

2

임상 실습 유형과 평가 방식에 따른 평가의 공정성과 정확성에 대한 학생들의 인식도

간추린 제목 : 임상 실습 평가에 대한 학생 인식도 조사

강은희¹, 박신영²

¹서울대학교 미래치의학센터, ²서울대학교 치의학대학원 치위학교육학교실

ORCID ID

Eunhee Kang,  <http://orcid.org/0000-0003-1128-0186>

Shin-Young Park,  <https://orcid.org/0000-0002-3776-4130>

ABSTRACT

Students' perceptions of fairness and accuracy in clinical practice evaluation

Eunhee Kang¹, Shin-Young Park²

¹Center for Future Dentistry, Seoul National University

²Department of Dental Education, School of Dentistry, Seoul National University

Objectives: This study aims to investigate dental students' perceptions regarding the evaluation of clinical practice and to explore the factors influencing these perceptions.

Methods: An online survey was conducted targeting third and fourth-year dental students from 11 nationwide dental schools. The survey focused on their perceptions of the accuracy and fairness of two clinical practices: 1) clinical clerkship in specialty clinics, and 2) direct patient care in pre-doctoral treatment centers. Additionally, the effects of different clerkship models (disciplinary block, longitudinal integrated), evaluation units (individual subject, integrated subject), and grading systems (grading, pass/fail) were assessed.

Results: The overall perception of fairness and accuracy in clinical practice evaluation was found to be low. Students in disciplinary block clinical practice settings reported lower perceptions of evaluation accuracy and fairness compared to those in longitudinal integrated or mixed schedule types. Furthermore, the grading system influenced the perception of fairness and accuracy in the evaluation of clinical practices. Regarding direct patient care in pre-doctoral treatment centers, the evaluation units significantly impacted students' perceptions, although the grading system did not. Groups with grading systems showed lower perceptions compared to those with pass/fail or mixed evaluation systems.

Conclusion: The study reveals that dental students perceive clinical practice evaluations as generally unfair and inaccurate, particularly within disciplinary block schedules. To improve future dental education, adopting longitudinal integrated schedules could be considered to enhance the perceived fairness and accuracy of clinical practice evaluations.

Key words : clinical practice, clinical clerkship model, perception, evaluation, fairness and accuracy

Corresponding Author

Shin-Young Park DDS,MS,PhD, Associate Professor
Department of Dental Education, Seoul National University School of Dentistry
101 Daehak-ro, Jongno-gu, Seoul, Republic of Korea
Tel : +82-2-740-8667 / E-mail : nalby99@snu.ac.kr

ACKNOWLEDGEMENT 본 설문조사 연구는 한국치과대학치의학전문대학원 협회 주관으로 수행되었습니다.

I. 서론

임상 실습은 의학, 치의학과 같은 의료 분야 교육의 핵심적인 교육과정이다. 임상 실습을 통해 학생들은 졸업 후 의료인으로서 접해야 할 다양하고 복잡한 임상 상황과 매우 유사한 경험을 함으로써 책을 통해 습득한 의료 지식을 현장에 적용하는 능력을 가지며, 의료인으로서의 정체성과 역할을 이해하고, 환자에 대한 책임감을 갖기를 기대한다^{1~3}. 특히, 임상 실습 성적은 졸업 이후 전공의 선발에 주요 근거 자료로 사용되어^{4~6} 학생들의 스트레스에 중요한 원인 가운데 하나이기도 하다^{7,8}.

이에 따라 임상 교육 현장의 교수자와 연구진들은 임상 실습에 많은 노력과 시간을 들이며 효과적인 임상 교육 환경을 위한 다양한 평가 방법을 고민하고 개발하고 있다. 그럼에도 많은 연구들에서는 보건 분야의 학생들이 임상 실습 평가 방법 및 채점의 공정성과 정확성에 대해 신뢰하지 못하고 있는 것으로 나타났다^{9~13}. 임상 실습에서 평가의 중요성을 고려할 때, 평가에 대한 부정적인 인식은 학생들의 동기부여, 학습 행동 및 결과에도 부정적인 영향을 미칠 수 있다¹⁴.

학생들의 임상 실습 평가에 대한 연구는 많이 이루어져 왔는데, 평가자에 대한 우려^{9,15,16}, 평가의 공정성^{9,10}, 평가의 정확성^{11,12}에 대한 학생들의 인식을 조사한 연구들이 있다. Bullock et al.은 임상 실습 평가 등급 체제에 관한 학생들의 인식이 임상 실습의 학습환경과 평가에 대한 인식과의 관계를 조사한 바 있다¹⁷. 국내 연구에서도 임상 실습 평가와 관련한 연구들이 많이 진행되어 왔으며, 임상 실습 평가 방식에 대한 현황을 조사하거나^{18,19} 학생들의 임상 성과와 졸업시험 성적과의 상관성 분석¹⁹과 같은 조사가 있었다. 김순구 등은 임상 실습 평가에 대한 간단한 질문을 통해 학생들의 인식을 조사하여 평가의 공정성과 신뢰성의 문제를 제기하였다²⁰. 그러나 치의학 임상 교육의 특성을 반영하여 임상 실습의 평가에 대한 학생들의

인식을 체계적으로 밝히거나, 이러한 인식에 영향을 주는 요인을 밝힌 연구는 찾기 어렵다.

이 연구에서는 교육의 수요자이면서 학습의 중심인 학생들이 가진 임상 실습 평가에 대한 인식을 구체적으로 조사하고, 이에 영향을 미치는 요인을 탐색하고자 한다. 이를 통해 치의학 분야의 임상 실습 평가 현황을 이해하고 개선할 수 있는 구체적인 방안을 제시할 수 있을 것이다.

연구 문제는 다음과 같다.

- (1) 임상 실습 유형에 따라 임상 실습 평가의 정확성과 공정성에 대한 학생들의 인식이 다른가?
- (2) 평가 방식에 따라 임상 실습 평가의 정확성과 공정성에 대한 학생들의 인식이 다른가?

II. 연구대상 및 방법

1. 연구 상황과 연구 대상

한국치과대학·치의학전문대학원협회는 국내 임상교육 환경을 개선하기 위해 2023년 임상교육 현황 조사를 의뢰하였고, 이 연구는 평가와 관련한 일부 설문조사 결과를 후향적으로 분석하여 정리한 것이다. 본 연구는 서울대학교 치의학대학원 연구윤리심의위원회의 승인을 받아 시행하였다(IRB NO: S-D20240011).

연구 대상은 연구의 목적에 맞게 임상 실습 경험이 있는 전국의 11개 치과대학, 치의학전문대학원의 본과 3, 4학년 학생을 대상으로 전수조사를 실시하였다. 설문은 학생들이 의견을 자유롭게 개진할 수 있도록 온라인 구글 설문을 이용하였다. 설문 문항을 시작하기 전 설문의 목적을 설명하는 안내문을 삽입하였으며 이후 설문을 이해하고 참여 동의를 묻는 문항을 추가하여 참여자가 자발적으로 설문에 참여하도록 하였다. 가능한 많은 학생들이 설문에 참여하도록 하기 위해 11개 학교의 대표 학생들에게 설

문의 목적을 설명하고 온라인 설문 링크를 전달하여 3, 4학년 학생들에게 설문에 대한 안내를 하였다. 학생들의 응답은 2023년 10월 9일부터 10월 31일까지 약 20일 동안 수집하였다. 이 기간 동안 11개 대학의 3, 4학년 학생 750여명 가운데 251명(3학년 25%, 4학년 10%)이 설문에 참여하여 33.5%의 참여율을 나타내었다.

2. 설문 도구

설문 문항은 1) 전문과에서 시행되는 진료 관찰 및 참여 임상 실습과 2) 학생진료 임상 실습으로 구분되었다. 진료 관찰 및 참여 임상 실습은 각 과목별 톨 실습 이후의 임상 실습 기간 동안의 실습을 의미함을 안내하였다. 학생들은 먼저 진료 관찰 및 참여 임상 실습에 관한 설문에 응답하였고 마지막 문항에 학생 진료 임상 실습에 대한 경험 여부를 확인하였다. 학생 진료 임상 실습 경험이 있다고 답한 학생의 경우 학생 진료 임상 실습 설문으로 이동하여 설문을 계속하도록 하였고, 학생 진료 임상 실습 경험이 없다고 답한 경우 설문을 바로 제출하도록 하여 각각의 임상 경험이 있는 학생만 설문에 참여할 수 있도록 하였다.

문항의 내용은 참여자의 배경 문항과 그리고 평가에 대한 인식을 묻는 문항으로 구분된다(Table1). 참여자의 임

상 실습 배경 문항은 진료 관찰 및 참여 임상 실습의 일정 유형(블록형, 장기통합, 혼합), 평가 부여 단위(개별과목, 통합과목, 혼합), 평가 등급(등급제, 급락제, 혼합)을 묻는 3개의 문항을 포함한다.

평가에 대한 인식을 묻는 문항은 Bullock et al.⁹⁾의 연구를 사용하였으며, 평가의 공정성과 정확성, 평가자에 관한 범주를 포함한다(Table1). 평가에 대한 인식 문항은 5점 척도(1점 전혀 아니다, 5점 매우 그렇다)를 사용하였다.

3. 자료 분석

수집한 자료는 각 문항의 기술적인 통계와 집단별 비교 분석을 하였다. 집단 변수는 임상 실습 일정 유형(블록형, 장기통합), 평가 부여 단위(개별과목, 통합과목)와 평가 등급(등급제, 급락제) 변수를 사용하였다. 즉, 임상 실습 일정 유형과 평가 부여 단위, 평가 등급에 따라 학생들의 임상 실습 평가에 대한 인식이 다른지 분석하였다.

임상 실습 배경 문항에서 학생들의 응답이 적은 보기의 경우 보기의 특성을 고려하여 다음과 같이 통합하였다(Table 2). 임상 실습의 일정 유형의 세 가지 보기(블록형, 장기통합, 혼합) 가운데, 장기통합의 응답 수가 블록형의 응답 수보다 확연히 적어 장기통합이 포함된 혼합의 보기

Table 1. 설문 문항 개요

항목	설명	문항 수
참여자 배경 정보	일반 배경 성별, 나이, 학년, 성적	4
	임상 실습 배경 진료 관찰 및 참여 임상 실습의 일정 유형, 평가 부여 단위, 평가 등급	3
평가에 대한 인식	평가 공정성 수행능력 평가 기준 이해, 평가자 기준의 이해, 임상 실습 성적 결정 방식 이해, 임상 실습 성적의 공정성, 임상 실습 성적 기준의 명확성, 임상 실습 성적의 상대 평가, 이익제기 가능성	7
	평가 정확성 평가자와 무관한 평가의 일관성, 치의학 지식 평가의 정확성, 임상 술기 역량 평가의 정확성, 의사소통능력 평가의 정확성, 전문성 평가의 정확성	5
	평가자 평가자의 평가점수 척도 이해도, 학생 이해도, 복수의 학생 관찰, 비관찰 부문 평가(부정문항)	4

Table 2. 연구 대상의 일반 및 임상 실습 배경 정보

변수		진료 관찰 및 참여		학생 진료	
		빈도(n)	비율(%)	빈도(n)	비율(%)
성별	남성	149	59.4	108	55.1
	여성	102	40.6	88	44.9
나이	20~25세	124	49.4	91	46.4
	26~30세	109	43.4	89	45.4
	31세 이상	18	7.2	16	8.2
일반 배경	학년				
	3학년	188	74.9	134	68.4
	4학년	63	25.1	62	31.6
학점	A (3.6이상/4.5, 3.40이상/4.3)	95	37.8	75	38.3
	B (2.6~3.5/4.5, 2.4~3.3/4.3)	132	52.6	105	53.6
	C (1.6~2.5/4.5, 1.4~2.3/4.3)	24	9.6	16	8.2
	D (1.5이하/4.5, 1.30이하/4.3)	0	0.0	0	0.0
임상 실습 일정 유형	블록형 (전공과 중심, Fix, Keep)	75	29.9	-	-
	장기통합 (자유실습)	43	17.1	-	-
	혼합 (전공과 중심과 장기통합의 혼합)	133	53.0	-	-
임상 실습 배경	평가 부여 단위				
	개별과목	193	76.9	140	71.4
	통합과목	16	6.4	25	12.8
	혼합 (일부는 개별과목, 일부는 통합과목으로 점수 부여)	42	16.7	31	15.8
평가 등급	등급제 (A/B/C/D)	203	80.9	120	61.2
	급락제1 Pass/fail	17	6.8	40	20.4
	급락제2 Honor/pass/fail	4	1.6	5	2.6
	혼합 (일부는 Pass/fail과 일부는 등급제)	27	10.8	31	15.8
전체		251	100.0	196	100.0

를 장기통합과 통합하였다. 마찬가지로 평가 부여 단위의 세 가지 보기(개별과목, 통합과목, 혼합) 가운데 통합과목과 혼합의 보기를 통합하였고, 평가 등급의 네 가지 보기(등급제 (A/B/C/D), 급락제1 Pass/fail, 급락제2 Honor/pass/fail, 혼합) 가운데 등급제를 제외한 세 가지의 보기를 하나로 통합하였다.

각 임상 실습 내의 집단별 비교 분석은 독립표본 t-검정을 이용하였다. 분석 도구는 SPSS version 26.0 (IBM

Co., Armonk, NY, USA)을 이용하였다.

III. 연구결과

1. 임상 실습 평가 현황

진료 관찰 및 참여 임상 실습과 학생 진료 임상 실습 설

문에 각각 251명, 196명이 참여하였다(Table 2). 참여자의 일반 배경에서, 진료 관찰 및 참여 임상 실습의 설문 참여자 가운데 남녀 비율은 남성(149명, 59.4%)이 여성(102명, 40.6%)보다 많았으며, 나이는 30세 이하(233명, 92.8%)가 대부분을 차지하였다. 학년은 3학년이 74.9%(188명)를 차지하여 4학년(63명, 25.1%)보다 많았으며, 학점은 B학점(132명, 52.6%), A학점(95명, 37.8%) 순으로 많았다. 이러한 비율은 학생 진료 임상 실습에서도 비슷하게 나타났다.

참여자의 임상 실습 배경 문항에서 진료 관찰 및 참여 임상 실습의 경우 사전에 계획된 일정에 따라 해당 전문 과목 진료실에 출석하여 진료에 참여하는 블록형(전공과 중심) 임상 실습은 29.9%(75명)를 차지하였고, 정해진 곳에 출석하지 않고 장기 추적하는 환자 스케줄에 따라 임상 실습에 참여하는 장기통합 임상 실습은 17.1%(43명), 두 모델을 혼합한 경우는 53.0%(133명)로 나타났다. 또한 평가 부여 단위에서 개별과목별로 점수를 부여한 경우가 76.9%(193명)로 가장 많았고, 전공을 통합한 통합과목으로 점수를 부여하거나, 앞의 두 경우를 혼합한 경우는 23.1%(58명)를 차지하였다. 평가 등급에서는 A/B/C/D로 학점을 부여하는 등급제가 80.9%(203명)로 대부분을 차지하였다.

학생 진료 임상 실습에서 평가 부여 단위와 평가 등급의 비율은 진료 관찰 및 참여 임상 실습과 크게 다르지 않았다.

2. 임상 실습 유형과 평가 방식에 따른 학생들의 평가에 대한 인식

참여자들이 임상 실습의 평가가 공정하고 정확하다고 생각하는지, 임상 실습의 평가자는 공정하다고 생각하는지 그리고 이러한 인식이 임상 실습의 일정 유형, 평가 부여 단위, 평가 등급 체계에 따라 차이가 있는지 분석하였다.

1) 진료 관찰 및 참여 임상 실습

진료 관찰 및 참여 임상 실습에서 평가의 세 가지 항목 모두 평균이 낮았으며, 특히 평가 정확성(2.50 ± 1.07)이 가장 낮게 나타났다(Table 3).

평가 공정성 항목(2.83 ± 0.96)에서는 학생들의 평가 기준 이해 항목(3.33 ± 1.07)이 가장 높았으며, 평가 성적에 대한 이의제기 가능성 문항(2.26 ± 1.29)이 가장 낮았다. 집단별로는 임상 실습 일정 유형 집단에서 장기통합 & 혼합 일정 집단(3.00 ± 0.87)이 블록형(전공과 중심) 일정 집단(2.45 ± 1.05)보다 모든 항목에서 높았는데 평가 기준 이해 문항을 제외하고 평가자 기준 이해, 성적의 공정성을 포함한 6개 문항에서 통계적으로 유의미한 차이를 보였다. 평가 부여 단위에서는 개별과목 부여 집단(2.85 ± 0.95)이 통합과목 & 혼합 점수 부여 집단(2.77 ± 1.00)보다 전체 평균은 약간 높았으나 통계적으로 유의미하지 않았다. 평가 등급에서는 급락제 & 혼합 등급 집단이 등급제 집단보다 모든 세부 문항에서 높은 평균을 보여주었고, 학생들의 평가 기준 이해와 임상 실습 성적의 상대 평가 문항을 제외하고 통계적으로 유의미하였다.

평가 정확성 항목에서는 5개의 평가 문항의 평균이 거의 비슷하게 나타났다. 집단별로 구분하였을 때, 장기통합 & 혼합 일정 집단(2.61 ± 1.05)이 블록형(전공과 중심) 집단(2.23 ± 1.06)보다 높았으며 임상실기 평가의 정확성 문항을 제외하고 통계적으로 유의미한 차이를 보였다. 평가 등급에서도 급락제 & 혼합 등급 집단이 등급제 집단보다 모든 세부 문항에서 높은 평균을 보여주었으며, 평가의 일관성, 지식 평가의 정확성을 포함한 5개 항목 모두에서 유의수준 0.005 수준에서 유의미한 차이를 보였다. 평가 공정성 항목과 마찬가지로 평가 부여 단위에 따른 두 집단의 유의미한 차이는 없었다.

평가자 공정성 항목에서는 평가자의 평가점수 척도 이해 문항에서 가장 높은 점수(2.90 ± 1.14)가, 학생 이해도 문항(2.57 ± 1.24)에서 가장 낮은 점수가 나타났다. 집단

Table 3. 진료 관찰 및 참여 임상 실습에서 임상 유형과 평가 방식에 따른 학생들의 평가에 대한 인식 비교

범주	평균	임상 실습 일정 유형			평가 부여 단위			평가 등급		
		블록형 (전공과 중심)	장기통합 & 혼합	p-value	개별과목	통합과목 & 혼합	p-value	등급제	급락제 & 혼합	p-value
평가 공정성	2.83±0.96	2.45±1.05	3.00±0.87	0.000	2.85±0.95	2.77±1.00	0.550	2.72±0.94	3.29±0.93	0.000
평가 기준 이해	3.33±1.07	3.17±1.12	3.40±1.04	0.128	3.37±1.04	3.19±1.16	0.253	3.29±1.07	3.52±1.05	0.171
평가자 기준 이해	2.88±1.26	2.39±1.38	3.09±1.14	0.000	2.96±1.25	2.62±1.24	0.072	2.77±1.26	3.35±1.12	0.003
성적 결정 방식 이해	3.02±1.24	2.37±1.34	3.30±1.08	0.000	3.06±1.24	2.91±1.22	0.440	2.92±1.25	3.46±1.07	0.007
성적의 공정성	2.61±1.27	2.27±1.28	2.75±1.24	0.005	2.59±1.27	2.66±1.26	0.735	2.43±1.23	3.35±1.18	0.000
성적 기준의 명확성	2.71±1.21	2.35±1.32	2.87±1.13	0.003	2.72±1.21	2.71±1.21	0.964	2.60±1.21	3.19±1.10	0.002
성적의 상대 평가	3.03±1.21	2.69±1.29	3.17±1.14	0.006	3.05±1.19	2.95±1.25	0.567	2.99±1.21	3.19±1.18	0.308
이의제기 가능성	2.26±1.29	1.93±1.25	2.40±1.28	0.008	2.24±1.27	2.34±1.36	0.581	2.09±1.24	2.98±1.23	0.000
평가 정확성	2.50±1.07	2.23±1.06	2.61±1.05	0.009	2.48±1.08	2.55±1.04	0.669	2.36±1.04	3.04±1.02	0.000
평가의 일관성	2.51±1.22	2.20±1.24	2.65±1.19	0.007	2.53±1.23	2.47±1.17	0.731	2.36±1.22	3.15±1.01	0.000
지식 평가의 정확성	2.47±1.18	2.19±1.16	2.59±1.18	0.013	2.46±1.17	2.25±1.23	0.730	2.36±1.15	2.92±1.24	0.003
임상술기 평가의 정확성	2.47±1.18	2.25±1.20	2.56±1.17	0.057	2.45±1.19	2.53±1.14	0.637	2.34±1.16	3.02±1.12	0.000
의사소통능력 평가의 정확성	2.50±1.19	2.25±1.21	2.61±1.17	0.030	2.49±1.20	2.53±1.16	0.813	2.37±1.16	3.06±1.17	0.000
전문성 평가의 정확성	2.52±1.20	2.24±1.18	2.64±1.19	0.015	2.47±1.67	2.69±1.29	0.225	2.40±1.19	3.04±1.11	0.001
평가자 공정성	2.74±0.76	2.49±0.81	2.85±0.70	0.000	2.74±0.76	2.74±0.76	0.997	2.66±0.76	3.10±0.62	0.000
평가점수 척도 이해	2.90±1.14	2.47±1.25	3.09±1.04	0.000	2.95±1.13	2.74±1.16	0.214	2.81±1.14	3.31±1.04	0.006
학생 이해도	2.57±1.24	2.19±1.18	2.73±1.23	0.001	2.53±1.23	2.69±1.27	0.400	2.43±1.22	3.15±1.15	0.000
복수의 학생 관찰	2.68±1.18	2.45±1.23	2.78±1.15	0.046	2.68±1.20	2.69±1.14	0.951	2.58±1.18	3.13±1.12	0.004
비관찰 부문 평가	2.81±1.15	2.84±1.31	2.80±1.09	0.796	2.80±1.18	2.84±1.07	0.787	2.81±1.18	2.81±1.05	0.980

별로는 장기통합 & 혼합 일정 집단(2.85±0.70)이 블록형 (전공과 중심) 집단(2.49±0.81)보다 평균이 높았고, 평가자가 비관찰 부문에서도 평가를 한다는 문항을 제외하고 평가점수 척도 이해, 학생 이해도, 복수의 학생 관찰 문항에서의 두 집단의 평균 차이는 통계적으로 유의미하였다. 평가 등급에서는 급락제 & 혼합 등급 집단(3.10±0.62)이 등급제 집단(2.66±0.76)보다 높았으며, 마찬가지로 비관찰 부문 평가를 제외한 3개의 문항에서 유의미한 차이를 보였다. 평가 부여 단위에 따라서는 집단별 차이가 나타나지 않았다.

2) 학생 진료 임상 실습

학생 진료 임상 실습에서 세 가지 평균 항목의 평균은 진료 관찰 및 참여 임상 실습보다 높았으며, 세 가지 평가 항목 가운데에서는 평가자 공정성(3.13±0.67)이 가장 낮았다(Table 4).

평가 공정성 항목에서는 진료 관찰 및 참여 임상 실습과 마찬가지로 평가 기준 이해(3.74±0.92)가 가장 높았으며 이의제기 가능성 문항(2.63±1.35)이 가장 낮았다. 집단별로 보았을 때, 급락제 & 혼합제 점수 부여 집단(3.59±0.72)이 등급제 부여 집단(3.26±0.90)보다 높은

Table 4. 학생 진료 임상 실습에서 임상 유형과 평가 방식에 따른 학생들의 평가에 대한 인식 비교

범주	평균	평가 부여 단위			평가 등급		
		개별과목	통합과목 & 혼합	p-value	등급제	급락제 & 혼합	p-value
평가 공정성	3.39±0.85	3.37±0.81	3.43±0.95	0.634	3.26±0.90	3.59±0.72	0.007
평가 기준 이해	3.74±0.92	3.76±0.85	3.71±1.07	0.790	3.68±0.94	3.86±0.88	0.180
평가자 기준의 이해	3.66±0.98	3.65±0.95	3.68±1.06	0.855	3.52±1.07	3.88±0.78	0.006
성적 결정 방식 이해	3.60±1.05	3.62±1.00	3.55±1.16	0.701	3.48±1.11	3.79±0.91	0.037
성적의 공정성	3.45±1.08	3.40±1.07	3.59±1.13	0.270	3.28±1.13	3.72±0.95	0.004
성적 기준의 명확성	3.38±1.09	3.42±1.04	3.27±1.20	0.373	3.23±1.17	3.61±0.91	0.013
성적의 상대 평가	3.25±1.15	3.22±1.13	3.32±1.22	0.584	3.20±1.21	3.33±1.05	0.447
이의제기 가능성	2.63±1.35	2.51±1.32	2.91±1.38	0.063	2.41±1.33	2.97±1.31	0.004
평가 정확성	3.22±1.02	3.18±1.00	3.31±1.05	0.379	3.05±1.05	3.48±0.92	0.004
평가의 일관성	3.19±1.18	3.12±1.16	3.36±1.21	0.206	3.06±1.25	3.39±1.02	0.051
지식 평가의 정확성	3.19±1.16	3.13±1.16	3.34±1.15	0.249	3.00±1.17	3.49±1.06	0.004
임상술기 평가의 정확성	3.27±1.16	3.21±1.13	3.41±1.22	0.284	3.10±1.17	3.54±1.09	0.009
의사소통능력 평가의 정확성	3.21±1.15	3.21±1.14	3.21±1.17	0.969	3.06±1.21	3.45±1.00	0.020
전문성 평가의 정확성	3.22±1.14	3.21±1.14	3.27±1.14	0.737	3.04±1.18	3.51±1.00	0.004
평가자 공정성	3.13±0.67	3.07±0.67	3.28±0.66	0.045	3.05±0.68	3.25±0.64	0.046
평가점수 척도 이해	3.40±1.12	3.34±1.13	3.57±1.08	0.183	3.28±1.11	3.59±1.11	0.059
학생 이해도	3.07±1.21	2.96±1.18	3.32±1.24	0.061	2.95±1.24	3.25±1.15	0.090
복수의 학생 관찰	3.28±1.20	3.21±1.19	3.46±1.21	0.175	3.17±1.25	3.46±1.09	0.094
비관찰 부문 평가	3.23±1.17	3.23±1.15	3.23±1.25	0.985	3.18±1.20	3.30±1.14	0.490

평균을 보였으며 평가 기준 이해와 성적의 상대 평가 문항을 제외하고 유의미한 차이를 보였다. 평가 부여 단위 집단별로는 큰 차이를 보이지 않았다.

평가 정확성 항목에서도 급락제 & 혼합 등급 부여 집단(3.48±0.92)이 등급제 부여 집단(3.05±1.05)보다 평균이 높았으며 평가의 일관성 1개 문항을 제외하고 이 차이는 통계적으로 유의미한 것으로 나타났다. 평가 부여 단위 집단 사이에서는 큰 차이가 없었다.

평가자 공정성 항목에서도 마찬가지로 급락제 & 혼합 등급 부여 집단(3.25±0.64)이 등급제 부여 집단(3.05±0.68)보다 평균이 높았으며 통계적으로 유의미한

수준으로 나타났다. 또한 통합과목 & 혼합 점수 부여 집단(3.28±0.66)이 개별과목 점수 부여 집단(3.07±0.67)보다 평균이 높았고 0.05 수준에서 이 차이는 통계적으로 유의미하였다.

IV. 논의 및 결론

전국의 치과대학, 치의학대학원생 학생수는 730명 내외로 본 설문조사에 참여한 인원은 3학년은 약 25%, 4학년은 약 10%이었다. 학생 진료에 대한 설문 조사는 3학

년 중 학생 진료를 시작하지 않은 54명은 학생 진료에 대한 설문은 답변하지 않아 전국 치과대학, 치의대학원생 중 약 18%, 4학년 10% 내외가 본 설문조사에 참여하였다. 본 설문조사는 취약한 피험자인 학생들의 자유로운 의견 개진을 위하여 학생의 개인 정보, 소속 등에 대한 정보를 포함하지 않고 무기명으로 수행되었으며 학생들의 임상 실습 유형에 따라 학생들의 평가에 대한 인식도를 분석하였다.

Wu 등의 온라인 설문조사에 관한 메타분석 결과에 따르면 설문조사 전 대상자에 대한 사전 미팅이나 사전 인센티브, 명확하게 정의된 연구대상자 설정 등이 설문조사 응답율을 높인다고 하였으나 그럼에도 불구하고 약 40% 이상의 응답율은 기대하기 어렵다고 하였다²¹⁾. 본 설문조사는 설문 주제의 성격상 학생들에게 민감한 주제로 설문조사 결과가 학생들의 평가에 영향을 미칠 수 있다고 생각할 수 있어 설문조사에 대한 무응답율이 높을 수 있다. 본 설문조사는 25% 내외의 응답율을 보였으며 전반적인 설문 응답자의 성별, 나이, 학업 성취도 등도 전반적인 학생들의 분포와 비슷하게 도출되었다²²⁾. 또한 진료 관찰 및 참여 실습에 관한 설문조사에 참여한 학생 중 사전에 계획된 임상 실습 일정에 따라 해당 전문과목 진료실에 출석하여 전공의 및 교수의 진료에 참여하는 학생이 약 30%, 학생이 장기추적하는 환자 스케줄에 따라 임상 실습에 참여하는 학생이 약 20%, 두 시스템을 혼합한 유형이 약 50%으로 이는 11개 대학 및 대학원의 학교별 임상 실습 유형의 비율과 비슷하였다. 따라서, 본 설문조사가 전체 학생들의 의견에 대한 대표성을 어느 정도 확보했다고 할 수 있다.

본 설문조사 결과 임상 실습 일정 유형은 진료 참여 및 참여 실습에 대한 평가에 대한 인식에 있어서 유의하게 영향을 미치는 요소였다. 블록형(전공과 중심)으로 진행되는 사전에 계획된 임상 실습 일정으로 실습하는 학생들이 장기통합 임상 실습을 경험한 학생들보다 평가의 공정성, 평가의 정확성 및 평가자 공정성 모두 유의미하게 낮았다. 각

각 개별 항목에서도 일관되게 나타나는 경향이였다. 이는 의학교육에서도 비슷한 경향성을 보고하기도 하였다. Teherani 등은 의학교육에서 임상 실습 유형에 따른 학생들의 성과를 분석하였는데 그 결과 역시 장기통합 실습에서 학생들의 교수자에 대한 만족도, 임상 술기 관찰 및 피드백 등이 블록형(전공과 중심)의 실습보다 더 높게 나왔다²³⁾.

블록형(전공과 중심) 실습은 일정기간동안 진료실에 내원한 환자의 진료를 학생들이 직접 관찰하거나 참여하여 배우는 도제식 학습 방법이다. Horst 등의 연구에서 50 시간 이상의 진료 관찰 및 참여는 실제 환자 진료 시 심적 부담을 완화시켜주고 진료에 대한 이해력을 증진시키는데 도움이 된다고 하고 있다²⁴⁾. 그러나, 진료실은 전공의, 교수들이 외래로 내원한 환자들을 진료하는 공간으로 학생들의 교육을 위해 설계되거나 배려되는 공간은 아니다²⁵⁾. 학생들은 임상 실습 기간 동안 배정된 담당 교수자의 환자의 진료를 보조하거나 참여하면서 도제식 학습을 하게 되는데, 학생들은 해당 환자의 병력, 치료의 경과 등에 대한 숙지 없이 해당 진료에 참여하게 된다. 즉, 환자의 상황에 따른 치료 계획의 수립 및 여러가지 예측불가능한 변수들에 대한 사전 고려 등에 대한 교육보다는 이미 사전에 결정된 치료계획에 따라 교수자가 진료를 수행하는 과정을 관찰하는 것 위주의 교육이 진행된다.

게다가 바쁜 외래 진료 스케줄로 인해 환자 진료 중 교육적 의미가 있는 내용에 대하여 전공의/교수자와 학생이 토의를 하기에 충분하지 않은 경우도 종종 있다. 그러다 보니 학생들은 자신의 진료실 실습 수행에 대하여 기준에 충족하게 수행하고 있는지 등의 피드백이 부족한 상태에서 본인의 역량에 대하여 정확하게 인지하지 못하고 있는 상황에 놓이게 되고 평가 결과를 납득하지 못하는 것으로 연결된다. 게다가 이의제기 가능성은 전체 평가 인식도 점수 중 최하위를 보여주었다. 이러한 폐쇄적인 문화에서 학생들은 자포자기적인 심정으로 수동적인 학습자의 모습을 보일 수 있다.

반면, 장기통합 실습의 경우 학생들은 자신이 관찰하고 참여하는 환자에 대한 병력 및 주소 인터뷰, 치료계획 수립 과정에서부터 참여하는 경우가 많아 환자에 대한 이해력이 높은 편이다^{26,27}. 이러한 상황에서 진료실에서 수행하는 진료에 대한 관찰 및 참여는 학생들에게 동기부여 측면에서 교육적 효과가 더욱 높으며 수동적인 학습자가 아니라 적극적인 학습자로서 진료에 참여하게 된다. 진료 종료 후 본인에 예측하지 못한 진료 과정 또는 진료 결정에 대하여 학생들은 의문점을 갖게 되고 이에 대하여 전공의/교수자에게 적극적으로 물어보기도 할 수 있다. 이러한 과정에서 전공의/교수는 학생의 지식 수준, 술기에 대한 이해도 등에 대한 피드백을 전달할 수 있으며 이는 학생이 스스로의 상태에 대한 인지력을 높이는 데 도움이 된다.

또한 장기통합 실습에서는 실습의 주체가 환자이다²⁸. 어떤 환자를 장기추적하느냐가 더 성적에 중요한 영향을 미치는 요소로 작용한다. 반면 블록형 진료실 실습에서는 담당 전공의/교수 중심으로 실습 교육이 이루어지다보니 학생들은 담당 전공의/교수에 대한 불만이 있는 경우 이는 성적에 대한 불만족으로 연결된다.

평가 부여 단위는 학생들의 평가에 대한 인식에 큰 차이가 없었으나 등급제와 급락제는 학생들의 평가에 대한 인식에 유의한 영향을 미쳤다. 급락제 학생에서 등급제 학생에 비해 임상 실습 평가에 대한 인식도가 모두 높은 편이었다. 평가기준의 측면에서는 급락제 및 등급제 학생 모두 큰 차이가 없었으나 등급제 학생들은 성적의 공정성 측면에서 낮은 점수를 보여주었다. 진료 관찰 및 참여 실습 뿐만 아니라 학생 진료에서도 급락제 학생들이 평가에 대한 공정성, 정확성에 대한 인식도가 낮게 나타났다. Bullock et al. 등도 의학교육에서도 등급제 또는 급락제 등 평가 시스템이 임상 실습에 영향을 미친다고 보고하였다⁹. 해당 논문에서 학생들은 투명한 성적 평가 시스템 및 평가자에 대한 훈련, 장기적 관찰에 의한 평가를 요구하였다.

본 설문조사는 11개 치과대학 치의학대학원 학생들의

임상 실습 중 실습 유형 및 평가 체계에 따른 평가 인식도 조사 결과이다. 평가에 대한 공정성 및 정확성에 대한 인식은 학생들의 학교에 대한 신뢰도에 중요한 영향을 미치며 이는 학생들의 학습 환경, 학습 동기 및 학습 효과에도 영향을 미친다. 효과적인 임상 교육을 위해 임상 실습 평가의 공정성, 정확성을 높이기 위해 평가 기준을 명확하게 수립하고 일관된 평가를 위한 평가자들에 대한 교육 및 투명한 평가 기준 공개, 이의제기를 위한 창구 개설 등이 필요하다.

본 연구는 학교 사이의 비교가 목적이 아니며, 또한 학생들의 자유로운 의견 개진을 위하여 학교를 포함하여 개인을 특정할 수 있는 정보를 수집하지 않았다. 따라서 개별 대학별로 학생들의 설문 참여 비율은 확인할 수 없다. 각 대학은 각자의 임상 실습 경험을 제공하기 때문에 특정 대학 학생들이 설문에 많이 참여한 경우 결과 해석은 제한적일 수 있다. 즉, 본 연구의 결과는 임상 실습에 대한 전국 치과대학 학생들의 전반적인 평가로 이해하기보다는 학생들의 임상 경험에 따른 임상 실습 평가에 대한 차이를 중심으로 이해할 필요가 있다.

결론적으로, 임상 실습 유형 및 평가 등급에 따라 학생들은 평가에 대한 공정성, 정확성에 대한 인식이 유의미하게 차이가 났다. 임상 실습 전 기간을 블록형(전공과 중심)으로 설계된 임상 실습 교육보다는 장기통합 실습에서 평가에 대한 인식도가 높게 나타났다. 등급제보다는 급락제 학생들이 평가에 대한 인식도가 높았다.

전국 11개 치과대학, 치의학대학원은 학교의 미션과 비전 및 환경에 따라 최적의 임상 실습 교육 과정을 선택하였으며 이를 성실하게 수행하기 위해 각 학교 당국자들은 사회의 변화 및 학생들의 학습 요구에 따라 학교 교육과정을 개선시키며 다양한 노력을 기울이고 있다. 본 연구는 향후 각 학교의 교육과정 개선 작업 시 참고자료로 활용할 수 있을 것이다.

참고문헌

1. 박주현, 김선. 일개 의과대학 임상 실습 교육과정 평가. *한국의학교육*. 2004;16(2):157-67.
2. Cooke M, Irby DM, O'Brien BC. Educating physicians: a call for reform of medical school and residency: John Wiley & Sons: 2010.
3. Lucey CR. Medical education: part of the problem and part of the solution. *JAMA Internal medicine*. 2013;173(17):1639-43.
4. Green M, Jones P, Thomas JX, Jr. Selection criteria for residency: results of a national program directors survey. *Acad Med*. 2009;84(3):362-7.
5. Watling CJ, Ginsburg S. Assessment, feedback and the alchemy of learning. *Med Educ*. 2019;53(1):76-85.
6. Hernandez CA, Daroowalla F, LaRochelle JS, Ismail N, Tartaglia KM, Fagan MJ, et al. Determining Grades in the Internal Medicine Clerkship: Results of a National Survey of Clerkship Directors. *Acad Med*. 2021;96(2):249-55.
7. Reed DA, Shanafelt TD, Satele DW, Power DV, Eacker A, Harper W, et al. Relationship of pass/fail grading and curriculum structure with well-being among preclinical medical students: a multi-institutional study. *Acad Med*. 2011;86(11):1367-73.
8. Dyrbye L, Shanafelt T. A narrative review on burnout experienced by medical students and residents. *Med Educ*. 2016;50(1):132-49.
9. Bullock JL, Lai CJ, Lockspeiser T, O'Sullivan PS, Aronowitz P, Dellmore D, et al. In pursuit of honors: A multi-institutional study of students' perceptions of clerkship evaluation and grading. *Academic Medicine*. 2019;94(11S):S48-S56.
10. Duffield K, Spencer J. A survey of medical students' views about the purposes and fairness of assessment. *Medical education*. 2002;36(9):879-86.
11. Schopper H, Rosenbaum M, Axelson R. 'I wish someone watched me interview:' medical student insight into observation and feedback as a method for teaching communication skills during the clinical years. *BMC medical education*. 2016;16:1-8.
12. Howley LD, Wilson WG. Direct observation of students during clerkship rotations: a multiyear descriptive study. *Academic Medicine*. 2004;79(3):276-80.
13. Hauer KE, Lucey CR. Core clerkship grading: the illusion of objectivity. *Academic Medicine*. 2019;94(4):469-72.
14. Mennin SP, Kalishman S. Student assessment. *ACADEMIC MEDICINE-PHILADELPHIA-*. 1998;73:S46-S54.
15. de Jonge LP, Timmerman AA, Govaerts MJ, Muris JW, Muijtens AM, Kramer AW, et al. Stakeholder perspectives on workplace-based performance assessment: towards a better understanding of assessor behaviour. *Advances in Health Sciences Education*. 2017;22:1213-43.
16. Gingerich A, Kogan J, Yeates P, Govaerts M, Holmboe E. Seeing the 'black box' differently: assessor cognition from three research perspectives. *Medical education*. 2014;48(11):1055-68.
17. Bullock JL, Seligman L, Lai CJ, O'Sullivan PS, Hauer KE. Moving toward mastery: changes in student perceptions of clerkship assessment with pass/fail grading and enhanced feedback. *Teaching and Learning in Medicine*. 2022;34(2):198-208.
18. 임은정, 장봉현. 의과대학 및 의학전문대학원 임상 실습 수업평가 현황 분석. *Korean Journal of Medical Education*. 2011;23(2):119.
19. 고석봉, 박혜진. 일개 의과대학 학생의 필기, 임상 실습, 임상 수행능력 평가 및 졸업시험 성적 간의 상관성 분석. *Korean J Med Educ*. 2009;21(4):347-52.
20. 김순구, 이애화, 황일선. 일개 의과대학에서의 임상 실습 교육과정 요구도 분석. *인문사회* 21. 2022;13(6):63-76.
21. Wu M-J, Zhao K, Fils-Aime F. Response rates of online surveys in published research: A meta-analysis. *Computers in Human Behavior Reports*. 2022;7:100206.
22. 전지은 임. 치과의사 인력 수급의 현황과 전망. *치과의료정책연구원 ISSUE Report*. 2023;2023-09호:1-11.
23. Teherani A, Irby DM, Loeser H. Outcomes of different clerkship models: longitudinal integrated, hybrid, and block. *Acad Med*. 2013;88(1):35-43.
24. Horst JA, Clark MD, Lee AH. Observation, Assisting, Apprenticeship: Cycles of Visual and Kinesthetic Learning in Dental Education. *Journal of Dental Education*. 2009;73(8):919-33.
25. Gordon J, Hazlett C, Ten Cate O, Mann K, Kilminster S, Prince K, et al. Strategic planning in medical education: enhancing the learning environment for students in clinical settings. *Medical Education*. 2000;34(10):841-50.
26. Gupta S, Howden S. Insights into post-longitudinal integrated clerkship experience: medical students' perceptions of transition and learning. *Education for Primary Care*. 2021;32(4):211-8.
27. Norris TE, Schaad DC, DeWitt D, Ogur B, Hunt DD. Longitudinal integrated clerkships for medical students: an innovation adopted by medical schools in Australia, Canada, South Africa, and the United States. *Academic Medicine*. 2009;84(7):902-7.
28. Hauer KE, O'Brien B, Poncelet AN. Longitudinal, integrated clerkship education: better for learners and patients. *Academic Medicine*. 2009;84(7):821.