The Journal of the Convergence on Culture Technology (JCCT)
Vol. 8, No. 1, pp.271-275, January 31, 2022. pISSN 2384-0358, eISSN 2384-0366

http://dx.doi.org/10.17703/JCCT.2022.8.1.271

JCCT 2022-1-30

교수와 대학생의 수업평가 결과

Instructional Ratings by Professors and College Students

길양숙*

Yangsook Kil*

요 약 이 연구는 수업경연대회에 심사자로 참여한 교수 심사자와 대학생 심사자들 간의 평가 결과의 차이와 상관을 검증하였다. 연구대상자는 한 사범대학의 수업경연대회에서 수업시연을 심사한 6명의 교수와 9명의 사범대 학생이었다. 이들은 각 과의 대표로 출전한 학생 11명의 수업시연을 10개 문항으로 구성된 심사평가지에 5단계 척도에 따라 응답하였고, 이 자료를 바탕으로 독립표본 t-검증, Pearson의 상관계수를 구하였다. 그 결과, 교수와 학생 심사자들 간의 평가 점수의 평균에는 통계적으로 의미 있는 차이가 없었다. 교수와 학생 심사자의 평가결과의 상관은 .78로 통계적으로 의미가 있었을 뿐 아니라, 상당히 높았다. 이 결과는 대학생의 수업평가 능력을 지지하며, 수업연습에 동료 학생들을 의미 있는 평가자로 활용할 수 있음을 시사한다. 이런 연구는 한 학기라는 오랜 기간 동안의 수업으로 인한 과정변인의 중재를 통제하고, 수업시연만을 대상으로 통제된 환경에서 연구하였다는 점에서 기존연구와 구별되는 의미가 있다. 연구대상자가 사범대에 국한되었다는 것은 제한점이 될 수 있으므로, 일반화를 위해서는 일반대학의 교수와 대학생 연구도 추가할 필요가 있다.

주요어 : 수업평가, 교수 평가자, 학생 평가자, t-검증, 상관 분석

Abstract We tested the difference and the relationship in instructional ratings by professors and college students. Subjects were 6 professors and 9 college students participated in teaching competition as a rater. They completed instructional rating forms consisted of 10 criteria for 11 participants according to Lickert Scale. Data were analyzed by independent t-test and Pearson correlation. Results were as follows: ① There was no statistical difference in mean scores between the professor and the student; ② Correlation between professors' and students' evaluations was .78 which is statistically meaningful and high. These results support the notion that students evaluate as reliably as professors. They also imply that pre-service college students can be utilized for teaching practice as meaningful peer evaluators. The findings are meaningful in that instructional ratings of instructors and students were compared in a controlled setting. However, replication studies involving wide range of instructors and students are needed for generalization since the subjects in this study were restricted to pre-service teachers and their instructors.

Key words : Instructional Rating, Professor Evaluator, Student Evaluator, T-test, Pearson Correlation

*정회원, 강원대학교 교육학과 교수 (제1저자)
접수일: 2021년 12월 30일, 수정완료일: 2022년 1월 5일
게재확정일: 2022년 1월 8일

Received: December 30, 2021 / Revised: January 5, 2022
Accepted: January 8, 2022
*Corresponding Author: yskil@kangwon.ac.kr
Dept. of Education, Kangwon National Univ, Korea

Ⅰ. 서 론

학생이 교수의 수업을 평가하는 것은 중세 대학으로 거슬러 올라갈 만큼 오래된 일이다. 2000년대 이후로는 우리나라의 대학에서도 학기말 수업평가가 보편화되어 있다. 학생에 의한 수업평가는 효율성을 강조하고, 학습자를 소비자로 보는 사회의 분위기에 힘입어 교육의 질을 확인하는 한 방법으로 자리 잡았다[1].

그럼에도 불구하고, 교수들은 학생에 의한 수업평가가 마냥 편하지는 않다. 학습자의 피드백을 교육능력 향상을 위한 형성평가 자료로 활용할 수 있다는 점에는 공감하지만, 수업평가의 작은 점수 차가 업적평가와 승진, 인센티브에 영향을 주는 현실에는 불만을 토로한다. 학생들에 의한 수업평가의 타당성과 신뢰성을 의심하기 때문이다. 교수들은 학생들이 교수의 교육 또는 수업능력 자체를 평가하기 보다는 인성을 평가한다고 생각하거나, 수업과정의 변수들-과제의 어려움(72%), 학점(68%), 사전 흥미(62%), 과제부담(60%), 학급의 크기(60%), 수강 이유(55%), 학생의 GPA(53%) 등-에 영향을 받는다고 믿는다[2]. 수업의 맥락이 다양한 데 비해, 강의평가 도구는 그에 맞는 타당성과 신뢰성을 확보하지 못했다고 본다[3]. 수업평가 실행도 공정하게 이루어지 않는다고 생각한다.

이런 인식에 근거가 없는 것은 아니다. 선행연구들은 수업평가가 학급의 크기, 학문 계열, 이수 구분, 평가가 이루어지는 환경(형성평가용 대 의사결정용), 학생이 열심히 공부한 정도, 사전 흥미, 예상학점, 수업환경에 따라 차이가 있음을 보여주었다. 즉, 학급의 크기가 작을수록, 인문사회 계열일수록, 선택과목일수록, 사전에 흥미가 높은 상태일수록, 예상학점이 높을수록, 승진이나 업적과 같이 수업평가에 의한 영향이 큰 상황일수록 수업평가 점수가 후한 경향이 있음을 보고했다[4-6]. 동일한 교수의 수업이라도 학부 과목 사이의 상관은 .66, 학부 수업에 대한 평가와 대학원 수업에 대한 평가는 상관이 .15에 불과하다는 연구결과도 있다[7]. 한 대학에서 1040명의 학생들을 대상으로 한 설문조사에서는 33%의 응답자가 문항을 보지도 않고 평정한다고 답하였고, 응답자의 절반 정도가 모든 문항에 같은 점수를 주기도 하였다[8]. 이런 결과는 수업평가결과를 의사결정의 목적으로 사용하려 할 때 쟁점들을 야기한다.

그러나 다른 한편에서는 학생들이 생각도 하지 않고, 낮은 점수를 주고 만다는 교수들의 우려와 달리 수업평가를 성실히 하며 긍정적인 점수(리커트 척도의 5점)를 일관되게 주는 비율(7.83%~18.79%)에 비해 부정적인 점수를(1점) 일관되게 주는 일은 매우 적음(전체의 0.46%~0.62%)을 보고한다[9]. 학생의 수업평가 점수와 그 학급에서의 학생들의 학습량의 상관이 .45에 이를 정도로 타당하다는 긍정적인 연구결과도 있다[10]. 이 상관계수는 고등학교에서의 성적과 대학에서의 성적 사이의 상관과 같은 크기로 학생의 학습 정도를 예언하는 변수로 의미가 있다. Murray의 연구에 의하면 학생들이 높이 평가한 교사들은 낮은 평가를 받은 교사와는 다른 교수행동을 보인다[11]. 즉, 3~8명의 훈련된 관찰자가 1명의 교수를 한 학기 동안 1시간 씩 3번, 총 18시간 이상 관찰한 결과 학생들이 높이 평가한 교수들은 중간 이하의 평가를 받은 교수들과 달리 어려운 부분은 반복했고(명료성 증진행동), 감정이입해서 말하거나 표현을 잘 했고(열의 요인), 학습자의 필요에 민감하였다(신뢰 형성). 상중하로 구분된 교수자들은 26개의 행동에서 달랐다. 이런 발견을 기초로 그는 학생이 교수의 인성을 보고 평가하는 것이 아니라 수업행동에 기초하여 평가하며, 학생의 수업평가와 훈련된 수업관찰자가 유사하게 평가한다고 보고하였다. 이탈리아 대학의 한 사례연구도 학생들의 수업에 대한 만족감은 교수자의 성격이 아닌 과목에서의 경험과 교수자의 능력에 근거한다고 결론짓는다[12]. 이런 연구들에 기초하여, 수업평가는 다른 변수의 측정치와 함께 사용한다면 교수의 승진 등을 결정할 때 유용한 정보라는 주장이 힘을 얻고 있다.

그럼에도 불구하고, 학기 말에 실시되는 수업평가는 수업의 맥락이 다양하고, 수업이 진행되는 시간이 길고 그 과정에 작용하는 변수들이 많아서 신뢰성이나 타당도를 높이는데 한계가 있다. 수업평가가 보편화되면서 여러 과목의 수업평가가 강제되는 점도 문제를 더하고 있다. 미국에서는 학생의 수업평가에 대한 연구가 1970-1980년대에 주로 이뤄졌고, 우리나라에서는 수업평가가 대학에 도입되던 2000년대 초에 얼마간 진행된 터라 학생에 의한 수업평가의 타당성과 신뢰성에 대한 검증은 이후 크게 진전되지 않은 모습이다. 타당성을 검증할 때도 수업평가 점수가 다른 변수들, 예컨대 학생들의 학업성취, 동료나 행정가, 졸업생, 훈련된 관찰자, 학생들의 자유 응답, 설문조사 등과의 관련 속에서

판정해야 하고, 신뢰도도 일관성, 안정성, 일반화 가능성 등을 고려해야 하는 상황이라 연구결과를 종합하여 결론을 내리기는 더욱 어려워지고 있다[13]. 이에 따라 연구자들은 수업평가 결과의 해석은 수업 과정 요소를 고려하여 신중할 것을 권고한다. 또한 수업평가를 하더라도, 신뢰성이 확보되도록 학급 학생수가 10명 이상인 경우, 3분의 2 이상이 평가에 참여한 경우에 한정하고, 한 교수 당 5코스 이상을 평가하도록 제안한다[14]. 교수의 기본 수업시수가 학기 당 6~9시간 정도이고, 그 중 일부는 대학원 수업을 맡는 현실을 고려하면 수업평가가 공정하게 되려면 적어도 3학기 이상의 기간에 평가를 해서 종합적으로 고려해야 한다는 결론이다. 그리해도 학부 수업과 대학원 수업 평가는 상관이 낮아서 각기 다른 상황에서 좋은 평가를 받는 교수를 어떻게 평가할 지의 문제는 남는다. 이런 저런 문제점들을 고려할 때 수업평가 결과의 해석은 전문가의 도움이 필요할 정도이다.

이 연구는 이 점에 착안하여 수업시연능력 외의 변수를 통제하고, 수업시연에 대한 평가만을 분리하여 대학생들이 교수들과 수업평가 능력에 있어서 차이가 있는지, 차이가 없다면 평가점수의 일치 정도는 어떤지를 검증하려 하였다. 이런 연구는 수업시연 능력에 기인하는 변량과 수업시연 외적인 과정 변수(수업 관리 능력, 교수 인성, 학생 특성 변수, 수업환경 변수 등등)에 의한 영향을 분리하여 보다 타당한 평가를 위한 방법을 모색한다는 점에서 의미가 있다.

II. 연구방법

이 연구는 한 국립대학교 사범대학에서 예비교사들의 수업시연 능력을 증진하고자 실시한 수업경연대회 자료를 분석하였다. 수업경연에는 11개 학과에서 각각 1개 팀이 참여하였고, 팀은 수업시연자 1명과 수업계획을 돕고, 학생 역할도 수행하는 4명으로 구성되었다. 수업경연 심사자로는 6개 학과에서 6명의 교수와 9개 학과에서 9명의 학생이 참여하였다. 공정한 심사를 위하여 상피제도도 도입하였다. 그리하여 경연에 참가한 각 1인의 수업시연에 대하여 교수 3인, 학생 3~5인이 심사를 하였다. 수업시연은 1인당 30분씩 진행하였다.

수업시연 평가는 효율적인 수업의 특징 10개를 평가하는 문항으로 구성된 평가지에 5단계 리커르트 척도에 따라 하도록 하고, 합산하여 50점 만점으로 총점을 구하였다. 이 점수들이 학생심사자들과 교수심사자들 사이에 평균에 차이가 있는지는 독립표본 t-검증을 통해 알아보았고, 이 두 집단의 수업시연 평가 점수간의 일치 정도는 Pearson의 상관계수를 구하여 알아보았다.

III. 연구결과

연구결과는 교수 심사자와 대학생 심사자의 수업시연 평가점수의 평균의 차이 검증 결과, 이들의 평가 점수 사이의 상관 검증 결과의 순서로 제시하였다.

1. 교수, 학생 심사자의 수업시연 평가의 차이

교수와 학생 심사자가 수업경연에 참여한 11명의 수업시연을 보고 평가한 점수의 평균이 집단 간에 차이가 있는지를 검증한 결과는 표 1과 표 2에 제시하였다.

표 1에 따르면 10개의 기준으로 이루어진 수업시연 평정표에 부여한 점수의 합산 평균은 교수가 34.30(리커르트 척도로 환산하면 3.43)이고, 대학생이 33.84(리커르트 척도로 환산하면 3.38)이다. 이 차이가 통계적으로 의미가 있는 것인지를 알아보기 위해 평균의 차이 검증을 실시한 결과, 교수와 학생 심사자가 심사한 사례 수는 같으므로 등분산 가정이 가능하고, t값 0.17은 자유도가 20일 때의 P값 0.87> 0.05여서 교수 심사자와 학생 심사자의 수업시연 평가 점수는 통계적으로 차이가 없다.

표 1. 교수, 학생 심사자의 수업시연 평가점수 집단 통계량
Table 1. Group statistics of teaching evaluation by professors and students

집단	사례 수	합산평균 (5단계 척도 환산)	표준편차	평균의 표준오차
교수	11	34.30(3.43)	4.95	1.49
학생	11	33.84(3.38)	7.54	2.27

표 2. 독립표본 t-검정 결과
Table 2. Result of independent t-test

Levene의 등분산 검정		평균의 동일성에 대한 T 검정						
F	유의확률	t	자유도	유의확률 (양측)	평균차이	차이의 표준오차	차이의 95% 신뢰구간	
							상한	하한
3.91	0.06	0.17	20	0.87	0.46	2.72	-5.21	6.14

2. 교수, 학생 심사자의 수업시연 평가 점수의 상관

교수와 학생 심사자의 수업시연 평가가 얼마나 같은지를 알아보는 또 다른 방법은 이 두 집단의 수업시연 평가결과의 상관을 구하는 것이다. 표 3은 이 검증 결과를 정리한 것이다. 표 3에 따르면 교수와 학생 심사자의 수업시연 평가 결과는 0.781로 통계적으로도 의미있는 높은 상관을 보인다.

표 3. 교수와 학생 심사자의 평가 결과의 상관
Table 3. Correlation of teaching evaluations by professors and students

		교수	학생
교수	Pearson 상관	1	0.781**
	유의확률(양측)		0.005
	N	11	11
학생	Pearson 상관	0.781**	1
	유의확률(양측)	0.005	
	N	11	11

**상관관계가 0.01 수준에서 유의함(양측)

Ⅳ. 논 의

이 연구는 학생에 의한 수업평가 능력에 대한 의심을 사범대학의 수업경연 상황에서 검증하였다. 그 결과 대학생도 교수와 차이 없이 수업시연을 신뢰롭게 평가할 수 있음이 밝혀졌다.

그러나 이런 결과를 수업평가 전반에 일반화하기 위해서는 연구가 이루어진 상황이 일반적인 수업평가의 상황과 다르다고 볼 수 있는 부분이 있어서 다음과 같은 주의와 추가 연구가 필요하다.

첫째, 학생들이 교수자의 수업을 평가하는 상황은 일반적으로 대학 수준에서, 한 학기의 수업을 경험한 후에 발생한다. 따라서 이때의 수업평가에는 수업시연 능력이나 수업기술 뿐만 아니라, 한 학기 동안의 수업관리 능력, 교수자의 인성, 과제의 양, 시험 난이도 등 다른 요소들도 작용할 수 있다. 이 연구는 수업시연 능력만을 평가했다는 점에서 일반적인 수업평가의 상황과 차이가 있다. 따라서 연구결과를 학기말 수업평가에 일반화하려면 한 학기의 수강 후, 같은 수업평가 문항으로 수업시연 능력만 평가하라는 지시와 함께 교수와 학생이 수업을 평가하는 연구가 필요하다.

둘째, 이 연구에 참여한 교수심사자는 사범대 교수이고, 학생 심사자도 교직과목을 수강했거나 수강하고 있다는 점에서 일반적인 대학생과 교수들의 특성과 다를 수 있다. 사범대 교수라도 모두가 교직과목을 수강하였거나 일반 대학의 교수들과 다른 교육적 소양을 가지는 것은 아니지만, 그 중 일부는 사범대학을 졸업하였고, 학생들도 교직과목을 일부라도 수강하였으므로 효율적인 수업을 판별하는 기준이나 개념을 더 잘 이해했을 수 있다. 그런 점에서 이들은 훈련된 관찰자로 분류될 수도 있다. 따라서 이 연구결과를 학기말 수업평가에도 일반화하려면 사범대 출신이 아닌 교수와 사범대를 제외한 학생들에게 같은 실험상황을 주고 수업시연을 평가하게 하는 연구가 필요하다.

Ⅴ. 결 론

이런 제한점들 때문에 이 연구결과가 학생의 수업평가 능력을 둘러싼 쟁점이나 의구심을 일거에 해결한다고 말할 수는 없다. 그렇지만, 제한점을 고려하더라도 사범대 학생들은 사범대 교수들만큼 수업을 타당하고, 신뢰롭게 평가할 수 있다는 점은 밝혀졌다. 적당한 교육이나 훈련으로 학생들도 교수심사자 만큼 수업을 잘 평가할 수 있는 것이다. 이것은 사범대처럼 수업능력을 증진하는 교육이 필요한 상황에서는 동료학생도 수업연습의 평가자로, 협력자로 유용한 자원임을 시사한다.

학습자를 만족시키려면 수업을 잘하는 것으로 충분하지 않을 수 있다. 대학생들도 지적 만족뿐만 아니라 정서적 욕구, 성장에 대한 관심을 필요로 하는 것처럼 보인다[15]. 학습자들이 수업평가에 코스 관리나 학생관리, 교수 인성에 대한 평가도 포함하는 것은 어쩌면 이런 필요 때문일 수 있다. 기왕에 이런 현상이 벌어지고 있다면, 앞으로는 수업평가에 이런 변수들을 포함하고, 그 효과를 분리해 보는 것도 필요해 보인다. 그것이 수업평가를 둘러싼 복잡성을 해소하는 방법일 수 있을 것이다.

References

[1] M. J. Kim, “An Analysis on the Educational Needs of College Faculty,” *Journal of the Convergence on Culture Technology(JCCT)*. Vol. 6, No. 2, pp. 239-250, 2020.

[2] H. W. Marsh, “Student Evaluation of University Teaching: Research Findings, Methodological

Issues, and Directions for Future Research," *International Journal of Educational Research*, Vol. 11, No. 3, pp. 253-388, 1987. https://doi.org/10.1016/0883-0355(87)90001-2
[3] K. S. Han, S.H. Choi, and J. C. Park, "Problems in Mandatory Course Evaluations," *Journal of the Korean Statistical Society*, Vol. 18, No. 1, pp. 35-45, 2011. DOI: 10.5351/CKSS.2011.18.1.035
[4] R. D. Shingles, "Faculty Rating: Procedures for Interpreting Student Evaluations," *American Educational Research Journal*, Vol. 14, No. 4, pp. 459-470, 1977. https://doi.org/10.2307/1162343
[5] K. A. Feldman, "The Significance of Circumstances for the Students' Ratings of Their Teachers and Courses," *Research in Higher Education*, Vol. 10, No. 2, pp. 149-172, 1979.
[6] J. A. Centra, *Reflective Faculty Evaluation*, Jossey-Bass Publishers, 1993.
[7] H. G. Murray, P. J. Ruston, and S. V. Paunonen, "Teacher Personality Traits and Student Instructional Ratings in Six Types of University Courses," J*ournal of Educational Psychology*, Vol. 82, No. 2, pp. 250-261, 1990.
[8] M. H. Kim, "Validity and Reliability of Student Ratings of Instruction," *Asian Journal of Education*, Vol. 6, No. 3, pp. 1-24, 2005.
[9] Y. S. Kil, "Results of Teaching Evaluation and the Issues," *Korean Journal of Educational Research*, Vol. 41, No. 3, pp. 303-323, 2003.
[10]C. A. Cohen, "Student Ratings of Instruction and Student Achievement: A Meta-Analysis of Multisection Validity Studies," *Review of Educational Research*, Vol. 51, No. 3, pp. 281-309, 1981.
[11]H. G. Murray, "Low Inference Classroom Teaching Behaviors and Student Ratings of College Teaching Effectiveness," *Journal of Educational Psychology*, Vol. 75, No. 1, pp. 138-149, 1983.
[12]M. Centoni and A. Maruoti, "Students' Evaluation of Academic Courses: An Exploratory Analysis to an Italian Case Study," *Studies in Educational Evaluation*, Vol. 70, September, 2021. https://doi.org/10.1016/j.stueduc.2021.101054
[13]B. Strumm, "Student Ratings of Instruction: A Summary of the Literature," In M. B. Paulsen (Ed.), *Higher Education: Handbook of Theory and Research*, Springer, pp.1-16, 2019.
[14]G. M. Gilmore, M. T. Kane, and R. W. Naccarato, "The Generalizability of Student Ratings of Instruction: Estimation of Teacher and Course Components," *Journal of Educational Measurement*, Vol. 15, No. 1, pp. 1-13, 1978.
[15]W. Liu and X. Jin, "The Effect of College Counselors' Authentic Leadership and Ethical Leadership on Students' School Satisfaction," *The International Journal Promotion Agency of Culture Technology(JCCT)*, Vol. 9, No. 3, pp. 202-211, 2021. https://doi.org/10.17703/IJACT.2021.9.3.202