

의과대학 강의평가의 형성적 활용 조건과 평가영역에 대한 교수 인식*

신혜경**

국문초록 본 연구는 의과대학 강의평가가 수업개선을 위한 형성적 환류 체계로 기능하기 위한 조건을 교수 인식의 관점에서 분석하는 데 목적이 있다. 이를 위해 전국 40개 의과대학 전임교수 356명의 설문자료를 바탕으로 강의평가의 필요성, 타당성, 신뢰성, 수업개선 기능에 대한 인식과 직급·교육경력별 차이, 적절한 평정자와 결과 공개 대상, 교수행동변화 관련 문항의 평가영역을 분석하였다. 분석 결과, 의과대학 교수들은 강의평가제도의 필요성과 수업개선 기능은 비교적 높게 인정하였으나 평가 방법의 타당성, 결과 신뢰성, 문항 적절성에 대해서는 낮은 동의 수준을 보였다. 직급과 교육경력에 따른 차이는 일부 문항에서만 나타났다. 평정자로는 학생이 가장 많이 선택되었으며, 결과 공개 대상으로는 교수 본인과 과목책임교수가 가장 많이 선택되었다. 교수행동변화 관련 24개 문항에 대한 탐색적 요인분석 결과, 최종 16개 문항이 설계, 분석, 평가, 실행, 개발의 다섯 평가영역으로 구성되었다. 이는 확정적 평가도구의 제시라기보다, 의과대학 강의평가를 교수자가 해석하고 실행할 수 있는 형성적 환류 자료로 재구성하기 위한 기초자료로 볼 수 있다.

주제어: 강의평가, 수업개선, 형성적 환류, 교수행동변화

목차

- I. 서론
- II. 이론적 배경
- III. 연구 방법
- IV. 연구 결과
- V. 결론

논문접수일: 2026.06.18.

논문수정일: 2026.07.16.

게재확정일: 2026.08.01.

I. 서론

대학에서 학생에 의한 강의평가(Student Evaluation of Teaching)는 교수자의 수업을 점검하고 교육의 질을 관리하기 위해 널리 활용되는 평가제도이다. 강의평가는 학생이 실제 수업을 경험한 당사자로서 교수·학습 과정에 대한 정보를 제공한다는 점에서 수업개선을 위한 중요한 자료가 될 수 있다(Marsh, 1987; Wachtel, 1998; Richardson, 2005). 그러나 최근 강의평가 연구에서는 학생평가 결과를 교수효과성이나 학습성과의 직접적 지표로 단순 해석하는 데 신중해야 한다는 논의가 지속되어 왔다(Spooren, Brockx, & Mortelmans, 2013; Uttl, White, & Gonzalez, 2017; Quansah, Ankoma-Sey, Dankyi, Nyarko, & Akoto, 2024). 특히 강의평가의 타당도와 신뢰도는 평가문항의 구성, 응답 맥락, 평가 시행 방식,

* 본 논문은 저자의 연세대학교 박사학위논문을 바탕으로 함.

** 제1저자 및 교신저자, 서울대학교 의과대학 객원부교수 / hkshin012@snu.ac.kr

결과 해석 및 활용 방식에 따라 달라질 수 있으므로, 강의평가가 수업개선에 기여하기 위해서는 평가도구 뿐 아니라 결과 해석과 환류 체계가 함께 검토되어야 한다(Constantinou & Wijnen-Meijer, 2022; Quansah et al., 2024).

의학교육에서도 강의평가는 교수자의 수업을 점검하고 교육과정의 질을 개선하기 위한 자료로 활용되어 왔다(Kogan & Shea, 2007; Constantinou & Wijnen-Meijer, 2022). 의과대학 교수는 정보 제공자, 역할모델, 학습촉진자, 평가자, 교육과정 설계자 등 다양한 역할을 수행하며(Harden & Crosby, 2000), 의과대학 교육은 강의, 실험·실습, 임상교육, 술기교육, 문제중심학습, 팀티칭 등 다양한 수업 형태로 구성된다(Spencer, 2003; Kogan & Shea, 2007; Hwang et al., 2017). 특히 통합교육과정에서는 여러 전공의 교수가 한 과목에 참여하는 경우가 많아, 단일한 강의평가 결과를 개별 교수의 수업행동으로만 해석하기 어렵다(Hwang et al., 2017). 이러한 특성은 의과대학 강의평가의 문항 구성, 결과 해석, 평정자 설정, 결과 공개 범위와 직접 관련된다(Kogan & Shea, 2007; Fluit, Bolhuis, Grol, Laan, & Wensing, 2010; Constantinou & Wijnen-Meijer, 2022).

강의평가가 수업개선으로 이어지기 위해서는 교수자가 평가 결과를 신뢰할 수 있는 자료로 받아들이고, 자신의 수업 맥락에서 해석하며, 실제 개선 행동으로 전환할 수 있어야 한다(Richardson, 2005; Hattie & Timperley, 2007). 국내 대학에서도 강의평가의 타당성, 신뢰성, 활용 방식에 대한 문제는 지속적으로 제기되어 왔다(김명화, 2005; 오은주, 2009; 한신일, 김혜정, 이정연, 2005). 교수자가 강의평가제도의 필요성을 인정하더라도 평가 방법이 타당하지 않거나 결과가 신뢰롭지 않다고 인식하면 강의평가 결과를 수업개선 자료로 적극적으로 활용하기 어렵다(김명화, 2005; 오은주, 2009; Spooren et al., 2013). 따라서 의과대학 교수들이 강의평가의 필요성, 타당성, 신뢰성, 수업개선 기능을 어떻게 인식하는지 살펴보는 것은 강의평가가 형성적 환류 자료로 기능할 수 있는 조건을 확인하는 데 필요하다. 또한 이러한 인식이 특정 직급이나 교육경력 집단에 국한된 것인지, 교수집단 전반에서 나타나는 경향인지 확인하는 것은 강의평가 개선 방향을 설정하는 데 중요한 근거가 될 수 있다.

의과대학 강의평가에서는 평정자와 결과 공개 범위도 중요한 쟁점이다. 학생은 실제 수업을 경험한 당사자로서 강의평가의 핵심 평정자이지만, 학생평가 결과에는 교수효과성 외에도 응답 맥락, 평가문화, 편향 요인이 개입될 수 있으므로 단일 지표로 활용하기보다 다른 자료와 함께 신중하게 해석할 필요가 있다(Berk, 2005; Boring, 2017; Fan et al., 2019; Stoesz, De Jaeger, Quesnel, Bhojwani, & Los, 2022). 의과대학 수업은 여러 교수가 한 과목에 참여하거나 강의와 실습, 임상교육이 결합되는 경우가 많기 때문에, 평가 결과를 교수 개인에게만 제공하는 방식으로는 과목 단위의 수업개선으로 연결되기 어려울 수 있다(Kogan & Shea, 2007; Hwang et al., 2017). 반대로 결과가 지나치게 광범위하게 공개될 경우 교수자는 강의평가를 수업개선 자료가 아니라 행정적 통제나 서열화의 수단으로 인식할 수 있다(Wachtel, 1998; Berk, 2005; Constantinou & Wijnen-Meijer, 2022). 따라서 의과대학 강의평가가 형성적 환류 체계로 작동하기 위해서는 누가 평가하고, 결과가 누구에게 어떤 범위로 공유되어야 하는지에 대한 교수 인식을 확인할 필요가 있다.

강의평가 문항의 내용과 구성 역시 강의평가의 형성적 활용을 좌우한다. 강의평가 문항은 교수자가 자신의 수업행동을 점검하고 개선할 수 있는 정보를 제공해야 하며, 교수자가 실제 수업 맥락에서 해석할 수 있는 평가영역으로 제시될 필요가 있다(Morrison, 2003; Richardson, 2005; Constantinou & Wijnen-Meijer, 2022). 교수설계 관점에서 보면 수업은 분석, 설계, 개발, 실행, 평가의 과정을 통해 구성되며, 강의평가 문항도 이러한 교수설계 과정과 연결될 수 있다(이성흙, 2001; 이성흙, 2005; Branch, 2009). 그러나 개별 문항이 많거나 문항이 서로 다른 수준의 정보를 혼재하여 제공할 경우, 교수자는 평가 결과를 수업개선 행동으로 전환하기 어렵다(Spooren et al., 2013; Quansah et al., 2024). 따라서 의과대학 교수들이 교수행동변화와 관련하여 강의평가 문항을 어떠한 평가영역으로 인식하는지 탐색하는 것은 강의평가 결과를 해석 가능한 피드백으로 전환하기 위한 기초 작업이 될 수 있다.

본 연구는 기존 강의평가 문항을 바탕으로 의과대학 교수들이 교수행동변화와 관련하여 어떤 문항을 의미 있는 정보로 인식하는지, 그리고 그러한 문항들이 어떠한 평가영역으로 구성될 수 있는지를 탐색적으로 분석하고자 한다. 이는 의과대학 강의평가가 단순한 점수 통보를 넘어 교수자가 해석하고 실행할 수 있는 형성적 환류 자료로 활용되기 위한 문항 구성 방향을 제안하는 데 의의가 있다.

이에 본 연구는 의과대학 강의평가가 수업개선을 위한 형성적 환류 체계로 기능하기 위한 조건을 교수 인식의 관점에서 분석하고자 한다. 구체적으로 강의평가의 필요성, 타당성, 신뢰성, 수업개선 기능에 대한 교수 인식과 직급·교육경력에 따른 차이를 살펴보고, 강의평가의 적절한 평정자와 결과 공개 대상에 대한 인식을 분석한다. 또한 교수행동변화와 관련하여 의과대학 교수들이 인식한 강의평가 문항이 어떠한 평가영역으로 구성되는지 탐색적으로 검토한다. 이를 통해 의과대학 강의평가가 교수자의 수업개선 행동을 지원하는 형성적 환류 자료로 활용되기 위한 방향을 제시하고자 한다. 본 연구의 연구문제는 다음과 같다.

첫째, 의과대학 교수는 강의평가의 필요성, 타당성, 신뢰성, 수업개선 기능을 어떻게 인식하며, 이러한 인식은 직급과 교육경력에 따라 차이가 있는가?

둘째, 의과대학 교수는 강의평가의 적절한 평정자와 결과 공개 범위를 어떻게 인식하는가?

셋째, 교수행동변화와 관련하여 의과대학 교수들이 인식한 강의평가 문항은 어떠한 잠정적 평가영역으로 묶일 수 있는가?

II. 이론적 배경

1. 학생 강의평가의 형성적 활용과 해석 조건

학생에 의한 강의평가는 대학 수업의 질을 점검하고 교수자에게 수업개선을 위한 정보를 제공하는

대표적인 평가제도이다. 강의평가는 학생이 경험한 교수·학습 과정에 대한 자료를 수집하여 교수자와 대학에 제공하는 절차로서, 수업내용, 교수방법, 수업운영, 상호작용, 평가방법 등에 대한 정보를 포함한다(원효현, 2002; Marsh, 1987; Wachtel, 1998). 이러한 자료는 교수자가 자신의 수업을 성찰하고 개선하는 데 활용될 수 있으며, 교수·학습 질 관리의 기초자료로도 사용될 수 있다(Richardson, 2005; Constantinou & Wijnen-Meijer, 2022).

강의평가는 활용 목적에 따라 총괄적 활용과 형성적 활용으로 구분할 수 있다. 총괄적 활용은 강의평가 결과를 교수자의 교육활동 평가, 인사, 승진, 보상, 책무성 판단 등에 사용하는 방식이다. 반면 형성적 활용은 강의평가 결과를 교수자가 자신의 수업을 이해하고 개선하는 데 활용하는 방식이다(Wachtel, 1998; Richardson, 2005). 피드백이 효과적으로 작동하기 위해서는 현재 수행과 목표 수행 사이의 차이를 인식하게 하고 이후의 행동 조정을 가능하게 해야 한다(Hattie & Timperley, 2007). 이러한 관점에서 강의평가가 형성적 평가로 기능하려면 단순한 점수 산출이나 만족도 통보를 넘어, 교수자가 이해하고 실행할 수 있는 피드백으로 제공되어야 한다(Richardson, 2005; Hattie & Timperley, 2007).

그러나 학생 강의평가의 해석에는 주의가 필요하다. 학생평가 자료는 수업 경험에 대한 중요한 정보를 제공하지만, 문항 구성, 응답 맥락, 평가 시행 방식, 결과 해석 방식에 따라 타당도와 신뢰도가 달라질 수 있다(Spooren et al., 2013; Quansah et al., 2024). Uttl 등(2017)은 학생 강의평가 점수와 학생 학습성과의 관계를 검토하면서 강의평가 점수를 교수효과성의 직접적 지표로 사용하는 데 한계가 있음을 제기하였다. 또한 학생 강의평가에는 성별, 문화적 배경, 교수자에 대한 고정관념 등 교수효과성 외적 요인이 개입될 수 있다는 연구도 보고되어 왔다(Boring, 2017; Fan et al., 2019; Stoesz et al., 2022). 따라서 학생 강의평가 결과는 교수효과성을 단일하게 판단하는 지표라기보다, 수업개선을 위해 활용할 수 있는 여러 정보 중 하나로 해석될 필요가 있다(Wachtel, 1998; Berk, 2005; Constantinou & Wijnen-Meijer, 2022).

최근 연구는 강의평가 개선이 평가도구의 문항 구성만으로 해결되기 어렵다는 점을 강조한다. Constantinou와 Wijnen-Meijer(2022)는 의학교육 맥락에서 학생 강의평가의 개선이 도구, 시행, 해석 차원에서 함께 이루어져야 하며, 교수효과성을 포괄적으로 이해하기 위해 학생평가를 다양한 자료와 함께 해석할 필요가 있다고 보았다. Quansah 등(2024) 역시 학생 강의평가의 타당도 검토에서 측정오차, 연구 맥락, 분석 방법, 응답자 특성 등 다양한 요인을 고려해야 함을 제시하였다. 이러한 논의는 강의평가를 수업개선 자료로 활용하기 위해 평가문항의 타당성뿐 아니라 결과 해석의 조건과 환류 체계가 함께 마련되어야 함을 보여준다.

국내 연구에서도 강의평가의 타당성, 신뢰성, 결과 활용 방식에 대한 문제는 지속적으로 제기되어 왔다(김명화, 2005; 오은주, 2009; 한신일, 김혜정, 이정연, 2005). 국내 대학에서 강의평가는 제도적으로 널리 운영되고 있으나, 강의평가 결과가 실제 수업개선으로 이어지기 위해서는 교수자의 결과 수용성, 평가도구의 적절성, 결과 활용 방식이 함께 고려되어야 한다(김명화, 2005; 오은주, 2009). 따라서 강의

평가에 대한 교수 인식은 강의평가의 형성적 활용 가능성을 판단하는 중요한 근거가 된다.

최근 국내 연구에서도 수업 맥락과 질 관리 목적에 맞추어 강의평가 문항과 활용 체계를 재구성하려는 논의가 이어지고 있다. 비대면·원격수업 연구는 변화된 수업환경에 맞는 평가영역과 문항을 별도로 구성할 필요가 있음을 제시하였고(송해덕, 이예찬, 2020; 한송이, 신윤미, 임은선, 2022), 교수·학습의 질을 중심으로 학습목표, 교수·학습 활동, 수업만족도와 성찰을 포함하는 문항체계도 제안되었다(이영태, 장선영, 2024). 또한 강의평가 결과를 교수법 특강, 수업컨설팅 및 교수지원 프로그램과 연계하려는 연구가 보고되었다(오선아, 김진영, 2024; 조미나, 윤나라, 김정현, 2021). 다만 국내 문헌에서 결과 공개 범위에 대한 교수의 선호가 실제 교수행동변화나 수업개선으로 이어지는지를 직접 검증한 연구는 상대적으로 제한적이다. 이는 교수 인식, 결과 공개 방식, 교수개발 지원과 실제 개선 행동의 관계를 후속 연구에서 검증할 필요가 있음을 보여준다.

2. 의과대학 교육 맥락과 강의평가 결과의 해석·환류

의과대학은 강의평가의 형성적 활용을 논의할 때 일반 대학과 다른 교육 맥락을 고려해야 한다. 의과대학 교수는 정보 제공자, 역할모델, 학습촉진자, 평가자, 교육과정 설계자 등 다양한 역할을 수행한다(Harden & Crosby, 2000). 이러한 교수 역할의 복잡성은 의과대학 수업의 설계, 실행, 평가 방식에도 영향을 미친다.

의과대학 교육은 일반 강의뿐 아니라 실험·실습, 임상실습, 술기교육, 문제중심학습, 팀티칭 등 다양한 형태로 구성된다(Spencer, 2003; Kogan & Shea, 2007). 임상교육은 실제 환자와 임상 문제를 중심으로 이루어지는 의학교육의 핵심 장면이며(Spencer, 2003), 통합교육과정에서는 여러 전공의 교수가 한 과목에 참여하는 팀티칭이 흔히 이루어진다(Hwang et al., 2017). 따라서 의과대학 강의평가 결과는 개별 교수의 수업행동뿐 아니라 과목 설계, 수업유형, 교육환경, 평가 방식, 학생의 학습경험 등 여러 요인이 복합적으로 반영된 자료로 이해될 필요가 있다(Kogan & Shea, 2007; Fluit et al., 2010). 국내 일개 의과대학 사례에서도 교수와 학생의 수업평가 인식 차이가 보고되어, 의과대학 맥락에서 평가자료의 해석 주체와 범위를 구분할 필요가 제기되었다(채수진, 임기영, 2008).

의학교육에서 강의평가는 교육과정 질 관리의 중요한 요소이지만, 의학교육의 고유한 특성은 평가 설계와 결과 해석에 영향을 준다(Kogan & Shea, 2007). 의학교육 평가인증에서도 교육과정 운영, 학생 평가, 교육성과, 지속적 질 개선은 중요한 평가 영역으로 다루어져 왔으며(Yoo, 2020), 국내 의학교육 평가인증 기준의 운영과 개선 가능성 역시 지속적으로 검토되고 있다(Lee et al., 2024). 이러한 맥락에서 의과대학 강의평가는 개별 수업에 대한 학생 반응을 수집하는 절차를 넘어 교육과정 개선과 질 관리 체계의 일부로 이해될 필요가 있다.

의과대학 강의평가에서 평정자와 결과 공개 범위는 형성적 환류 체계와 직접 관련된다. 학생은 실제 수업을 경험한 당사자로서 강의평가의 핵심 평정자이다. 그러나 학생평가 결과가 교수자의 수업

개선 행동으로 이어지기 위해서는 결과의 의미를 해석하고 개선 방향으로 연결하는 지원이 필요하다(Berk, 2005; Kogan & Shea, 2007; Constantinou & Wijnen-Meijer, 2022). 의학교육에서 피드백은 단순히 정보를 전달하는 과정이 아니라, 피드백을 주고받는 관계, 신뢰, 해석 가능성, 학습문화의 영향을 받는 과정으로 논의되어 왔다(Telio, Ajjawi, & Regehr, 2015; Watling & Ginsburg, 2019). 따라서 강의평가 결과도 교수자에게 단순히 통보되는 점수가 아니라, 교수자가 신뢰하고 해석하며 후속 행동으로 연결할 수 있는 피드백 자료로 구성되어야 한다.

의과대학 수업은 다수 교수 참여와 과목 단위 운영이 빈번하므로 교수 본인과 과목책임자 또는 교육평가전문가가 결과를 함께 해석하는 방식은 과목 단위 개선을 위한 하나의 운영 대안으로 검토할 수 있다(Hwang et al., 2017; Constantinou & Wijnen-Meijer, 2022). 반면 강의평가 결과의 공개 범위가 지나치게 넓어질 경우 교수자는 평가 결과를 수업개선 자료가 아니라 교수자 통제나 비교의 수단으로 인식할 수 있다(Wachtel, 1998; Berk, 2005). 특히 학생 강의평가에는 교수효과성 외적 요인과 편향 가능성이 보고되어 왔으므로(Boring, 2017; Fan et al., 2019; Stoesz et al., 2022), 강의평가 결과를 광범위하게 공개하거나 단독 책무성 지표로 활용하는 방식은 신중하게 검토할 필요가 있다. 따라서 의과대학 강의평가는 학생평가 결과를 수집하는 절차에 그치지 않고, 결과가 누구에게 제공되고 어떤 범위에서 공유되어야 형성적 환류의 운영 대안이 될 수 있는지를 함께 고려해야 한다.

또한 강의평가 결과가 실제 교수개선으로 이어지려면 교수개발 체계와 연결될 필요가 있다. Steinert 등(2016)은 의학교육 교수개발 연구를 검토하면서 교수효과성 향상을 위해서는 단순한 정보 제공을 넘어 교수자가 교육 실천을 성찰하고 변화시킬 수 있는 구조화된 지원이 중요하다고 보았다. 이러한 논의는 의과대학 강의평가가 단순한 평가 결과 통보에 머무르지 않고, 과목책임교수, 교육평가전문가, 교수개발 프로그램과 연결되는 환류 체계 안에서 운영될 필요가 있음을 뒷받침한다.

3. 교수설계 기반 강의평가 문항과 평가영역

강의평가가 수업개선으로 이어지기 위해서는 문항의 내용과 구성이 중요하다. 교육평가가 교육개선에 기여하려면 교육내용, 교수방법, 학습환경, 평가와 피드백 등 교육과정의 핵심 요소에 관한 정보를 제공해야 한다(Morrison, 2003). 강의평가 문항 역시 교수자가 자신의 수업행동을 점검하고 개선할 수 있는 정보를 제공해야 하며, 교수자가 실제 수업 맥락에서 해석 가능한 형태로 제시될 필요가 있다(Richardson, 2005; Constantinou & Wijnen-Meijer, 2022).

교수설계 관점은 강의평가 문항을 체계적으로 이해하는 데 유용하다. ADDIE 모형은 분석(Analysis), 설계(Design), 개발(Development), 실행(Implementation), 평가(Evaluation)의 절차를 통해 교수·학습 과정을 체계적으로 개발하고 점검하는 틀을 제공한다(Branch, 2009). 이성흠(2001)은 교수설계 이론에 근거하여 대학 강의평가 도구 개발의 이론적 기초를 탐색하였고, 이성흠(2005)은 교육·훈련 프로그램 반응평가의 관점에서 평가도구 구성과 활용을 논의하였다. 이러한 논의는 강의평가 문항을 단순 만족

도 문항이 아니라 교수설계 과정과 연결된 수업개선 정보로 이해하는 근거가 된다.

ADDIE 관점에서 강의평가 문항은 수업의 여러 단계와 연결될 수 있다. 분석 단계에서는 학습목표와 학습자 특성, 수업환경 등을 확인하고, 설계 단계에서는 수업목표, 수업계획, 난이도, 내용 구성을 구체화한다. 개발 단계에서는 교재와 매체 등 수업자료를 준비하고, 실행 단계에서는 교수자의 설명, 수업준비, 수업방법, 수업진행, 학습동기 유발이 이루어진다. 평가 단계에서는 과제, 성적평가, 피드백, 수업만족도, 개선 의견이 다루어진다(이성흠, 2001; 이성흠, 2005; Branch, 2009). 이러한 교수설계 기반 문항 구성은 강의평가 결과를 수업의 어느 과정과 연결하여 해석할 것인지 판단하는 데 도움을 줄 수 있다.

다만 교수설계 단계에 따라 강의평가 문항을 나열하는 것만으로는 교수자가 평가 결과를 쉽게 활용하기 어렵다. 강의평가 문항이 많거나 문항이 서로 다른 수준의 정보를 제공할 경우, 교수자는 어떤 결과를 우선적으로 해석하고 개선 행동으로 연결해야 하는지 판단하기 어려울 수 있다(Spooren et al., 2013; Quansah et al., 2024). 예를 들어 수업내용, 설명 명료성, 수업준비, 수업방법, 평가와 피드백, 교육환경, 학생 참여와 같은 문항들은 모두 수업개선과 관련될 수 있지만, 각 문항이 교수 개인의 행동, 과목 운영, 학습환경, 학생 노력 요인 중 어느 수준의 정보를 제공하는지는 서로 다르다.

따라서 강의평가 문항은 개별 문항 단위로만 제시되기보다 교수자가 해석할 수 있는 평가영역으로 조직될 필요가 있다. 평가영역은 강의평가 결과를 수업개선 행동으로 전환하기 위한 해석 단위가 될 수 있다. 예컨대 수업계획과 운영 관련 문항이 하나의 영역으로 묶이면 교수자는 수업 설계와 운영의 일관성을 점검할 수 있고, 실행 관련 문항이 하나의 영역으로 묶이면 교수자의 설명 방식, 수업준비, 수업진행을 성찰할 수 있다. 평가 관련 문항은 성적평가 기준과 공정성, 피드백 방식의 개선으로 연결될 수 있다. 이처럼 문항을 평가영역으로 구조화하는 것은 개별 점수의 나열을 넘어 교수자가 이해하고 실행할 수 있는 피드백을 제공하는 데 도움이 된다(Hattie & Timperley, 2007; Watling & Ginsburg, 2019).

의과대학 강의평가 문항은 이러한 평가영역 구성의 필요성이 더욱 크다. 의과대학 수업은 강의, 실습, 임상교육, 팀티칭이 혼재되어 있고, 동일 과목 안에서도 여러 교수자의 수업행동과 과목 운영 요인이 함께 작용한다(Kogan & Shea, 2007; Hwang et al., 2017). 따라서 개별 문항 점수만을 제공할 경우 교수자는 평가 결과가 자신의 수업행동, 과목 설계, 교육환경 중 어느 영역과 관련되는지 판단하기 어려울 수 있다. 교수설계 단계에 기반한 평가영역은 강의평가 결과를 수업계획, 수업내용, 수업자료, 수업실행, 평가와 피드백 등으로 구분하여 해석하게 함으로써, 의과대학 강의평가의 형성적 환류 가능성을 높이는 기초가 될 수 있다.

Ⅲ. 연구 방법

1. 연구대상

본 연구는 의과대학 강의평가의 형성적 활용 조건과 평가영역에 대한 교수 인식을 분석하기 위한 횡단적 설문조사 연구이다. 모집단 규모는 양은배, 이태선, 조명자(2019)가 2017년 전국 40개 의과대학을 조사하여 보고한 교원 11,111명을 기준으로 설정하였다. 이 가운데 각 의과대학의 기초의학 및 임상의학 교수 수에 비례하여 대학별 조사대상 인원을 배정하고, 총 1,111명을 선정하였다. 따라서 본 연구의 표집은 대학 단위의 규모에 비례하여 조사 인원을 배분하는 비례할당 방식에 해당한다.

자료 수집은 2019년 9월 16일부터 2019년 10월 30일까지 이루어졌다. 설문은 지필 설문과 온라인 설문을 병행하였다. 총 396명이 응답하였으며 회수율은 35.6%였다. 응답 자료 중 누락 응답이 있는 26명, 최근 1년간 강의 및 실습 교육시간이 모두 0시간인 8명, 강의교육시간 또는 실습교육시간이 극단값에 해당하는 4명, 직급을 기타로 응답한 2명을 제외하였다. 최종 분석에는 356명의 자료를 사용하였다.

최종 분석 대상자의 일반적 특성은 <표 1>과 같다. 본 연구에서는 교수 배경변인에 따른 강의평가 인식 차이를 확인하기 위해 직급과 교육경력을 주요 배경변인으로 활용하였다. 직급은 조교수, 부교수, 정교수로 구분하였고, 교육경력은 10년 이하, 11~20년, 21년 이상으로 구분하였다. 본 연구의 분석에는 개인 식별 정보를 사용하지 않았으며, 분석 결과는 집단 수준에서 제시하였다.

<표 1> 연구대상자의 일반적 특성 (N=356)

	구분	n	%
소재지	수도권	179	50.3
	강원·충청권	81	22.8
	영남권	66	18.5
	호남권	30	8.4
직급	조교수	91	25.6
	부교수	108	30.3
	정교수	157	44.1
교육경력	10년 이하	178	50.0
	11~20년	140	39.3
	21년 이상	38	10.7
연간강의교육시간	10시간 이하	213	59.8
	11~20시간	66	18.5
	21시간 이상	77	21.6

2. 측정도구

본 연구의 측정도구는 의과대학 교수의 강의평가 인식, 강의평가 평정자와 결과 공개 대상, 교수행동 변화와 관련하여 인식된 강의평가 문항으로 구성하였다. 강의평가 인식 문항은 오은주(2009), 이희원, 강호선, 정연순, 허은녕(2005), 장영주(2012)의 연구를 참고하여 구성하였으며, 의과대학 교수 3인의 검토를 거쳐 의과대학 맥락에 맞게 수정·보완하였다. 강의평가 인식은 강의평가제도 필요성, 강의평가 방법의 타당성, 강의평가 결과 신뢰도, 현행 강의평가제도 개선 필요성, 강의평가 결과 확인 관심도, 강의평가제도 만족도, 강의평가 결과의 수업개선 도움, 강의평가 결과 활용도, 강의평가 문항의 적절성 등 총 9문항으로 측정하였다. 각 문항은 1점 ‘전혀 그렇지 않다’에서 5점 ‘매우 그렇다’까지의 5점 Likert 척도로 구성하였다. 점수가 높을수록 각 문항 내용에 대한 동의 정도가 높음을 의미한다. 다만 현행 강의평가제도 개선 필요성 문항은 내적 일관성 산출 시 역환산하였다. 강의평가 인식 9문항의 내적 일관성은 Cronbach's $\alpha = .78$ 로 나타났다.

강의평가 평정자와 결과 공개 대상은 복수응답 문항으로 측정하였다. 강의평가 평정자는 학생, 교수 본인, 동료교수, 교육평가전문가 중 교수에게 도움이 된다고 인식하는 대상을 모두 선택하도록 하였다. 강의평가 결과 공개 대상은 교수 본인, 수강 학생, 학장 및 주요 보직교수, 교실주임교수, 과목책임교수, 의과대학 전체교수, 대학 내 모든 구성원, 대학 외부 중 적절하다고 인식하는 공개 대상을 모두 선택하도록 하였다. 이 문항들은 복수응답 방식의 명목문항이므로 신뢰도 계수 산출 대상에 포함하지 않고 항목별 선택 빈도와 비율을 중심으로 분석하였다.

교수행동변화와 관련하여 인식된 강의평가 문항은 본 연구에서 이성흠(2005)의 강의평가 문항을 교수설계의 분석, 설계, 개발, 실행, 평가 단계에 대응시켜 구성하였다. 분석 6문항, 설계 4문항, 개발 2문항, 실행 5문항, 평가 7문항 등 총 24문항이다. 설문에서는 “다음 문항에 대하여 학생들로부터 긍정적 또는 부정적 결과를 받는다면, 이러한 결과는 교수님의 행동변화에 어느 정도 영향을 줄 것이라고 생각하십니까?”라고 질문하였다. 각 문항은 1점 ‘전혀 영향을 미치지 않는다’에서 5점 ‘매우 영향을 미친다’까지의 5점 Likert 척도로 측정하였으며, 점수가 높을수록 해당 문항이 교수행동변화와 관련 있는 정보로 인식되는 정도가 높음을 의미한다. 이 문항들은 일반 강의평가 문항의 교수설계 기반 구성이 의과대학 교수 인식 자료에서 어떠한 잠정적 평가영역을 형성하는지 탐색하기 위해 사용하였다.

3. 분석방법

자료 분석은 연구문제에 따라 실시하였다. 첫째, 의과대학 교수의 강의평가 인식을 확인하기 위해 문항별 평균과 표준편차를 산출하였다. 또한 ‘그런 편이다’와 ‘매우 그렇다’ 응답을 합산하여 동의율을 제시하였다. 추가적으로 강의평가 인식이 교수 배경변인에 따라 차이를 보이는지 확인하기 위해 직급과 교육경력에 따른 일원분산분석을 실시하였다. 직급은 조교수, 부교수, 정교수로 구분하였고, 교육경

력은 10년 이하, 11~20년, 21년 이상으로 구분하였다. 집단 간 차이가 통계적으로 유의한 경우 Bonferroni 사후검증을 실시하였다. 둘째, 강의평가 평정자와 결과 공개 대상에 대한 인식을 확인하기 위해 빈도분석을 실시하였다. 해당 문항은 복수응답 방식으로 구성되었으므로 각 항목별 선택 빈도와 동의율을 산출하였다. 셋째, 교수행동변화와 관련하여 인식된 24개 강의평가 문항이 어떠한 잠정적 평가영역을 형성하는지 탐색하기 위해 탐색적 요인분석을 실시하였다. 요인분석에 앞서 KMO 표본적합도와 Bartlett 구형성 검정을 통해 자료의 적합성을 확인하였다. 요인 추출은 주축요인추출을 사용하고, 요인 간 상관 가능성을 고려하여 직접 오블리민 회전을 적용하였다. 요인 수는 교수설계의 분석, 설계, 개발, 실행, 평가 5단계를 이론적 준거로 하여 5개로 지정하였다. 공통성 .40 미만인 문항은 제외하였고, 최종 요인별 내적 일관성은 Cronbach's α , 요인 간 관련성은 요인 간 상관계수로 확인하였다. 독립 표본을 이용한 확인적 요인분석과 외적 준거타당도 검증은 실시하지 않았으므로, 이 분석의 목적은 특정 평가도구의 개발이 아닌 교수행동 변화에 영향을 미칠 것으로 예상되는 문항의 탐색으로 한정하였다.

IV. 연구 결과

1. 강의평가에 대한 의과대학 교수의 인식

강의평가에 대한 의과대학 교수의 인식을 평가한 결과는 <표 2>와 같다. 전체 응답자의 동의수준이 가장 높은 문항은 1번 “강의평가제도는 필요하다”(4.29±.65)로 나타났다. 다음으로 4번 “현행 강의평가 제도에 개선이 필요하다”(3.90±.72)였으며, 5번 “강의평가 결과에 관심을 가지고 확인한다”(3.82±.88), 7번 “강의평가 결과는 수업 개선에 도움을 준다”(3.70±.77)의 순으로 나타났다.

한편, 평균이 가장 낮은 문항은 9번 “강의평가 문항의 내용은 수업의 질을 평가하는데 적절하다”(3.08±.75)였으며, 3번 “강의평가 결과에 대해 신뢰한다”(3.19±.72), 6번 “강의평가제도에 대해 전반적으로 만족한다”(3.22±.71), 2번 “강의평가 방법은 타당하다”(3.23±.68)으로 나타났다.

교수의 특성에 따른 집단간 차이는 1번 문항과 4번 문항에서만 유의하였다. 먼저 강의평가의 필요성(1번 문항)에 대해서는 정교수 집단(4.40±.61)이 조교수 집단(4.13±.67)에 비해 동의수준이 높았고($F=5.048, p=.007$), 교육경력이 21년 이상인 집단(4.66±.48)이 10년 이하인 집단(4.18±.65)과 11년에서 20년인 집단(4.34±.67)에 비해 동의수준이 높았다($F=9.401, p<.001$). 다음으로, 현행 강의평가 제도의 개선 필요성(4번 문항)에 대해서는 정교수 집단(3.98±.72)이 조교수 집단(3.75±.72)에 비해 동의수준이 높았다($F=3.050, p=.049$). 이외의 문항들에 대한 동의정도는 집단간 유의한 차이가 없었다.

〈표 2〉 의과대학 교수의 특성에 따른 강의평가 인식 비교

문항	전체 집단 (<i>n</i> =356)	직급			<i>F</i>	사후 검증	교육경력			<i>F</i>	사후 검증
		1. 조교수 (<i>n</i> =91)	2. 부교수 (<i>n</i> =108)	3. 정교수 (<i>n</i> =157)			1. 10년이하 (<i>n</i> =178)	2. 11~20년 (<i>n</i> =140)	3. 21년이상 (<i>n</i> =38)		
1	4.29±.65	4.13±.67	4.28±.68	4.40±.61	5.048**	3)1	4.18±.65	4.34±.67	4.66±.48	9.401***	3)1, 2
2	3.23±.68	3.26±.68	3.24±.61	3.20±.73	.239		3.25±.62	3.17±.74	3.37±.75	1.358	
3	3.19±.72	3.13±.69	3.20±.72	3.22±.73	.480		3.15±.68	3.21±.75	3.34±.75	1.151	
4	3.90±.72	3.75±.72	3.90±.71	3.98±.72	3.050*	3)1	3.83±.71	3.98±.71	3.92±.78	1.786	
5	3.82±.88	3.71±.91	3.81±.94	3.89±.82	1.102		3.76±.90	3.86±.86	3.89±.86	.671	
6	3.22±.71	3.16±.69	3.25±.70	3.24±.74	.435		3.21±.68	3.20±.77	3.39±.64	1.219	
7	3.70±.77	3.55±.82	3.75±.82	3.76±.69	2.430		3.66±.79	3.74±.74	3.74±.80	.464	
8	3.35±.85	3.32±.87	3.31±.84	3.40±.85	.496		3.31±.82	3.39±.88	3.39±.89	.331	
9	3.08±.75	3.01±.71	3.04±.70	3.14±.80	1.066		3.06±.68	3.02±.78	3.34±.88	2.834	

p* < .05, *p* < .01, ****p* < .001

주. 모든 값은 평균±표준편차, 사후검증은 Bonferroni 검증 결과임.

문항 1: 강의평가제도는 필요하다, 2: 강의평가 방법은 타당하다, 3: 강의평가 결과에 대해 신뢰한다, 4: 현행 강의평가제도에 개선이 필요하다, 5: 강의평가 결과에 관심을 가지고 확인한다, 6: 강의평가제도에 전반적으로 만족한다, 7: 강의평가결과는 수업개선에 도움을 준다, 8: 강의평가 결과를 잘 활용하고 있다, 9: 강의평가 문항의 내용은 수업의 질을 평가하는데 적절하다.

2. 강의평가 평정자와 결과 공개 대상에 대한 인식

강의평가 평정자와 결과 공개 대상에 대한 의과대학 교수의 인식을 복수응답 방식으로 분석한 결과는 <표 3>과 같다. 강의평가 평정자와 관련하여 학생을 선택한 교수는 327명(91.9%)으로 가장 많았다. 다음으로 교육평가전문가 214명(60.1%), 동료교수 145명(40.7%), 교수 본인 119명(33.4%) 순으로 나타났다.

〈표 3〉 강의평가 평정자와 결과 공개 대상의 적절성에 대한 의과대학 교수의 인식 (복수응답)

구분	항목	선택빈도(<i>n</i>)	동의율(%)
평정자	학생	327	91.9
	교수본인	119	33.4
	동료교수	145	40.7
	교육평가전문가	214	60.1
결과공개대상	교수본인	340	95.5
	과목책임교수	267	75.0
	교실주임교수	170	47.8
	학장 및 주요 보직 교수	162	45.5
	수강 학생	146	41.0
	의과대학 전체교수	25	7.0
	대학 내 모든 구성원	17	4.8
	대학 외부	7	2.0

주. 복수응답 문항으로 선택 빈도와 동의율을 제시함.

강의평가 결과 공개 대상에 대해서는 교수 본인을 선택한 응답이 340명(95.5%)으로 가장 많았다. 다음으로 과목책임교수 267명(75.0%), 교실주임교수 170명(47.8%), 학장 및 주요 보직교수 162명(45.5%), 수강 학생 146명(41.0%) 순으로 나타났다. 의과대학 전체교수는 25명(7.0%), 대학 내 모든 구성원은 17명(4.8%), 대학 외부는 7명(2.0%)으로 나타났다.

3. 교수행동변화 관련 강의평가 문항의 잠정적 평가영역 탐색

본 연구에서 이성흠(2005)의 강의평가 문항을 교수설계 5단계에 대응시켜 구성한 24개 문항이 교수 인식 자료에서 어떠한 잠정적 평가영역을 형성하는지 탐색하기 위해 탐색적 요인분석을 실시하였다. 요인 수는 교수설계의 분석, 설계, 개발, 실행, 평가 5단계를 이론적 준거로 하여 5개로 지정하였다. 공통성 .40 미만인 4, 24, 23, 5, 9, 3, 18, 19번 문항을 제외하였다. 최종 16개 문항은 다섯 영역으로 묶였고 누적설명량은 64.47%였다(KMO=.86, Bartlett's $\chi^2=3088.77$, $df=120$, $p<.001$). 요인명은 각 요인에 포함된 문항 내용과 교수설계 단계의 관련성을 함께 고려하여 부여하였다. 수업시간 준수와 수업계획서의 충실성·이행을 포함한 6, 7, 8번 문항은 '설계', 교육목표와 수업내용을 포함한 1, 2번 문항은 '분석', 성적평가의 내용·기준·공정성을 포함한 20, 21, 22번 문항은 '평가', 수업내용·설명·준비·방법·진행·학습동기를 포함한 10, 13, 14, 15, 16, 17번 문항은 '실행', 교재와 수업매체를 포함한 11, 12번 문항은 '개발'로 명명하였다. 요인별 내적 일관성은 설계 $\alpha=.83$, 분석 $\alpha=.82$, 평가 $\alpha=.86$, 실행 $\alpha=.87$, 개발 $\alpha=.76$ 이었고, 요인 간 상관계수는 모두 .50 미만이었다. 요인 수를 이론에 따라 사전 지정했고 개발 요인의 고유값도 .72였으므로, 다섯 영역은 후속 확인적 타당화가 필요한 잠정적 결과로 해석하였다(표 4, 표 5).

〈표 4〉 교수행동변화 관련 강의평가 문항의 탐색적 요인분석 결과(N=356)

문항	영향정도 (<i>M</i> ± <i>SD</i>)	공통성	요인적재량				
			1	2	3	4	5
분석 (Analysis)							
1. 교육목표는 분명하게 진술되었다	3.83±.83	.98	.01	1.01	.03	-.11	.03
2. 수업내용은 학습목표 달성에 도움이 되었다	4.06±.70	.59	.08	.62	-.01	.22	.00
3. 수업에 흥미를 가지고 열심히 참여하였다	4.04±.72						
4. 수업을 위한 학생의 주간 자율학습시간은 몇 시간이었습니까	2.82±.79						
5. 수업방법에 적절한 기자재가 강의실에 배치되어 있었다	3.13±.84						
6. 수업시간이 전반적으로 잘 준수되었다	3.80±.78	.44	.52	.01	-.00	-.06	.29
설계 (Design)							
7. 수업계획서의 내용은 충실하였다	3.91±.72	.99	1.01	.01	.01	.08	-.14
8. 수업은 수업계획서에 따라 진행되었다	3.87±.74	.69	.69	.12	.07	.06	.05
9. 수업 난이도는 적절하였다	3.99±.73						
10. 수업내용이 관련분야의 이해에 많은 도움이 되었다	4.25±.66	.52	.01	.04	-.05	.70	.06
개발 (Development)							
11. 수업에 사용된 교재는 과목을 이해하는데 적절하였다	3.67±.78	.57	.12	.11	.09	.12	.56
12. 수업에 활용된 매체(technologies)는 적절하였다	3.65±.74	.62	.05	.04	.14	.08	.65
실행 (Implementation)							
13. 교수의 설명이 이해하기 쉽게 진행되었다	4.22±.69	.43	.00	-.06	.00	.53	.27
14. 교수는 수업준비가 잘 되었다	4.21±.69	.55	.09	-.08	.05	.68	.07
15. 수업방법이 적절하여 학습효과가 있었다	4.16±.69	.72	.01	-.00	.04	.86	-.12
16. 수업진행이 체계적이었다	4.06±.72	.67	.09	.05	.11	.74	-.11
17. 수업활동을 통하여 학습동기가 향상되었다	4.06±.76	.48	-.05	.10	.00	.66	.03
평가 (Evaluation)							
18. 학습과제는 수업내용을 이해하는데 도움이 되었다	3.82±.81						
19. 과제물에 대한 사후지도(feedback)는 충실하였다	3.57±.84						
20. 성적평가는 수업의 내용을 적절하게 반영하였다	3.85±.80	.58	-.09	.08	.69	.06	.10
21. 성적평가의 기준(방법)이 적절하였다	3.76±.84	.84	.05	-.05	.94	-.05	-.01
22. 성적평가는 공정하였다	3.78±.90	.64	.03	-.01	.80	.02	-.04
23. 전반적으로 나는 이 수업에 만족한다	3.93±.91						
24. 이 수업의 좋은점과 개선점을 구체적으로 기술해 주십시오	4.08±.83						
Eigen Value			3.72	1.53	2.94	1.40	.72
Explained variance (%)			23.25	9.55	18.39	8.75	4.52
Accumulative variance (%)			23.25	32.80	51.20	59.95	64.47
KMO = .86, Bartlett's test of sphericity: $\chi^2=3088.77$, $df=120$, $p<.001$							

KMO = .86, Bartlett's test of sphericity: $\chi^2=3088.77$, $df=120$, $p<.001$

주. 주축요인추출 · 직접 오블리민 회전을 적용함. 적재량은 패턴행렬 계수(요인 간 상관을 허용하므로 1 초과 가능)이며, 음영은 공통성 .40 미만 제외 문항임. 행 구분은 ADDIE 이론 분류, 요인 1~5는 EFA 잠정요인(설계 · 분석 · 평가 · 실행 · 개발)이며, 각 문항은 최대 적재 요인으로 해석함.

〈표 5〉 요인 간 상관관계 행렬

	1	2	3	4	5
1. 설계	1.00				
2. 분석	.39	1.00			
3. 평가	.36	.34	1.00		
4. 실행	.42	.37	.45	1.00	
5. 개발	.40	.21	.37	.36	1.00

V. 결론

1. 요약 및 논의

본 연구는 의과대학 강의평가가 수업개선을 위한 형성적 환류 체계로 기능하기 위한 조건을 교수 인식의 관점에서 분석하고자 하였다. 이를 위해 강의평가의 필요성, 타당성, 신뢰성, 수업개선 기능에 대한 교수 인식과 직급·교육경력에 따른 차이, 강의평가의 평정자와 결과 공개 대상에 대한 인식, 교수행동변화 관련 강의평가 문항의 교수설계 단계별 평가영역을 살펴보았다. 주요 결과를 바탕으로 논의하면 다음과 같다.

첫째, 의과대학 교수들은 강의평가제도의 필요성과 수업개선 기능은 비교적 높게 인정하였으나, 강의평가 방법의 타당성, 결과의 신뢰성, 문항의 적절성에 대해서는 상대적으로 낮은 동의 수준을 보였다. 이는 교수들이 강의평가 제도 자체를 부정한다기보다, 현행 강의평가 결과를 신뢰할 수 있는 수업개선 자료로 활용하는 데에는 유보적 인식을 가지고 있음을 보여준다. 이러한 결과는 학생 강의평가가 수업개선의 중요한 자료가 될 수 있지만, 그 결과를 교수효과성이나 학습성과의 단일 지표로 해석하기보다 문항 구성, 응답 맥락, 결과 해석 방식과 함께 검토해야 한다는 선행연구의 논의와 연결된다(Wachtel, 1998; Berk, 2005; Uttl et al., 2017; Constantinou & Wijnen-Meijer, 2022; Quansah et al., 2024). 따라서 의과대학 강의평가의 개선은 평가 실시 자체를 확대하는 데 그치기보다, 교수자가 평가 결과를 신뢰하고 해석할 수 있도록 문항의 의미, 평가단위, 결과 산출 방식, 환류 절차를 함께 정교화하는 방향으로 이루어질 필요가 있다.

둘째, 직급과 교육경력에 따른 차이는 일부 문항에서만 나타났다. 정교수 집단은 조교수 집단보다 강의평가제도의 필요성과 현행 강의평가제도의 개선 필요성에 더 높은 동의 수준을 보였고, 교육경력 21년 이상 집단도 강의평가제도의 필요성을 더 높게 인식하였다. 그러나 강의평가 방법의 타당성, 결과 신뢰성, 제도 만족도, 결과 활용도, 문항 적절성 등에서는 직급과 교육경력에 따른 유의한 차이가 나타나지 않았다. 이는 강의평가의 타당성, 신뢰성, 문항 적절성에 대한 유보적 인식이 특정 직급이나

경력 집단에만 국한된 문제가 아니라 교수집단 전반에서 공유되는 인식일 가능성을 보여준다. 따라서 강의평가 개선은 특정 교수집단의 인식 개선에만 초점을 두기보다, 의과대학 강의평가 체계 전반에서 교수자가 수용할 수 있는 타당하고 해석 가능한 평가정보를 제공하는 방향으로 접근할 필요가 있다.

셋째, 의과대학 교수들은 학생을 강의평가의 핵심 평정자로 인식하면서도, 교육평가전문가의 역할을 비교적 높게 인식하였다. 이는 학생평가가 수업 경험에 관한 중요한 자료라는 점을 교수들이 인정하고 있음을 보여준다. 동시에 학생평가 결과가 교수자의 수업개선 행동으로 이어지기 위해서는 결과의 의미를 해석하고 개선 방향으로 연결하는 전문적 지원이 필요하다는 인식도 함께 존재한다고 볼 수 있다. 의학교육에서는 피드백이 단순한 정보 전달이 아니라 신뢰, 해석 가능성, 학습문화의 영향을 받는 과정으로 논의되어 왔다(Telio et al., 2015; Watling & Ginsburg, 2019). 이러한 관점에서 강의평가 결과도 학생 응답을 수치화하여 전달하는 데 그치지 않고, 교수자가 자신의 수업 맥락에서 이해하고 실행 가능한 개선 방향으로 전환할 수 있도록 지원되어야 한다.

넷째, 강의평가 결과 공개 대상으로는 교수 본인과 과목책임교수가 상대적으로 많이 선택되었고, 의과대학 전체교수, 대학 내 모든 구성원, 대학 외부와 같은 광범위한 공개에 대한 동의는 낮았다. 이 결과는 의과대학 교수들이 강의평가 결과의 전면적 공개보다 교수 본인과 과목책임교수 중심의 제한적·형성적 공유 방식을 더 적절하게 인식하고 있음을 보여준다. 의과대학 수업은 여러 교수가 한 과목에 참여하거나 강의, 실습, 임상교육이 결합되는 경우가 많으므로 교수 본인, 과목책임교수, 교육평가전문가가 결과를 함께 해석하는 방식은 과목 단위의 수업개선으로 연결될 수 있는 운영 대안이 될 수 있다(Kogan & Shea, 2007; Hwang et al., 2017). 다만 본 연구는 공개 범위에 대한 교수 인식을 분석한 것이므로, 공개 범위별 결과 수용성, 개선계획 수립, 교수행동변화와 수업성과에 미치는 영향은 후속 연구에서 검토될 필요가 있다.

2. 의과대학 강의평가의 형성적 활용을 위한 시사점

본 연구의 결과는 의과대학 강의평가가 수업개선을 위한 형성적 환류 자료로 활용되기 위해 다음과 같은 조건이 고려될 필요가 있음을 시사한다.

첫째, 의과대학 강의평가는 수업유형과 평가단위를 고려한 평가도구로 정교화되어야 한다. 의과대학 교수들은 강의평가의 필요성과 수업개선 기능은 인정하였지만, 평가 방법의 타당성, 결과의 신뢰성, 문항의 적절성에 대해서는 낮은 동의 수준을 보였다. 이는 의과대학 강의평가 문항이 일반 강의식 수업에만 맞추어 구성될 경우 교수자가 자신의 수업 맥락에서 평가 결과를 해석하기 어렵다는 문제와 관련될 수 있다. 의과대학 수업은 강의, 실습, 임상교육, 팀티칭, 통합교육과정 등 다양한 형태로 운영되므로, 강의평가 문항은 교수 개인의 수업행동, 과목 운영, 교육환경, 학생 참여 요인을 구분하여 해석할 수 있도록 설계될 필요가 있다. 특히 교수 개인 수준에서 해석할 문항과 과목책임교수 또는 교육과정 차원에서 다루어야 할 문항을 구분하여 제시하는 방식이 요구된다.

둘째, 교수 본인, 과목책임교수, 교육평가전문가가 강의평가 결과를 제한적으로 공유하고 함께 해석하는 방식은 형성적 환류 체계의 중요한 운영 조건이 될 수 있다. 본 연구에서 교수들은 교수 본인과 과목책임교수에 대한 공개를 상대적으로 적절하게 인식하였다. 교수 개인에게는 자신의 수업을 성찰할 자료를, 과목책임교수에게는 과목의 구성과 운영을 점검할 자료를 제공하고, 교육평가전문가가 결과 해석과 개선방안 도출을 지원하는 방식은 강의평가 결과를 수업개선 자료로 전환하는 데 도움이 될 수 있다. 이때 결과 공개는 교수자 간 비교나 서열화가 아니라, 수업개선과 과목 운영 개선을 위한 제한적·형성적 공유 원칙에 기반해야 한다.

셋째, 본 연구에서 탐색된 다섯 평가영역은 강의평가 결과를 구조화하는 잠정적 틀로 활용될 수 있다. 개별 문항 점수를 나열하는 대신 영역 단위로 결과를 제시하면, 교수자가 어떤 영역을 우선 검토해야 하는지 판단하고 개선계획을 구체화하는 데 도움이 될 수 있다. 예를 들어 설계, 분석, 평가, 실행, 개발 영역별 결과를 제공하면 교수자는 자신의 수업에서 수업계획, 학습목표와 수업내용, 성적평가, 수업 실행, 교재와 매체 중 어느 영역을 우선적으로 점검해야 하는지 파악할 수 있다. 다만 실제 적용에 앞서 문항의 내용타당도와 요인구조를 별도 표본에서 추가 확인할 필요가 있다.

넷째, 강의평가 결과는 교수개발 체계와 연결될 필요가 있다. 교수자가 평가 결과를 자신의 수업 맥락에서 해석하고 개선 목표를 설정할 수 있도록 교수개발 프로그램, 과목책임교수의 과목 운영 지원, 교육평가전문가의 컨설팅이 연계될 필요가 있다. 의학교육 교수개발 연구에서도 단순한 정보 제공보다 교육 실천을 성찰하고 변화시킬 수 있는 구조화된 지원의 중요성이 논의되어 왔다(Steinert et al., 2016). 따라서 의과대학 강의평가는 교수자가 평가 결과를 확인하는 절차에 머무르지 않고, 해석, 성찰, 개선계획 수립, 교수개발 연계로 이어지는 환류 체계 안에서 운영될 때 수업개선 도구로서의 기능을 강화할 수 있다.

3. 연구의 의의와 한계

본 연구는 전국 40개 의과대학 교수 356명의 자료를 바탕으로 강의평가의 필요성·타당성·신뢰성에 대한 인식, 평정자와 결과 공개 범위에 대한 인식, 교수설계 기반 평가영역을 하나의 분석틀 안에서 함께 다루었다는 점에서 의의가 있다. 특히 의과대학 강의평가를 단순한 점수 통보나 행정적 평가가 아니라, 교수자가 해석하고 실행할 수 있는 형성적 환류 자료로 재구성해야 한다는 방향을 제시하였다.

다만 본 연구 자료는 2019년에 수집되었으므로 코로나19 이후 온라인·혼합수업 경험의 확대와 AI 기반 교육지원 도구의 확산 등 최근 의학교육 환경 변화를 직접 반영하지는 못한다. 따라서 본 연구 결과를 현재 의과대학 강의평가 체계에 그대로 일반화하기보다는, 급격한 교육환경 변화 이전 전국 의과대학 교수들이 강의평가의 필요성, 타당성, 신뢰성, 결과 공개, 평가문항을 어떻게 인식하였는지를 보여주는 기초자료로 해석할 필요가 있다. 이러한 점에서 본 연구는 향후 변화된 의학교육 환경에서 강의평가 문항과 환류 체계를 재검토하기 위한 비교 기준으로 활용될 수 있다.

또한 본 연구는 교수 인식에 관한 횡단적 조사연구이므로 강의평가 결과가 실제 교수행동변화와 수업개선으로 이어지는 과정을 직접 확인하지는 못하였다. 자료는 익명화된 개인 수준 자료로 분석되었기 때문에 대학별 교육문화나 평가제도 차이를 충분히 반영하지 못한 제한도 있다. 후속 연구에서는 강의평가 결과 제공 방식, 교수자의 결과 해석 과정, 교수개발 프로그램 연계, 실제 수업개선 행동을 중단적으로 검토할 필요가 있다.

마지막으로 본 연구에서 제시한 다섯 개 평가영역은 교수설계 5단계를 이론적 준거로 하여 탐색적 요인분석을 통해 확인한 잠정적 틀이다. 따라서 이를 확정된 평가도구나 검증된 요인구조로 해석하기 보다는, 후속 문항 개발과 확인적 타당화를 위한 기초 결과로 이해할 필요가 있다. 그럼에도 본 연구는 의과대학 강의평가의 형성적 환류 체계를 설계할 때 고려해야 할 운영 조건과 평가영역을 교수 인식에 근거하여 제시하였다는 점에서 의의가 있다.

참고문헌

- 김명화 (2005). 강의평가의 타당도와 신뢰도. **아시아교육연구**, 6(3), 1-24.
- 송해덕 · 이예찬 (2020). 비대면 대학강의평가 영역 및 문항 탐색 연구. **교육과 학습: Global Creative Leader**, 10(4), 245-273. <https://doi.org/10.34226/gcl.2020.10.4.245>
- 양은배 · 이태선 · 조명자 (2019). 의과대학 교원 현황과 업적평가제도 특징 분석. **의학교육논단**, 21(1), 41-50. <https://doi.org/10.17496/kmer.2019.21.1.41>
- 오선아 · 김진영 (2024). 대학 교수지원 프로그램의 참여가 강의평가에 미치는 영향: A대학 사례를 중심으로. **한국과 국제사회**, 8(3), 475-496. <https://doi.org/10.22718/kg.2024.8.3.017>
- 오은주 (2009). 강의평가 실태조사를 통한 강의평가 개선방향 연구. **교육방법연구**, 21(2), 1-20.
- 원효현 (2002). 『수업평가의 이해와 적용』. 서울: 교육과학사.
- 이성흠 (2001). 교수설계 이론에 근거한 대학 강의평가 도구개발을 위한 기초연구. **교육공학연구**, 17(1), 81-108.
- 이성흠 (2005). 『교육 · 훈련프로그램 반응평가』. 서울: 교육과학사.
- 이영태 · 장선영 (2024). 대학 수업의 교수 · 학습 질 제고를 위한 강의평가 문항 개발. **교육방법연구**, 36(4), 461-480. <https://doi.org/10.17927/tkjems.2024.36.4.461>
- 이희원 · 강호선 · 정연순 · 허은녕 (2005). 대학 강의 질 개선을 위한 강의 평가 문항 분석: 서울대학교 이공계열 교양과목을 중심으로. **공학교육연구**, 8(4), 52-63.
- 장영주 (2012). 대학 강의평가제도에 대한 교수와 학생의 인식 분석. 석사학위논문, 고려대학교 교육대학원.
- 조미나 · 윤나라 · 김정현 (2021). 강의평가 하위 교수자를 위한 교수지원 프로그램 개발 연구. **대학**

- 교수-학습 연구**, 14(2), 21-61. <https://doi.org/10.23122/kactl.2021.14.2.002>
- 채수진 · 임기영 (2008). 수업평가에 대한 교수와 학생 인식 비교: 일개 의과대학 사례. **Korean Journal of Medical Education**, 20(2), 163-167.
- 한송이 · 신윤미 · 임은선 (2022). 실시간 원격교육 강의평가 문항 개발 연구 -A대학의 사례를 중심으로-. **문화와융합**, 44(7), 137-154.
- 한신일 · 김혜정 · 이정연 (2005). 한국대학의 강의평가실태 분석. **교육행정학연구**, 23(3), 379-403.
- Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1), 48-62.
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27-41. <https://doi.org/10.1016/j.jpubeco.2016.11.006>
- Branch, R. M. (2009). *Instructional design: The ADDIE approach*. New York, NY: Springer. <https://doi.org/10.1007/978-0-387-09506-6>
- Constantinou, C., & Wijnen-Meijer, M. (2022). Student evaluations of teaching and the development of a comprehensive measure of teaching effectiveness for medical schools. *BMC Medical Education*, 22, 113. <https://doi.org/10.1186/s12909-022-03148-6>
- Fan, Y. et al. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PLOS ONE*, 14(2), e0209749. <https://doi.org/10.1371/journal.pone.0209749>
- Fluit, C. R. M. G., Bolhuis, S., Grol, R., Laan, R., & Wensing, M. (2010). Assessing the quality of clinical teachers: A systematic review of content and quality of questionnaires for assessing clinical teachers. *Journal of General Internal Medicine*, 25, 1337-1345. <https://doi.org/10.1007/s11606-010-1458-y>
- Harden, R. M., & Crosby, J. (2000). AMEE Guide No. 20: The good teacher is more than a lecturer: The twelve roles of the teacher. *Medical Teacher*, 22(4), 334-347. <https://doi.org/10.1080/014215900409429>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hwang, J. E. et al. (2017). Individual class evaluation and effective teaching characteristics in integrated curricula. *BMC Medical Education*, 17, 252. <https://doi.org/10.1186/s12909-017-1097-7>
- Kogan, J. R., & Shea, J. A. (2007). Course evaluation in medical education. *Teaching and Teacher Education*, 23(3), 251-264. <https://doi.org/10.1016/j.tate.2006.12.020>
- Lee, Y. et al. (2024). Challenges and potential improvements in the Accreditation Standards of the Korean Institute of Medical Education and Evaluation 2019 (ASK2019) derived through meta-evaluation: A cross-sectional study. *Journal of Educational Evaluation for Health Professions*, 21, 8. <https://doi.org/10.3352/jeehp.2024.21.8>

- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11(3), 253-388. [https://doi.org/10.1016/0883-0355\(87\)90001-2](https://doi.org/10.1016/0883-0355(87)90001-2)
- Morrison, J. (2003). ABC of learning and teaching in medicine: Evaluation. *BMJ*, 326(7385), 385-387. <https://doi.org/10.1136/bmj.326.7385.385>
- Quansah, F., Ankoma-Sey, V. R., Dankyi, L. A., Nyarko, J., & Akoto, C. O. (2024). Validity of student evaluation of teaching in higher education: A systematic review. *Frontiers in Education*, 9, 1329734. <https://doi.org/10.3389/feduc.2024.1329734>
- Richardson, J. T. E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education*, 30(4), 387-415. <https://doi.org/10.1080/02602930500099193>
- Spencer, J. (2003). Learning and teaching in the clinical environment. *BMJ*, 326(7389), 591-594. <https://doi.org/10.1136/bmj.326.7389.591>
- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), 598-642. <https://doi.org/10.3102/0034654313496870>
- Steinert, Y. et al. (2016). A systematic review of faculty development initiatives designed to enhance teaching effectiveness: A 10-year update: BEME Guide No. 40. *Medical Teacher*, 38(8), 769-786. <https://doi.org/10.1080/0142159X.2016.1181851>
- Stoesz, B. M., De Jaeger, A. E., Quesnel, M., Bhojwani, D., & Los, R. (2022). Bias in student ratings of instruction: A systematic review of research from 2012 to 2021. *Canadian Journal of Educational Administration and Policy*, 201, 39-62. <https://doi.org/10.7202/1095482ar>
- Telio, S., Ajjawi, R., & Regehr, G. (2015). The “educational alliance” as a framework for reconceptualizing feedback in medical education. *Academic Medicine*, 90(5), 609-614. <https://doi.org/10.1097/ACM.0000000000000560>
- Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22-42. <https://doi.org/10.1016/j.stueduc.2016.08.007>
- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 23(2), 191-212. <https://doi.org/10.1080/0260293980230207>
- Watling, C. J., & Ginsburg, S. (2019). Assessment, feedback and the alchemy of learning. *Medical Education*, 53(1), 76-85. <https://doi.org/10.1111/medu.13645>
- Yoo, H. H. (2020). Changes in the accreditation standards of medical schools by the Korean Institute of Medical Education and Evaluation from 2000 to 2019. *Journal of Educational Evaluation for Health Professions*, 17, 2. <https://doi.org/10.3352/jeehp.2020.17.2>

Abstract

Faculty Perceptions of the Conditions for Formative Use and Evaluation Domains of Course Evaluation in Medical Schools

Shin, Hyekyung¹

¹*Visiting Associate Professor, Seoul National University College of Medicine*

This study aimed to analyze the conditions under which course evaluation in medical schools can function as a formative feedback system for instructional improvement from the perspective of faculty perceptions. Survey data were collected from 356 full-time faculty members at 40 medical schools in Korea. The analysis examined faculty perceptions of the necessity, validity, reliability, and instructional improvement function of course evaluation; differences in these perceptions by academic rank and teaching experience; appropriate evaluators and disclosure targets; and evaluation domains of course evaluation items related to teaching behavior change. The results showed that faculty members relatively strongly recognized the necessity of course evaluation and its function for instructional improvement. However, they showed lower levels of agreement regarding the validity of evaluation methods, reliability of evaluation results, and appropriateness of evaluation items. Differences by academic rank and teaching experience were found only in some items. Students were most frequently selected as appropriate evaluators, while individual instructors and course directors were most frequently selected as appropriate disclosure targets. An exploratory factor analysis of 24 items related to teaching behavior change identified 16 items grouped into five evaluation domains: design, analysis, evaluation, implementation, and development. Rather than presenting a finalized evaluation instrument, these findings provide preliminary evidence for reorganizing course evaluation in medical schools as formative feedback that faculty members can interpret and translate into instructional improvement.

Key Words: course evaluation, instructional improvement, formative feedback, teaching behavior change