평가에 대한 문헌고찰을 시행하였다. 2022년 4월에 2018년 3월부터 2022년 3월 사이에 발간된 문헌을 대상으로 국내외 데이터베이스를 검색하였으며, 국외 문헌은 PubMed (www.ncbi.nlm.nih.gov/pubmed), 국내 문헌은 학술연구정보서비

스(<https://www.riss.kr>)를 활용하였다. 주요 검색어로는 연구대상인 “의학교육(medical education)”과 “의과대학(medical school),” 중재방법인 “프로그램 평가(program evaluation)”와 “질 향상(quality improvement)”을 기본으로 하였고, “평가모형(evaluation model)”과 “질 향상 체계(quality improvement framework)”를 추가적으로 포함하였다. 이를 통해 총 155건의 문헌이 도출되었다. 문헌 선정은 1저자가 2차에 걸쳐 시행하였으며 1차는 제목과 초록을 중심으로, 2차는 원문을 검토하였다. 영어나 한국어 외 다른 언어로 기술된 연구, 원문 이용이 불가능한 경우, 프로그램 개발 또는 교수개발 프로그램 평가 등에 중점을 둔 연구는 제외하였고,

최종적으로 영어논문 26건, 한글논문 9건을 포함한 총 35개의 문헌을 선정하였다. 선별된 문헌은 선행연구에서 적용된 평가모형과 주요 평가요소를 이해하고, 평가체계 설계의 이론적 토대를 마련하기 위한 기초자료로 활용하였다(Appendix 1).

3) 교육프로그램 평가체계 초안 수립

교육프로그램 평가체계 틀의 초안을 구성하기 위하여 연구진 3인은 해당 자료를 심층적으로 검토하고 11차례의 논의를 통해 초안 작성과 수정작업을 시행하였다. 프로그램 평가 설계절차에 따라 평가활동의 정당화 및 근거를 확인하기 위해 평가영역을 분류하고, 각 영역별 평가목적에 도출하였다. 이어서 평가의 주요 목적을 반영하면서 평가영역 및 내용을 다루기에 적절한 이론적 평가모형을 구성하고, 이에 적합한 세부적인 평가질문을 수립하였다. 또한 평가모형의 구성요소를 고려하고 평가질문에 대한 근거 기반의 결과를 제시할 수 있도록 구체적이고 세분화된 평가지표를 개발하였다. 이를 통해 '평가목적-평가모형-평가질문-평가지표'의 정합을 이루는 평가체계 틀을 구성하였다. 이는 프로그램의 효과성 및 효율성을 평가하기 위해 평가계획을 구조화하는 일반적인 절차로 초기에 넓은 범위의 목적을 제시한 뒤 이를 이론적 모형을 통해 구체화하여 주요 평가질문과 단계, 필요한 자원을 명확히 정의하는 과정을 포함한다[30,31]. 이러한 접근은 평가의 초점을 정교화하고 각 평가의 목적과 평가지표, 자료수집활동 간의 논리적 연결성과 일관성을 유지하고, 효과적 평가설계와 실행에 도움이 될 수 있다. 이후 교육평가센터에서 활동하는 의과대학 교원 5인의 추가적인 검토와 합의를 거쳐 교육프로그램 평가체계 틀의 최종적인 초안을 수립하였다.

3. 델파이 조사

평가지표 초안의 타당도를 검증하고 합의를 도출하기 위해 본 연구에서는 Murry와 Hammons [32]가 제안한 수정 델파이 기법(modified Delphi technique)을 적용하였다. 수정 델파이 기법은 문헌고찰, 사례연구 등 기초연구를 통해 구성한 체계적이고 구조화된 설문지를 패널에게 제시하므로 조사과정의 견고하다는 특징이 있다[33].

1) 델파이 패널 구성

델파이 패널은 전문성을 기반으로 구성하였으며, 델파이 조사의 타당성 확보를 위해 필요한 참여자 수를 고려하여 총 55명을 대상으로 하였다. 패널 구성의 기준은 연세대학교 의과대학 교육과정 및 맥락에 대한 전문적인 이해와 실증적 경험을 보유한 대상자를 포함하는 데 중점을 두어 평가체계 설계의 실효성을 높이 고자 하였다. 주요 패널은 교육전문가인 의과대학 교수, 주요 보

직자, 교육 관련 위원회 위원으로 구성하였다. 또한 교육에 대한 다각적인 이해와 관점을 반영하기 위해서 학생 대표, 행정 직원, 타 의과대학 교수, 본교 보직자, 인턴/전공의 대표, 지역사회 의사, 학부모 대표 등 다양한 이해관계자도 포함하였다. 총 55명의 패널은 2차에 걸친 델파이 설문조사에 최소 한 번 이상 참여하였으며, 각 평가영역별로 16-18명의 패널이 배정되었다. 이 중 17명은 전문성의 범위와 평가영역 간 연관성을 고려하여 두 개 이상의 평가영역에 배정하였다. 델파이 패널에 대한 세부적인 정보는 Table 1에 제시하였다.

2) 델파이 조사절차

2022년 5월 2일부터 11일까지 1차 델파이 조사, 2022년 5월 26일부터 6월 6일까지 2차 델파이 조사를 시행하였다. 평가영역별 설문조사지는 온라인 설문조사도구인 구글 폼(Google Forms;

Table 1. Sociodemographic characteristics of the participants in the modified Delphi panel (N=55)

Characteristic	No. (%)
Age group (yr)	
20-29	9 (16.4)
30-39	7 (12.7)
40-49	20 (36.4)
≥50	19 (34.5)
Gender	
Female	14 (25.5)
Male	41 (74.5)
Educational level	
Undergraduate medical student	7 (12.7)
Bachelor's degree	7 (12.7)
Master's degree	3 (5.5)
Doctoral degree	38 (69.1)
Types of stakeholders	
Professors from Yonsei University College of Medicine	38 (69.1)
Employees from Yonsei University College of Medicine	2 (3.6)
Professors from other Medical Schools	2 (3.6)
Director of Institute for Higher Education Innovation Yonsei University	1 (1.8)
Director of Community Health Center	1 (1.8)
Graduates from Yonsei University College of Medicine	3 (5.5)
Students from Yonsei University College of Medicine	7 (12.7)
Parent	1 (1.8)
Years of teaching experience ^{a)}	
≤5	10 (21.3)
6-10	9 (19.1)
11-15	12 (25.5)
16-20	7 (14.9)
>20	9 (19.1)

^{a)} Relevant to 47 participants, excluding students and parents.

Google LLC, Mountain View, CA, USA)을 활용하여 제작하였고, 1, 2차 동일하게 (1) 평가질문의 우선순위 기입, (2) 각 평가지표의 중요도(교육프로그램 평가에서 중요한 정도), 실행 가능성(평가지표에 대한 자료를 수집하고 조사하는 데 실행이 용이한 정도), 변화 필요성(교육의 지속적 질 향상을 위해 현재의 교육에서 변화가 필요한 정도)을 5점 리커트 척도로 평정, (3) 평가지표에 대한 수정, 보완 의견 제시, (4) 추가되어야 할 평가지표 제안 항목을 포함하였다. 델파이 패널들에게 델파이 조사에 대한 상세한 설명문과 함께 구글 폼 링크가 첨부된 이메일을 발송하여 조사를 진행하였다.

1차 조사에서 수집된 피드백을 바탕으로 연구진이 평가지표를 수정 및 보완하고 새로운 지표를 추가하였다. 1차 조사는 평가체계의 틀과 구성요소에 대한 패널들의 이해를 촉진하고, 피드백을 수집하여 평가지표를 정교하게 보완하는 것에 초점을 맞추었기 때문에 1차 조사에서 합의기준에 도달하지 못한 지표라도 제외하지 않고 2차 조사를 통해 재평가될 수 있도록 하였다. 2차 설문조사지에는 1차 응답의 기술통계값(평균, 중앙값, 점수 범위, 리커트 척도 점수별 응답자 비율)과 피드백을 함께 제공하였다.

델파이 조사 이후의 합의(post-group consensus) 단계로 평가영역별 회의를 개최하여 델파이 패널이 자신의 관점을 정당화하는 논의를 통해 최종 평가체계 결정에 도달할 수 있도록 패널 간 상호작용의 기회를 제공하였다. 이는 델파이 기법을 적용할 때 논의 부족으로 인한 의사결정의 불확실성이 초래되는 문제를 보완할 수 있다[34].

4. 통계분석

평가질문 항목의 상대적 중요도를 산출하기 위해 항목별 평균 순위에 기초한 역순위 가중치(rank reciprocal weights) 방법을 적용하였다[35]. 이는 다기준 의사결정(multi-criteria decision making) 상황에서 활용되는 순위 기반 가중치로, j 개의 항목으로 구성된 평가질문의 우선순위를 결정하기 위해 각 항목에 대해 1부터 j 사이의 순위를 중복 없이 부여하는 조건부 평정과정에 의해 우선순위를 결정하는 방식을 의미한다[36]. 본 연구에서는 평가질문의 중요도를 결정하기 위해서 1순위부터 j 순위까지를 묻는 응답자료에 각 j 에서 1까지의 값을 부여하고, 각 항목별 순위 총합의 비율로 계산하여 가중치를 산출하였다. 평가질문의 가중치 합은 1로 설정하였으며, 가중치 값이 클수록 상대적 중요도가 높은 평가질문으로 간주하여 평가의 우선순위를 결정하였다.

한편, 2차에 걸친 델파이 조사에서 평가지표의 중요도, 실행 가능성, 변화 필요성에 대한 응답을 분석하여 평가지표의 타당성을 검토하였다. 평가지표의 타당성은 평가영역을 구성하고 있는 평가지표의 내용타당도를 검증하기 위한 정량적 방법으로 내용

타당도 비율(content validity ratio, CVR)을 비교하였다. CVR은 평가도구 개발에 있어 각 항목, 즉 평가지표가 포함되어야 하는지에 대한 전문가 패널의 합의 비율에 근거한 방법이며, 패널의 수가 작을 경우 우연에 의해 포함되는 항목이 발생할 수 있기 때문에 동원된 패널 수에 따른 CVR 최소값의 기준에 따라 내용타당도를 검증한다[37]. CVR 산출에 있어 1차 델파이 조사에서는 5점 척도로 구성된 검사에서 3점 이상의 평점을 합의의 기준으로 설정하였다. 이는 패널들이 평가지표에 대한 충분한 이해를 바탕으로 의견을 제공할 시간이 부족할 수 있음을 감안하였으며, 상대적으로 낮은 기준을 적용함으로써 초기 단계에서 평가지표에 대한 다양한 패널 의견을 수집하여 평가지표 설계의 방향성을 보완하고자 함이었다. 반면, 2차 델파이 조사에서는 4점 이상의 평점을 합의의 기준으로 설정하였다. 차이를 거듭하면서 참여 패널들이 평가지표의 의미와 중요성에 대해 점진적으로 더 깊이 이해할 수 있었으므로, 보다 엄격한 기준을 적용하여 지표의 정밀성을 높이고, 패널 간의 의견 일치를 강화하기 위함이었다. 본 연구에서 평가영역별 1, 2차 델파이 조사에 응답한 패널의 수는 11-15명이었으며, CVR 값은 0.49-0.59를 기준으로 하였다. 한편, 합의의 일관성을 검증하기 위해 패널 간 의견의 편차를 고려한 합의도(degree of consensus), 수렴도(degree of convergence) 값을 고려하였다. 본 연구에서는 합의도가 0.75 이상, 수렴도가 0.50 이하일 때를 의견 합치의 기준으로 정하였다. 평가지표의 타당성은 내용타당도를 중점으로 평가하였으며, 합의도와 수렴도는 부가적인 참고자료로 활용하였다.

5. 윤리적 고려

본 연구는 연세의료원 기관생명윤리위원회(institutional review board, IRB)의 승인을 받았으며 통상적인 교육의 과정에서 발생한 자료를 활용한 연구로 대상자 동의 면제를 득하였다(IRB no., 4-2022-1620).

결과

1. 교육프로그램 평가체계 틀 초안

1) 평가영역 및 평가목적

연세대학교 의과대학 교육의 목적과 성과, 교육과정, 입학에서 졸업까지의 교육의 과정, 학습환경, 교육 전반의 시스템과 자원이 교육프로그램 평가에서 다루어 질 수 있도록 5개의 평가영역을 구성하였다. 해당 영역은 교과목 단위의 개선을 위한 '1. 기본 의학교육과정'과 '2. 임상실습교육과정' 영역, 교육프로그램 내 다양한 체계의 적절성 확보와 질 향상을 위한 '3. 교육환경, 자원

및 운영체계' 영역, 재학생 및 졸업생의 성과를 점검하고 지속적 개선을 위한 '4. 학생 역량 및 교육프로그램 성과' 영역, 전체 교육과정의 계획, 운영, 성과를 평가하는 '5. 교육과정의 운영 및 성과' 영역이다. 각 평가영역은 교과목 단위의 개선에서 시작하여 점차 교육과정 전반과 그 성과에 대한 종합적인 평가로 확장되는

구조를 가지며, 교육프로그램의 효과적인 개선 및 성과분석을 위한 점진적인 접근을 위하여 설계되었다. 평가영역의 특성에 따라 평가목적은 수립하였으며 영역별 평가목적은 Table 2에서 확인할 수 있다.

Table 2. Constructive alignment of the purposes of evaluation, evaluation questions, and evaluation model components

Domain	Purpose of evaluation	Evaluation questions	Evaluation model components	Weights	
				1st	2nd
1. Preclinical curriculum	To identify the types and content of the phases/courses/lessons that require improvement in the Yonsei University College of Medicine curriculum and provide feedback based on the evaluation results to promote quality enhancement.	Q1. Has an appropriate plan been established for each subject/clerkship to achieve the intended educational outcomes?	(Logic) Input	0.25	0.31
		Q2. Are the resources for operating the subject/clerkship adequate?	(Logic) Input	0.13	0.14
		Q3. Was the subject/clerkship implemented as planned?	(Logic) Activity/output	0.20	0.19
		Q4. Were student assessments valid and reliable?	(Logic) Activity/output	0.18	0.15
		Q5. Did the subject/clerkship achieve its intended educational outcomes?	(Logic) Outcome: (Kirkpatrick) reaction/learning	0.24	0.20
2. Clinical curriculum	To identify the types and content of the phases/courses/lessons that require improvement in the Yonsei University College of Medicine curriculum and provide feedback based on the evaluation results to promote quality enhancement.	Q1. Has an appropriate plan been established for each subject/clerkship to achieve the intended educational outcomes?	(Logic) Input	0.22	0.21
		Q2. Are the resources for operating the subject/clerkship adequate?	(Logic) Input	0.14	0.07
		Q3. Was the subject/clerkship implemented as planned?	(Logic) Activity/output	0.18	0.18
		Q4. Were student assessments valid and reliable?	(Logic) Activity/output	0.22	0.25
		Q5. Did the subject/clerkship achieve its intended educational outcomes?	(Logic) Outcome: (Kirkpatrick) reaction/learning	0.24	0.29
3. Educational environment, resources, and system	To identify areas requiring quality improvement throughout the entire educational process from admission to graduation at Yonsei University College of Medicine and provide feedback based on evaluation results to promote quality enhancement.	Q1. Are changes in the current and future medical environment reflected in the educational program?	(CIPP) Context	0.17	0.19
		Q2. Are the mission and outcomes of the educational program appropriately established?	(CIPP) Context	0.20	0.20
		Q3. Are students' educational needs and characteristics reflected in the educational program?	(CIPP) Context	0.17	0.17
		Q4. Are the resources for operating the educational program adequate?	(CIPP) Input	0.12	0.12

(Continued on next page)

Table 2. Continued

Domain	Purpose of evaluation	Evaluation questions	Evaluation model components	Weights	
				1st	2nd
		Q5. Are the student support systems adequate? (academic, social, financial, mental health, etc.)	(CIPP) Input/process	0.09	0.08
		Q6. Is the student selection system appropriate?	(CIPP) Input/process	0.09	0.11
		Q7. Is the educational program evaluation system appropriate?	(CIPP) Input/process	0.08	0.07
		Q8. Is the quality improvement system for education adequate?	(CIPP) Input/process	0.09	0.05
4. Performance of students and graduates and program outcomes	To assess whether the educational program goals pursued by Yonsei University College of Medicine are being adequately achieved and to provide feedback on the evaluation results to promote quality enhancement	Q1. Does the educational program have adequate expertise necessary to achieve program outcomes?	(CIPP) Input/process	0.15	0.14
		Q2. Is the student competency evaluation system appropriate? (principle, methods, areas, alignment with outcome/educational content, standard setting, feedback, etc.)	(CIPP) Input/process	0.19	0.22
		Q3. Is the level of academic progress of students appropriate? (whether there is improvement or achievement of phase/course outcomes)	(CIPP) Product: (Kirkpatrick) learning	0.14	0.14
		Q4. Is the students' performance adequate?	(CIPP) Product: (Kirkpatrick) learning/behavior	0.16	0.20
		Q5. Is the graduates' performance adequate?	(CIPP) Product: (Kirkpatrick) behavior	0.13	0.12
		Q6. Is the medical school educational program achieving its mission and goals?	(CIPP) Product: (Kirkpatrick) result	0.12	0.10
		Q7. Are continuous improvements being made to the medical school educational program?	(CIPP) Product: (Kirkpatrick) result	0.11	0.10
5. Implementation and outcome of the curriculum	To assess whether the goals pursued by CDP2013/CDP2023 have been achieved.	Q1. Has the curriculum been planned/organized according to SORI principles?	(CIPP) Input	0.35	0.33
		Q2. Has the curriculum been implemented in accordance with SORI principles?	(CIPP) Process	0.35	0.30
		Q3. What achievements have been made in each area of SORI?	(CIPP) Product: (Kirkpatrick) reaction/learning/behavior/result	0.31	0.37

CIPP, context, input, process, and product evaluation model; SORI, student-oriented, outcome-based, research-oriented, and integrated learning (acronym for the four principles of the educational program at Yonsei University College of Medicine).

2) 평가모형 및 평가질문

평가목적에 적합한 평가요소를 구체화하고 이에 따른 평가지표를 개발하기 위해 평가모형을 설계하였다. 단계별 교육과정을 대상으로 하는 1, 2영역은 logic 모형과 Kirkpatrick 모형을 통합한 평가모형을 구성하였고, 교육프로그램 전반의 시스템과 자원, 운영과 성과를 다루는 3-5 영역은 context, input, process, and prod-

uct evaluation model (CIPP 모형)과 Kirkpatrick 모형을 통합한 평가모형을 설계하였다. Logic 모형은 프로그램의 구성요소로 투입(input), 활동(activity), 산출(output), 단기 및 장기 성과(outcome) 간의 논리적 관계를 개념화해주는 모형으로 프로그램의 계획, 실행, 평가에 이르는 선형적 흐름이 강조된다[4]. CIPP 모형은 평가영역을 상황(context), 투입(input), 과정(process), 산출(product)로

제시하며, 평가요소 간 역동적이고 비선형적인 관계를 고려하기 때문에 장기적으로 운영되는 교육프로그램의 전 과정을 평가하고 상황과 맥락을 적극적으로 반영하는 데 유용하다[4]. Logic 모형이나 CIPP 모형과는 달리 Kirkpatrick 모형은 교육프로그램의 성과 평가에 초점을 맞추고 있으며, 교육성과를 반응(reaction), 학습(learning), 행동(behavior), 결과(result)와 같이 4단계 수준으로 구분하여 평가한다. Figure 2에 제시된 것처럼, 본 연구에서는 logic 모형의 성과요소, CIPP 모형의 산출요소에 각각 Kirkpatrick 모형을 결합한 통합적 접근을 통해 교육프로그램의 효과성을 체계적으로 측정할 수 있는 평가모형을 구성하였다. 이는 단일 모형이 지닐 수 있는 평가의 편향성과 한계점을 보완할 수 있으며, 교육의 각 단계를 선형적으로 평가하는 것을 넘어 각 과정에서 요소 간 상호작용을 면밀히 분석하고, 지속적인 조정을 통해 유연한 평가가 가능한 설계이다.

다음으로 영역별 평가목적의 달성 여부를 확인하고 통합 평가 모형의 구성요소를 반영할 수 있는 평가질문을 수립하였다. 1, 2 영역의 경우 5개의 동일한 평가질문으로 이루어져 있으며, 3영역은 8개, 4영역은 7개, 5영역은 3개의 평가질문으로 이루어져 있다. 1, 2영역의 평가질문을 예시로 들면, ‘1. 각 교과목은 의도한 교육성과를 달성하기에 적합한 계획을 수립하였는가?’, ‘2. 교과목 운영을 위한 자원이 적절한가?’, ‘3. 교과목은 계획한 대로 실행/운영되었는가?’, ‘4. 학생 평가는 타당하고 신뢰롭게 시행되었는가?’, ‘5. 교과목이 의도한 교육성과를 달성하였는가?’와 같다. Table 2에서 5개 평가영역의 평가질문과 함께, 평가영역별 평가 목적, 평가질문, 평가모형의 구성요소가 연결된 구조적, 논리적

체계를 확인할 수 있다.

3) 평가지표 초안

연구진은 2021년 10월부터 2022년 4월까지 11회의 반복적인 논의를 통해 총 219개의 평가지표 초안을 개발하였다. 평가영역별 지표의 개수는 1영역 34개, 2영역 39개, 3영역 49개, 4영역 72개, 5영역 25개였다. 평가질문별로는 적게는 2개, 많게는 17개의 평가지표가 개발되었다.

2. 평가지표 타당화 및 최종 선정

1) 1차 델파이 조사결과

1차 델파이 조사는 총 52명의 패널이 참여하였고, 평가영역별로는 13-15명의 패널이 참여하였다. 평가영역별 평가질문 간 상대적 중요도에 대한 복합가중치를 백분율로 산정한 결과는 Table 2에 제시되어 있다. 1차 델파이 조사결과, ‘기본의학교육과정’ 영역에서는 교과목 계획수립의 적절성(0.25)이, ‘임상실습교육과정’ 영역에서는 교과목의 교육성과 달성 정도(0.24)가 가장 중요한 질문으로 나타났다. 또한 ‘교육환경, 자원 및 운영체계’ 영역에서는 교육프로그램의 사명과 성과 수립의 적절성(0.20), ‘학생 역량 및 교육프로그램 성과 달성’ 영역은 학생 역량 평가체계의 적절성(0.19), 마지막으로 ‘교육과정의 운영과 성과’ 영역은 교육과정의 계획, 편성 및 실행(0.35)이 중요한 평가질문으로 선정되었다.

평가지표의 타당성을 검증한 결과, 전체 219개의 지표 중 204

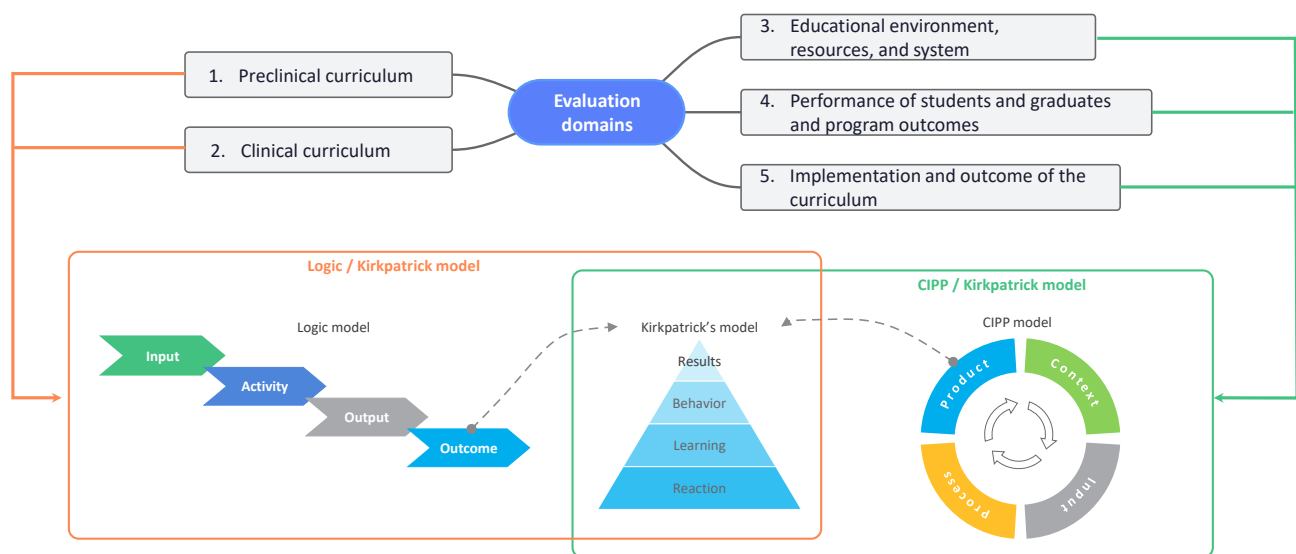


Figure 2. Integrated evaluation models applied to each of the five evaluation domains.

개의 지표가 중요도, 210개의 지표가 실행 가능성, 195개의 지표가 변화 필요성 측면에서 내용타당도 기준을 통과하였다. 중요도, 실행 가능성, 변화 필요성 중 하나 이상의 내용타당도 기준을 만족한 지표는 전체 219개 중 217개로 나타났다. 또한 중요도, 실행 가능성, 변화 필요성 기준 중 하나 이상의 합의도를 달성하지 못한 지표는 18개, 수렴도를 달성하지 못한 지표는 5개로 나타났다. 그러나 1차 조사에서는 델파이 패널 간 이해 정도가 다른 평가지표가 있을 수 있기 때문에 합의기준을 충족하지 못한 평가지표라도 제외하지 않고 2차 조사에서 재평가 받을 수 있도록 하였다.

1차 델파이 조사에서 평가지표에 대한 패널의 의견을 수집한 결과, 주요 의견으로는 교육과정의 적절성을 정성적으로 평가할 수 있는 지표 추가(가령, 과목운영 사전 회의 시 운영계획 및 의사결정 논의의 질을 측정할 수 있는 항목, 학생 및 교육자로부터 수집된 피드백의 질을 평가하는 항목, 졸업 포트폴리오의 주기적 성과를 측정하는 항목 등), 평가지표에 대한 이해를 돕고 구체적인 응답을 유도하기 위해 실제 활동을 포함한 예시 포함(가령, 교육자원의 충분성 정도를 평가하기 위해 과목 운영비, 교과목 운영을 지원하는 행정지원의 인력비율, 지원체계 등의 세부항목을 포함하여 지표를 개선), 평가 관련 용어에 대한 설명을 보완하여 평가지표의 의미를 명료화, 개념 측면이나 교육과정 실행상 연계될 수 있는 평가항목은 통합하여 제시하는 것 등이 있었다. 이와 같은 패널들의 제안을 바탕으로 2차 델파이 조사에서는 일부 평가항목의 추가, 삭제, 통합, 분리가 이루어졌으며, 최종적으로 10개의 평가지표 추가, 35개의 평가지표 수정 및 보완, 6개의 평가지표가 3개의 지표로 통합되었다.

2) 2차 델파이 조사결과

2차 델파이 조사는 총 44명의 패널이 참여하였고, 평가영역별로는 11-14명의 패널이 참여하였다. 2차 조사에서는 1차 조사와 동일하게 평가질문의 상대적 중요도 및 평가지표의 중요도, 실행 가능성, 변화 필요성에 대해 질문하였으며, 개발한 평가지표의 수정이 필요하거나 또는 추가해야 할 사항이 있는 경우 자유롭게 의견을 기입할 수 있도록 하였다. 2차 델파이 조사에는 1차 조사를 통해 수집된 패널의 피드백을 반영하여 총 226개의 평가지표가 포함되었다.

2차 조사에서 평가영역별 평가질문의 상대적 중요도 순서는 1차 조사와 유사하게 나타났으나, '교육과정의 운영과 성과' 영역에서는 1차 조사와 달리 교육과정의 성과(0.37)가 교육과정의 계획, 운영과 비교 시 가장 중요한 평가질문으로 나타났다(Table 2). 평가질문별 산출된 가중치 값은 최종 평가지표 선정에서 활용되었으며, 가중치 값이 높은 평가질문에 최대한 평가지표를 많이 포함할 수 있도록 고려하였다.

평가지표의 타당성을 검증한 결과, 전체 평가지표 중 중요도는 109개, 실행 가능성은 86개, 변화 필요성은 20개가 내용타당도 기준을 충족하였다. 중요도, 실행 가능성, 변화 필요성 측면에서 하나 이상의 내용타당도 기준을 만족한 평가지표는 총 146개로 나타났다. 또한 중요도, 실행 가능성, 변화 필요성 기준 중 하나 이상의 합의도를 달성하지 못한 지표는 7개였으며, 수렴도를 달성하지 못한 지표는 없었다.

3) 최종 평가지표 선정

평가지표의 선정을 위해 2차 델파이 조사결과를 바탕으로 각 평가영역별로 전문가 패널을 대상으로 한 협의회를 개최하였다. '기본의학교육과정' 영역에는 8명, '임상실습교육과정' 영역에는 12명, '교육환경, 자원 및 운영체계' 영역에는 7명, '학생 역량 및 교육프로그램 성과 영역'에는 8명, '교육과정의 운영 및 성과' 영역에는 7명이 참여하였다. 평가영역별 전문성을 고려하여 일부 패널은 복수 영역에 해당하는 협의회에 참석하였으며, 평가목적 달성을 위한 지표 설정의 균형 및 연계성을 유지하기 위해 제1저자는 5개 영역 협의회에 모두 참석하여 논의과정을 지원하였다. 또한 제2저자는 4개 영역, 제3저자는 2개 영역에 참석하였다. 협의회에서는 평가지표의 세부사항에 대해 자유롭게 의견을 교환하고, 각 지표 간 중복성, 개선 가능성, 중요도를 심층적으로 다루었다. 특히 특정 평가지표가 실제 교육프로그램에 적용될 때 교육적 영향력에 대해 실질적인 사례를 공유하여 논의를 보다 심화하였다. 협의회 결과를 바탕으로 연세대학교 의과대학 교육프로그램의 목표 성취를 효과적으로 평가할 수 있는 평가지표를 선정하였으며, 이를 평가질문에 맞게 재조정하고, 구체적인 기준 및 예시를 추가하여 수정 및 보완하였다. 도출된 최종 평가지표의 목록은 Table 3과 같다. '기본의학교육과정'과 '임상실습교육과정' 영역에서는 각 영역별 총 13개의 평가지표가 선정되었으며, 선정기준은 중요도, 실행 가능성, 변화 필요성 기준 중 하나 이상의 내용타당도가 충족된 평가지표가 포함되었다. '교육환경, 자원 및 운영체계' 영역에서는 총 18개의 평가지표가 선정되었다. '36. 의학과 4학년 <특성화선택과정> 교류협정기관 실습 견수'와 '39. 학년별 정신건강 실태조사율' 평가지표는 2차 델파이 조사에서 중요도, 실행 가능성, 변화 필요성에 대한 내용타당도 기준을 통과하지는 못하였으나, 교육의 질 관리를 위해서 꾸준히 추적관찰이 필요한 항목으로 논의되어 최종 평가지표에 포함되었다. '학생역량 및 교육프로그램 성과 달성' 영역에서는 총 34개의 평가지표가 선정되었으며, 수년간에 이르는 교육자료의 수집 및 분석, 장기적 성과의 조사가 필요한 평가영역의 특성상 대부분의 평가지표에서 실행 가능성을 평가하는 내용타당도가 낮게 나타났으며, 중요도 측면에서 내용타당도가 높은 평가지표가 위

Table 3. Second-round Delphi survey results for final evaluation indicators

Evaluation indicators	Importance				Feasibility				Necessity for change			
	Mean±SD	CVR	CSS	CVG	Mean±SD	CVR	CSS	CVG	Mean±SD	CVR	CSS	CVG
Domain 1: Preclinical curriculum												
1. (EQ1) Holding course committee pre-meetings for course implementation	3.50±0.76	0.29	0.75	0.50	3.93±0.83	0.57	1.00	0.00	3.42±0.79	-0.17	0.67	0.50
2. (EQ1) Establishment of plans for setting pass scores for each course	4.21±0.58	0.86	0.81	0.38	3.79±0.58	0.43	0.81	0.38	4.50±0.67	0.83	0.80	0.50
3. (EQ1) Establishment of plans for remediation in each course	4.21±0.80	0.57	0.75	0.50	4.14±0.66	0.71	0.81	0.38	4.17±0.72	0.67	0.75	0.50
4. (EQ1) Development of test blueprints containing content areas, item difficulty, number of items, and scoring	4.14±0.53	0.86	1.00	0.00	4.00±0.39	0.86	1.00	0.00	4.00±0.85	0.67	0.94	0.13
5. (EQ2) Sufficiency of administrative support	4.36±0.63	0.86	0.75	0.50	3.71±0.99	0.43	0.81	0.38	4.08±0.67	0.67	0.94	0.13
6. (EQ3) Application of various teaching and learning methods besides lectures	4.21±0.58	0.86	0.81	0.38	4.14±0.66	0.71	0.81	0.38	3.92±0.79	0.33	0.69	0.63
7. (EQ3) Implementation of pre-defined standard-setting methods stated in the course syllabus	4.14±0.66	0.71	0.81	0.38	3.71±0.47	0.43	0.81	0.38	3.75±0.75	0.17	0.75	0.50
8. (EQ4) Notification of students about the method for setting pass scores before the start of the course	4.07±0.92	0.57	0.75	0.50	4.21±0.80	0.57	0.75	0.50	3.92±0.67	0.50	0.94	0.13
9. (EQ4) Implementation of pre-planned methods of determining pass scores	4.36±0.74	0.71	0.78	0.50	4.00±0.68	0.57	1.00	0.00	3.67±0.78	0.33	0.75	0.50
10. (EQ4) Adequacy of course assessment and the percentage of inappropriate items in the exam	4.14±0.77	0.57	0.75	0.50	3.79±0.80	0.14	0.75	0.50	3.67±0.65	0.17	0.75	0.50
11. (EQ4) Establishment of a system for providing feedback and facilitating interaction between students and teachers	4.07±0.47	0.86	1.00	0.00	3.71±0.73	0.14	0.75	0.50	4.08±0.67	0.67	0.94	0.13
12. (EQ5) Course evaluation results from students	4.14±0.53	0.86	1.00	0.00	4.29±0.61	0.86	0.75	0.50	3.42±0.67	0.00	0.71	0.50
13. (EQ5) Predictive validity of course grades about academic achievement assessed through other examinations	4.00±0.68	0.57	1.00	0.00	3.86±0.36	0.71	1.00	0.00	3.83±0.72	0.33	0.75	0.50
Domain 2: Clinical curriculum												
14. (EQ1) Distribution of graduation competencies covered in the clerkship curricula	4.36±0.50	1.00	0.75	0.50	4.09±0.54	0.82	1.00	0.00	3.36±0.67	-0.45	0.83	0.25
15. (EQ1) Establishment of plans for setting pass scores for each clerkship	4.00±0.45	0.82	1.00	0.00	4.27±0.65	0.82	0.75	0.50	3.90±0.88	0.20	0.56	0.88
16. (EQ2) Sufficiency of administrative support	4.55±0.52	1.00	0.80	0.50	4.00±0.63	0.64	1.00	0.00	4.10±0.74	0.60	0.81	0.38
17. (EQ3) Percentage of essential diseases listed in syllabi that students were given learning opportunities during clerkship	4.55±0.52	1.00	0.80	0.50	3.82±0.60	0.45	0.88	0.25	3.90±0.88	0.60	1.00	0.00
18. (EQ3) Percentage of OSCE/CPX items listed in syllabi that were actually executed during clerkship	4.64±0.50	1.00	0.80	0.50	4.18±0.60	0.82	0.88	0.25	3.20±0.63	-0.40	0.75	0.38
19. (EQ4) Notification of students about the method for setting pass scores before the start of clerkship	4.36±0.50	1.00	0.75	0.50	4.45±0.52	1.00	0.75	0.50	4.00±0.47	0.80	1.00	0.00
20. (EQ4) Validity of the criteria for selecting students eligible for remediation	4.55±0.52	1.00	0.80	0.50	3.73±0.65	0.27	0.75	0.50	4.00±0.67	0.60	1.00	0.00
21. (EQ4) Descriptions of objective criteria for Honors grade	4.45±0.52	1.00	0.75	0.50	4.00±0.45	0.82	1.00	0.00	3.90±0.74	0.40	0.81	0.38
22. (EQ4) Implementation of assessment methods and grading criteria presented in clerkship syllabi	4.45±0.52	1.00	0.75	0.50	4.00±0.45	0.82	1.00	0.00	3.50±0.71	-0.20	0.67	0.50

(Continued on next page)

Table 3. Continued

Evaluation indicators	Importance				Feasibility				Necessity for change			
	Mean±SD	CVR	CSS	CVG	Mean±SD	CVR	CSS	CVG	Mean±SD	CVR	CSS	CVG
23. (EQ5) Clerkship course evaluation results from students	4.09±0.54	0.82	1.00	0.00	4.27±0.47	1.00	0.88	0.25	3.20±0.42	-0.60	1.00	0.00
24. (EQ5) Patient care competency demonstrated to standardized patients	4.55±0.52	1.00	0.80	0.50	4.27±0.47	1.00	0.88	0.25	3.30±0.82	-0.40	0.75	0.38
25. (EQ5) Results of the Comprehensive Clinical Medicine Examination	4.45±0.52	1.00	0.75	0.50	4.55±0.52	1.00	0.80	0.50	3.10±0.57	-0.60	1.00	0.00
26. (EQ5) Results of the Comprehensive Clinical Performance Examination	4.09±0.54	0.82	1.00	0.00	4.45±0.52	1.00	0.75	0.50	3.20±0.79	-0.60	1.00	0.00
Domain 3: Educational environment, resources, and system												
27. (EQ1) Regular analysis of changes in the medical/educational environment	3.75±0.62	0.33	0.75	0.50	3.92±0.67	0.50	0.94	0.13	3.82±0.40	0.64	1.00	0.00
28. (EQ1) Modification of the educational program reflecting changes in the medical/educational environment	4.00±0.95	0.50	0.69	0.63	3.83±0.72	0.33	0.75	0.50	4.00±0.63	0.64	1.00	0.00
29. (EQ2) Establishment of hierarchical learning outcomes, including the exit-phase-course-lesson outcomes	4.67±0.49	1.00	0.80	0.50	3.83±0.58	0.50	0.94	0.13	4.30±0.67	0.80	0.75	0.50
30. (EQ2) Degree of awareness among students and teachers regarding the mission and outcomes of the institution	4.00±0.43	0.83	1.00	0.00	4.00±0.74	0.50	0.88	0.25	3.80±0.63	0.80	1.00	0.00
31. (EQ3) Use of the results of the analysis of student characteristics and performance for program improvement	4.00±0.74	0.83	1.00	0.00	4.08±0.67	0.67	0.94	0.13	3.60±0.70	0.00	0.71	0.50
32. (EQ3) Execution rate of student committee agendas proposed by student representatives	4.00±0.43	0.83	1.00	0.00	4.17±0.58	0.83	0.94	0.13	3.50±0.85	0.00	0.71	0.50
33. (EQ4) Sufficiency of educational facilities and student satisfaction	4.50±0.67	0.83	0.80	0.50	4.25±0.62	0.83	0.75	0.50	4.30±0.82	0.60	0.78	0.50
34. (EQ4) Policies and systems to guarantee protected hours for teaching	4.58±0.67	0.83	0.80	0.50	3.92±0.51	0.67	1.00	0.00	4.00±0.94	0.20	0.50	1.00
35. (EQ4) Number of primary and secondary healthcare facilities affiliated with the medical school for clinical clerkship	4.17±0.72	0.67	0.75	0.50	4.50±0.80	0.67	0.80	0.50	3.60±0.70	0.00	0.71	0.50
36. (EQ4) Number of exchange agreement institutions and exchange students for the <Specialization Elective Course> in 4th year	3.58±0.51	0.17	0.75	0.50	4.25±0.87	0.50	0.72	0.63	3.50±0.53	0.00	0.71	0.50
37. (EQ5) Scholarship payment ratio compared to total tuition income	4.25±0.75	0.67	0.75	0.50	4.42±0.67	0.83	0.78	0.50	3.40±0.52	-0.20	0.67	0.50
38. (EQ5) Effectiveness of remediation programs for underachieving students	4.00±1.04	0.33	0.50	1.00	4.25±0.62	0.83	0.75	0.50	3.50±0.85	0.00	0.71	0.50
39. (EQ5) Screening for mental health status for students in all years	4.17±0.83	0.50	0.69	0.63	3.92±0.79	0.33	0.69	0.63	3.60±0.70	0.00	0.71	0.50
40. (EQ6) Student performance results analyzed by medical school admission type	3.92±0.67	0.83	1.00	0.00	4.00±0.95	0.50	0.69	0.63	3.70±0.82	0.40	0.81	0.38
41. (EQ6) Number of faculty development programs for admission officers and interviewers	4.25±0.62	0.83	0.75	0.50	4.33±0.98	0.67	0.80	0.50	3.90±0.74	0.40	0.81	0.38
42. (EQ7) Percentage of ASK2019 accreditation standards reflected in the program evaluation indicators	4.42±0.79	0.67	0.80	0.50	4.33±0.78	0.67	0.78	0.50	3.60±0.52	0.20	0.75	0.50
43. (EQ7) Systematic and regular analysis of data collected in monitoring the educational process	4.25±0.75	0.67	0.75	0.50	4.25±0.87	0.50	0.72	0.63	3.70±0.67	0.20	0.75	0.50

(Continued on next page)

Table 3. Continued

Evaluation indicators	Importance				Feasibility				Necessity for change			
	Mean±SD	CVR	CSS	CVG	Mean±SD	CVR	CSS	CVG	Mean±SD	CVR	CSS	CVG
44. (EQ8) Feedback of program evaluation results to relevant departments	4.25±0.62	0.83	0.75	0.50	4.25±0.62	0.83	0.75	0.50	3.90±0.74	0.40	0.81	0.38
Domain 4: Performance of students and graduates and program outcomes												
45. (EQ1) Percentage of full-time faculty members who participated in faculty development programs on education for over 3 hours	3.82±0.87	0.45	0.88	0.25	3.18±1.08	-0.09	0.33	1.00	3.27±0.79	-0.09	0.67	0.50
46. (EQ1) Diversity of topics covered in faculty development programs on education	4.18±0.40	1.00	1.00	0.00	3.82±0.60	0.45	0.88	0.25	4.00±0.77	0.45	0.75	0.50
47. (EQ1) Current status of the educational track faculty system by department	4.18±0.60	0.82	0.88	0.25	3.45±1.13	0.27	0.75	0.50	3.91±0.54	0.64	1.00	0.00
48. (EQ2) Number of evaluator training programs on determining valid, reliable pass scores	4.00±0.63	0.64	1.00	0.00	3.73±0.79	0.45	0.88	0.25	4.00±0.63	0.64	1.00	0.00
49. (EQ2) Development of assessment tools by graduation competency	4.18±0.60	0.82	0.88	0.25	3.45±0.82	0.27	0.75	0.50	4.09±0.83	0.45	0.63	0.75
50. (EQ2) Implementation of comprehensive assessments on phase learning outcomes	4.45±0.52	1.00	0.75	0.50	4.00±0.45	0.82	1.00	0.00	3.55±0.52	0.09	0.75	0.50
51. (EQ2) Use of various methods to provide feedback on learning	4.36±0.50	1.00	0.75	0.50	3.73±0.47	0.45	0.88	0.25	4.18±0.60	0.82	0.88	0.25
52. (EQ2) Provision of feedback based on assessment results by subject and phase	4.27±0.47	1.00	0.88	0.25	3.82±0.60	0.82	1.00	0.00	3.91±0.70	0.45	0.88	0.25
53. (EQ3) Percentage of students reaching phase outcomes	4.73±0.47	1.00	0.90	0.25	4.18±0.40	1.00	1.00	0.00	3.40±0.70	0.00	0.71	0.50
54. (EQ3) Results of the Comprehensive Basic Medicine Examination	3.91±0.54	0.64	1.00	0.00	4.18±0.60	0.82	0.88	0.25	3.18±0.60	-0.45	0.83	0.25
55. (EQ3) Results of the Comprehensive Clinical Medicine Examination	4.27±0.65	0.82	0.75	0.50	4.27±0.47	1.00	0.88	0.25	3.18±0.60	-0.45	0.83	0.25
56. (EQ3) Continuous increase in progress test scores	4.00±0.45	0.82	1.00	0.00	4.09±0.70	0.64	0.88	0.25	3.10±0.57	-0.60	1.00	0.00
57. (EQ4) Graduation competency levels self-reported by graduating medical students	4.27±0.65	0.82	0.75	0.50	4.18±0.40	1.00	1.00	0.00	3.36±0.67	-0.09	0.67	0.50
54. (EQ4) Results of the Comprehensive Basic Medicine Examination	3.91±0.54	0.64	1.00	0.00	4.18±0.60	0.82	0.88	0.25	3.18±0.60	-0.45	0.83	0.25
55. (EQ4)) Results of the Comprehensive Clinical Medicine Examination	4.27±0.65	0.82	0.75	0.50	4.27±0.47	1.00	0.88	0.25	3.18±0.60	-0.45	0.83	0.25
58. (EQ4) Patient care competency demonstrated to actual patients in clinical clerkships	4.27±0.65	0.82	0.75	0.50	3.82±0.60	0.45	0.88	0.25	3.27±0.79	-0.09	0.67	0.50
59. (EQ4) Number of students who published papers as the main authors during their medical school years	3.00±1.00	-0.64	1.00	0.00	4.00±0.45	0.82	1.00	0.00	3.18±1.08	-0.27	0.67	0.50
60. (EQ4) Results of assessment on interpersonal communication skills and collaboration	4.27±0.47	1.00	0.88	0.25	3.45±0.52	-0.09	0.67	0.50	3.55±0.52	0.09	0.75	0.50
61. (EQ4) Results of assessment on professionalism	4.18±0.60	0.82	0.88	0.25	3.55±0.52	0.09	0.75	0.50	3.36±0.67	-0.09	0.67	0.50
62. (EQ5) Pass rate of Korean Medical Licensing Examination by year	4.45±0.52	1.00	0.75	0.50	4.64±0.50	1.00	0.80	0.50	3.27±0.65	-0.27	0.67	0.50
63. (EQ5) Performance analysis results of interns	3.91±0.83	0.64	1.00	0.00	4.00±0.00	1.00	1.00	0.00	3.70±0.48	0.40	0.81	0.38
64. (EQ5) Performance analysis results of resident doctors	4.18±0.60	0.82	0.88	0.25	3.64±0.50	0.27	0.75	0.50	3.82±0.60	0.45	0.88	0.25

(Continued on next page)

Table 3. Continued

Evaluation indicators	Importance				Feasibility				Necessity for change			
	Mean±SD	CVR	CSS	CVG	Mean±SD	CVR	CSS	CVG	Mean±SD	CVR	CSS	CVG
65. (EQ5) Number of interns and residents who published papers as main authors during their training years	3.91±0.70	0.45	0.88	0.25	4.09±0.54	0.82	1.00	0.00	3.00±0.63	-0.64	1.00	0.00
66. (EQ5) Participation in degree programs for becoming physician-scientists	3.55±1.04	0.27	0.75	0.50	3.91±0.54	0.64	1.00	0.00	3.00±0.63	-0.64	1.00	0.00
60. (EQ5) Results of assessment on interpersonal communication skills and collaboration	4.27±0.47	1.00	0.88	0.25	3.45±0.52	-0.09	0.67	0.50	3.55±0.52	0.09	0.75	0.50
61. (EQ5) Results of assessment on professionalism	4.18±0.60	0.82	0.88	0.25	3.55±0.52	0.09	0.75	0.50	3.36±0.67	-0.09	0.67	0.50
67. (EQ5) Leadership roles of alums in professional societies and healthcare-related organizations	4.09±0.54	0.82	1.00	0.00	3.55±0.93	0.45	0.88	0.25	3.36±1.03	0.27	0.75	0.50
68. (EQ6) Self-reported achievement levels of institution's educational goals by graduating medical students	4.00±0.45	0.82	1.00	0.00	4.18±0.40	1.00	1.00	0.00	3.27±0.47	-0.45	0.83	0.25
57. (EQ6) Graduation competency achievement levels self-reported by graduating medical students	4.27±0.65	0.82	0.75	0.50	4.18±0.40	1.00	1.00	0.00	3.36±0.67	-0.09	0.67	0.50
69. (EQ6) Diversity of career pathways in graduates	4.09±0.70	0.64	0.88	0.25	3.64±0.67	0.45	0.88	0.25	4.00±0.63	0.64	1.00	0.00
70. (EQ6) Major domestic and international academic awards won by graduates	4.09±0.70	0.64	0.88	0.25	4.00±0.00	1.00	1.00	0.00	3.55±0.52	0.09	0.75	0.50
71. (EQ6) Collaboration research status with other institutions and fields	3.91±0.54	0.64	1.00	0.00	3.27±0.65	-0.27	0.67	0.50	3.36±0.50	-0.27	0.67	0.50
72. (EQ6) Status of global development projects on health care and medical education conducted by the institution	3.91±0.54	0.64	1.00	0.00	3.09±0.54	-0.64	1.00	0.00	3.18±0.60	-0.45	0.83	0.25
73. (EQ7) Use of the results of the analysis of the educational environment, resources, and admission policies for program improvement	4.18±0.40	1.00	1.00	0.00	3.82±0.40	0.64	1.00	0.00	3.91±0.54	0.64	1.00	0.00
74. (EQ7) Use of the results of the analysis of curricula for educational quality assurance and continuous quality improvement	4.55±0.52	1.00	0.80	0.50	3.82±0.40	0.64	1.00	0.00	3.91±0.54	0.64	1.00	0.00
75. (EQ7) Use of the results of the analysis of student performance enhancement for program improvement	4.27±0.47	1.00	0.88	0.25	3.55±0.52	0.09	0.75	0.50	3.64±0.50	0.27	0.75	0.50
76. (EQ7) Use of teacher and student feedback for program improvement	4.18±0.40	1.00	1.00	0.00	3.82±0.60	0.45	0.88	0.25	3.73±0.47	0.45	0.88	0.25
77. (EQ7) Notification of students and teachers about program improvement based on their feedback	3.64±0.81	0.27	0.75	0.50	3.27±0.47	-0.45	0.83	0.25	3.09±0.54	-0.64	1.00	0.00
78. (EQ7) Systematic seeking of quality feedback from students and teachers	4.09±0.70	0.64	0.88	0.25	3.73±1.19	0.45	0.63	0.75	3.73±0.65	0.64	1.00	0.00
Domain 5: Implementation and outcome of the curriculum												
79. (EQ1) Percentage of courses with horizontal integration of associated sciences and disciplines	3.67±0.49	0.33	0.75	0.50	4.08±0.67	0.67	0.94	0.13	3.90±0.99	0.40	0.63	0.75
80. (EQ1) Provision of structured and systematic opportunities for reflection for self-directed learning	4.17±0.72	0.67	0.75	0.50	4.00±0.43	0.83	1.00	0.00	4.10±0.99	0.60	0.75	0.50

(Continued on next page)

Table 3. Continued

Evaluation indicators	Importance				Feasibility				Necessity for change			
	Mean±SD	CVR	CSS	CVG	Mean±SD	CVR	CSS	CVG	Mean±SD	CVR	CSS	CVG
81. (EQ1) Provision of sufficient learning opportunities for achieving each graduation competency	4.42±0.67	0.83	0.78	0.50	3.92±0.67	0.50	0.94	0.13	4.20±0.63	0.80	0.81	0.38
82. (EQ1) Systematic organization of curriculum for student research and scholarly activities	4.25±0.75	0.67	0.75	0.50	4.25±0.75	0.67	0.75	0.50	4.00±0.82	0.40	0.63	0.75
83. (EQ1) Percentage of courses with horizontal integration of associated sciences and disciplines	3.92±1.00	0.33	0.50	1.00	4.00±0.85	0.67	0.94	0.13	3.50±0.85	0.00	0.71	0.50
84. (EQ1) Percentage of courses with vertical integration of basic, clinical, and medical humanities disciplines	3.58±0.79	0.17	0.75	0.50	3.42±0.79	0.17	0.75	0.50	3.70±0.82	0.00	0.71	0.50
85. (EQ2) Percentage of courses applying student-centered instructional strategies	3.75±0.87	0.33	0.75	0.50	4.17±1.03	0.50	0.72	0.63	4.00±0.94	0.20	0.50	1.00
86. (EQ2) Constructive alignment between educational outcomes – learning contents – teaching and learning methods – assessment methods	4.33±0.78	0.67	0.78	0.50	3.50±0.67	0.17	0.75	0.50	3.80±0.79	0.20	0.75	0.50
87. (EQ2) Application of assessment tools focusing on graduation competency	3.83±0.58	0.50	0.94	0.13	3.58±0.51	0.17	0.75	0.50	3.70±0.82	0.00	0.71	0.50
88. (EQ2) Monitoring of research mentoring program satisfaction and need for improvement from students	3.92±0.67	0.50	0.94	0.13	3.92±0.67	0.50	0.94	0.13	3.20±0.63	−0.40	0.75	0.38
89. (EQ2) Gathering feedback from research mentoring supervisors on program satisfaction and the need for improvement	3.83±0.58	0.50	0.94	0.13	3.83±0.58	0.50	0.94	0.13	3.10±0.57	−0.60	1.00	0.00
90. (EQ2) Percentage of examination items requiring application of integrated knowledge	3.75±0.87	0.33	0.75	0.50	3.67±0.65	0.17	0.75	0.50	3.80±0.79	0.20	0.75	0.50
91. (EQ3) Number of students participating in various self-directed activities such as entrepreneurship, volunteering, and extracurricular activities	3.58±0.67	0.00	0.71	0.50	3.83±0.72	0.33	0.75	0.50	3.20±0.79	−0.20	0.67	0.50
92. (EQ3) Graduation outcome achievement levels self-reported by graduating medical students	4.33±0.78	0.67	0.78	0.50	3.67±0.49	0.33	0.75	0.50	3.70±0.67	0.20	0.75	0.50
93. (EQ3) Comparison of performance of cohorts of CDP2004 and cohorts of CDP2013	4.50±0.67	0.83	0.80	0.50	4.50±0.80	0.67	0.80	0.50	3.30±0.67	−0.20	0.67	0.50
94. (EQ3) Percentage of students who participated in research at on-campus, domestic, or international institutions	3.75±0.45	0.50	0.94	0.13	4.17±0.72	0.67	0.75	0.50	2.90±0.57	−0.80	1.00	0.00
95. (EQ3) Percentage of graduates who chose careers in basic science	3.42±1.00	−0.17	0.67	0.50	4.00±1.04	0.33	0.50	1.00	3.30±0.67	−0.20	0.67	0.50
96. (EQ3) Evaluation results of Graduation Portfolios	3.67±0.78	0.33	0.75	0.50	3.42±0.67	−0.33	0.67	0.50	3.30±0.67	−0.20	0.67	0.50
97. (EQ3) Student research outcomes in terms of quantitative and qualitative aspects	-	-	-	-	-	-	-	-	-	-	-	-
98. (EQ3) Number of applicants and competition rate for the “Seoksan-Yonsei Physician-Scientist Training Program”	-	-	-	-	-	-	-	-	-	-	-	-

SD, standard deviation; CVR, content validity ratio; CSS, degrees of consensus; CVG, degrees of convergence, EQ, evaluation question.

주로 선정되었다. 다만, '45. 교육 관련 교수개발 프로그램에 3시간 참여한 전임교원의 비율' 평가지표의 경우 내용타당도 기준을 만족하지는 못하였으나, 교육의 질 향상 및 평가인증기준의 강화를 위해 반드시 필요한 항목이라는 합의를 거쳐 최종 평가지표에 포함되었다. 또한 '77. 피드백 분석결과에 따른 개선사항을 학생 및 교육자에게 고지한 횟수' 평가지표도 내용타당도 기준을 만족하지는 못하였으나, 피드백을 바탕으로 개선사항을 고지하는 과정은 교육프로그램 개선절차의 투명성과 신뢰성을 높이는 데 중요한 역할을 하고, '76. 교육과정 및 교육의 과정 모니터링을 통해 수집된 교육자와 학생의 피드백을 반영한 개선율' 평가지표와 함께 사용할 때 유용한 보완적 정보를 제공할 수 있는 평가지표로 협의되어 최종적으로 선정되었다. '교육과정의 운영과 성과' 영역에서는 총 20개의 평가지표가 선정되었다. 이 중 '95. 기초의학 전공자의 비율'은 내용타당도를 충족하지 못하였으나, 기초의학 교육의 강화를 위한 특정 비율 이상의 전공자 유지 및 증대 필요성으로 본 의과대학의 교육전략 및 목표와 밀접하게 연관되어 최종적으로 선정하게 되었다. '97. 학생연구 성과의 양적·질적 평가'와 '98. 석산·연세 의과대학 육성사업의 지원자 수 및 경쟁률'은 협의회를 거쳐 새롭게 신설된 평가지표로 학생들의 연구역량을 추적하고 강화하기 위한 필요성을 반영하였으며, 의과대학 육성을 위한 프로그램의 운영성과와 경쟁력을 관리하기 위한 중요 평가지표로 추가되었다.

고찰

본 연구는 의과대학 교육 전반에 걸쳐 교육 여건을 개선하고 교육의 질을 체계적으로 관리할 수 있는 평가체계의 개발을 목표로 하였다. 이를 위해 교육프로그램의 다양한 측면을 포괄할 수 있는 평가영역을 설정하고, 평가모형에 적합한 평가지표를 개발하였으며, 델파이 조사방법을 활용하여 평가지표의 타당성을 검토하고 교육전문가 및 다양한 이해관계자의 의견을 반영하였다. 그 결과, 5개 영역에 걸쳐 총 98개의 평가지표를 도출하였으며, 이는 의과대학의 교육프로그램 전반을 종합적으로 평가할 수 있는 도구로서 중요한 의미를 가진다.

평가영역별 최종 선정된 평가지표의 특징은 다음과 같다. 첫째, '기본의학교육과정'과 '임상실습교육과정' 영역에서는 의과대학에서 설정한 교육목표에 도달하기 위한 교육과정 내 전략수립 및 운영을 평가하는 지표로 구성되었다. 이 영역에는 매 학년도 시행되는 교과목 단위에서 교육과정의 계획 및 실행에 대한 자료를 수집하여 교육과정 이행의 적절성을 확인하고, 자원 및 지원의 충분성, 평가기준의 객관성, 피드백 제공 외 행정적 지원과 관련한 측면을 측정할 수 있는 13개의 지표가 포함되었다. 둘

째, '교육환경, 자원 및 운영체계' 영역에서는 의료환경 변화, 성과체계 및 인지 정도, 자원 확보와 활용, 교수-학습 환경의 적절성, 평가 및 개선활동의 성과를 분석할 수 있는 18개의 지표가 선정되었다. 셋째, '학생 역량 및 교육프로그램 성과' 영역은 재학생과 졸업생의 역량, 교육프로그램의 효과, 교수 개발, 평가 및 피드백 체계, 피드백 활용 및 개선 내역과 관련하여 총 34개의 평가지표가 선정되었다. 넷째, '교육과정의 운영과 성과' 영역에서는 교육프로그램의 설계와 구성, 교수-학습 방법 및 평가, 자기주도적 활동 및 성과와 관련된 20개의 평가지표가 선정되었으며, 이는 교육과정의 전반적인 실행을 효과적으로 분석하고 개선할 수 있는 근거를 제공하는 데 중점을 두었다.

본 연구의 의의는 기관의 고유한 맥락과 교육과정의 특징을 반영하여, 교육의 각 단계에서 발생하는 복합적인 상호작용을 종합적으로 평가할 수 있는 평가체계를 제공했다는 데 있다. 이를 통해 교육프로그램의 전반적인 질을 평가하고 지속적인 개선을 위한 기초자료를 제공함으로써, 교육과정의 발전을 위해 각 단계에서 발생할 수 있는 다양한 변수를 체계적으로 평가할 수 있을 것이다. 해외 주요 의과대학들도 기관의 교육철학과 교육목표에 맞추어 구조화된 평가체계를 독자적으로 구축하여 교육프로그램의 질을 관리하고 있으며, 본 연구에서 개발된 평가지표도 이와 같은 대학의 사례와 유사점을 가진다. 예를 들어, 미국의 하버드 의과대학(Harvard Medical School)은 자체 교육프로그램의 질 관리를 위해 평가체계를 통해 지속적인 개선을 도모하고 있으며, 정기적으로 교육프로그램의 성과를 분석하고 개선방향을 제시하고 있다. 하버드 의과대학에서 적용하는 평가체계에는 교육과정뿐만 아니라 교수진의 연구 및 교육역량에 대한 평가, 학생들의 교육성과의 추적 등을 포괄하는 다양한 평가항목을 포함하고 있다[38]. 또한 캐나다의 토론토 의과대학(The University of Toronto Temerty Faculty of Medicine)은 교육의 질적 개선을 위해 정기적인 평가와 교수진 및 학생들의 피드백을 중요하게 고려하고, 이를 바탕으로 교육과정의 변화를 추진하고 있다[39]. 국내에서도 일부 의과대학은 자체적인 평가체계를 도입하여 교육의 질적 개선을 추구하고 있으며[15,20,21,40], 이를 통해 교육프로그램의 지속적인 질 향상을 추구하고 있다.

특히 본 연구에서 제안된 평가체계는 교육프로그램의 전반적인 과정을 아우를 수 있도록 구조화되어 교육프로그램의 성과뿐만 아니라 과정 및 실행단계에서 발생하는 교육의 요소들을 포괄적으로 평가할 수 있다는 특징이 있다. 또한 기존의 단일 평가모형이 가지는 한계점을 보완하여 평가영역 및 평가질문에 적합한 통합모형을 제시함으로써 교육의 질이나 성과에 이르는 기전을 체계적으로 평가할 수 있게 하였다. 더 나아가 이러한 평가가 일회성에 그치지 않고, 주기적인 평가를 통해 교육의 개선을 유도

하는 데 적합한 지표에 비중을 두고 설계되어 지속적인 질 향상 활동을 지원할 수 있는 기반을 마련할 수 있도록 하였다.

본 연구의 한계점은 다음과 같다. 첫째, 연구에서 개발한 평가 지표는 델파이 조사를 통해 이루어졌으며, 이는 제한된 전문가의 의견에 기반하고 있다. 델파이 조사에서 패널의 전문성은 연구의 신뢰성과 타당성에 중요한 영향을 미치며, 연구결과는 조사참여자의 주관적 견해에 영향을 받을 수 있다[41]. 또한 일부 평가지표는 협의회를 개최하여 전문가들의 합의를 거쳐 추가되거나 수정되었으나 제한된 시간으로 인해 각 평가지표에 대한 심층적이고 다면적인 검토가 충분히 이루어지지 못했을 가능성이 있다. 따라서 실제 교육프로그램에서 평가지표를 적용한 결과를 바탕으로 평가의 신뢰도 및 타당도를 검토하고, 평가도구의 실효성에 대한 추가적인 피드백을 수집하는 등 향후 모니터링을 통한 평가체계의 개선이 필요할 것이다. 둘째, 본 연구에서는 양적 자료를 중심으로 교육성과를 측정하는 평가지표가 다수 포함되었으므로, 교육의 질적 측면을 충분히 반영하지 못할 가능성이 있다. 특히 학생들의 학습경험이나 교수-학생 간 상호작용과 같은 교육의 질적 요소를 평가하는 데 한계가 있을 수 있다. 따라서 후속연구를 통해 교육의 질적인 측면을 강화할 수 있는 평가지표로 범위를 확장할 필요가 있다. 셋째, 본 연구에서 개발된 평가지표는 특정 의과대학의 교육환경과 목표를 반영한 것이므로, 일반화에 한계가 있을 수 있다. 각 의과대학은 교육철학과 운영방식, 자원이 다를 수 있기 때문에 본 연구에서 도출된 평가지표의 적용 가능성은 대학의 맥락과 교육에 따라 달라질 수 있다.

적절한 교육프로그램 평가체계의 수립은 교육과정의 개발과 실행에 있어 필수적인 요소이다. 본 연구에서 개발된 평가체계는 교육프로그램 전반을 평가하고 개선하기 위한 기초자료를 제공할 뿐만 아니라, 주기적인 평가를 통해 의과대학 교육프로그램의 성취 정도를 파악하고, 개선이 필요한 영역을 진단하는 데 중요한 역할을 할 것으로 기대된다. 평가체계의 적용은 단순히 현재 상태를 진단하는 데 그치는 것이 아니라, 교육의 질 향상을 위한 전략적 기초를 마련하는 데 기여할 수 있을 것이다. 향후 연구에서는 개발된 평가체계를 실제 교육현장에 적용하여 평가결과의 타당도와 신뢰도를 확인하고 그 실효성을 검증하는 것이 필요하다. 한편, 변화하는 교육환경과 의료시스템에 대응하여 평가지표가 지속적으로 유용하게 사용될 수 있도록 주기적인 재평가와 보완이 필요하다. 이를 통해 교육프로그램의 성과와 한계점을 체계적으로 분석하고 교육과정의 설계와 실행에 필요한 구체적인 방향을 제시할 수 있을 것이다.

ORCID

Yulim Kang	https://orcid.org/0009-0007-3180-9565
Hae Won Kim	https://orcid.org/0000-0002-9321-8361
Jun Yong Choi	https://orcid.org/0000-0002-2775-3315

Conflict of interest

이 연구에 영향을 미칠 수 있는 기관이나 이해당사자로부터 재정적, 인적 자원을 포함한 일체의 지원을 받은 바 없으며, 연구윤리와 관련된 제반 이해상충이 없음을 선언한다.

Authors' contribution

연구설계: 강유림, 김혜원, 최준용; 자료분석: 강유림; 원고작성 및 수정: 강유림, 김혜원; 원고검토 및 수정: 최준용

References

- Boelen C, Pearson D, Kaufman A, Rourke J, Woollard R, Marsh DC, et al. Producing a socially accountable medical school: AMEE guide no. 109. *Med Teach*. 2016;38(11):1078-91. <https://doi.org/10.1080/0142159X.2016.1219029>
- Vassar M, Wheeler DL, Davison M, Franklin J. Program evaluation in medical education: an overview of the utilization-focused approach. *J Educ Eval Health Prof*. 2010;7:1. <https://doi.org/10.3352/jeehp.2010.7.1>
- Bendermacher GW, De Grave WS, Wolfhagen IH, Dolmans DH, Oude Egbrink MG. Shaping a culture for continuous quality improvement in undergraduate medical education. *Acad Med*. 2020;95(12):1913-20. <https://doi.org/10.1097/ACM.0000000000003406>
- Frye AW, Hemmer PA. Program evaluation models and related theories: AMEE guide no. 67. *Med Teach*. 2012;34(5):e288-99. <https://doi.org/10.3109/0142159X.2012.668637>
- Musick DW. A conceptual model for program evaluation in graduate medical education. *Acad Med*. 2006;81(8):759-65. <https://doi.org/10.1097/00001888-200608000-00015>
- Ruhe V, Boudreau JD. The 2011 Program Evaluation Standards: a framework for quality in medical education programme evaluations. *J Eval Clin Pract*. 2013;19(5):925-32. <https://doi.org/10.1111/j.1365-2753.2012.01879.x>
- Allen LM, Hay M, Palermo C. Evaluation in health professions edu-

- cation-Is measuring outcomes enough? *Med Educ.* 2022;56(1):127-36. <https://doi.org/10.1111/medu.14654>
8. Ragsdale JW, Berry A, Gibson JW, Herber-Valdez CR, Germain LJ, Engle DL, et al. Evaluating the effectiveness of undergraduate clinical education programs. *Med Educ Online.* 2020;25(1):1757883. <https://doi.org/10.1080/10872981.2020.1757883>
9. Liaison Committee on Medical Education. Functions and structure of a medical school: standards for accreditation of medical education programs leading to the MD degree. Washington (DC): Liaison Committee on Medical Education; 2024.
10. Korean Institute of Medical Education and Evaluation. Accreditation Standards of Korean Institute of Medical Education and Evaluation 2019 [Internet]. Seoul: Korean Institute of Medical Education and Evaluation; 2019 [cited 2024 Mar 20]. Available from: <https://kimee.or.kr/medical-education/criteria/>
11. Haji F, Morin MP, Parker K. Rethinking programme evaluation in health professions education: beyond 'did it work?'. *Med Educ.* 2013; 47(4):342-51. <https://doi.org/10.1111/medu.12091>
12. Shrivastava SR, Shrivastava PS. Program evaluation in the field of medical education: need and approaches. *Curr Med Issues.* 2023; 21(3):181-2. https://doi.org/10.4103/cmi.cmi_31_23
13. Moreau KA. Has the new Kirkpatrick generation built a better hammer for our evaluation toolbox? *Med Teach.* 2017;39(9):999-1001. <https://doi.org/10.1080/0142159X.2017.1337874>
14. Colliver JA. Effectiveness of problem-based learning curricula: research and theory. *Acad Med.* 2000;75(3):259-66. <https://doi.org/10.1097/00001888-200003000-00017>
15. Lee SY, Lee SH, Shin JS. Development of evaluation criteria for competency-based medical education curriculum in medical schools. *J Learn Cent Curr Instr.* 2023;23(15):727-55. <https://doi.org/10.22251/jlcci.2023.23.15.727>
16. Ju H, Oh M, Lee JT, Yoon BY. Adapting an integrated program evaluation for promoting competency-based medical education. *Korean Med Educ Rev.* 2021;23(1):56-67. <https://doi.org/10.17496/kmer.2021.23.1.56>
17. Lee Ah, Park HJ, Kim SG, Kim JY, Kang YN, Lee SY, et al. Implementing medical education continuous quality improvement using design-based research. *Korean Med Educ Rev.* 2020;22(3):189-97. <https://doi.org/10.17496/kmer.2020.22.3.173>
18. Yune SJ, Lee SY, Im S. Educational program evaluation system in a medical school. *Korean Med Educ Rev.* 2020;22(2):131-42. <https://doi.org/10.17496/kmer.2020.22.2.131>
19. Van Melle E, Frank JR, Holmboe ES, Dagnone D, Stockley D, Sherbino J, et al. A core components framework for evaluating implementation of competency-based medical education programs. *Acad Med.* 2019;94(7):1002-9. <https://doi.org/10.1097/ACM.0000000000002743>
20. Cha B, Shon M. Validation of the curriculum in medical education using CIPP evaluation model. *J Educ Cult.* 2020;26(5):105-30. <https://doi.org/10.24159/joec.2020.26.5.105>
21. Lee A. A study regarding improvement of the CQI program in medical school based on the CIPP evaluation model. *Korean J Teach Educ.* 2022;38(6):131-55. <https://doi.org/10.14333/KJTE.2022.38.6.07>
22. Yoo HH, Kim YJ. The implementation and evaluation of learning experience-based professionalism program in medical school. *J Korea Contents Assoc.* 2018;18(1):164-72. <https://doi.org/10.5392/JKCA.2018.18.01.164>
23. Noble RA, McKinnell JC, Shaw S, Bassett S, Woods L, Asrar M, et al. Evaluating a process of academic detailing in primary care: an educational programme for acute kidney injury. *BMC Med Educ.* 2019; 19(1):253. <https://doi.org/10.1186/s12909-019-1659-y>
24. Yazdimoghaddam H, Samadipour E, Ghardashi F, Borzoei F, Akbarzadeh R, Zardosht R, et al. Designing a comprehensive clinical competency test for operating room technology student: using Delphi technique and CIPP model evaluation. *J Educ Health Promot.* 2021;10:240. https://doi.org/10.4103/jehp.jehp_1563_20
25. Gibson KA, Boyle P, Black DA, Cunningham M, Grimm MC, McNeil HP. Enhancing evaluation in an undergraduate medical education program. *Acad Med.* 2008;83(8):787-93. <https://doi.org/10.1097/ACM.0b013e31817eb8ab>
26. Rooholamini A, Amini M, Bazrafkan L, Dehghani MR, Esmaeilzadeh Z, Nabeiei P, et al. Program evaluation of an integrated basic science medical curriculum in Shiraz Medical School, using CIPP evaluation model. *J Adv Med Educ Prof.* 2017;5(3):148-54.
27. Lee SY, Lee SH, Shin JS. Evaluation of medical humanities course in college of medicine using the context, input, process, and product evaluation model. *J Korean Med Sci.* 2019;34(22):e163. <https://doi.org/10.3346/jkms.2019.34.e163>
28. Jorm C, Roberts C. Using complexity theory to guide medical school evaluations. *Acad Med.* 2018;93(3):399-405. <https://doi.org/10.1097/ACM.0000000000001828>
29. Poland B, Frohlich KL, Cargo M. Context as a fundamental dimension of health promotion program evaluation. In: Potvin L, McQueen DV, Hall M, de Salazar L, Anderson LM, Hartz ZM, editors.

- Health promotion evaluation practices in the Americas: values and research. New York (NY): Springer New York; 2009. p. 299-317.
30. Newcomer KE, Hatry HP, Wholey JS. Planning and designing useful evaluations. In: Newcomer KE, Hatry HP, Wholey JS, editors. Handbook of practical program evaluation. 4th ed. Hoboken (NJ): John Wiley & Sons Inc.; 2015. p. 7-35.
31. Condelli L. Quality indicators, measures and performance standards. Washington (DC): Pelavin Associates Inc.; 1992.
32. Murry Jr JW, Hammons JO. Delphi: a versatile methodology for conducting qualitative research. *Rev High Educ*. 1995;18(4):423-36. <https://doi.org/10.1353/rhe.1995.0008>
33. Avella JR. Delphi panels: research design, procedures, advantages, and challenges. *Int J Dr Stud*. 2016;11:305-21. <https://doi.org/10.28945/3561>
34. Von Der Gracht HA. Consensus measurement in Delphi studies: review and implications for future quality assurance. *Technol Forecast Soc Change*. 2012;79(8):1525-36. <https://doi.org/10.1016/j.techfore.2012.04.013>
35. Stillwell WG, Seaver DA, Edwards W. A comparison of weight approximation techniques in multiattribute utility decision making. *Organ Behav Hum Perform*. 1981;28(1):62-77. [https://doi.org/10.1016/0030-5073\(81\)90015-5](https://doi.org/10.1016/0030-5073(81)90015-5)
36. Liu D, Li T, Liang D. An integrated approach towards modeling ranked weights. *Comput Ind Eng*. 2020;147:106629. <https://doi.org/10.1016/j.cie.2020.106629>
37. Lawshe CH. A quantitative approach to content validity. *Pers Psychol [Internet]*. 1975 [cited 2024 Mar 20];28:563-75. <https://caepnet.org/~media/Files/caep/knowledge-center/lawshe-content-validitypdf>
38. Sullivan AM, Krupat E, Dienstag JL, McSparron JI, Pelletier SR, Fazio SB, et al. The Harvard Medical School Pathways curriculum: a comprehensive curricular evaluation. *Med Teach*. 2022;44(11):1268-76. <https://doi.org/10.1080/0142159X.2022.2081142>
39. Tait GR, Kulasegaram KM. Assessment for learning: the University of Toronto Temerty Faculty of Medicine MD program experience. *Educ Sci*. 2022;12(4):249. <https://doi.org/10.3390/educsci12040249>
40. Yune SJ, Lee SY, Im S. Educational program evaluation system in a medical school. *Korean Medical Educ Rev*. 2020;22(2):131-42. <https://doi.org/10.17496/kmer.2020.22.2.131>
41. Stewart D, Gibson-Smith K, MacLure K, Mair A, Alonso A, Codina C, et al. A modified Delphi study to determine the level of consensus across the European Union on the structures, processes and desired outcomes of the management of polypharmacy in older people. *PLoS One*. 2017;12(11):e0188348. <https://doi.org/10.1371/journal.pone.0188348>

APPENDICES

Appendix 1. 문헌검토에 따른 프로그램 평가모형 및 평가요소별 평가지표

프로그램 평가모형	평가지표		참고문헌
CIPP	Context	구체적인 교육목표 설정, 역량 규정, 인증기준 파악, 요구조사 실시 여부, 교육목표 및 성과체계, 졸업역량에 따른 교육과정의 적절성, 과정성과 설계 및 수행, 학습자 요구 파악, 선수지식 및 입학생 특성, 학습동기 검사 및 상담, 교수자 다양성, 교수자의 요구 반영, 교수자 교육업적 및 규정, 학생평가의 종류와 주기, 평가결과의 통보 및 공유, 자료수집 체계, 평가체계 및 환류체계, 교육자원 요구도 파악, 시설 및 장비, 학습환경 및 문화(교외활동, 동아리, 상담 포함), 재정 및 규정의 충실성, 조직의 체계성, 구성원 합의, 이해관계자 요구 분석, 사명 진술과 사회적 책무성, 지역사회 파트너십 및 의료환경 변화 반영, 목표 달성의 장애요소, 프로그램 개선체계	A1-A13
	Input	교육시간, 실습항목, 강의실, 교육시설 및 설비, 자발적인 참여도, 프로그램 수행 난이도, 참여인원, 교육내용 및 교수학습법 적합성, 학생평가방법, 평가주기, 교수자의 전문성, 교수능력 개발 지원, 교육관련 지원 인력, 재정 운영 효율성, 학습환경, 교육지원 프로그램 및 교수강화 프로그램, 임상실습환경, 교육기본시설, 학생복지시설, 교육과정 설계원칙, 강의계획서의 면밀성, 학습량, 교육연계의 적절성, 과정·과목의 균형 조정, 교육편성, 실제 환자 진료 참여 기회 제공, 발달평가, 포트폴리오 평가, 학습동기 검사·상담, 다양성 및 형평성	A1-A13
	Process	교육과정 운영, 교육운영자원, 학생평가, 강의방법, 수업 운영방식, 위원회 활동, 교육 운영지침, 전공 간 교류, 실습기관과의 협력, 피드백의 적절성, 평가 적절성, 목표 제시, 구성체계, 사전 안내, 일정, 운영계획에 따른 진행, 교육과정 준수율, 교육성과 모니터링 실적, 교육계획 이행의 충실성, 교육진행의 원활성, 재학습 및 재평가 실행, 교육과정 구성 및 조화, 수직통합, 선택과목과 필수과목 균형, 수업 시수 및 임상실습 교육기간 적절성, 과학적 방법 및 의학연구 교육, 교수학습방법의 다양성, 성과에 맞는 평가방법, 형성평가 및 시험 유형, 수업 운영계획 이행, 실습교육과정 다양성, 일차진료 실습, 환자안전 강조, 임상실습 일정 및 방법, 실습 피드백, 임상실습 성적 공개	A1-A13
	Product	학생 성장 역량, 전문직업성, 진료역량, 사회공헌역량, 연구역량, 소통, 협력역량, 교수역량, 강의계획 능력, 가르치는 역량, 학생평가 역량, 평가의 정기적 실시, 평가항목 구성, 교수자 참여, 학생 참여, 평가결과 차기 교육운영 반영, 평가결과 공지, 의학교육 역량 강화, 리더 활동 도움, 연구활동 도움, 프로그램 추천 의사, 평가 모니터링 체계의 적절성, 교육과정 평가방법의 적절성, 교육과정 성공요인의 지속 가능성, 교육과정 평가결과 피드백, 과정성과 및 만족도, 교육만족도, 학습성과 달성도, 진급, 낙제, 중도탈락 비율, 의사국가시험 합격률, 다양한 진로 진출, 사회적 책무 관련 교육, 졸업성과별 수행능력 향상, 교육과정 및 투입자원에 따른 수행능력 차이, 졸업 후 역량과 의과대학 성적 상관	A1-A8, A10-A13
Logic	Input	재정적 지원, 교수인력 및 교육보호시간, 시설, 인력 및 자원, 이해관계자 참여, 사회적 요구	A14-A19
	Activity	요구분석, 프로그램 개발, 워크숍, 교육과정 개발, 인증기준, 인증절차, 교육과정 검토, 교수개발 과정	A14-A19
	Output	참가자 수, 참석률, 긍정적인 경험 비율, 교육 모듈 및 평가도구 생성, 세션 제공횟수, 지식 및 태도 변화, 행동 변화	A16-A19
	Outcome	자기효능감, 자기존중감, 프로그램 강점 및 개선 기회, 지식과 기술의 실제 적용, 환자 결과 향상, 소통 및 협력(그룹 응집력, 공감적 소통, 커뮤니케이션 실패 감소), 지속적인 질 개선 문화발전, 학습경험 개선, 문제 학생 식별, 의료오류 감소 및 환자 치료결과 향상	A14-A15, A17-A19
	Impact	교육적 영향(행동 변화, 수업 및 실천 공유), 의료시스템 영향(치료과정 변화, 환자 결과 변화), 프로그램이 이루어지는 환경의 변화(구조적, 정치적, 문화적 변화), 의도된 또는 의도되지 않은 성과의 변화	A16, A19
Kirkpatrick	Reaction	학습과정 및 교육내용에 대한 만족도, 교수법 만족도, 교육 만족요인 파악의 충분성, 졸업생 설문지, 학생평가, 임상실습 평가, 교수 평가, 동료 평가	A2, A9, A20-A25

(Continued on next page)

Appendix 1. Continued

프로그램 평가모형		평가지표	참고문헌
	Learning	지식/술기/태도 역량 달성의 충분성, 경험, 성찰, 봉사활동 내용 및 이수 시간, 동료평가 점수, 국가/지역 평가(USMLE, OSCE, 기초 및 임상의학지식 시험 등), 교육 이후 지식/술기/태도 역량의 변화	A2, A9, A20-A25
	Behavior	졸업역량의 수행, 학습전이(기본의학교육의 임상의학교육에의 적용, 이전학습 내용의 임상실습 적용, 졸업 후 의료현장에서의 적용, 지역사회로 적용), 학습자의 행동 변화, 졸업 이후 전문분야 선택	A2, A9, A21-A24
	Result	독립적인 졸업역량 수행, 졸업생의 수련과정에서의 변화, 졸업 후 진로 및 사회 기여	A2, A9, A21
기타			
Process evaluation	Reach	그룹(학생, 교수자, 부모)의 목표 도달 여부	A26
	Does/delivery Context	수업과 활동의 전달 및 학생 참여의 적합성 맥락적 요인 및 개입이론과의 일관성	
Complexity theory	Diversity	다양한 배경을 가진 교수진, 다양한 임상경험 노출	A27
	Nesting	의료시스템과 의과대학의 연계, 정부의 역할	
Self-report assessment	Self-organizing	학생들의 능동적 학습, 임상현장에서 새로운 학습기회 탐색	
		과정 및 수업의 유용성, 세션 목표 달성 여부, 자료의 적합성, 안전한 학습환경, 상호작용 적 교수 전략, 프로젝트 시작 및 완료의 장애물, 평가목표의 인식 정도	A28-A30
Pre- and post-survey		지식, 태도, 행동의 변화, 프로그램의 유용성과 강점, 개선점, 프로그램에 대한 인식 및 자기보고, 프로그램 목적과 내용, 강의계획서, 지식 습득, 강의내용과 활동의 난이도, 교육방법과 매체 활용, 시간, 강의, 자료, 시청각 보조물, 토의방법, 교수자 상호작용, 교수법, 학습평가와 환경, 과제물, 평가방식, 학습환경의 적절성, 종합평가, 전반적인 만족도	A31-A34
Mixed method		연례 프로그램 평가(APE) 템플릿, SMART 목표 설정, 대시보드 프로그램 및 기관 지표, 각 항목에 대한 PDSA 사이클, 설문조사, 인터뷰, 문서 검토	A15, A35

USMLE, United States Medical Licensing Examination; OSCE, objective structured clinical examination; APE, annual program evaluation; SMART, specific, measurable, accountable, realistic, timely; PDSA, plan-do-study-act.

문헌검토에 포함된 참고문헌 목록

- A1. Yazdimoghaddam H, Samadipour E, Ghardashi F, Borzoei F, Akbarzadeh R, Zardosht R, et al. Designing a comprehensive clinical competency test for operating room technology student: using Delphi technique and CIPP model evaluation. *J Educ Health Promot.* 2021;10:240. https://doi.org/10.4103/jehp.jehp_1563_20
- A2. Lee SY. Developing evaluation criteria for competency-based curriculum in medical colleges and exploring ways to apply them [dissertation]. Seoul: Seoul National University; 2021. 423 p. <https://hdl.handle.net/10371/175993>
- A3. Barber C, van der Vleuten C, Leppink J, Chahine S. Social accountability frameworks and their implications for medical education and program evaluation: a narrative review. *Acad Med.* 2020;95(12):1945-54. <https://doi.org/10.1097/ACM.00000000000003731>
- A4. Schwartz AR, Siegel MD, Lee AI. A novel approach to the program evaluation committee. *BMC Med Educ.* 2019;19(1):465. <https://doi.org/10.1186/s12909-019-1899-x>
- A5. Ashghali-Farahani M, Ghaffari F, Hoseini-Esfidarjani SS, Hadian Z, Qomi R, Dargahi H. Neonatal intensive care nursing curriculum challenges based on Context, Input, Process, and Product evaluation model: a qualitative study. *Iran J Nurs Midwifery Res.* 2018;23(2):111-8. https://doi.org/10.4103/ijnmr.IJNMR_3_17
- A6. Burke E, Hennessy M. Evaluation of an early career clinical academic training programme using the CIPP model. *BMJ Open.* 2021;11(11):e052965. <https://doi.org/10.1136/bmjopen-2021-052965>
- A7. Cha B, Shon M. A study on the development of evaluation indicators for the education process in medical school. *J Educ Cult.* 2019;25(4):239-61. <https://doi.org/10.24159/joec.2019.25.4.239>

- A8. Lee A. A study regarding improvement of the CQI program in medical school based on the CIPP evaluation model. *Korean J Teach Educ.* 2022;38(6):131-55. <https://doi.org/10.14333/KJTE.2022.38.6.07>
- A9. Ju H, Oh M, Lee JT, Yoon BY. Adapting an integrated program evaluation for promoting competency-based medical education. *Korean Med Educ Rev.* 2021;23(1):56-67. <https://doi.org/10.17496/kmer.2021.23.1.56>
- A10. Cha B, Shon M. Validation of the curriculum in medical education using CIPP evaluation model. *J Educ Cult.* 2020;26(5):105-30. <https://doi.org/10.24159/joc.2020.26.5.105>
- A11. Lee SY, Shin JS, Lee SH. How to execute Context, Input, Process, and Product evaluation model in medical health education. *J Educ Eval Health Prof.* 2019;16:40. <https://doi.org/10.3352/jeehp.2019.16.40>
- A12. Cha B. Development of evaluation indicators for the curriculum in medical education using the CIPP evaluation [dissertation]. Incheon: Inha University; 2021. <http://inha.dcollection.net/common/orgView/200000368781>
- A13. Yune SJ, Lee SY, Im S. Educational program evaluation system in a medical school. *Korean Med Educ Rev.* 2020;22(2):131-42. <https://doi.org/10.17496/kmer.2020.22.2.131>
- A14. Blouin D, Tekian A. Accreditation of medical education programs: moving from student outcomes to continuous quality improvement measures. *Acad Med.* 2018;93(3):377-83. <https://doi.org/10.1097/ACM.0000000000001835>
- A15. Hamza DM, Ross S, Oandasan I. Process and outcome evaluation of a CBME intervention guided by program theory. *J Eval Clin Pract.* 2020;26(4):1096-104. <https://doi.org/10.1111/jep.13344>
- A16. Balayah Z, Khadjesari Z, Keohane A, To W, Green JS, Sevdalis N. National implementation of a pragmatic quality improvement skills curriculum for urology residents in the UK: application and results of 'theory-of-change' methodology. *Am J Surg.* 2021; 221(2):401-9. <https://doi.org/10.1016/j.amjsurg.2020.12.007>
- A17. Balmer DF, Rama JA, Simpson D. Program evaluation models: evaluating processes and outcomes in graduate medical education. *J Grad Med Educ.* 2019;11(1):99-100. <https://doi.org/10.4300/JGME-D-18-01084.1>
- A18. Preis H, Bojsza E, Lindenfeld L, Gan TJ, Pati S. Process evaluation of a medical improvisation program for healthcare communication training. *J Commun Healthc.* 2022;15(4):260-6. <https://doi.org/10.1080/17538068.2021.2012750>
- A19. Bavare AC, Thomas JK, Yeppeze D, Gazzaneo MC, Guffey D, Thammasitboon S. Evaluation of a theory-informed pediatric rapid response training program using the logic model. *Hosp Pediatr.* 2021;11(12):1395-403. <https://doi.org/10.1542/hpeds.2021-005895>
- A20. Azar AJ, Khamis AH, Naidoo N, Lindsbro M, Boukhaled JH, Gonuguntla S, et al. Design, implementation and evaluation of a distance learning framework to expedite medical education during COVID-19 pandemic: a proof-of-concept study. *J Med Educ Curric Dev.* 2021;8:23821205211000349. <https://doi.org/10.1177/23821205211000349>
- A21. Johnson SB, Fair MA, Howley LD, Prunuske J, Cashman SB, Carney JK, et al. Teaching public and population health in medical education: an evaluation framework. *Acad Med.* 2020;95(12):1853-63. <https://doi.org/10.1097/ACM.0000000000003737>
- A22. Dumenco L, Monteiro K, George P, Rougas S, Dollase R. Outcomes of a longitudinal quality improvement and patient safety preclerkship curriculum. *Acad Med.* 2019;94(12):1980-7. <https://doi.org/10.1097/ACM.0000000000002898>
- A23. Ragsdale JW, Berry A, Gibson JW, Herber-Valdez CR, Germain LJ, Engle DL, et al. Evaluating the effectiveness of undergraduate clinical education programs. *Med Educ Online.* 2020;25(1):1757883. <https://doi.org/10.1080/10872981.2020.1757883>
- A24. Pallari E, Khadjesari Z, Biyani CS, Jain S, Hodgson D, Green JS, et al. Pilot implementation and evaluation of a national quality improvement taught curriculum for urology residents: lessons from the United Kingdom. *Am J Surg.* 2020;219(2):269-77. <https://doi.org/10.1016/j.amjsurg.2019.11.011>
- A25. Kim SJ. Development and effectiveness evaluation of interprofessional communication education in healthcare [master's thesis]. Seoul: Seoul National University; 2019. <https://hdl.handle.net/10371/161564>
- A26. Bonde AH, Stjernqvist NW, Sabinsky MS, Maindal HT. Process evaluation of implementation fidelity in a Danish health-promoting school intervention. *BMC Public Health.* 2018;18(1):1407. <https://doi.org/10.1186/s12889-018-6289-5>
- A27. Jorm C, Roberts C. Using complexity theory to guide medical school evaluations. *Acad Med.* 2018;93(3):399-405. <https://doi.org/10.1097/ACM.0000000000001828>
- A28. Payne AS, Pavuluri P, Sten MB, Corriveau C, Berkowitz DA, Sarnacki R, et al. A multimodal quality improvement curriculum for

- pediatric GME program directors. *Am J Med Qual.* 2021;36(2):110-4. <https://doi.org/10.1177/1062860620929419>
- A29. Greenfield J, Traboulsi EI, Lombardo-Klefos K, Bierer SB. Best practices for building and supporting effective ACGME-Mandated Program Evaluation Committees. *MedEdPORTAL.* 2020;16:11039. https://doi.org/10.15766/mep_2374-8265.11039
- A30. Berger JS, Ho G, Paul C, Cioletti A, Marko K. A novel, web-based quality improvement platform to address ACGME CLER requirements. *J Grad Med Educ.* 2020;12(1):80-5. <https://doi.org/10.4300/JGME-D-19-00339.1>
- A31. Shah KP, Goyal S, Ramachandran V, Kohn JR, Go JA, Wiley Z, et al. Efficacy of quality improvement and patient safety workshops for students: a pilot study. *BMC Med Educ.* 2020;20(1):126. <https://doi.org/10.1186/s12909-020-1982-3>
- A32. Baxter K, Petz C, Middleton JL, Chan M. Re-thinking how we teach quality improvement: adding meaning to an ACGME requirement. *J Gen Intern Med.* 2019;34(8):1662-7. <https://doi.org/10.1007/s11606-019-05059-7>
- A33. Noble RA, McKinnell JC, Shaw S, Bassett S, Woods L, Asrar M, et al. Evaluating a process of academic detailing in primary care: an educational programme for acute kidney injury. *BMC Med Educ.* 2019;19(1):253. <https://doi.org/10.1186/s12909-019-1659-y>
- A34. Yoo HH, Kim YJ. The implementation and evaluation of learning experience-based professionalism program in medical school. *J Korea Contents Assoc.* 2018;18(1):164-72. <https://doi.org/10.5392/JKCA.2018.18.01.164>
- A35. Stansfield RB, Markova T, Baker R. Integration of continuous quality improvement methods into annual program and institutional evaluation. *J Grad Med Educ.* 2019;11(5):585-91. <https://doi.org/10.4300/JGME-D-19-00145.1>