

A Study on Controlling the External Effect in Student Evaluation of Teaching¹⁾

S. W. Lee²⁾ and K. H. Lee³⁾

Abstract

Student evaluation of university teaching is a common tool for measuring the educational contribution of a professor and improving the quality of classes. There, however, exist external factors in the beyond of control of a instructor, which affect the result of Student's rating to prevent practical use of evaluation for administrative purpose. This paper investigates the factors that spoil the validity and the reliability of student evaluation and proposes a method to control the effect by the statistical analysis of evaluation data of Jeonju University for two years.

Keywords : student evaluation of teaching, reliability, validity

1. 연구의 목적

대학의 경쟁력이 기업의 경쟁력 나아가 국가의 경쟁력을 결정한다는 사실은 잘 알려진 사실이다. 이에 따라 정부나 기업에서는 더 좋은 졸업생을 양성하기 위해 대학이 더욱 많은 투자를 하도록 요구하고 있으며 대학들도 자체의 생존과 경쟁력의 강화를 위해 교육의 질의 향상에 많은 노력을 기울이고 있다. 그 결과 많은 대학들이 교육업적을 교수의 업적평가에 반영하고 있는데 이를 위해 강의평가가 시행되고 있다.

강의평가는 평가의 목적에 따라 형성적(formative) 평가와 종합적(summative) 평가로 나뉜다(Cohen, 1980). 형성적 평가는 학기중간에 학생들이 의견을 반영하여 강의의 질을 향상시키기 위해 사용되며(Crumbley와 Fliedner, 2002), 종합적 평가는 승진이나 연봉의 산정, 종신재직권의 부여 등 관리적 목적으로 주로 사용된다. 강의평가가 강의의 질을 개선하고 이 결과가 교수의 업적에 반영되어야 한다는 점에서 본다면 형성적 평가와 종합적 평가는 강의평가의 두 축이 될 것이다. 그러나 많은 학교의 경우 종합적 평가에 중심을 두어 강의평가를 교수의 평가수단으로 이용하는 경우가 대부분인데 Crumbley와 Fliedner(2002)의 연구에 의하면 미국의 경우 90.6퍼센트의 학교에

1) 본 논문은 전주대학교 교내 연구비를 지원받아 연구되었음

2) Professor, School of Business, Jeonju University, Jeonju, Chonbuk, 560-759, Korea.
E-mail : minaa@jj.ac.kr

3) Professor, School of Business, Jeonju University, Jeonju, Chonbuk, 560-759, Korea.

서 강의평가결과가 승진에 이용되며, 69.3퍼센트의 학교에서 연봉에 강의평가를 반영하고 있다.

이렇게 강의평가가 단순히 교수의 강의개선을 위한 참조자료가 아니라 교수를 직접적으로 평가한다는 점에서 강의평가결과의 신뢰성에 대한 논란이 많이 있는데(김영진, 1994) 이러한 강의평가에 대한 논쟁은 강의평가를 실시한지 60년이 넘는 미국의 경우에도 예외는 아니며 지금도 계속적인 연구가 실시되고 있다(Ahmadi, Helms와 Ralszadeh, 2001). 강의평가에 대한 논란의 중심에는 강의평가결과의 타당성에 대한 문제가 존재한다. 즉 강의평가결과가 학생수나 과목의 난이도 등 강의외적인 환경에 의해서가 아닌 오직 강의개선을 위한 교수의 노력에 의해서만 결정되는가의 문제이다.

우리는 주위에서 “수강학생이 많아 강의평가점수가 낮다”라든가 “저학년을 담당하기 때문에 강의평가점수가 낮다”는 등의 불만을 흔히 접할 수 있다. 만약 수강생의 수라던가 수강학년 등 교수의 노력이외의 요소가 교수의 강의평가에 영향을 미친다면 강의평가의 타당성은 떨어지며, 이러한 결과를 토대로 한 교수평가는 부작용을 가져올 가능성이 크다. 따라서 교수의 노력이외의 어떠한 요소들이 강의평가에 영향을 미치는지를 규명하고 이러한 요소들을 적절히 통제하여, 강의개선을 위한 교수의 노력을 적절히 평가할 수 있는 메커니즘의 구축이 절실히 요구된다.

우리나라에서는 많은 학교들이 강의평가를 실시하고 이를 교수의 업적에 반영함에도 불구하고 강의평가제도의 문제점을 인식하고 이를 개선하려는 노력은 크게 이루어지지 않고 있다(한신일, 2002, 류춘호와 이정호, 2003) 현재까지의 논문은 대부분 자신의 학교를 대상으로 어떠한 요인이 강의평가에 영향을 주는가에 대한 기술적(descriptive) 분석에 그치고 있는데 비해 본 논문은 보다 적극적으로 강의평가에 영향을 주는 요인을 통제하는 방법을 제시하기로 한다. 물론 이러한 결과는 저자들이 재직 중인 대학의 자료를 이용한 것으로 일반성을 부여하기는 어렵지만 이러한 모형을 통해 타 대학에서도 강의평가에 대한 시행착오를 줄이고 좀 더 정교한 강의평가모형을 수립하고, 이를 교수업적에 반영할 수 있는 계기가 될 것으로 기대한다. 이에 따라 강의평가로 인한 부작용을 최소화하며, 강의평가가 교수의 강의능력을 개발하고, 나아가 대학의 경쟁력을 제고할 수 있는 수단으로 자리매김할 수 있게 될 것이다.

2. 이론적 배경

강의평가에 대한 연구는 강의평가의 신뢰성과 타당성에 관한 연구로 나누어 볼 수 있다. 강의평가의 신뢰성은 대개 평가결과의 내적 일치성과 안정성을 통하여 확인한다. 여기서 내적 일치성이란 동일한 학생집단이 여러 가지 형식의 평정도구를 이용하여 특정 강의를 평가했을 때 각 평정도구에서 얻은 결과간의 일치정도를 의미하는 것이다. 그리고 안정성은 동일 교수의 강의를 시기를 달리하여 두 번 이상 평가한 결과들이 어느 정도 일치하는가에 따라 결정된다(이종승, 1993). 대체로 학생에 의한 교수강의평가는 내적 일치성과 안정성이라는 관점에서 볼 때 비교적 만족할 만한 신뢰도를 나타내고 있는 것으로 보고되고 있다(Kulik과 McKeachie, 1975).

그러나 강의평가의 타당성에 대한 연구 결과는 매우 다양하다. Cohen(1981)이나 Dunkin(1986)은 학생들이 수업과정에서 벌어지는 다양한 교수행동의 적절성 여부를 변별할 수 있는 능력을 갖추었으며, 강의평가가 상당히 높은 타당성을 가지고 있다고 주장한다. 하지만 많은 실증연구들은 강의평가결과가 교수법의 효율성(instructional effectiveness)에 의해서 영향 받을 뿐 아니라 강의의 질과 무관한 여러 가지 요소에 의해 영향 받는다는 사실을 말해주고 있는데(Marsh와 Roche,

1993, Langbein, 1994, Koh와 Tan, 1997) 본 장에서는 강의에 영향을 주는 요소 중에서 관리적 목적에서 통제할 수 있는 요소를 중심으로 살펴보기로 한다.

교수의 노력 외에 강의평가에 영향을 주는 요소로 대표적으로 제시되는 것이 학생의 성적이다. 많은 연구들에 의하면, 강의평가의 결과는 교수의 학점에 비례하고 이에 따라 교수가 좋은 강의평가결과를 얻기 위해 높은 학점을 주는 경향이 있는데, Marsh(1984)는 이러한 현상을 교수가 학생에게 보다 높은 성적을 부여함으로써 강의평가결과를 구입한다(buy)고 말하고 있다.

학점 이외에도 많은 요인이 교수의 강의평가에 영향을 미친다고 보고되고 있는데 이러한 요인들은 크게 강좌의 특징(course characteristics)과 수강생의 특징(student characteristics)에 관한 것으로 나누어 볼 수 있다.

강좌의 특징에 관한 요소로는 수강인원, 강의의 수준, 과목의 계량/비계량 여부, 강의의 시간배치 등을 들 수 있으며 수강생의 특징으로는 수강생의 나이나 학년 등을 들 수 있다.

수강인원과 강의평가의 결과를 고찰한 선행연구에 의하면(Feldman, 1984) 수강인원이 적을수록 평가점수는 높아지며, 수강인원과 강의평가점수는 U자형 모양을 이룬다. 즉 수강인원이 많을수록 강의평가점수는 어느 정도까지는 감소하다 그 후에는 증가하는 모습을 보이게 된다(한신일, 2002).

강좌의 수준이 강의평가결과에 영향을 미친다는 연구들이 있는데 대부분의 경우에는 강좌의 수준이 높을수록 강의평가결과는 상향된다(Aleamoni와 Hexner, 1980, Goldberg와 Callahan, 1991, Langbein, 1994, Cranton과 Smith, 1986, Koh와 Tan, 1997).

과목이 계량과목인가 비계량과목인가에 의해 강의평가가 영향을 받는다는 연구도 있다(DeBerg와 Wilson, 1990, Langbein, 1994, Marsh와 Overall, 1981). 또한 강제로 수강해야 하는 과목과 그렇지 않은 과목 간에도 강의평가점수는 차이가 있을 수 있다(Aleamoni, 1981).

그리고 강의의 시간배치도 영향을 줄 수 있는데 몇몇 연구자에 의하면 오후 늦게 개설되는 강좌의 경우 강의평가점수가 낮게 나타나는 경향이 있다고 한다(DeBerg와 Wilson, 1990, Husbands와 Fosh, 1993).

학생의 특징을 나타내는 요소로 학생의 연령이나 학년이 강의평가에 영향을 줄 수 있다는 연구가 있다. Klann과 Hoff(1976)는 상위 수준의 과목에서는 학생의 연령이 많고 성숙할수록, 그리고 강의주제의 수준이 높을수록 강의평가점수가 높게 나타난다고 말하고 있다. 이들은 연구결과에서 20세 이상의 학생의 강의평가결과가 20세 미만의 학생의 강의평가보다 높은 것으로 보고하고 있다. 또한 Arreola(1995)는 학년이 높아갈수록 관대화되는 경향이 있다는 사실을 보여주고 있는데 우리나라에서도 류춘호와 이정호(2003)는 학년이 높고 학년이 높을수록 강의평가점수가 상향된다는 사실을 보고하고 있다.

3. 연구모형 및 방법

3.1. 가설의 설정

본 논문에서는 강의평가의 타당성을 검증하기 위해 강의의 효율성과 관계없는 요소가 강의평가에 영향을 미치는가를 알아본다. 유사한 연구가 류춘호와 이정호(2003), 한신일(2002) 등에서 진행되었지만 기존의 연구들은 대부분 특정 요인이 강의평가결과에 어떠한 영향을 미치는가를 단편적으로 검정하고 있는데 반해 본 연구에서는 모든 변수를 함께 포함한 모형을 고려하였으며 강의평가에 영향을 주는 외적요인을 통제하는 방법을 제시하였다는 점에서 의미를 갖는다.

본 연구의 목적과 이론적 배경을 토대로 본 논문에서는 다음과 같은 가설을 설정하도록 한다.

가설 1 : 성적이 높은 과목일수록 강의평가점수(만족도)가 높을 것이다.

가설 2 : 수강인원이 적을수록 강의평가점수(만족도)가 높을 것이다.

가설 3 : 주 수강생의 학년이 높을수록 강의평가점수(만족도)가 높을 것이다.

가설 4 : 과목의 유형에 따라 강의평가점수(만족도)가 다를 것이다.

가설 4-1 : 전공과목의 강의평가점수(만족도)가 교양과목에 비해 높을 것이다.

가설 4-2 : 선택과목의 강의평가점수(만족도)가 필수과목에 비해 높을 것이다.

가설 4-3 : 인문사회·예체능계열보다는 자연과학·공학계열의 강의평가점수(만족도)가 낮을 것이다.

마지막으로 본 논문에서는 강의평가에 영향을 미치는 요소를 통제하기 위한 방법에 대해 제안하기로 한다.

3.2. 변수의 정의 및 측정

강의는 다차원적인 것이며, 이를 하나의 수치로 평가하기는 어렵다. 따라서 강의평가지에는 다양한 문항을 통해 강의의 내용을 평가하게 된다. 그러나 교수의 승진이나 업적평가 등의 자료로 사용하기 위해서는 평가된 결과를 하나의 수치로 나타내야 하는데 이 경우에는 개별 변수들에 대해 중요도에 따라 가중치를 부여한 가중평균이 사용되는 것이 권장된다(Marsh와 Roche, 1997). 그러나 가중치를 어떻게 부여할 것인가에 대해서는 합의된 결론이 없기 때문에 실제로 이러한 이론을 그대로 사용하기는 어렵다. 그런데 Harrison, Douglas와 Burdsal(2004)의 연구에 의하면 강의효율성을 나타내는 여러 가지 요소를 단순 평균한 값이나 중요도에 따라 가중평균한 값 그리고 강의만족도를 직접적으로 평가한 값들이 거의 동일한 것으로 나타나고 있다. 따라서 본 논문에서는 종속변수인 강의평가결과를 각 항목에 대한 단순평균인 강의평가점수와 만족도점수 두 가지로 사용하도록 한다.

전주대학교에서는 10개의 항목으로 나누어 질문을 하고 있는데 1~9번 문항은 강의에 대한 다양한 교수의 노력을 평가하는 것이며, 10번 문항은 학생들의 강의에 대한 만족도를 직접적으로 평가하는 문항이다(<표 1> 참조). 따라서 강의평가점수는 1~9번 문항의 산술평균으로, 만족도는 10번 문항의 값으로 정의하기로 한다.

독립변수로는 일반적으로 강의평가에 영향을 주는 수강생의 성적, 학년, 선택/필수, 전공/교양, 계량/비계량 등의 과목분류로 정하였다. 그런데 학교 전체적으로 계량과목여부를 판단하기는 어려우므로 본 연구에서는 계량/비계량 구분 대신, 인문사회·예체능/자연과학·공학 과목으로 분류하기로 한다.

4. 실증분석

4.1. 표본의 선정과 기초통계분석

수강학생이 적은 경우에 평가의 익명성이 보장되기 어렵고, 극단치에 의해 평가성적이 크게 좌우되며, 일부 과목에서 학점 인플레이션이 나타날 수 있기 때문에 전주대학교에서는 20명 미만의

강좌에 대해서는 평가점수를 전체평균으로 대체하고, 또한 20명 이상의 강좌에 대해서는 성적산출에 상대평가를 실시하고 있다. 20명 미만의 과목을 제외한 이유는 신뢰성의 결여 이외에도 교양세미나, 전공체험학습, 예체능 실기 등 다른 과목과 차별화되는 과목이 많기 때문이다. 본 논문에서는 2002년 2학기부터 2004년 1학기까지 수강인원의 수가 20명 미만의 강좌를 제외한 5,853과목을 분석하였다. 자료의 기초통계량을 보면 수강인원의 평균이 약 42명(최소 20, 최대 510명)이고, 강의평가 평균이 3.73(표준편차 0.24), 만족도 평균이 3.68(표준편차 0.30) 등이었다.

분석에 사용한 강의평가 설문에 대한 신뢰도를 검증하고자 본 연구에서는 변수를 구성하고 있는 항목의 내적 일관성을 나타내는 크론바하(Cronbach) 알파계수(α)를 사용하였다. 5점 척도를 이용한 측정항목의 신뢰도를 분석한 결과는 <표 1>에 요약되어 있다. 이에 의하면 크론바하 알파가 0.9에 가까워 본 설문의 신뢰성이 매우 우수한 것으로 판단된다. 개별 항목에 대한 알파를 분석한 결과 1번 문항은 삭제되었을 때 신뢰도가 높아지므로 신뢰성을 낮추는 항목으로 판단되지만 그 정도가 미미하기 때문에 교수 강의의 충실도 측정을 위하여 설문에 포함시키고 있다.

<표 1> 강의평가 설문과 크론바하(Cronbach) 알파계수

문항	설문내용	α if Item Deleted	α
1	교수는 공식 행사 외에 휴강을 몇 번 정도 하였습니까	.9002	.8925
2	강의는 강의 계획서에 맞추어 진행되었습니까	.8818	
3	교재(학습자료)는 학습에 도움이 되었습니까	.8802	
4	교수는 강의 준비를 철저히 하였습니까	.8796	
5	교수는 강의내용을 이해하기 쉽게 전달하였습니까	.8782	
6	강의의 성격에 맞게 적절한 수업방법을 사용하였습니까	.8779	
7	교수는 학생들의 관심과 참여를 유도했습니까	.8811	
8	강의의 난이도는 적당하였습니까	.8820	
9	성적의 평가기준은 합리적이고 미리 제시되었습니까	.8820	
10	강의에 대하여 전반적으로 만족합니까	.8737	

<표 2> 변수들 간의 상관관계

	평균점수	만족도	수강인원	평균성적	학년	선택여부 ^{a)}	전공여부 ^{b)}
평균점수	1.000	0.885	-0.247	0.181	0.183	0.132	0.120
만족도	0.898	1.000	-0.221	0.181	0.177	0.153	0.110
수강인원	-0.162	-0.148	1.000	-0.110	-0.123	-0.008 ^{c)}	-0.159
평균성적	0.194	0.199	-0.032	1.000	0.347	0.224	0.385
학년	0.177	0.170	-0.087	0.322	1.000	0.402	0.453
선택여부	0.126	0.146	0.012 ^{c)}	0.240	0.402	1.000	0.589
전공여부	0.111	0.103	-0.161	0.375	0.453	0.589	1.000

a) 필수과목 = 0, 선택과목 = 1

b) 교양과목 = 0, 전공과목 = 1

c) 유의성이 없음, 나머지는 모두 1% 유의수준에서 유의함

<표 2>에는 각 변수들 간의 상관계수가 나타나 있다(대각선 하단은 피어슨 상관계수이고 상단

은 스피어만 상관계수임). 이를 살펴보면 수강인원이나 강좌의 성적, 학년, 선택과목여부, 전공여부가 모두 강의평가점수나 만족도와 관련이 있는 것으로 나타나고 있다. 그러나 강의평가점수에 영향을 주는 독립변수들 간에도 유의한 상관관계가 있어 이들 사이에 교호작용이 강의평가에 영향을 줄 가능성이 있다.

평가자료의 타당성을 측정하기 위해서 본 논문에서 제시하지는 않았지만 구성타당성 검증에 해당하는 요인분석을 실시하였다. 그 결과 10번 만족도를 제외한 9개 문항이 한 개의 요인으로 추출되었고 1번 문항을 제외한 각 변수의 요인에 적재된 부하치가 모두 .7이상으로 나타났다. 또한 강의평가 요인과 평균성적을 회귀분석한 결과 결정계수가 0.563으로 높은 강의만족이 높은 학업성취를 의미함으로써 강의평가 문항의 타당성이 상당히 높다고 할 수 있다.

4.2. 가설검정

가설 1의 평균성적이 강의평가점수에 미치는 영향을 회귀분석을 통하여 고려해 보았다. 여기서 자세한 수치를 제시하지는 않겠지만 독립변수를 과목의 평균성적만으로 하였을 때 강좌의 평균학점이 강의평가점수나 만족도에 미치는 영향은 1%의 유의수준에서 유의하게 나타나 강좌의 성적이 좋으면 학생들의 강의평가결과가 높아지는 것으로 판명되었다. 이는 기존의 연구(Marsh, 1984, Nimmer와 Stone, 1991, 류춘호와 이정호, 2003)와 동일한 결과이다. 이 경우에 White 검정결과 카이제곱값이 각각 0.58 및 0.96로 나타나 이분산성은 존재하지 않는 것으로 판단된다.

다음으로 가설 2의 수강인원이 강의평가점수에 미치는 영향을 살펴보기 위하여 수강인원을 10명씩 급간으로 구분하여 강의평가점수와 만족도의 추이를 조사한 결과 20명대의 수강인원 과목의 평가가 가장 높으며, 인원이 증가하면서 평가점수는 계속 하락함을 알 수 있었다. 평가는 51~70명 사이의 강좌에서 가장 낮은 점수를 보이다 수강인원 90명까지는 약간 증가하여 Feldman(1984)의 연구에서와 같은 U자형 모습을 보이고 있었다. 이들의 관계는 선형의 관계가 아닐 수 있고 수강인원이 증가하면서 수강인원이 강의평가점수 및 만족도에 미치는 영향은 감소할 가능성이 있는데 이를 고려하기 위해 수강인원에 로그를 취한 값을 독립변수로 사용하여 회귀분석하였다. 이 결과는 <표 3>에 나타나 있다. 이를 살펴보면 로그를 취한 학생수와 강의평가점수 및 만족도는 매우 밀접한 관계를 가짐을 알 수 있다. 또한 잔차의 이분산성을 체크하기 위해 White 검정을 실시한 결과 독립변수로 수강인원을 사용한 모형에서는 Chi-Square 값이 8.29로 나타나 5% 유의수준에서 잔차의 이분산성이 존재하는 것으로 나타났으나 독립변수로 수강인원에 로그를 취한 값을 사용한 경우에는 Chi-Square 값이 1.97로 나타나 이분산성이 없는 것으로 판명되었다.

<표 3> 로그를 취한 학생수와 평가점수간의 회귀분석

종속변수	독립변수	자유도	회귀계수	표준오차	t 값	R ²
강의평가점수	상수항	1	4.19784	0.02705	155.17**	0.050
	ln수강인원	1	-0.13033	0.00741	-17.60**	
만족도	상수항	1	4.20594	0.03377	124.55**	0.042
	ln수강인원	1	-0.14724	0.00924	-15.93**	

** : 1% 유의수준에서 유의함

수강인원, 학년, 과목의 특성 등이 강의평가점수나 만족도에 영향을 미치는가에 관한 가설을 검정하기 위해 분산분석을 실시하였다(<표 4>). 이를 보면 학년이 높아질수록 강의평가점수 및 만

족도가 높아진다는 사실을 알 수 있는데 이는 가설 3을 지지하는 것이며 학년이 높아지면서 관대 화경향이 발생한다는 사실을 입증하고 있다. 교양 및 전공으로 나누어 이들 교과목간의 강의평가 점수의 차이가 있는지를 비교해 보았을 때(가설 4-1) 전공과목의 강의평가 성적이 높은 것으로 나타났다. 과목의 선택권 유무에 따른 강의평가의 결과차이에 관한 가설 4-2를 살펴보면 매우 유의적인 차이를 보여 선택과목의 강의평가점수와 만족도가 필수과목에 비해 높다는 결론을 내릴 수 있다. 교과목의 계열별 특성에 따라 강의평가결과가 달라지는가를 분석한 결과를 보면 수리과목이 많은 이공계열의 과목이 평가가 낮을 것이라는 예상과는 달리 오히려 높게 나왔고 만족도에서는 계열 간에 차이가 있다고 볼 수 없기에 가설 4-3은 지지되지 않는다.

<표 4> 분산분석표

요인	종속변수: 강의평가					종속변수: 만족도				
	제곱합	자유도	평균제곱	F	유의 확률	제곱합	자유도	평균제곱	F	유의 확률
모형	28.758	11	2.614	48.831	.000	41.434	11	3.767	45.283	.000
절편	29459.8	1	29459.8	550243.5	.000	28572.3	1	28572.3	343498.2	.000
학년	4.279	3	1.426	26.640	.000	5.870	3	1.957	23.522	.000
선택필수	1.833	1	1.833	34.234	.000	5.062	1	5.062	60.859	.000
교양전공	.274	1	.274	5.113	.024	.534	1	.534	6.424	.011
인원계급	15.668	4	3.917	73.160	.000	20.997	4	5.249	63.107	.000
과목계열	.784	2	.392	7.319	.001	.375	2	.188	2.257	.105
오차	312.725	5841	5.3E-02			485.857	5841	8.3E-02		
합계	81548.9	5853				79428.4	5853			

R² = .084 (수정된 R² = .082)

R² = .079 (수정된 R² = .077)

4.3. 교호작용을 고려한 관련성 검정

<표 2>에서 보았듯이 강의평가에 영향을 주는 대부분의 변수들이 유의적인 상관관계를 갖고 있다. 따라서 이 경우에는 개별변수 뿐 아니라 두 변수 혹은 그 이상 변수의 교호작용이 강의평가 결과에 영향을 줄 가능성이 있다. 다른 논문에서는 이러한 변수의 교호작용에 대한 고찰을 생략하고 있으나 본 논문에서는 이를 고려하기로 한다.

강의평가점수와 변수간의 통계적 모형은 다음과 같다.

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2 + \beta_3x_3 + \beta_{13}x_1x_3 + \beta_{23}x_2x_3 + \beta_{123}x_1x_2x_3 + \beta_4x_4 + \beta_{14}x_1x_4 + \beta_{24}x_2x_4 + \beta_{34}x_3x_4 + \epsilon,$$

여기서 x_1 은 수강인원에 로그를 취한 값, x_2 는 수강생들의 평균성적이다. x_3 는 학년변수로 던킨의 분석을 통해 두 그룹으로 나누어 저학년(1, 2학년)은 0, 고학년(3, 4학년)은 1의 값을, x_4 는 과목의 계열변수로 인문사회계열과 예체능계열은 0, 이공계열은 1의 값을 부여하였다.

개별 변수들과 이들의 교호작용을 고려한 회귀분석 결과는 <표 5>에 나타나 있다. 이 결과를 살펴보면 개별 변수들은 로그를 취한 수강인원을 제외하고는 두 종속변수 모두에서 유의한 변수는 없음을 알 수 있다. 즉 변수의 유의성 검정 결과 상수항을 제외한 다음의 귀무가설만이 1% 유의수준에서 기각되었다.

$$H_0 : \beta_1 = 0 \text{ 대 } H_1 : \beta_1 \neq 0.$$

개별변수만을 고려하였을 때 여러 변수가 강의평가에 영향을 주는 것으로 나타났지만 모든 교호작용을 고려하였을 때 가장 큰 영향을 주는 변수는 수강인원임을 확인할 수 있다.

<표 5> 강의평가점수와 변수들 간의 회귀분석^{a)}

변수명	종속변수: 강의평가				종속변수: 만족도			
	추정치	표준오차	t 값	유의수준	추정치	표준오차	t 값	유의수준
상수항	4.661	0.409	11.37	<.0001	4.460	0.512	8.71	<.0001
인원 ^{b)}	-0.393	0.114	-3.45	0.0006	-0.395	0.142	-2.77	0.0056
성적	-0.005	0.005	-1.11	0.2657	-0.003	0.006	-0.51	0.6111
인원*성적	0.003	0.001	2.25	0.0246	0.003	0.001	1.71	0.0873
학년	-1.492	0.982	-1.52	0.1286	-1.442	1.227	-1.18	0.2400
인원*학년	0.4902	0.274	1.79	0.0741	0.515	0.343	1.50	0.1331
성적*학년	0.016	0.011	1.38	0.1672	0.015	0.014	1.06	0.2870
인원*성적*학년	-0.005	0.003	-1.61	0.1069	-0.005	0.004	-1.35	0.1757
계열	-2.084	1.752	-1.19	0.2345	-2.590	2.191	-1.18	0.2372
인원*계열	0.465	0.475	0.98	0.3274	0.594	0.593	1.00	0.3170
성적*계열	0.023	0.021	1.08	0.2798	0.029	0.026	1.09	0.2758
학년*계열	3.749	3.314	1.13	0.2581	3.565	4.144	0.86	0.3896

a) 기타 교차 교호작용항 생략

b) 수강인원에 로그를 취한 값

5. 관리적 목적을 위한 변수의 통제

5.1. 수강인원의 통제

앞 절에서 여러 가지 요인이 강의평가결과에 영향을 미치지만 교호작용을 고려하는 경우 유익적으로 강의평가결과에 영향을 미치는 요소는 수강인원이라는 사실을 알 수 있었다. 여기에서 나타난 결과는 수강인원이 증가하게 되면 담당교수는 동일한 노력으로 강의를 하더라도 더 낮은 강의평가점수를 얻게 된다는 것이다.

실제로 전공이나 교양, 선택이나 필수 등의 구분으로 교수의 기여도를 평가하기 어렵지만 많은 인원을 대상으로 강의하는 교수는 더 많은 노력을 투입하며, 재정적인 면을 비롯하여 여러 측면에서 학교에 공헌하고 있는 것은 분명한 사실이다. 그러나 강의평가결과를 그대로 이용하게 되면 이들이 상대적으로 불이익을 받게 되므로 교수평가라는 관리적 목적으로 강의평가결과를 사용하기 위해서는 인원수를 통제하는 것이 바람직한 것으로 생각되는데 <표 5>의 결과는 이러한 생각을 뒷받침하고 있다. 강의평가점수에서 인원수를 통제된 수정된 강의평가점수는 다음과 같이 표현할 수 있다.

$$y = x - (\log(n) - \overline{\log(n)})\beta. \tag{5.1}$$

여기서, x 는 원 강의평가점수, n 은 수강인원, $\overline{\log(n)}$ 는 수강인원 로그값의 평균, β 는 강의평가

점수를 인원수의 로그값으로 회귀분석한 경우의 회귀계수이다.

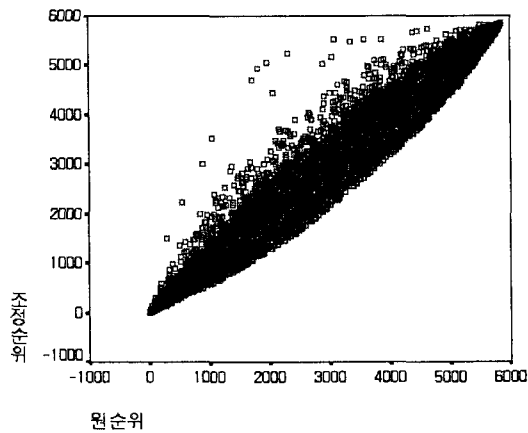
전주대학교의 경우 수강인원 로그값의 평균은 3.63이고, 회귀계수는 -0.133이므로 조정된 강의평가점수는 $y_0 = x_0 + (\log(n_0) - 3.63)0.133$ 과 같이 계산된다.

이를 통해 새로운 점수(y)를 구하고 새로운 순위(R_y)를 계산하였다. 새로운 점수와 새로운 순위의 원점수(x) 및 원순위(R_x)에 대한 상관계수는 <표 6>과 같다. 이를 보면 새로운 강의평가점수는 과거의 강의평가점수와 매우 높은 상관관계를 가짐을 알 수 있다. 그러나 새로운 강의평가점수는 수강인원과 상관관계를 갖지 않으며, 새로운 순위도 수강인원과 관계가 없음을 알 수 있다. 과거의 순위와 새로운 강의평가점수의 순위를 표시한 <그림 1>을 보면 통계에 의한 새로운 순위가 원 순위에 비해 많이 높아지는 경우는 있지만 순위가 급격히 하락하는 경우는 없어 관리적인 측면에서 교수들에게 공감을 얻기에 무리가 없을 것으로 보인다.

<표 6> 원점수 및 새로운 점수의 상관관계

변수명	x	y	y^*	R_x	R_y	R_{y^*}	$\log(n)$
x	1.00000						
y	0.97454**	1.00000					
y^*	0.96802**	0.98946**	1.00000				
R_x	0.96547**	0.93652**	0.93064**	1.00000			
R_y	0.93944**	0.96276**	0.95255**	0.96930**	1.00000		
R_{y^*}	0.93305**	0.95219**	0.96310**	0.96195**	0.98762**	1.00000	
$\log(n)$	-0.22421**	0.00000	-0.01674	-0.23545**	-0.00531	-0.02275	1.00000

** : 1% 유의수준에서 유의함, y^* 는 식 (5.2)에 정의되어 있음



<그림 1> 원순위와 새로운 순위의 산포도

주 : 순위가 높을수록 강의평가점수가 높음

이론적으로 보면 식 (5.1)과 같은 방법에 의해 점수를 조정하는 것이 가장 자연스러울 것이다. 그러나 수강인원에 로그를 취하고 회귀계수 추정값이 포함된 점수산정식은 그 복잡성과 점수의

하향조정 등으로 인하여 교수들의 불만요인이 될 수 있다. 따라서 보다 간단한 통제방법을 찾기 위해 수강인원을 10명 단위로 구분하여 Duncan의 방법에 의해 급간을 구분하여 다음과 같이 가산점을 부가하는 방식으로 수강인원을 통제할 수도 있다.

$$y^* = \begin{cases} x, & \text{수강인원} < 30 \\ x + 0.12, & 30 \leq \text{수강인원} < 90. \\ x + 0.20, & 90 \leq \text{수강인원} \end{cases} \quad (5.2)$$

수정된 강의평가점수(y^*)는 가장 수강인원이 적은 집단을 기준으로 하여 수강인원이 많은 집단에 가산점을 부여하는 형식으로 강의평가점수를 산정했기 때문에 대부분의 교수가 원점수에 비해 상향하게 된다. 하지만 <표 6>을 보면 수정된 강의평가점수와 이를 통해 구해진 순위도 통제된 강의평가점수나 그로 인한 순위와 유사한 역할을 함을 할 수 있다.

5.2. 수강생 성적의 통제

교수가 학생들에게 높은 성적을 보장함으로써 높은 강의평가결과를 얻을 수 있는 개연성에 관해서는 앞에서 설명한 바 있다. 이러한 문제점을 최소화하기 위해 전주대학교는 성적산출에 있어 상대평가비율을 지키도록 하고 있다(20명 미만 과목 제외). 상대평가가 실시된 이후 3년간 수강인원 20명 이상의 과목들의 학과성적 평균을 산출해본 결과 강의평가결과를 구입(buy)하기 위한 학점 인플레이션이나 기타 변화의 모습은 보이지 않았다. 그리고 학과 성적과 강의평가점수 만을 포함한 회귀분석에서는 두 변수 간에 유의한 관계를 보였지만 기타 변수와 교호작용을 포함한 <표 5>에서는 그 유의성을 찾지 못하였기 때문에 학과성적에 대한 강의평가 점수 통제는 따로 하지 않고 성적산출시 상대평가 원칙만 적용하고 있다.

그러나 개별적인 평가자의 성적을 고려할 때, 학업성취도가 낮은 학생의 평가의 신뢰성에 의문을 가질 수 있다. 강좌의 만족도가 낮으면 성취도가 낮은 것은 당연하지만 낮은 만족도를 갖는 학생들이 객관성을 갖고 평가하는지를 판단하기 위해서 강의평가 설문지(1)번 문항의 강의충실도를 2002학년도부터 정성적 평가에서 정량적 평가로 바꾸어 측정하였다. 즉, “교수는 강의시간을 준수 하였습니다니까”에서 “교수는 공식 행사 외에 휴강을 몇 번 정도 하였습니다니까”로 바꾸었다. <표 7>에 평가자의 성적 분류에 따른 강의평가 점수를 표시하였다.

<표 7> 성적 구분에 의한 강의평가 점수(2003학년도 2학기)

성적	강의 시간	계획서	교재	강의 준비	수업 전달	수업 방법	학생 참여	난이도	성적 평가	만족도
70이상	4.52	3.62	3.66	3.78	3.68	3.69	3.66	3.57	3.66	3.70
70미만	4.27	3.27	3.30	3.46	3.34	3.38	3.36	3.27	3.38	3.31
차 이	0.25	0.35	0.36	0.32	0.34	0.31	0.30	0.30	0.28	0.39

학업성적이 높고 낮음에 관계없이, 강의시간과 같은 정량적 평가설문은 동일한 값을 보여주어야 하는데 <표 7>에서 보는 바와 같이 0.25의 차이를 보이는 것은 학업성적이 낮은 학생들의 평가의 객관성에 의문을 갖게 한다. 학생이 교수의 강의를 평가할 자격이 있는가에 대한 의문은 계속 제

기되어 왔었기에 이에 대한 대응방안으로 전주대학교에서는 평가의 객관성이 부족하다고 판단되는 학과성적 70점 미만의 학생의 평가자료는 제외하고 있다(70점 이상의 경우에는 성적에 따른 강의시간준수항목의 점수차이가 크게 발생하지 않았다).

5.3. 기타 요인의 통제

전주대학교에서는 수강인원이 20명 미만인 경우 성적산출에 상대평가를 적용하지 않은 뿐더러 적은 평가인원에 의한 오류를 없애기 위해 원 평가점수 대신 전체평균으로 대체하고 있다. 그러나 이러한 과목의 경우 평균보다 높은 평가를 받는 교수의 불만과 소규모강의의 강의의욕 저하라는 문제를 야기하고 있어 전체평균 보다는 수강인원 수와 20명과의 거리에 따른 다음과 같은 축소추정량(shrinkage estimator)의 형태를 제안하고 있다.

$$x^* = (1 - \frac{n-10}{10})(\bar{x} - x) + x; 10 \leq n < 20,$$

여기서 x^* 는 축소추정량, x 는 원 평가점수, \bar{x} 는 전체 평균, n 은 수강인원이다. 즉, 수강인원이 10명인 경우는 x^* 는 전체평균이 되고 수강인원이 19명이면 원 점수에서 10%만큼 전체평균으로 축소(shrinkage)한 값이 된다. 그러나 이러한 값을 사용하면 원점수가 전체평균보다 낮은 경우에 불이익을 받는 경우가 발생하므로 실제로는 $\max(x^*, \bar{x})$ 를 사용할 것을 제안한다.

6. 결론

강의평가가 교수의 교육업적을 공정하게 평가하는 도구가 되기 위해서는 교수의 전공지식, 강의 능력 등과 관계없는 요인이 평가에 영향을 주지 않아야 한다. 본 논문에서는 학생들의 강의평가에 대한 수강생 규모, 수강생의 학년, 과목의 평균 성적, 과목의 특성(전공/교양, 선택/필수) 등의 변수의 영향을 검증하고 있고, 이러한 영향을 통제하는 방법을 제안하였다. 그 결과 수강생 규모가 강의만족도에 가장 유의한 영향력을 갖고 있었는데, 수강생 수에 따라 평가점수에 가산점을 부여하여 수강생규모에 의한 외적 요인의 통제방법을 제안하였다. 또한 다른 외적 요인에 관해서는 직접적인 통제가 아닌 강의평가의 합리적 운영방법을 통하여 그 영향력을 줄일 수 있도록 하는 방안 등을 제시하였다.

참고문헌

- [1] 김영진 (1994). 교수강의평가제-과연 생산성이 있는가, 「생산성논집」, 제8권 2호, 252-235.
- [2] 이종승(1993). 교수의 강의평가에 관한 연구, 「교육발전논총」, 제14권 1호, 83-95.
- [3] 한신일 (2002). 강좌규모와 강의평가의 관계분석, 「고등교육연구」, 제13권, 155-173.
- [4] 류춘호, 이정호 (2003). 대학의 강의평가에 영향을 미치는 학생관련 요인에 관한 연구, 「경영학연구」, 제32권 3호, 789-807.
- [5] Ahmadi, M., Helms, M. and Ralszadeh, F.(2001). Business Students' Perceptions of Faculty Evaluations, *The International Journal of Educational Management*, Vol. 15, no. 1,

12-22.

- [6] Aleamoni, L.M.(1981). *Student ratings of instruction, Handbook of Teaching Evaluation*, Sage Publishing, Beverly Hills, CA.
- [7] Aleamoni, L.M. and Hexner, P.Z.(1980). A Review of the Research on Student Evaluation and a Report on the Effect of Different Sets of Instructions on Student Course and Instructor Evaluation, *Instructional Science*, Vol. 9, 67-84.
- [8] Arreola, R.A.(1995). *Developing a comprehensive faculty evaluation system: a handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system*. Bolton, Mass.: Anker.
- [9] Cohen, P.A.(1980). Effectiveness of Student Rating Feedback For Improving College Instruction : A Meta-Analysis of Findings, *Research in Higher Education*, Vol. 13, 321-341.
- [10] Cohen, P. A. (1981). Student Ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies. *Review of Educational Research*, Vol. 51, 281-309.
- [11] Cranton, P.A. and Smith, R.A.(1986). A New Look at the Effect of Course Characteristics on Student Ratings of Instructions, *American Education Research Journal*, Vol. 23, no. 1, 117-128.
- [12] Crumbley, D.L. and Fliedner, E.(2002). Accounting Administrators' Perceptions of Student Evaluation of Teaching(SET) Information, *Quality Assurance in Education*, Vol 10, no. 4, 213-222.
- [13] DeBerg, C.L. and Wilson, J.R.(1990). An Empirical Investigation of the Potential Confounding Variables in Student Evaluation of Teaching, *Journal of Accounting Education*, Vol. 8. no. 1, 37-62.
- [14] Dunkin, J.M.(1986). *Research on Teaching in Higher Education, Handbook of Research in Teaching*, 754-777.
- [15] Feldman, K.A.(1984). Class Size and College Students' Evaluations of Teachers and Courses : A Closer Look, *Research in Higher Education*, Vol. 21, no. 1, 45-116.
- [16] Goldberg, G. and Callahan, J.(1991). Objectivity of Student Evaluation of Instructors, *Journal of Education for Business*, Vol. 66, 377-378.
- [17] Harrison, P.D., Douglas, D.K. and Burdsal, C.A.(2004). The Relative Merits of Different Types of Overall Evaluations of Teaching Effectiveness, *Research in Higher Education*, Vol. 45, no 3, 311.
- [18] Husbands, C.T. and Fosh, P.(1993). Students' evaluation of teaching in higher education: experiences from four European countries and some implications of the practice, *Assessment and Evaluation in Higher Education*, Vol. 18, no. 2, 95-114.
- [19] Klann, W.E. and Hoff, E.(1976). The Use of Judgment Analysis in Analyzing Student Evaluation of Teachers, *Mathematical Association of Two-Year College Journal*, Vol. 10, 137-139.
- [20] Koh, H.C. and Tan, T.M.(1997). Empirical investigation of the factors affecting SET

- results, *International Journal of Educational Management*, Vol. 11, no. 4, 170-178.
- [21] Kulik, J.A. and McKeachie, W.J.(1975). The Evaluation of Teachers in Higher Education, *Review of Research in Education*, Vol. 3, 210-230.
- [22] Langbein, L.I.(1994). The Validity of Student Evaluations of Teaching, *Political Science and Politics*, Vol. 27 no. 3, 545-553.
- [23] Marsh, H.W.(1984). Students' Evaluation of University Teaching : Dimensionality, Reliability, Validity, Potential Biases, and Utility, *Journal of Educational Psychology*, Vol. 76, 707-754.
- [24] Marsh, H.W. and Overall, J.U.(1981). The Relative Influence of Course Level, Course Type, and Instructor on Students' Evaluations of College Teaching, *American Educational Research Journal*, Vol. 18, no. 1, 103-112.
- [25] Marsh, H.W. and Roche, L.A.(1993). The Use of Students' Evaluations and Individually Structured Intervention to Enhance University Teaching Effectiveness, *American Educational Research Journal*, Vol. 30, no. 1, 217-251.
- [26] Marsh, H.W. and Roche, L.A.(1997). Making Students' Evaluations of Teaching Effectiveness Effective: The Critical Issues of Validity, Bias and Utility, *American Psychologist*, Vol. 52 no. 11, 1187-1197.
- [27] Nimmer, J.G. and Stone, E.F.(1991). Effects of Grading Practices and Time of Rating on Student Ratings of Faculty Performance and Student Learning, *Research in Higher Education*, Vol. 32, 195-215.

[2005년 6월 접수, 2005년 8월 채택]