

Factors that influence student ratings of instruction

Su Jin Chae^{1,2}, Yun Hoon Choung³ and Yoon Sok Chung^{2,4}

¹Department of Medical Humanities & Social Medicine, ²Office of Medical Education, Departments of ³Otolaryngology and ⁴Endocrinology & Metabolism, Ajou University School of Medicine, Suwon, Korea

학생에 의한 강의평가에 영향을 미치는 요인 분석

아주대학교 의과대학 ¹인문사회의학교실, ²의학교육실, ³이비인후과학교실, ⁴내분비대사내과학교실

채수진^{1,2}, 정연훈³, 정윤석^{2,4}

Purpose: The purpose of this study was to examine the validity of student ratings of instruction by analyzing their relationships with several variables, including gender, academic rank, specialty, teaching time, and teaching method, at a medical school.

Methods: This study analyzed the student ratings of 297 courses at Ajou University School of Medicine in 2013. SPSS version 12.0 was used to analyze the data and statistics by t-test, analysis of variance, and Scheffe test.

Results: There were no statistically significant differences in student ratings between gender, rank, and specialty. However, student ratings were significantly influenced by teaching times and methods ($p < 0.05$). Student ratings were high for teaching times of 10 hours or more and small-group learning, compared with lectures. There was relatively mean differences in students ratings by teaching times, specialty and rank, although the difference in ratings was not statistically significant.

Conclusion: Student ratings can be classified by teaching time and method for summative purposes. To apply student ratings to the evaluation of the performance of faculty, further studies are needed to analyze the variables that influence student ratings.

Key Words: Student rating, Faculty evaluation, Medical students, Teaching

서론

우리나라에서 교수업적평가가 본격적으로 시작된 것은 1990년대 중후반이다. 1994년부터 시행한 대학종합평가인정제를 기점으로 정부의 대학평가를 강화하는 흐름은 각 대학들의 교수업적평가의 도입을 촉발하였다[1]. 연구, 교육, 봉사업적으로 구분하여 시행하는 교수업적평가 중 연구 업적은

쉽게 정량화 할 수 있는 영역인 반면, 교육 업적은 수업 시간 외의 정량적으로 평가하기가 쉽지 않다. 이러한 상황에서 학생에 의한 강의평가는 중요하게 부각되었다. 의과대학도 예외가 아니어서 2000년도에 들어 본격적으로 강의평가가 실시되었고[2], 2005년 2주기 평가인증에서 강의평가 시행이 기본 기준으로 포함되면서 모든 의과대학이 강의평가를 시행하고 있다[3].

강의평가 결과의 활용과 관련하여 대부분의 의과대학이 수

Received: January 6, 2015 • Revised: January 26, 2015 • Accepted: January 26, 2015

Corresponding Author: Yoon Sok Chung (<http://orcid.org/0000-0003-0179-4386>)

Department of Endocrinology and Metabolism, Ajou University School of Medicine, Suwon 443-721, Korea

Tel: +82.31.219.5127 Fax: +82.31.219.4497 email: yschung@ajou.ac.kr

Korean J Med Educ 2015 Mar; 27(1): 19-25.

<http://dx.doi.org/10.3946/kjme.2015.27.1.19>

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.

업개선을 위한 강의평가의 형성적 기능을 수행하고 있지만, 일부 대학에서는 인센티브, 승진, 우수 교수 선정 등의 총합적 목적으로 활용하고 있다[2,4,5,6]. 강의평가 점수를 교수업적 평가로 활용하는 대학의 경우 학생에 의한 강의평가 점수만으로 교수를 서로 비교하고 순위를 매기는 평가 방법에 대해서는 반대하는 의견도 많다[7]. 왜냐하면 학생들의 평가가 ‘강의’ 자체와는 무관한 다른 요인에 의해 영향을 받을 수 있기 때문이다. Chae & Lim의 연구[8]에 의하면, 일부 교수들은 수업시간이 많을수록, 새로운 수업 방법을 적용할수록, 수업 내용이 어려울수록, 임상의학 교수보다 기초의학 교수가 강의평가 점수가 낮게 나온다고 생각한다. 이러한 생각은 강의평가에 영향을 줄 수 있는 여러 요인을 고려하지 않고 일률적인 기준으로 평가하고 그 점수를 업적평가에 반영하는 것에 대한 불만이라고 볼 수 있다.

본 연구는 의과대학에서 학생에 의한 강의평가 점수를 교수업적평가에 반영하는 정책을 시행함에 있어 그 정당성을 뒷받침할 수 있는 실증적인 근거가 부족하다는 문제에서 시작한다. 의과대학은 다른 일반대학과 교수환경의 차이가 있다. 말하자면, 의과대학 교수들은 일반적으로 학생교육을 담당하는 수업시간의 수가 일반대학 교수들에 비해 현저하게 적다. 의과대학 교수는 기초의학과 임상의학으로 나누어져 있으며 임상의학 교수들은 교육, 연구, 봉사 이외의 진료라는 특수 임무를 띠고 있다. 의과대학은 수업의 형태가 매우 다양하다. 대부분이 실습과 체험을 중요시하는 교과목이며 학과 공동 컨퍼런스, 소집단 교육 등 양적으로 평가하기 어려운 다양한 형태의 교육활동이 많다. 의과대학은 problem-based learning (PBL), team-based learning (TBL), case-based learning (CBL), objective structured clinical examination, clinical performance examination 등 일반대학에 비해 교육 방법이 다양하다. 일반대학은 한 명의 교수가 한 과목을 한 학기동안 수업하지만, 의과대학은 한 개의 과정(course)을 여러 명의 교수가 함께 하는 팀티칭을 시행한다[9].

그 동안 의과대학은 강의평가에 대한 대학들의 자체적인 기초연구를 여러 편 수행해 왔으나[10,11,12,13], 여전히 강의평가의 신뢰성, 타당성, 효과성 등에 관한 연구나 논의는 크게 부족한 실정이다. 본 연구는 일반대학과 의과대학의 대학 환경의 차이가 있다는 것을 인정하고 의과대학 수업환경에서

교수 요인은 성별, 직급, 전공, 그리고 수업 요인은 수업 시간과 수업 방법 등을 선정하여 이 요인들이 의과대학 강의평가에 미치는 영향력을 실증적으로 분석하여 강의평가 결과를 객관적으로 해석할 수 있는 근거자료를 마련하는 데에 목적이 있다.

대상 및 방법

1. 연구 대상

본 연구는 아주대학교 의과대학에서 실시한 2013학년도 의학과 1학년과 2학년 통합교육을 수업한 교수들의 강의평가 결과를 분석 대상으로 하였다. 1년 동안 통합교육은 1학년 14개, 2학년 12개로 총 26개의 과정이 진행되었으며 총 319개의 강좌가 진행되었다. 분석은 교수업적평가 점수를 시스템에 입력하지 않는 특임교수와 외부강사의 22개 강좌를 제외하고 297개의 강좌를 최종 대상으로 하였으며, 297개 강좌에 대한 교수의 강의평가 점수를 분석하였다. 성별, 직급, 전공별로 구분한 강좌의 교수 현황은 Table 1과 같다.

26개 통합교육과정 수업은 총 2,139시간이었으며 그 중 강의를 제외한 다양한 수업, 예를 들어 실습이나 PBL, TBL 등 증례를 이용한 소그룹 수업이 440시간(21%) 진행되었다. 수업시간은 1년 동안 교수가 수업한 시간을 1~3시간, 4~9시간, 10~19시간, 20시간 이상 등 4개 구간으로 범주화하였다. 연구를 위해 구분한 수업 시간과 수업 방법 현황은 Table 2와 같다.

Table 1. Faculty Characteristics

Factor	No. (%)
Gender	
Male	214 (72.1)
Female	83 (27.9)
Academic rank	
Professor	159 (53.5)
Associate professor	64 (21.5)
Assistant professor	74 (24.9)
Specialty	
Basic science	83 (27.9)
Clinical science	214 (72.1)

Table 2. Course Characteristics

Factor	Times (hr)	Course	Basic science area	Clinical science area
Duration of teaching	1-3	115 (38.7)	20 (24.1)	95 (44.4)
	4-9	129 (43.4)	34 (41.0)	95 (44.4)
	10-19	43 (14.5)	19 (22.9)	24 (11.2)
	20 or more	10 (3.4)	10 (12.0)	0
Method of teaching	Lecture	204 (68.7)	45 (54.0)	159 (74.3)
	Experience ^{a)}	34 (11.4)	32 (39.0)	2 (0.9)
	Small group learning ^{b)}	54 (18.2)	3 (3.6)	51 (23.8)
	Variety of teaching method ^{c)}	5 (1.7)	3 (3.6)	2 (0.9)

Data are presented as number (%).

^{a)}Include experiment and laboratory course, ^{b)}Include problem-based learning (PBL), team-based learning (TBL), and case-based learning (CBL), etc., ^{c)}Mean teaching method of more than three types (CBL, PBL, TBL, lecture, practice, etc.).

Table 3. Mean Differences in Students of Ratings by Gender and Specialty

Factor	No.	Mean	SD	t	p-value
Gender				-0.77	0.44
Male	214	4.05	0.38		
Female	83	4.09	0.34		
Sum	297	4.07	0.37		
Specialty				-0.62	0.53
Basic	83	4.04	0.44		
Clinical	214	4.07	0.33		
Sum	297	4.07	0.37		

SD: Standard deviation.

아주대학교 의과대학 교수평가는 과정을 모두 마치고 실시하는 과정평가와 교수의 강의를 마치고 개별로 실시하는 강의 평가로 나뉜다. 과정평가는 의학교육실이 개발한 10개의 공통 문항을 사용하여 학생 참여, 학습 내용 수준, 학습 내용 양, 통합 정도, 수업 간 유기성, 강의자료, 자율학습, 시험, 교수준비도, 전체만족도 등을 평가하며, 과정평가의 결과는 책임교수와 강의교수 전체에게 전달된다. 반면 교수의 수업을 마치고 개별적으로 실시하는 강의평가 설문지는 교수의 열의 정도와 전달효과를 묻는 2개의 문항으로 구성되어 있으며 5점 척도로 이루어져 있다(5점, 매우 긍정; 1점, 매우 부정). 강의평가 결과는 교수 개인에게 전달되며, 교수업적평가에는 교수의 열의 정도와 전달효과 두 문항의 평균점수가 반영된다.

2. 자료 분석

연구의 자료는 SPSS version 12.0 (SPSS Inc., Chicago,

USA) 프로그램을 이용하여 분석하였다. 교수의 성별, 직급, 전공, 수업 시간, 수업 방법 등에 따라 강의평가 점수의 차이를 알아보기 위해서 기술통계, independent t-test, analysis of variance 분석을 실시하였고 집단 간 차이를 알아보기 위해서 사후 검증(Scheffe test)과 교차설계를 하였다. 통계적 유의수준은 모두 0.05 수준에서 검증하였다.

결과

1. 교수관련 요인 분석 결과

교수의 성별과 전공에 따른 강의평가 결과의 차이를 알아보기 위해서 t-검정을 실시한 결과, t 통계값이 각각 -0.77, -0.62로 유의수준 0.05에서 유의한 차이가 없었다(Table 3).

직급에 따라 강의평가 점수의 차이가 있는지 알아보기 위하여 분산분석을 실시한 결과, 교수, 부교수, 조교수 집단의 평균 차이에 대한 F 통계값이 1.61, 유의확률이 0.20으로 이 또한 통계적으로 유의한 차이가 없었다(Table 4).

2. 수업관련 요인 분석 결과

수업 시간에 따라 교수평가 점수에 차이가 있는지를 알아보기 위하여 교수의 수업 시간을 1~3시간, 4~9시간, 10~19시간, 20시간 이상 등 4개로 구분하고 분산분석을 실시한 결과,

F 통계값이 9.46, 유의확률은 0.00으로 수업 시간에 따라 교수점수에 유의한 차이가 있었다($p < 0.05$). Scheffe 사후비교 분석 결과, 1~3시간과 10~19시간 이상, 4~9시간과 10~19시간 구간에서 통계적으로 유의한 차이가 나타났다(Table 5).

수업 방법에 따라 강의평가 점수에 차이가 있는지를 알아보기 위하여 분산분석을 한 결과, F 통계값이 3.36, 유의확률의 0.01로 통계적으로 유의한 차이가 있었다($p < 0.05$). Scheffe 사후비교분석 결과, 강의와 소그룹수업 간에만 통계적으로 유의한 차이가 있을 뿐($p < 0.05$), 다른 방법들 간에는 유의미한

Table 4. Mean Differences in Students of Ratings by Academic Rank

Academic rank	No.	Mean	SD	F	p-value
Professor	159	4.09	0.33	1.61	0.20
Associate professor	64	4.03	0.44		
Assistant professor	74	4.01	0.37		
Sum	297	4.06	0.37		

SD: Standard deviation.

Table 5. Mean Differences in Students of Ratings by Duration of Teaching

Duration of teaching (hr)	No.	Mean	SD	F	p-value	Scheffe
1-3 ^{a1}	115	4.00	0.35	9.46	0.00	
4-9 ^{b1}	129	4.04	0.39			b) < c)
10-19 ^{c1}	43	4.25	0.29			a) < c)
20 or more ^{d1}	10	4.35	0.19			a) < d)
Sum	297	4.06	0.37			

SD: Standard deviation.

Fig. 1. (A, B) Mean Differences in Students of Ratings by Duration of Teaching, Specialty, and Rank

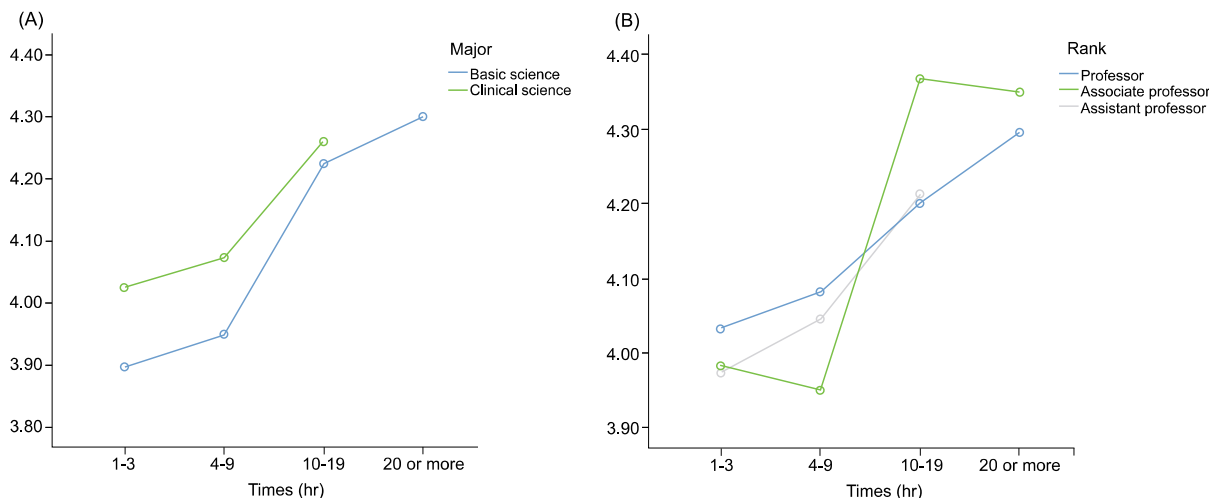


Table 6. Mean Differences in Students of Ratings by Method of Teaching

Method of teaching	No.	Mean	SD	F	p-value	Scheffe
Lecture ^{a)}	204	4.01	0.38	3.36	0.01	a)<c)
Experience ^{b)}	34	4.15	0.35			
Small group learning ^{c)}	54	4.17	0.31			
Variety of teaching method ^{d)}	5	4.15	0.26			
Sum	297	4.07	0.37			

SD: Standard deviation.

차이가 없었다(Table 4).

3. 교수와 수업관련 요인 교차분석 결과

교수의 전공과 직급과 강의평가 점수를 교차분석하여 상호 작용 효과를 구체적으로 나타낸 결과는 Fig. 1과 같다. 기초 의학 분야 교수가 임상의학 교수에 비해 통계적으로 유의미하지는 않지만 상대적으로 모든 시간에서 낮은 점수를 보였다. 직급의 경우, 4~9시간 구간에서는 교수, 조교수, 부교수 순으로 강의평가 점수를 보였으며, 10~19시간에서는 부교수가 조교수나 교수보다 통계적으로 유의미하지는 않지만($p > 0.05$) 상대적으로 높은 점수를 나타냈다(Table 6).

고찰

수업의 질은 강의교수가 얼마나 열성적으로 강의에 임하는가에 달려있다고 해도 과언이 아니다. 그 중요성에 부합하도록 많은 대학들이 학생들에 의한 강의평가 결과를 교수 업적에 반영하고 있다[14,15,16]. 그러나 교육을 담당하고 평가의 대상이 되는 교수들은 우선적으로 강의평가가 제대로 이루어져야 함을 지적한다. 본 연구는 의과대학에서 시행하는 강의평가에 대한 합리적인 정책을 결정하고 평가 결과를 객관적으로 해석할 수 있는 근거자료를 마련하기 위해서 한 대학의 1년간의 강의평가 점수와 교수의 성별, 직위, 전공, 수업 시간, 수업 방법 등의 요인들 간의 관계를 분석하여 학생에 의한 강의평가의 타당성을 검증하고자 하였다.

강의평가 결과에 영향을 미치는 교수와 수업 요인에 대한 선행 연구들의 결과에 대해서는 연구자들마다 상이하다. 교수 요인과 관련하여 Cashin [17]은 2,000개 이상의 논문을 메

타 분석한 결과, 나이, 경력, 성별, 인종, 성격, 연구 성과 등의 교수 요인과 강의평가 결과는 유의미한 관련이 없다고 결론 내렸다. 이 밖의 연구들[7,4]에서도 교수의 성별이나 경력과 강의평가 결과들에는 상관관계가 통계적으로 유의하지 못하다고 설명하고 있다. 반면 교수의 직위가 높아질수록, 교수 경력이 많을수록, 여자 교수보다는 남자 교수가, 공학보다는 인문사회 분야 교수의 강의평가 점수가 더 높다는 연구도 있다 [4,15].

수업 요인과 관련해서 Aleamoni & Yimer [18]는 교수의 교수법 활용과 강의평가 결과는 긍정적인 상관관계가 있음을 밝혔으며, Cranton & Smith [19]는 수업의 특성 중 수강 학생 수, 이론과 실습수업, 과제물 중요도, 시험의 중요도 등이 강의평가 결과와 유의미한 상관관계가 있다고 보고하였다. 그러나 의과대학 교수를 대상으로 강의평가 결과와 수업 시간과 수업 방법을 연계해서 분석한 연구는 거의 없다.

본 연구 결과, 강의평가 점수는 교수들의 성별, 직위, 전공에 따라 통계적으로 유의미한 차이를 보이지 않았지만, 수업 시간이나 수업 방법에서는 통계적으로 유의미한 차이가 나타났다. 수업시간은 10시간 이상에서, 강의가 아닌 증례를 이용한 소그룹 수업에서 교수들의 강의평가 점수가 높게 나타났다. 그리고 강의평가 점수와 교수 요인을 교차 분석한 결과, 기초의학 교수가 임상의학 교수에 비해 모든 수업시간 구간에서 상대적으로 낮은 점수를 받았으며, 직급의 경우에는, 교수는 4~9시간 구간에서 가장 낮은 점수를 받은 반면, 10~19시간 구간에서는 가장 높은 점수를 얻음으로써 시간에 따라 교수의 직급별 강의평가 점수가 상대적으로 차이가 있음을 알 수 있었다. 이러한 결과는 강의 자체 이외의 요인이 강의평가를 실행하는 과정에서 평가 결과에 영향을 주고 있다는 것을 시사한다.

본 연구가 일개 의과대학의 임상실습을 제외한 통합교육의 강의평가 결과에 대한 분석이라는 점에서 한계를 가질 수 있다. 본 연구에서 임상실습에 대한 교수평가를 제외한 이유는 임상실습에 대한 교수평가는 강의와는 전혀 다른 임상실습의 고유한 특성을 반영한 평가 도구가 별도로 마련되어 시행되어야 한다고 보았기 때문이다. 이러한 한계에도 불구하고 강의평가 결과에 대한 본 연구의 결과들은 의과대학의 교육업적평가 정책을 수립할 때 강의평가 점수를 어떻게 반영할 것인가에 대한 시사점을 줄 수 있을 것이다. 말하자면, 강의평가 결과는 교수의 성별, 직급, 전공보다는 교수의 강의를 구성하는 수업 시간이나 수업 방법 등 수업 요인에 의해 더 영향을 받을 수 있다. 따라서 강의평가 점수가 총합적 목적으로 활용될 때에는 수업의 시간이나 방법에 따라 구별할 필요가 있다.

의과대학에서 학생에 의한 강의평가를 시행하는 데 있어서 의과대학들이 공통으로 갖는 가장 큰 고민은 강의평가의 타당도와 신뢰도의 문제이다[2,20]. 낮은 타당도와 신뢰도의 평가 결과를 교수업적평가로 활용하는 것은 문제가 되며 잘못된 의사결정을 할 수도 있다. 강의평가 결과를 해석함에 있어서 학생들의 성실성에만 원인을 돌릴 것[20]이 아니라 강의평가의 여러 요인, 특히 의과대학의 특수성을 반영한 외적 요인들을 체계적으로 연구하여 다양한 평가지를 사용한다거나 교육업적 산정 시 요인에 따라 가중치를 두는 등의 합리적인 대책이 마련되어야 할 것이다.

Acknowledgements: None.

Funding: None.

Conflicts of interest: None.

REFERENCES

1. A nominal students' ratings for teaching [Internet]. Daily UNN; c2014 [updated 2014 September 22; cited 2014 September 23]. Available from: <http://news.unn.net/>.
2. Chae SJ, Lim KY. An analysis of course evaluation programs at Korean medical schools. *Korean J Med Educ* 2007; 19: 133-142.
3. Medical College Accreditation Standards [Internet]. Korean Institute of Medical Education and Evaluation; c2012 [updated 2012 April 16; cited 2014 September 14]. Available from: <http://www.kimee.or.kr/>.
4. Centra JA. Reflective faculty evaluation: enhancing teaching and determining faculty effectiveness. 1st ed. San Francisco, USA: Jossey-Bass; 1993. p 47.
5. Han SI, Kim HJ, Lee JY. A comprehensive study of Korean students' evaluations of university teaching. *J Educ Adm* 2005; 23: 379-403.
6. Kim SY, Park JY, Kim JC, Kang HS. Daehag hagsa gwajeong ganguipeonggae siltae bunseogeul tonghan gyoyugeobjeog pyeongga mohyeong yeongu. Seoul, Korea: Ministry of Education & Human Resources Development; 2001.
7. Feldman KA. Identifying exemplary teachers and teaching: evidence from student ratings. In: Perry RP, Smart JC, eds. *Effective teaching in higher education: research and practice*. New York, USA: Agathon Press; 1997. p 368-395.
8. Chae SJ, Lim KY. A comparison of student and faculty perspectives on course evaluation in a medical school. *Korean J Med Educ* 2008; 20: 163-167.
9. Ahn DS, Park KW, Lee YM, Lee YH, Jung HY, Chae SJ. Gugnaeoe uigwadaehag gyoyug teugseong gyoyug bunseog yeongu. Seoul, Korea: Higher Education Policy Research Institute; 2014.
10. Lim SD, Lee J, Park HS, Yu JR, Lee KY, Sohn IS, Lee R. Experience and consideration on online course evaluation by medical students. *Korean J Med Educ* 2008; 20: 367-371.
11. Chae SJ, Chang KH, Kang HS, Kim WS. The study on the validity and reliability of an instruction evaluation questionnaire. *Korean J Med Educ* 2002; 14: 287-292.
12. Im EJ, Chang BH. Analysis on teaching evaluation of clinical clerkship at Korean medical schools. *Korean J*

- Med Educ 2011; 23: 119-126.
13. Park SG, Seo DH, Seo SY, Seo YS, Song SK, Shin KH. Student's response to current lecture evaluation method in one medical school. Korean J Med Educ 2003; 15: 233-240.
 14. Yum S. Validating students' ratings of teaching scale and analyzing multilevel models. J Educ Eval 2008; 21: 25-52.
 15. Kim MH. Validity and reliability of lecture evaluation. Asian J Educ 2005; 6: 1-24.
 16. Han SY. Research on student ratings of university teaching: analysis of determinants related to professor, student and class. J Educ Adm 2001; 19: 247-266.
 17. Cashin WE. Student ratings of teaching: the research revisited. Manhattan, USA: Kansas State University, Center for Faculty Evaluation and Development; 1995.
 18. Aleamoni LM, Yimer M. An investigation of the relationship between colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness. J Educ Psychol 1973; 64: 274-277.
 19. Cranton PA, Smith RA. Reconsidering the unit of analysis: a model of student ratings of instruction. J Educ Psychol 1990; 82: 207-212.
 20. Han KS, Choi SH, Park JC. Problems in mandatory course evaluations. Commun Stat Appl Methods 2011; 18: 35-45.