

Peer assessment of small-group presentations by medical students and its implications

Sunmi Yoo¹, Kayoung Lee¹, Sang Heon Lee², HyeRin Roh³, Jong Tae Lee⁴, Byoung Doo Rhee² and Ikseon Choi⁵

Departments of ¹Family Medicine, ²Internal Medicine, ³Medical Education, ⁴Occupational and Environmental Medicine, Inje University College of Medicine, Busan, Korea, and ⁵Learning, Design, and Technology Program, The University of Georgia College of Education, Athens, GA, USA

의과대학생들의 소그룹 발표에 대한 동료평가의 의의

인제대학교 의과대학 ¹가정의학교실, ²내과학교실, ³의학교육학교실, ⁴예방의학교실, ⁵Learning, Design, and Technology Program, The University of Georgia College of Education

유선미¹, 이가영¹, 이상현², 노혜린³, 이종태⁴, 이병두², 최익선⁵

Purpose: The purpose of this study was to explore the relationships among medical students' assessments on peers' group presentations, instructors' assessments of those presentations, and students' educational achievements in other assignments and tests.

Methods: A total of 101 first-year students from a medical school participated in the study. The students' educational achievements in a 4-week long integrated curriculum were analyzed. Student's final grades were comprised of the following education criteria: two written tests (60%), 15 group reports (25%), one individual report (7%), and four group presentations (15%). We compared scores of the group presentation assessed by the peers and the two instructors. Furthermore, we compared peers' assessment scores with each component of the evaluation criteria.

Results: Pearson correlation analysis showed significant correlaton for the assessments between peers and instructors ($r=0.775$, $p<0.001$). Peer assessment scores also correlated significantly with scores for the group assignments ($r=0.777$, $p<0.001$), final grades on the curriculum ($r=0.345$, $p<0.001$), and scores for individual assignments ($r=0.334$, $p<0.001$); however, no significant correlation was observed between the peer-assessed group presentation scores and the two written test scores.

Conclusion: Peer assessments may be a reliable and valid method for evaluating medical students' performances in an integrated curriculum, especially if the assessments are used to academic processes, such as presentations, with explicit evaluation and judgment criteria. Peer assessments on group presentations might assess different learning domains compared to written tests that primarily evaluate limited medical knowledge and clinical reasoning.

Key Words: Peer review, Self-evaluation programs, Teaching, Group processes

Received: January 18, 2014 • Revised: February 10, 2014 • Accepted: February 10, 2014

Corresponding Author: Sunmi Yoo (<http://orcid.org/0000-0001-7407-8423>)

Department of Family Medicine, Inje University Haeundae Paik Hospital, 875 Haeun-daero, Haeundae-gu, Busan 612-862, Korea

Tel: +82.51.797.3220 Fax: +82.51.703.0434 email: syoo@paik.ac.kr

Korean J Med Educ 2014 Mar; 26(1): 31-40.

<http://dx.doi.org/10.3946/kjme.2014.26.1.31>

pISSN: 2005-727X eISSN: 2005-7288

© The Korean Society of Medical Education. All rights reserved.

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

서론

의과대학 교육을 통해 학생이 어느 정도의 수행능력을 가지게 되었는지 평가하는 방법은 다양하다. 학생을 지도하는 임상교육자에 의한 주관적인 평가, 지식과 문제해결능력을 평가하는 선다형 시험, 진찰 및 의사소통 기술을 주로 평가하는 표준화 환자 평가법 등이 흔히 사용된다[1]. 의사의 전문적인 수행능력은 의학적 지식, 환자 진료, 직업전문성, 의사소통 및 대인관계기술, 임상에 근거를 둔 학습 및 개선활동, 업무 체계 중심의 진료 등 다양한 영역을 포괄하기 때문에[1], 의과대학생의 수행능력을 평가할 때도 지식이나 기술뿐 아니라 대인관계 및 인간적인 자질에 대한 평가가 중요하다. 동료나 환자 등 다양한 사람으로부터 피드백을 받는 360도 평가가 대인관계 및 인간적인 자질에 대한 평가에 도움이 된다고 알려져 있다[1,2]. 이 중 동료평가(peer assessment)는 학습자가 비슷한 상태에 있는 다른 학습자의 결과물이나 수행의 수준, 가치, 질 등을 평가하는 것으로[3], 학습자의 책무성을 강화하며 학습자들이 소그룹 학습에 이바지하도록 독려하여 효과적인 구성원이 되는 데 도움이 된다[4].

의사로서 동료 의료인에 대한 객관적인 평가는 필수적인 역량이지만 우리나라의 문화에서는 직관적으로 동료의 성취도를 평가하는 데 저항이 크며 의과대학생들이 동료평가에 대한 수련을 받을 기회는 별로 없다. 이 때문에 지금까지 국내에서 시행된 동료평가 결과에 따르면 학생들이 심리적 부담감 때문에 변별력 없이 동일한 점수를 부여하거나, 객관적인 기준이 아닌 전반적인 인상에 기초하여 동료들을 평가하는 등 결과의 신뢰성과 실효성에 대한 우려의 목소리가 높았다[5,6]. 이러한 우려를 줄이기 위해서는 팀별 발표와 같은 특정 활동에 제한한 동료평가를 시행하되 미리 정형화된 평가 기준을 제시하고 이 결과가 실제 학생들의 학업성취도에 도움을 주도록 피드백을 제공하는 연습 과정이 도움이 될 것이다.

외국의 경우 적절한 시기에 익명성이 보장된 상태로 제공한 동료평가는 학습자들에게 유익하고 교육적으로 의미있는 피드백이 된다는 보고가 많이 있다. 동료평가 결과는 최종 학업성취도나 감독교육자의 평가를 잘 예측하며 시간이 경과하여 학년이 바뀌어도 일정하다고 알려져 있다[1]. 국내에서는

일부 의과대학에서 문제바탕학습이나 일개 과목의 실습에서 부분적으로 동료평가를 적용한 결과가 보고되었으나[5,7,8], 동료평가의 신뢰도를 평가하거나 다른 학습성취도 평가법 간의 연관성에 대한 분석은 별로 없었다. 또 임상실습 후에 감독 교육자로부터 받는 총평이나 표준화 환자가 학생의 업무수행능력을 총평 점수로 평가하는 방법은 의학교육에서 유용하게 사용되고 있으나[1] 동료평가에서 총평 점수를 이용한 평가의 유용성은 별로 연구되지 않았다.

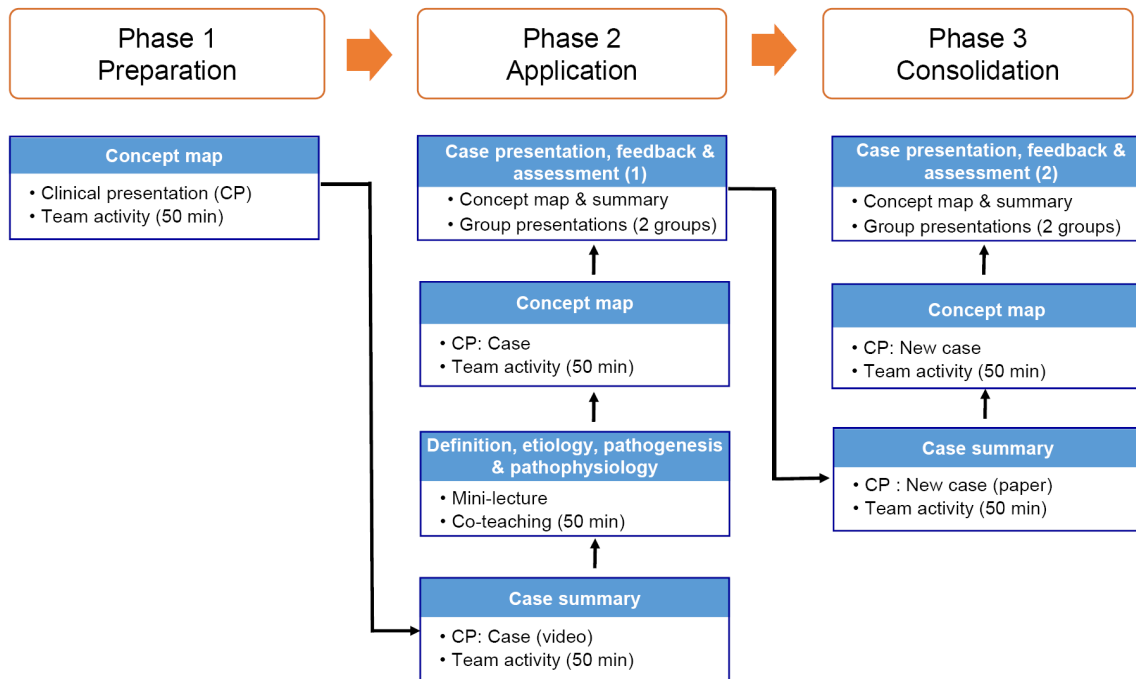
따라서 본 연구에서는 인제대학교 의과대학 1학년을 대상으로 시행한 <임상표현과 병태생리> 통합과정에서 소그룹 활동 결과물 발표에 대한 동료평가와 다양한 학습성취도 평가의 상관관계를 알아보고자 한다. 구체적인 연구 목표는 첫째, 소그룹 활동 결과 발표에 대한 학생들의 동료평가는 교수평가와 비교하여 신뢰도를 확보하는지 확인하고, 둘째, 발표에 대한 동료평가에서 총평 점수는 구체적인 항목에 대한 평가 점수의 총점과 상관성이 있는지 알아보고, 셋째, 발표에 대한 학생들의 동료평가가 학생들의 수행능력을 다면적으로 평가하는 다른 평가결과와 어느 정도 일치하는지 알아보는 것이다.

연구 대상 및 방법

1. 연구 대상

본 연구는 2013년 1학기에 개설된 <임상표현과 병태생리> 통합과정에 참가한 의학과 1학년 101명을 대상으로 하였다. 4주간 진행된 통합과정은 미리 정해진 15개의 임상표현을 다루도록 계획되었고 매일 한 개의 임상표현에 대해 준비-적용-근거기 과정을 통해 학습이 진행되었다(Fig. 1). 준비과정에서 학생들은 6~7명으로 구성된 15개 조로 나누어 조별로 자율 토의를 하면서 주어진 임상표현에 대한 기초적인 개념지도를 작성하였다. 적용과정에서는 비디오로 제시된 증례를 시청한 뒤 해당 임상표현의 정의, 원인, 발병기전 및 병태생리에 대한 강의를 들었다. 이후 조별로 증례를 요약하고 임상표현의 병태생리를 설명하는 개념지도를 작성하는 자율학습을 시행하고 무작위로 선정된 2개 조가 학습결과를 발표하고 교수 및 다른 조 학생들의 피드백을 받았고, 각 발표가 끝나면 발표하는 조

Fig. 1. Daily Process of Three Phases of Learning and Peer Evaluation in the Curriculum



외의 모든 학생과 담당 교수 2인이 미리 배포한 평가지를 이용하여 발표 결과를 평가하였다. 굳히기과정에서는 같은 임상표현에 대한 다른 증례가 제시되었고 이에 대해 다시 한 번 조별 토의를 하면서 증례 요약 및 개념지도를 수정 보완하는 자율 학습을 시행한 뒤 무작위로 선정된 2개 조가 이를 발표하였고, 발표에 대한 평가는 같은 방식으로 이루어졌다.

2. 조사 도구 및 분석 방법

1) 학습성취도 평가

통합과정의 학습성취도 평가는 의학적 지식 및 임상추론 능력을 평가하는 지필시험(2회, 각 30%), 임상 정보를 요약하고 해석하는 능력을 평가하는 구조화된 개인별 보고서(7%), 전문적인 자질, 대인관계, 팀워크 등의 태도를 반영하는 조별 보고서(25%)와 조별 발표 점수(15%)를 합산하고 결석이나 지각을 한 경우 총점에서 감점하는 방식으로 시행하였다. 각 조의 구성원은 조별 보고서와 조별 발표 결과에 따라 동일한 점수를 받았다.

2) 발표 평가

각 조는 통합과정 기간 동안 4회 발표하였고 조별 발표의 점수는 전체 학업성취도 평가 점수의 15%를 반영하였으므로,

1회당 3.75%의 발표 점수를 합산하여 조별로 동일한 발표 점수를 받았다. 증례 요약 및 개념지도를 평가하는 평가지는 증례 요약, 발표자의 이해도, 개념지도에 대한 평가, 발표 준비 및 시간 엄수, 토의 유도 등을 묻는 Likert식 4점 척도 문항 8개와 조별 발표를 전반적으로 평가하는 총평 문항 1개(5점 척도)로 이루어져 있다(Appendix 1). 사용한 평가지의 신뢰도(Cronbach α 계수)는 0.84였다. 통합과정의 첫 2주간은 평가지의 평가 문항을 전산화한 프로그램에 개인별 컴퓨터를 이용하여 평가 점수를 입력하였고, 강의실 사정 때문에 개인별 컴퓨터를 사용할 수 없었던 마지막 2주간은 개인별로 부여 받은 키패드에 평가 점수를 입력하였다.

평가지의 9개 평가 문항의 점수를 합산하고 이를 3.75%에 대한 상대 점수로 환산한 뒤, 발표하는 조 외의 모든 학생이 평가한 점수의 평균을 학생들이 평가한 조별 발표에 대한 동료평가 점수로 하였다. 이와는 별도로 담당교수들은 같은 평가지를 이용하여 발표에 대한 평가를 하였고, 학생들이 평가한 발표평가 점수와 교수들이 평가한 발표평가 점수의 평균을 최종 조별 발표 점수로 부여하였다. 통합과정 기간 동안 15개 팀이 4회씩 발표를 하였고 발표하는 조는 평가에서 제외하였으므로 학생들은 총 56회의 동료평가에 참여하였다.

3) 통계 분석

평가 문항의 점수를 합산하여 점수를 구하였으므로 한 항목이라도 응답하지 않은 경우는 분석에서 제외하였다. Pearson 상관분석을 통하여 조별 발표에 대한 동료평가와 교수평가 점수의 연관성을 분석하였다. 발표에 대한 평가 점수의 총점과 발표를 전반적으로 평가하는 1개 문항과의 연관성을 Pearson 상관관계 계수를 구하여 분석하였다. 조별 발표에 대한 동료평가 점수와 통합과정의 다른 학습성취도 결과 사이의 Pearson 상관관계 계수를 구하여 연관성을 분석하였다. 통계 분석에는 MedCalc version 12.7.2 (MedCalc Software, Mariakerke, Belgium)를 이용하였고 유의수준 0.05를 기준으로 양측검정을 실시하였다.

군 10.18이었고, 최댓값 11.27, 최솟값 9.75로 범위는 1.51이었다. 두 교수들이 평가한 발표 점수는 평균 10.74였고, 최댓값 12.20, 최솟값 9.46으로 범위는 2.74였다. 상관분석 결과 조별 발표에 대한 학생들의 평가 점수와 교수 평가의 상관계수는 0.775로 통계적으로 유의하였다(Table 2).

동료평가 중에서 학생들이 조별 발표를 전반적으로 평가하는 총평 점수의 평균은 13.38이었고, 최댓값 15.00, 최솟값 12.55로 범위는 2.45였다. 이 문항에 대한 교수들의 평가 점수는 평균 13.78이었고, 최댓값 15.34, 최솟값 12.39, 범위는 2.95였다. 상관분석 결과, 동료평가 중 총평 점수는 조별 발표에 대한 평가 점수의 총점과 상관계수 0.984의 유의한 상관관계를 보였고, 교수들의 총평 점수와도 상관계수 0.637로 유의한 상관관계를 보였다(Table 2). 평가 점수의 총점 및 총평 점수를 이용한 동료평가에서 최댓값을 받은 조는 교수평가에서도 최댓값을 받았다. 평가 점수의 총점을 이용한 동료평가에서 최솟값을 받은 조는 학생들의 총평 점수와 교수평가에서도 최솟값을 받았지만, 평가 점수의 총점을 이용한 교수평가에서는 최솟값을 받지 않았다(Table 1).

결과

1. 조별 발표에 대한 동료평가 결과

총 4회에 걸친 발표에 대한 동료평가 결과는 Table 1과 같다. 학생들이 평가한 조별 발표에 대한 평가 점수의 총점은 평

Table 1. Summed Scores and Global Assessment with One-Item Question for Group-Presentations in an Integrated Course of Medical Students

Group	Summed score for presentation by students	Summed score for presentation by instructors	Global assessment by students	Global assessment by instructors
1	10.01	10.56	13.15	13.59
2	10.42	11.44	13.65	15.18
3	10.37	11.22	13.69	13.95
4	10.20	11.35	13.54	14.45
5	10.09	11.26	13.36	14.67
6	10.01	10.25	13.13	13.30
7	10.26	10.76	13.60	13.16
8	9.94	10.78	12.81	14.11
9	9.94	9.46	12.94	12.39
10	10.33	10.86	13.67	13.74
11	11.27	12.20	15.00	15.34
12	9.95	10.68	13.02	14.26
13	10.07	9.84	13.17	12.91
14	10.14	10.44	13.40	13.10
15	9.75	10.05	12.55	12.49
Mean ± SD	10.18 ± 0.35	10.74 ± 0.70	13.38 ± 0.56	13.78 ± 0.90
Possible highest score	15.0	15.0	20.0	20.0

SD: Standard deviation.

Table 2. Correlations of Summed Scores and Global Assessment with One-Item Question for Group-Presentations between Peers and Instructors

	Summed score for presentation by students (n=15)		Summed score for presentation by instructors (n=15)		Global assessment by students (n=15)		Global assessment by instructors (n=15)	
	Correlation coefficients	p-value	Correlation coefficients	p-value	Correlation coefficients	p-value	Correlation coefficients	p-value
Summed score for presentation by students	1		0.775	0.001	0.984	<0.001	0.644	0.010
Summed score for presentation by instructors			1		0.784	0.001	0.917	<0.001
Global assessment by students					1		0.637	0.011
Global assessment by instructors							1	

The unit of this analysis was each group and the analysis used the final summed scores for each group from total 60 presentations.

Table 3. Correlation between Peer Assessment Scores of the Group-Presentation and Other Components of Assessment

	Group presentation		Group assignment		Individual assignment		First written examination		Second written examination		Final grade	
	Correlation coefficients	p-value	Correlation coefficients	p-value	Correlation coefficients	p-value	Correlation coefficients	p-value	Correlation coefficients	p-value	Correlation coefficients	p-value
Group presentation	1		0.777	<0.001	0.334	0.001	0.168	0.092	0.189	0.059	0.345	<0.001
Group assignment			1		0.360	<0.001	0.221	0.026	0.243	0.014	0.402	<0.001
Individual assignment					1		0.365	<0.001	0.369	<0.001	0.574	<0.001
First written examination							1		0.762	<0.001	0.904	<0.001
Second written examination									1		0.911	<0.001
Final grade											1	

The unit of this analysis was each student (n=101).

2. 발표에 대한 동료평가와 다른 학습성취도의 연관성

이 통합과정의 학습성취도는 2회의 지필시험, 조별 보고서, 개인별 보고서, 조별 발표 점수를 합산하여 평가하였다. 상관분석 결과 발표에 대한 동료평가 점수는 조별 보고서 점수와 연관성이 높았고($r=0.777, p<0.001$), 총점($r=0.345, p<0.001$), 개인별 보고서($r=0.334, p<0.001$)와도 유의한 상관관계를 보였다. 반면에 2회의 지필시험은 조별 발표에 대한 동료평가 점수와 낮은 상관성을 보였고 이는 통계적으로 유의하지 않았다 (Table 3).

한편 학습성취도 총점은 2회의 지필시험과 가장 높은 상관성을 보였지만($r=0.904, r=0.911$), 발표에 대한 동료평가를 비롯한 조별 보고서($r=0.402, p<0.001$), 개인별 보고서($r=0.574, p<0.001$)와도 유의한 상관성을 보였다.

고찰

본 연구 결과, 의과대학 1학년이 참여한 일개 통합과정에서 소그룹 토의학습 활동 후 발표에 대한 학생들의 동료평가 결과는 교수들의 평가 결과와 높은 상관관계를 보였다. 또한 발

표에 대한 동료평가는 2회의 지필고사 결과와는 상관관계가 높지 않았지만, 통합과정 성적, 개인과제물, 조별 과제물 평가 결과와 높은 상관관계를 보였다. 따라서 발표에 대한 학생들의 동료평가는 신뢰성 있는 평가 도구로 생각되며, 의학 지식 및 임상추론능력 등에 관한 개인적인 인지능력 혹은 암기 능력 등만을 평가할 수 있는 지필 고사와는 달리, 전문지식과 이를 적절하게 소통하며 협업할 수 있는 보다 포괄적인 수행 능력을 예측할 수 있는 도구로 사용될 수 있다고 생각된다.

의과대학에서 동료평가를 학생의 학습성취도 평가에 이용하려면 동료평가의 정확도가 확보되어야 한다. 동료평가의 정확도를 평가하는 간편한 방법은 교수들의 평가 결과와 비교하여 상관관계를 보는 것인데 외국의 연구에서는 일정한 결과를 보이지 않았다. 예를 들어 의과대학 1학년에서 이루어진 소그룹 증례중심 교육에서 동료평가는 교수들의 평가와 유의한 상관관계를 보였고 학생들은 동료평가에 대해 호의적이었다[9]. 그러나 problem-based learning tutorial에서 여러 차례에 걸쳐 시행한 동료평가 점수는 같이 시행한 교수평가에 비해 계속 높은 경향을 유지하였으며 동료평가와 자기평가의 상관계수($r=0.806$)에 비해 동료평가와 교수평가의 상관계수($r=0.456$)가 더 낮았다는 연구도 있다[10]. 동료평가의 정확도에 대한 메타분석에 따르면 동료평가와 교수평가의 상관관계는 0.14~0.99로 매우 이질적이나 평균 상관계수는 0.69로 동료평가는 전반적으로 교수평가와 일치하였다. 특히 임상 술기와 같은 전문적인 기술을 평가하는 경우(평균 상관계수 0.54)보다 학문적인 성과물(에세이, 과제물 등; 평균 상관계수 0.75)이나 학문적인 과정(구두 발표, 팀 학습 과정에 참여도 등; 평균 상관계수 0.83)을 평가하는 경우에 상관계수가 더 높았다. 평가 도구 면에서 볼 때, 구체적인 기준을 별도로 제시한 경우(평균 상관계수 0.53) 보다는 구체적인 기준을 제시하면서 전반적인 평가를 하거나(평균 상관계수 0.85) 구체적인 기준을 제시하지 않고 전반적인 평가를 할 때(평균 상관계수 0.72) 상관계수가 높았다[11]. 본 연구는 학문적인 과정을 평가하는 구두 발표에 대하여 구체적인 기준을 제시하면서 전반적인 평가를 하였기 때문에 동료평가와 교수평가의 상관계수가 높았던 것으로 생각된다. 본 연구에서는 총 60회의 동료평가 결과를 학습성취도 평가에 반영하여 점수를 계산하는 과정이 복잡하였는데 9개 문항을 이용한 학생들의 동

료평가 결과가 총평 점수의 평가 결과와 연관성이 매우 높았던 점($r=0.984$)을 감안하여 총평 점수만으로 동료평가를 한다면 과정을 단순화하는 동시에 평가의 정확도도 유지할 수 있을 것이다.

외국의 연구에서 동료평가는 전통적으로 학생 평가에 사용되었던 지필고사 결과와 약한 상관관계를 보이는 데 비해 [12,13], 임상수행능력 평가와는 유의한 상관관계를 보였다 [13,14]. 이는 지필고사가 개인의 지식과 인지능력을 주로 평가하는 데 비해 동료평가는 개인적인 특성이나 팀 프로세스를 반영하기 때문이다. 팀바탕학습에서 동료평가를 적용한 국내 연구에서도 동료평가 점수는 팀바탕학습 점수와 유의한 상관관계를 보이며 동료평가의 결과가 성적에 반영되어야 한다는 것에 긍정적으로 인식하고 있는 학생들이 많았고 이들이 좀 더 솔직하고 성실한 태도로 평가에 임하는 경향이 있었다[8]. 본 연구에서도 발표에 대한 동료평가는 2회에 걸친 지필고사와는 유의한 상관관계를 보이지 않았지만 조별 보고서와 높은 상관관계를 보였고, 개인별 보고서나 총점과도 유의한 상관관계를 보였다. 따라서 발표에 대한 학생들의 동료평가는 전통적인 지필고사로 평가할 수 없는 학습 과정에 대한 평가에 유용하게 사용될 수 있다고 생각된다. 본 연구에서는 발표에 대한 결과물만 평가하였지만 팀 수행 결과에 대한 개인의 기여도는 학습 과정에 대한 평가 외에도 개인의 학습 책임성, 팀의 응집력 등에 영향을 주고 나아가 대인 관계 능력이나 전문성 발달에도 연관이 있다고 알려져 있으므로 앞으로 발표에 대한 개인의 기여도를 별도로 평가하는 연구가 필요하다.

본 연구에서 발표 평가에 사용한 평가서는 증례 요약, 발표자의 이해도, 스키마에 대한 평가, 발표 준비 및 시간 엄수, 토의 유도 등을 4점 척도로 평가하게 하였다. 이 평가 항목은 학습자가 발표를 통하여 성취해야 할 분야를 제시한 것으로 일종의 평가 준거(criterion)가 된다. 이에 대한 내용 타당도(content validity), 구인 타당도(construct validity), 준거 관련 타당도(criterion-related validity) 등을 구하지 않은 점은 본 연구의 제한점이다. 동료평가에서 사용된 도구가 이러한 psychometric assessment를 거치지 않아 정확도나 신뢰도 면에서 문제가 된다는 점은 다른 메타분석에서도 지적되었다[15]. 그러나 평가지의 신뢰도(Cronbach α 계수)는

0.84로 높았고, 통합과정을 시행하는 15일간 평가지의 신뢰도를 일별로 구해 보았을 때, 시간이 경과함에 따라 신뢰도 계수가 점점 증가하는 경향을 보이는데(0.74에서 0.91로 향상)(Appendix 2), 학생들의 동료평가 능력도 경험 축적에 따라서 향상되어 시간이 경과함에 따라 평가의 일관성이 더 높아졌다고 해석된다. 동료평가를 할 때 학생들이 심리적 부담감 때문에 동료 전체에게 동일한 점수를 부여하거나, 객관적인 기준이 아닌 평소의 인상에 근거하여 동료들을 평가하는 것도 흔히 지적되는 동료평가의 한계점이다[5]. 본 연구에서 동료에게 일괄적으로 같은 점수를 부여하는 경우는 없었지만 동료평가에 의한 점수의 범위는 1.51, 표준편차는 0.35로 차이가 크지 않아 변별력이 크지는 않았다. 학생들은 통합과정이 끝난 뒤 실시한 학생들의 강의 만족도 평가에서 발표 평가를 할 때 동료의 눈치가 보인다고거나 동료평가가 비효율적이라는 문제를 제기하기도 하였다(자료는 제시하지 않음). 그러나 동료평가의 점수 차이가 크지 않았기 때문에 학생들이 비교적 공정하게 평가를 했을 가능성도 있다. 학생들은 동료들이 자신에게 내린 평가에 대하여 신뢰성이 떨어진다고 생각하거나 평가자로서 자신의 능력과 과정의 공정성에 의문을 제기하기도 한다. 학습의 내용이나 수업 중의 기여도에 따라 다르겠지만 국내 의과대학의 의료윤리 과정에서 조사한 바에 따르면 학생들은 동료평가의 결과를 성적에 반영하는 경우 전체 성적의 24.5% 정도가 적당하다고 생각하였다[16]. 본 연구에서 동료평가는 전체 성적에서 15%를 차지하여 학생들은 평가에 대한 부담이 크지 않았다고 생각된다. 동료평가를 시행하는 학생들이 부담을 적게 느끼는 방법은 어떤 것인지, 동료평가를 통해 학생들이 구체적으로 어떤 이득을 얻게 되었는지 등에 대한 의견을 듣는다면 이러한 제한점을 극복하는데 도움이 될 것이다.

결론적으로 의과대학생을 대상으로 한 통합과정에서 소그룹 토의학습활동 후 발표에 대한 학생들의 동료평가 결과는 신뢰성 있는 평가 도구로 생각되며, 개인적 지식과 제한된 문제해결능력을 평가하는 선다형 시험과는 달리, 좀 더 포괄적인 수행능력을 평가하는 유용한 대안적 도구로 사용될 수 있다고 생각된다. 따라서 최근 의학교육의 방향이 결과 중심의 수행능력 강화에 초점을 두고 있다는 점을 감안할 때 동료평가의 중요성이 크다고 생각되며, 동료평가의 신뢰도와 타당

도 및 그 실용성을 높이는 추가적인 연구가 지속적으로 요구된다.

Acknowledgements: Thanks are due to Mi Young Hong and Su Jin Kim in the Office of Medical Education for their administrative assistance during the curriculum and in the preparation of this paper.

Funding: None.

Conflicts of interest: None.

REFERENCES

1. Epstein RM. Assessment in medical education. *N Engl J Med* 2007; 356: 387-396.
2. Dannefer EF, Henson LC, Bierer SB, Grady-Weliky TA, Meldrum S, Nofziger AC, Barclay C, Epstein RM. Peer assessment of professional competence. *Med Educ* 2005; 39: 713-722.
3. Topping KJ. Peer assessment. *Theory Pract* 2009; 48: 20-27.
4. Michaelsen LK, Parmelee DX, McMahon KK, Levine RE, editors. *Team-based learning for health professions education: a guide to using small groups for improving learning*. Kim S, Park JH, Yoo NJ, Lee SJ, translators. Seoul, Korea: Academy Press; 2007. p 137-156.
5. Kim JY, Son HJ, Lee JH, Kim JH, Hong KP. Educational implication of peer evaluation in problem-based learning. *Korean J Med Educ* 2005; 17: 275-284.
6. Huh S. How to administer the peer evaluation in team-based learning. *Korean J Med Educ* 2012; 24: 359-361.
7. Han HS, Lee R, Sohn IS, Lee JM, Lee KY. A study on students' academic achievement and perception of team-based learning during a medical dermatopathology lecture. *Korean J Med Educ* 2007; 19: 257-263.
8. Cho AR, Han SI, Yoon SH, Park JH, Yoo NJ, Kim S.

- Methods of effective team-based learning administration and expected effects on medical education. *Korean J Med Educ* 2010; 22: 47-55.
9. Ferguson KJ, Kreiter CD. Assessing the relationship between peer and facilitator evaluations in case-based learning. *Med Educ* 2007; 41: 906-908.
 10. Machado JL, Machado VM, Grec W, Bollela VR, Vieira JE. Self- and peer assessment may not be an accurate measure of PBL tutorial process. *BMC Med Educ* 2008; 8: 55.
 11. Falchikov N, Goldfinch J. Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. *Rev Educ Res* 2000; 70: 287-322.
 12. Levine RE, Kelly PA, Karakoc T, Haidet P. Peer evaluation in a clinical clerkship: students' attitudes, experiences, and correlations with traditional assessments. *Acad Psychiatry* 2007; 31: 19-24.
 13. Arnold L, Willoughby L, Calkins V, Gammon L, Eberhart G. Use of peer evaluation in the assessment of medical students. *J Med Educ* 1981; 56: 35-42.
 14. Lurie SJ, Lambert DR, Nofziger AC, Epstein RM, Grady-Weliky TA. Relationship between peer assessment during medical school, dean's letter rankings, and ratings by internship directors. *J Gen Intern Med* 2007; 22: 13-16.
 15. Speyer R, Pilz W, Van Der Kruis J, Brunings JW. Reliability and validity of student peer assessment in medical education: a systematic review. *Med Teach* 2011; 33: e572-e585.
 16. Cheong YS. Experiences using peer assessment on small group discussions in medical ethics. *Korean J Med Ethics* 2012; 15: 70-77.

Appendix 2. The Reliability Indices (Cronbach α) of the Assessment Form to Evaluate Students' Group-Presentation on Daily Basis

Day	No. of students who participated in the assessments per day	Cronbach α
1	369	0.739
2	351	0.806
3	357	0.754
4	364	0.789
5	359	0.838
6	357	0.799
7	368	0.804
8	290	0.815
9	264	0.839
10	256	0.847
11	252	0.887
12	284	0.879
13	264	0.892
14	261	0.902
15	245	0.905