

강의평가의 타당도와 신뢰도

김명화(金明和)*

논문 요약

본 연구는 강의평가에 대한 자료를 분석하고 대학 사회 구성원의 의견을 조사하여 강의평가에 대한 기초 자료를 제공하고 개선안을 제시하려 하였다. 본 연구에서는 강의평가의 타당도와 신뢰도의 측면에서 선행연구를 분석하고, 강의평가 결과를 분석하고, I 대학 교수와 학생을 대상으로 설문조사를 하였다. 강의평가에 대한 미국의 연구는 연구의 역사가 깊고, 체계적인 연구가 이루어지고 있는데 비하여 우리나라는 강의평가에 대한 체계적인 연구가 부족하고, 이런 연구를 통하여 강의평가를 개선하는 예는 매우 드물다.

I 대학의 강의평가 분석결과 강의평가의 타당도와 신뢰도는 상당히 낮은 편으로 시급한 개선이 요구된다고 할 수 있다. 김성열의 연구에 따르면 우리나라의 많은 대학의 사정이 유사하므로 본 연구의 결과는 한 대학의 사례이나 다른 대학도 비슷하리라고 유추할 수 있을 것이다. 따라서 강의평가 도구의 타당도와 신뢰도에 대한 체계적인 연구를 통하여 강의평가지를 개발하고, 개발된 강의평가지는 주기적으로 평가되고, 개선되어야 할 것이다.

■ 주요어 : 강의평가, 강의평가의 타당도, 강의평가의 신뢰도, 강의평가연구

I. 서론

1. 강의평가 연구의 필요성

강의평가제도는 미국에서 1960년대에 본격적으로 도입되어 강의평가의 타당도와 신뢰도, 활용도 등에 관한 수많은 연구를 거쳐 1980년대에 정착되었다. 미국 교육부 조사에 따르면 미국 대학의 97%가 교수평가를 위하여 강의평가를 활용하고 있다고 보고하고 있다(김성열, 2001). 이제는 논란의 대상의 아니라 수강신청, 학점제와 같이 대학의 당연한 절차가 되었다(조벽, 1997). 우리나라에서도 90년대 초에 강의평가가 도입되기 시작하였고, 97년을 기점으로 거의 대부분 대학에 도입되었고 90년대 중반 이후에는 강의평가제를 거의 모든 대학들이 기정사실화하고 있다. 김성열 외(2001)의 연구에 따르면 전국의 181개 대학중 153(84.5%)개교가 교수업적평가 제도를 하고 있고, 이에 강의평가를 활용하고 있다. 초기에 도입당시 학생이 교수를 평가한다는 것에 대한 거부감으로 인하여 많은 반대를 겪었으나 대학평가에서 강의평가제도가 평가항목에 포함되면서 싫든 좋든 이제는 많은 대학에서 거의 의무적으로 시행하고 있다.

이렇게 많은 대학에서 강의평가제도가 널리 사용하고 있으나 내실 있게 운영되는 학교는 드물다. 김성열(2001)의 연구에 따르면 대부분의 대학이 20여개 안팎의 객관식 문항과 2개 정도의 주관식 문항을 사용하여 학기 말에 인터넷 조사를 시행하고 있다. 미국의 연구를 보면 강의평가제도는 많은 문제점을 내포하고 있고 아직도 그 타당성이나 신뢰성에 대하여 의문이 제기되고 있으나 우리나라에서는 강의평가의 타당도나 신뢰도에 대한 심층적인 연구는 많지 않다. 선행연구들은 타당하고 신뢰로운 강의평가 도구를 사용하였을 때 강의평가는 교수 효과성에 긍정적인 영향을 미친다고 하였다. 강의평가에 대한 타당도나 신뢰도에 대한 심층적인 연구 없이 강의평가의 결과가 교수의 업적평가에 반영될 때 공정성의 측면에서 문제가 있고, 또한 강의평가 결과가 교수의 효과성에 대한 정보를 제공하여 주지 못하므로 쓸데없이 시간과 비용을 낭비하는 것이라고 할 수 있다.

우리의 대학은 미국과 문화와 대학환경의 차이가 있다. 미국의 강의평가 연구 결과가 교수효과성에 긍정적인 영향을 미친다고 나왔다 하더라도 우리의 현실에서는 그 결과가 다를 수 있다. 따라서 이제는 미국에서 개발한 문항을 빌려다 쓰는 형태가 아닌 우리의 환경과 문화 상황에 맞는 강의평가를 할 필요가 있다. 이제 우리 대학도 강의평가를 시행하지 10년 가까이 되므로 기초 자료도 어느 정도 축적되었고, 이런 자료를 바탕으로 강의평가에 대한 연구를 하여 좀더 현실적이고 실제 교육에 도움이 되는 강의평가를 개발할 필요가 있다. 본 연구는 제한적이기는 하지만 이런 측면에서 강의평가에 대한 자료를 분석하고 대학

사회 구성원의 의견을 조사하여 강의평가에 대한 기초 자료를 제공하고 개선안을 제시하려 한다.

2. 연구의 목적

본 연구는 현재의 강의평가의 문제점으로 제기되는 문제를 검증하고, 합리적이고 효과적인 강의평가제도 정책을 수립하기 위한 기초 자료를 제공하기 위하여 강의평가의 타당도와 신뢰도를 분석하려 한다. 본 연구에서는 강의평가가 좀더 효과적으로 활용되기 위한 기초연구로서 선행연구 분석과 강의평가에 대한 의견 조사, 그리고 강의평가 결과의 분석을 통하여 강의평가의 문제점을 제시하고 강의평가의 문제점을 극복할 수 있는 대안을 제시하려 한다.

먼저 대학에 적절한 강의평가 제도의 문제점을 탐색하기 위하여 강의평가에 대한 대학 구성원인 학생과, 교수의 강의평가에 대한 의견 조사를 시행하였다. 또한 현재 대학에서 시행하고 강의평가 결과를 분석하였다. 강의평가의 타당도와 신뢰도를 중심으로 최근 4학기에 시행된 강의평가 결과를 분석하고 강의평가의 타당도와 신뢰도를 제고할 수 있는 방안을 탐색하였다.

II. 강의평가 연구

강의평가는 교수들에게 수업 효과성에 대한 진단적 피드백을 하고, 교수에 관한 의사결정(승진이나 재임용등)을 할 때 사용할 수 있는 교수효과성에 대한 정보를 주고, 학생들이 수업이나 강사를 선택할 때 사용할 수 있는 정보를 제공하고, 수업에 연구에 도움이 되는 자료를 제공한다. 최근에 미국뿐만 아니라 우리나라에서도 강의평가제도가 일반화되어 사용되고 있으나 아직도 우리나라 내에서는 강의평가의 타당도, 신뢰도, 그리고 강의평가의 활용 등에 대한 종합적이고 체계적인 연구가 부족하다. 미국의 연구는 60년대 70년대 도입과정에서 수많은 찬반양론에 대한 연구를 거쳐 최근에 강의평가는 기정사실화 되었고, 강의평가에 대한 연구는 강의평가 도구의 개발, 관련된 구인의 해석, 자료 수집, 논리성에 대한 연구에 집중되고 있다. 강의평가를 할 것인가 말 것인가에 대한 논의보다는 교수효과성 구인을 어떻게 구체화할 것인가, 합리적인 강의평가도구의 개발, 강의평가 결과의 합리적인 해석과 활용에 논의의 초점이 맞추어지고 있다.

강의평가 연구는 크게 강의평가의 타당도와 신뢰도, 강의평가의 유용성, 그리고 강의평가

의 편파성(잠재적인 혼동요소)에 대한 연구로 나누어 볼 수 있다. 강의평가에 대한 연구는 수천 개에 달하지만 실제로 강의평가의 타당도, 신뢰도, 유용성 등에 대한 이런 연구의 결론은 명확하지 않다(Marsh, 1984; Cohen, 1981; Whitely & O.Doyle, 1979).

1. 강의평가의 타당도

다른 평가 도구와 마찬가지로 강의평가에서도 타당도의 문제가 가장 중요하다. 타당도는 평가목적과 관련하여 평가도구가 평가목적에 부합되는지의 정도를 나타낸다. 강의평가 도구를 타당하려면 강의평가 목적이 무엇인지 명백히 기술하고, 강의평가의 구인을 구체화하여 조작적으로 정의하여야 한다. 강의평가의 목적은 크게 세 가지로 대별할 수 있다. 첫째, 강의평가 결과를 각 교수들에게 피드백을 주어 강의의 개선을 하려는 것이다. 둘째, 교수의 능력(교수 효과성)을 정확히 측정하여 보상과 처벌을 위한 기초 자료로 활용하고자 함이다. 셋째, 학생들이 과목 선택을 할때 도움이 되고자 교수와 교과목에 대한 정보를 학생들에게 제공하고자 함이다. 강의평가 도구가 타당하려면 강의평가 결과가 각 교수들에게 피드백 되어 수업의 개선이 있었는지에 대한 자료가 필요하고, 또한 강의평가 결과가 피드백 되어서 학생들의 학업성취도가 증가되었는지에 대한 자료가 필요하다. 또한 교수효과성에 대한 조작적인 정의를 내리고 강의평가 결과가 그런 구인을 정확히 측정하는지에 대한 자료가 있어야 한다. 이런 타당도를 입증하는 방법은 크게 내용타당도(논리적 타당도), 구인타당도, 준거관련 타당도가 있다.

내용타당도는 강의평가의 내용이 강의평가의 목적에 부합하는지, 교수효과성을 적절히 평가하는지에 대한 논리적인 판단에 근거하여 타당도를 입증한다. 내용타당도는 강의평가와 관련된 전문가들이 논리적으로 판단하여 적절성을 입증하기도 하고, 강의평가의 당사자들인 교수와 학생들이 강의평가 도구가 강의평가의 목적에 맞게 구성되었는가를 판단하게 하기도 한다(이런 경우는 안면타당도라고 하기도 한다). 이런 연구는 교수효과성이 무엇인가, 그리고 이런 교수효과성을 나타내는 지표는 무엇인가에 대한 논의에서 시작되었다(Hoyt & Cashin, 1977). Soloman(1966), Whitlock(1966)등의 초기 연구는 이상적인 교수의 태도와 행동을 기술하는 174개의 문항을 개발하고 타당화 하려 하였다. 그러나 이런 노력은 교수의 개인차를 고려하지 않고, 강의 목표나 특성에 따른 차이를 고려하지 않았다는 측면에서 비판을 받았다. 이런 비판은 우리대학의 강의평가가 일률적으로 똑같이 적용된다는 측면에서 우리에게도 해당된다. 최근의 연구는 효과적인 교수의 지표를 학생에게 미치는 효과, 즉 학습효과, 학습발전에 초점을 맞추고 강의 주제, 강의내용의 수준, 교수목표 등에 따라 다른 도구 개발

에 힘쓰고 있다. 특히 IDEA¹⁾는 교수효과성을 나타내는 지표를 학습발전에 두고, 교수가 선택한 수업목표에 따라 다른 강의평가 도구를 개발하는 연구에 집중하고 있다.

타당도를 입증하는 또 다른 방법은 구인타당도이다. 강의평가의 타당도는 강의평가가 과연 교수효과성을 실제로 측정하는가의 문제이다. McKeachie(1979)는 교수효과성을 교수가 학생의 학업 성취를 얼마나 조장(facilitate)가의 정도로서 정의하였다. 특정의 수업에서 학생이 학습한 양으로 구체화 될 수 있다(Cohen, 1981). 구인타당도를 입증하기 위하여 실제로 강의평가 결과가 과연 학습량의 증가와 관련이 있는가에 대한 연구가 진행되었다. 구인타당도의 연구는 강의평가가 과연 일차원성의 가정을 충족하는가 아니면 다차원성(multidimensionality)인가, 즉 강의평가가 하나의 구인을 측정하는가 다양한 구인을 측정하는가의 문제에 집중되었다. 대부분의 연구는 교수효과성을 하나의 구인이 아닌 다양한 구인으로 구성되어 있다는 것을 밝히고 있다(Marsh, 1984). 교수효과성에 관련된 구인을 조작적으로 정의하고 요인분석을 통하여 이런 구인의 적절성을 검증하는 연구가 이루어지고 있다. 요인분석은 학생들이 교수효과성의 다른 구인을 구별할 수 있는지, 경험적 자료가 실제로 의도한 구인과 일치하는지에 대한 자료를 제공해 준다. 최상일 외(2001)는 포항공대의 강의평가의 양호도를 검증하기 위하여 효과적인 수업의 구인을 수업조직 및 전반, 교수-학생관계, 학습결과 및 만족, 수업진행, 시험/과제의 적절성으로 정의하고 요인분석을 통하여 이런 구인의 타당성을 검증하였다. 김명화(2001)의 연구도 인하대학교의 강의평가 도구의 타당도를 검증하기 위하여 요인분석을 하여 구인의 적절성을 검증하였다.

또 다른 구인타당도에 대한 연구는 강의평가가 교수효과성 이외에 다른 요소 학생특성, 교수특성, 과목 특성 등의 외적 변인에 의해 영향을 받는가에 대한 수많은 연구가 이루어졌다. Ting(2000)은 이런 외적변인으로 교수차원의 결정요소, 학생차원의 결정요소, 수업차원의 결정요소, 강좌내용차원의 결정요소를 꼽았다(성균관대학교 대학교육개발센터, 2002에서 재인용). Cashin(1995)가 2000개 이상의 강의평가 관련 연구를 메타분석하고 교수의 나이, 교수경험, 교수의 성별, 인종, 성격, 연구 성과와 같은 교수요인은 강의평가와 관련이 없다고 결론지었다. 또한 학생의 나이, 학생의 성별, 학생의 평점, 학년과 같은 학생요인은 강의평가와 관련이 없고, 수업에서 학생수, 강의의 시기, 강의평가가 시행되는 시기 등은 강의평가와 관련이 없다고 하였다. 강의평가와 관련이 있는 요인은 교수의 열정과 에너지, 교수의 자아존중감, 학생의 학습동기, 학생이 예상하는 성적, 학문의 성격, 과목의 수준, 수업의 난이도이라고 결론지었다. 그러나 실제로 많은 연구는 이런 Cashin(1995)의 연구와 다른 결과를 내기도 한다. 성균관대학교 대학개발센터(2002)의 연구와 김명화(2001)의 연구에서는 교수의 연령

1) IDEA(Individual Development and Educational Assessment)는 캔자스 주립대학에 있는 교수평가·개발센터이다.

에 따라 다른 강의평가 결과를 얻었고, 학생의 학년이 높아질수록 강의평가 결과가 더 높게 나타나는 경향이 있으며, 학생수가 많은 대규모 강좌보다는 소규모 강좌가 더 강의평가 결과가 높았다.

타당도를 입증하는 방법에는 강의평가 결과와 교수효과성을 나타내는 다른 지표들 또는 준거와 적절한 상관을 통하여 검증하는 준거관련 타당도가 있다. 강의평가의 타당도를 검증하는 연구들은 강의평가와 동료교수들의 평가, 행정가의 평가, 교수의 자기평가, 동창회의 평가, 학생의 학업성취도와의 상관관계를 준거타당도를 입증하려 하였다.

교수의 자기평가와 강의평가 결과와의 상관은 Centra(1973)의 연구에서는 .3, Marsh(1982)의 연구에서는 .62, 최상일의(2001)의 연구에서는 평균 상관계수 .36을 보이고 있다. 주로 50명 이상의 교수자를 사용한 대규모 연구에서는 상관이 더 높아지는 경향이 있다(Marsh, 1984). 동료평가와 강의평가의 상관은 동료교수가 강의를 관찰하고 하는 것과 관찰하지 않고 평가한 결과 두가지를 사용하는데 교수가 강의를 관찰하고 동료교수를 평가한 결과는 강의평가와 유의미한 상관을 보이지 않았다(Marsh, 1984). Morsh 외(1956)의 연구는 동료평가, 행정가평가의 결과는 서로 유의미한 상관을 보이지만 강의평가 결과나 학생의 학업성취도와는 유의미한 상관을 보이지 않는다는 것을 보고했다.

2. 강의평가의 신뢰도

신뢰도는 강의평가 결과가 얼마나 믿을만 한가, 얼마나 오차가 적으냐의 정도를 나타낸다. 신뢰도를 검증하는 방법에는 검사결과와 안정성을 검증하는 검사-재검사 신뢰도, 검사의 동형성을 측정하는 동형검사 신뢰도, 문항간의 일관성을 측정하는 문항내적 일치도가 있다. 강의평가에서 신뢰도는 시간이나 환경요소에 따라 변하지 않는 검사의 안정성과 얼마나 일관되게 피험자가 응답을 하는가를 나타내는 일관성이 중요하다.

강의평가의 안정성은 학생들이 같은 교수에 대해서 시간이 지난 후에도 얼마나 같은 평정을 하는가에 대한 정도이다. 이런 안정성은 한번 평정을 하고 일정한 시간이 지난 후에 평정을 하게 하여 두 평정점수의 상관계수로 검증한다. 강의평가에서 안정성은 강의 중간에 학생들에게 강의평가를 하고, 강의가 끝날 때 쯤 다시 한번 강의평가를 시행하여 두 평가의 상관계수로 표현한다. Costin(1971)은 심리학, 사회과학, 인류학, 물리학, 생물학 분야에서 학기 중간과 학기말에 강의평가를 두 번 실시하여 상관계수를 계산하였는데 $r = .70-.87$ 이었다. Guthrie(1954)의 연구는 1년 후에 다시 강의평가를 시행하였는데 상관계수가 .87-89이었다. Marsh외(1984)는 강의평가의 장기적인 안정성을 검증하기 위하여 현재 강의를 들은 학생과

졸업생과의 강의평가 결과의 상관계수를 계산하였는데 상당히 높은 상관을 얻었다. 이런 안정성에 대한 연구는 강의평가가 시간의 변화에도 꽤 안정적이라는 것을 보여주고 있다 (Marsh, 1984).

문항의 일관성을 검증하는 방법은 문항간의 내적일치도 계수인 Cronbach α 를 계산하거나, 급내 상관계수(intraclass correlation)를 계산하는데 잘 개발된 강의평가지는 높은 상관계수가 나온다(최상일외, 2001). Hoyt & Cashin(1977)는 IDEA 시스템의 신뢰도를 검증하기 위하여 184개의 수업에 참가한 학생을 홀짝으로 나누고 홀수의 강의평가 결과와 짝수의 강의평가 결과를 비교하였는데 .81-.94로 평균 상관계수는 .86이었다.

Marsh(1984)는 강의평가 결과의 일반화가능도(generalizability)를 알아보기 위하여 같은 교수가 같은 과목을 가르칠 때, 같은 교수가 다른 과목을 가르칠 때, 다른 교수가 같은 과목을 가르칠 때, 다른 교수가 다른 과목을 가르칠 때의 상관계수를 계산하였다. 같은교수가 같은 과목을 가르칠때의 상관은 .63-.73이고, 같은교수가 다른과목을 가르칠때의 상관은 .40-.61이고, 다른교수가 같은과목을 가르칠때의 상관은 -.051-.39이고, 다른교수가 다른과목을 가르칠 때의 상관은 -.06-.22이었다.

Ⅲ. 연구방법

강의평가의 타당도와 신뢰도를 분석하기 위하여 설문조사와 I 대학의 강의평가 결과를 분석하였다.

1. 설문조사

강의평가에 대한 의견 및 개선안을 조사하기 위하여 I 대학에 재학중인 학생과 I 대학에 재직 중인 교수를 대상으로 설문조사하였다.

1) 강의평가에 대한 학생들의 의견 및 개선안 조사

(1) 설문조사 방법

설문대상은 I대학의 재학생 1060명으로 단과대 및 학년별 빈도는 아래의 <표 1>과 같다. 설문지는 주로 주요 단과대들이 공통으로 수강하는 대규모 강좌의 수업 후에 설문지를 배포하고

<표 1> 단과대 및 학년별 응답자 분포

	1학년	2학년	3학년	4학년	합계
공대	64	190	211	4	469
이과대	9	57	24	6	96
경상대	11	31	10		52
경영대	27	41	37	2	107
사범대	28	6	9		43
법대	3	24	12		39
사회과학대	18	48	12	1	79
문과대	28	76	14	5	123
생활과학대	3	34	6	1	44
의대		8			8
합계	191	515	335	19	1060

설문에 응답하도록 하였다. 설문조사 시기는 2003년 11월15일부터 2003년 11월 30일까지 설문지를 배포하고 수집하였다. Spss10.0을 사용하여 각 응답의 빈도와 분포를 분석하였다.

(2) 설문조사 내용

- 강의평가가 수업 향상에 미치는 효과
- 강의평가와 예상 성적과의 관계
- 학습동기와 강의평가의 관계
- 학생수와 강의평가의 관계
- 수업방법의 차이에 따른 강의평가 결과의 차이
- 수업의 난이도와 강의평가와의 관계
- 교과목의 특성(전공, 교양)과 강의평가의 관계
- 현재 강의평가의 문제점
- 강의에 대하여 깊이있게 생각한 후 평정하는지의 여부
- 강의평가 문항을 자세히 읽고 평정하는지의 여부

2) 강의평가에 대한 교수들의 개선방안 조사

(1) 설문조사 방법

I대학에 재직 중인 교수 98명이 응답하였다. 단과대 및 교수 직급별 빈도는 아래의 <표 2>에 제시하였다. 각 학과에 설문지를 배포하고 설문에 응답후 조교들이 수집하여 교수학습센터에 전달하도록 하였다. 설문조사 시기는 2003년 11월15일부터 2003년 11월 30일까지 설

<표 2> 단과대 및 교수직급별 빈도분포

	교수		부교수		조교수		전임강사	
	빈도	비율	빈도	비율	빈도	비율	빈도	비율
공대	19	39.6	6	28.6	7	33.3	3	37.5
이과대	4	8.3	2	9.5	1	4.8		
경상대	1	2.1	1	4.8				
경영대	4	8.3						
사범대	7	14.6	3	14.3	6	28.6	2	25.0
법대	2	4.2	1	4.8				
문과대	10	20.8	7	33.3	6	28.6	2	25.0
생활과학대	1	2.1						
의대			1	4.8	1	4.8	1	12.5
합계	48	100.0	21	100.0	21	100.0	8	100.0

문지를 배포하고 수집하였다. Spss10.0을 사용하여 각 응답의 빈도와 분포를 분석하였다

(2) 설문조사 내용

- 강의평가 결과를 교수가 수업개선에 활용하는지의 여부
- 수강생의 수에 따라 강의평가에 가중치를 두어야 하는지의 여부
- 수업방법에 따라 다른 가중치를 두어야 하는지의 여부
- 현재 강의평가의 문제점

2. 강의평가 결과 분석

1) 강의평가지의 구성

각 문항은 5단계 Likert 척도에 따라 ‘매우긍정적’, ‘긍정적’, ‘보통’, ‘부정적’, ‘매우부정적’으로 구성되어 있고, 해당강좌에 없거나 부적절한 문항은 ‘해당없음’란에 표하도록 하였다.

강의 평가지는 객관식 21문항, 주관식 학생 의견 수렴 내용으로 구성되어 있다. 객관식 문항은 교수방법에 대한 평가 12문항, 학습성취에 대한 평가 5문항, 강의 전반에 대한 평가 4문항으로 구성되어 있으며, 주관식 문항은 5개의 문항으로 수업 전체에 대한 의견과 이유를 직접 기술하도록 하였다. 강의평가지는 강좌의 특징과 교수방법, 수업환경 등을 고려하여 A형부터 J형까지 10개의 유형으로 분류되어 있다. 객관식 문항은 평가지 유형에 따라 다른 질문이 제시되고 주관식 문항은 유형에 상관없이 똑같은 문항이 제시된다. 본 연구에서는 객관식 문항 21개만을 분석하였다.

- A형은 인문사회 일반강좌에서 사용
- B형은 이공계 일반강좌에서 사용
- C형은 토론 중심의 수업
- D형은 이공계 전공강좌에서 사용
- E형은 실습중심의 수업에서 사용
- F형은 중단위 이하의 강좌중 강의 중심의 수업
- G형은 중단위 이상의 강좌중 강의 중심의 수업
- H형은 조별 활동을 활용하는 수업
- I형은 실험실습 중심의 수업용
- J형은 영어회화 수업

본 연구에 사용된 강의평가지는 총 세 부분으로 나뉘어져 있다. 교수방법 관련 문항 12, 학습성취 관련 문항 5, 강의 전반에 관한 문항 4문항이다. 10개의 유형으로 구분되어 있으나 기본적인 틀은 이 세 구조로 이루어져 있다.

2) 분석방법

강의평가 결과의 타당도 및 신뢰도를 분석하기 위하여 2001년 2학기, 2002년 1학기, 2002년 2학기, 2003년 1학기의 I 대학의 강의평가 결과를 비교 분석하였다. 총 분석인원은 102,192 명, 2002년 1학기 109,313명, 2002년 2학기 104,920 명, 2003년 1학기 119,522, 강좌수는 5195 이고, 총 4546명의 교수를 분석하였다.

강의평가의 내용타당도를 검증하기 위하여 교수와 학생들을 대상으로 강의평가 내용의 적절성을 조사하였다. 구인타당도를 검증하기 위하여 요인분석을 통하여 강의평가의 요인구조와 자료의 요인구조의 일치성을 검증하였다. 강의평가의 신뢰도를 검증하기 위하여 학생들의 반응의 일관성과 성실성을 분석하고, 강의평가도구의 안정성을 분석하였다. 학생들의 반응의 일관성을 알아보기 위하여 급내 상관계수를 계산하고 문항 동질성을 알아보기 위하여 Cronbach α 를 계산하였다. 학생들의 반응의 성실성을 알아보기 위하여 학생들이 강의평가에 얼마나 성실하게 응답하는지에 대한 학생들의 의견조사를 하였다. 21개의 문항에 대하여 모든 문항을 같은 평정을 학생들의 비율을 조사하였다. 강의평가 도구의 안정성을 검증하기 위하여 같은 교수가 같은 과목을 가르쳤을때(107명의 교수, 112개의 교과목을 분석) 두과목의 상관계수와 다른 교수가 같은 교과목을 가르쳤을때(196명의 교수, 98개의 강좌를 분석) 상관계수를 계산하여 비교하였다.

IV. 연구결과

강의평가의 타당도와 신뢰도를 검증하기 위하여 설문조사 결과와 강의평가 결과를 통계분석 하였다.

1. 강의평가의 타당도

1) 내용타당도

(1) 설문조사 결과

① 학생설문조사결과

강의평가의 내용타당도를 알아보기 위하여 학생설문조사를 하여 <표 3>, <표 4>, <표 5>와 같은 응답결과를 얻었다.

강의평가를 시행한 후에 수업이 향상되었는지를 묻는 질문에는 431명(40.7%)이 그렇다고 응답하여 많은 학생들이 강의평가 이후에도 수업의 질이 좋아졌는지에 대해서는 부정적인 시각을 갖고 있었다. 강의평가의 기본적인 목적이 수업의 질 개선에 있으므로 이런 결과는 강의평가 도구가 과연 타당한가에 의문을 제기하게 한다.

그 강의에서 본인 받으리라고 예상되는 성적에 따라 강의평가에서 다른 평정을 하는지를 묻는 질문에 학생들은 485명(45.7%)의 학생들이 그렇다고 응답하여 약 절반의 학생들이 본

<표 3> 학생 설문 조사 결과

질문문항	매우그렇다		그렇다		그렇지 않다		전혀 그렇지 않다		합계
	빈도	비율	빈도	비율	빈도	비율	빈도	비율	
1. 강의평가 이후 수업 향상	24	2.3	407	38.4	488	46.0	126	11.9	1045
4. 예상성적과 강의평가의 관계	50	4.7	435	41.0	489	46.1	60	5.7	1044
5. 학습동기와 강의평가의 관계	151	14.2	659	62.2	213	21.2	16	1.5	1039
6. 학생수와 강의평가의 관계	18	1.7	185	17.5	697	65.8	141	13.3	1041
9. 수업방법의 차이	35	3.3	352	33.2	520	49.1	130	12.3	1037
10. 수업 난이도와 강의평가의 관계	59	5.6	417	39.3	485	45.8	79	7.5	1040

<표 4> 강의평가에서 교양수업과 전공수업이 차이

	전공수업을 잘준다	교양수업을 잘준다	전공, 교양과목 상관없이 준다	합계
빈도	71	35	935	1041
비율	6.7	6.8	88.2	100

<표 5> 학생들이 지적한 강의평가의 문제점

항목	빈도	비율	전체
강의평가 문항의 부적절	151	14.2	1060
강의평가 방법의 부적절	101	9.5	1060
강의평가 시기의 부적절	103	9.7	1060
학생이 자격 부족	31	2.9	1060
수업개선에 활용되지 않음	663	62.5	1060
너무 많은 과목을 평가해야 한다.	288	27.2	1060
기타	37	3.5	1060

인의 예상성적과 강의평가의 평정과 관련이 있음을 보여주고 있다.

교과에 대한 흥미가 있을 경우 강의평가 점수를 높게 주는 경향이 있는지를 묻는 질문에는 810명(76.4)의 학생이 그렇다고 응답하였다. 선행연구들과 같이 학생들의 교과목에 대한 흥미 또는 학습동기가 강의평가와 관련이 있음을 나타내고 있다.

학급의 크기(과목당 학생수)가 클수록 강의평가 점수를 낮게 주는 경향이 있느냐는 질문에는 203명(19.2%)만이 그렇다고 응답하여 학급의 크기는 학생들의 강의평정 점수에 큰 영향을 미치지 않는다고 할 수 있다.

수업방법이 차이(강의식, 실습수업, 세미나 수업등)에 따라 다른 점수를 주는 경향이 있는지를 묻는 질문에는 387명(36.5%)가 그렇다고 응답하였다.

수업부담과 난이도에 따라 다른 점수를 주는지를 묻는 질문에는 476명(44.9%)이 그렇다고 응답하여 학생들은 수업의 난이도에 따라 다른 점수를 주는 경향이 있는 것으로 나타났다.

강의평가에서 교양수업과 전공수업의 차이에 따라 다른 점수를 주는 지를 묻는 질문에는 전공, 교양과목 상관없이 평정한다는 학생이 935명(88.2%)로 교양수업과 전공수업에 대하여는 큰 차이가 없었다.

현재 강의평가의 문제점을 묻는 질문에 학생들은 강의평가 결과가 수업개선에 활용되지 않는 것이 가장 큰 문제점으로 지적하였다. 또한 한꺼번에 너무 많은 과목을 평가해야 한다는 것과 강의평가 문항이 부적절하다는 것 순으로 문제점으로 지적하였다.

강의평가에 대한 교수들의 의견조사를 하여 <표 6>, <표 7>과 같은 응답결과를 얻었다.

<표 6> 교수들이 수업개선에 활용여부 조사 결과

질문문항	매우그렇다		그렇다		그렇지 않다		전혀 그렇지 않다		합계
	빈도	비율	빈도	비율	빈도	비율	빈도	비율	
1. 수업개선에 활용여부	12	12.0	67	67.0	15	15.0	3	3.0	97

질문사항	가중치를 둬		가중치를 두지 않아야 함		기타		합계
	빈도	비율	빈도	비율	빈도	비율	
5. 수강생의 수에 따라 가중치를 둬	64	66.7	29	30.2	3	3.1	96
6. 교양수업과 전공수업에 따라 다른 가중치	65	67.0	29	29.9	3	3.1	97
7. 수업방법에 따라 다른 가중치를 둬	53	55.2	42	44.8	1	1.0	96

<표 7> 교수들이 지적한 강의평가의 문제점

항목	빈도	비율	전체
강의평가 문항의 부적절	27	27.4	100
강의평가 방법의 부적절	8	8.0	100
강의평가 시기의 부적절	15	15.0	100
학생이 자격 부족	10	10.0	100
수업개선에 활용되지 않음	20	20.0	100
너무 많은 과목을 평가해야 한다.	52	52.0	100
기타	4	4.0	100

교수들이 강의평가 결과를 수업 개선을 위한 자료로서 활용하는지를 묻는 질문에 79명(79%)가 그렇다고 응답하여 많은 교수들이 강의평가 결과를 수업 개선에 활용하고 있는 것으로 나타났다. 이런 교수들의 응답은 학생들의 응답과 차이를 보이고 있는데 교수들은 강의평가 결과를 수업 개선에 활용하고 있으나 학생들은 교수들의 이런 노력을 피부로 느끼지 못하는 것으로 나타났다.

수강생의 수에 따라 강의평가 결과에 가중치를 두어야 하는지를 묻는 질문에 64명(66.7%)가 가중치를 두어야 한다고 응답하였다. 교양수업과 전공수업에 다른 가중치를 두어야 하는가를 묻는 질문에 65명(67%)이 가중치를 두어야 한다고 응답하였다. 또한 수업방법에 따라 다른 가중치를 두어야 하는가를 묻는 질문에도 53명(55.2%)가 가중치를 두어야 한다고 응답하여 교수들은 수업의 규모, 교과목의 특성, 수업방법 등이 강의평가 결과에 영향을 미친다고 생각하고 있었다.

교수들이 지적한 현재 강의평가의 문제점은 한꺼번에 너무 많은 과목을 평가해야 하기 때문이 가장 많고, 다음이 강의평가 문항의 부적절이 많았다. 이런 결과는 학생들의 의견과의 차이가 있는데 교수들은 수업개선에 활용되지 않는 것보다는 강의평가의 신뢰도에 의문을

갖고 있는 듯하다.

2) 구인타당도

강의평가지의 구조를 파악하기 위하여 탐색적 요인분석(주성분분석)을 하였다. 요인분석 결과 한개의 요인만이 추출되었다. 강의평가지는 교수방법, 학습성취, 강의전반으로 세 개의 요인구조를 가정하고 있으나 실제 자료는 한개의 요인만이 추출되었다. 한개의 요인의 설명량은 75.70으로 하나의 요인이 전체 분산을 설명하는 양이 상당히 많다. 그러므로 본 강의평가지는 구인타당도가 낮다고 할 수 있다.

<표 8> 강의평가 요인분석결과

문항번호	1.요인
전1	.936
전4	.912
전2	.906
학1	.900
교2	.879
학2	.875
교12	.873
교4	.865
교3	.861
학3	.856
전3	.855
교5	.854
교1	.850
교9	.832
교8	.830
교6	.826
학5	.809
교7	.797
교11	.794
학4	.772
교10	.771
고유값	15.62
설명된 총 분산	75.70

주: 전- 강의전반, 교-교수학습, 학-학습성취를 나타냄

2. 강의평가의 신뢰도

강의평가의 신뢰도를 알아보기 위하여 설문조사와 강의평가 결과를 분석하였다.

1) 설문조사 결과

학생들을 대상으로 한 설문조사 결과는 <표 9>과 같다.

<표 9> 학생들이 제시한 강의평가의 신뢰도 및 성실성

질문문항	매우그렇다		그렇다		그렇지 않다		전혀 그렇지 않다		합계
	빈도	비율	빈도	비율	빈도	비율	빈도	비율	
강의에 대한 숙고후 평정	64	6.0	639	60.3	305	28.8	31	2.9	1047
강의평가 문항을 자세히 읽고 평정	84	8.0	646	61.8	289	27.6	27	2.6	1046

강의에 대하여 깊이 있게 생각한 후에 강의평가의 평정을 하는가라는 질문에 703명(66.3%)가 '그렇다' 하고 응답하였다. 약 33%의 학생들이 강의에 대한 생각없이 그냥 평정을 하고 있는 것으로 나타났다. 그리고 강의평가 문항을 자세히 읽고 평정을 하는가 에 대한 응답은 730명(69.8%)이 '그렇다'라고 응답하였다. 약 30%의 학생들이 강의평가 문항을 읽지 않고 평정을 함을 알 수 있다.

2) 강의평가 결과분석 결과

강의평가의 신뢰도를 검증하기 위하여 학생들의 반응의 일관성과 성실성을 분석하고, 강의평가도구의 안정성을 분석하였다. 학생들의 반응의 안정성을 알아보기 위하여 같은 과목에 같은 교수에 대하여 얼마나 일관된 평정을 하는가를 분석하였다. 강의평가 결과의 일관성을 알아보기 위하여 문항내적 일치도인 Cronbach α 와 급내 상관계수를 조사하였다. 학생들의 반응의 성실성을 알아보기 위하여 21개의 문항에 대하여 모든 문항을 같은 평정을 학생들의 비율을 조사하고, 학생들이 강의평가에 얼마나 성실하게 응답하는지에 대한 의견조사를 하였다.

(1) 안정성

안정성계수는 같은 피험자가 일정한 시간이 지난 후에 얼마나 같은 평정을 하는가를 측정하는 지수이다. 그러나 여기서 제시한 상관계수는 학기별로 다른 피험자가 평정한 결과이므로

로 상관계수가 낮게 추정될 것이라고 예상된다. 그러나 Marsh(1984)의 연구에서 다른교수가 다른과목을 가르칠때의 상관도 $-.06 \sim .22$ 인데 비하여 우리의 연구결과는 상당히 낮은 상관계수를 보여주고 있다. 같은 과목을 같은 교수가 담당할 경우에 2002년 2학기과 2003년 1학기 강의평가 결과의 상관계수는 $.068$ 로 상당히 낮았고, 2002년 2학기과 2003년 1학기에 같은 과목을 다른 교수가 가르친 경우에도 $.042$ 로 상관계수는 상당히 낮았다. 따라서 강의평가 결과의 안정성은 매우 낮다고 할 수 있다.

(2) 일관성

학생들의 반응의 일관성을 알아보기 위하여 2003년 1학기 강의평가 결과를 사용하여 과목별로 문항들간의 상관정도 지수인 Cronbach α (문항 내적일치도)와 급내 상관계수를 계산하였다. 2003년 1학기 총 강좌수는 5195 이고, 총 4546명의 교수의 수업을 요인별로 분석하였다. 각 요인별 문항내적 일치도 계수 및 문항별 기초통계는 <표 10>와 같다.

<표 10> 각 요인별 문항내적 일치도 계수 및 문항별 기초통계

요인 및 문항	문항평균	문항표준편차	문항과 총점의 상관	문항제거시 α
I. 교수방법 Cronbach $\alpha = .975$				
1. 강의자료	3.89	.90	.92	.97
2. 교수의 설명	3.88	.87	.90	.97
3. 교수의 예시 제공	3.80	.91	.86	.97
4. 질문과 문제점 제기	3.86	.92	.91	.97
5. 질문유도	3.83	.93	.91	.97
6. 교수의 답변	3.80	.93	.89	.97
7. 강의열의	3.76	.94	.87	.97
8. 학습곤란 대처	3.69	1.09	.77	.97
9. 시험 및 과제 처리	3.78	1.00	.81	.97
10. 면담기회	3.84	.92	.85	.97
11. 강의시간 준수	3.85	.95	.87	.97
12. 학생에 대한 관심	3.83	.96	.84	.97
II. 학습성취 Cronbach $\alpha = .934$				
13. 지식의 성취도	3.83	.97	.84	.92
14. 학습내용의 적절성	3.82	.98	.85	.91
15. 평가방법	3.80	1.01	.83	.92
16. 과제물의 도움정도	3.83	.98	.79	.93
17. 시험문제와 강의내용의 연관성	3.83	.97	.81	.92
III. 강의전반 Cronbach $\alpha = .923$				
18. 전체적인 평가	3.80	.99	.84	.89
19. 내용	3.73	1.04	.80	.91
20. 교재 선택	3.87	.95	.80	.91
21. 강의방법	3.86	.94	.86	.89

<표 11> 2003년 1학기 요인별 내적일치도 및 급내 상관계수의 범위

요인	Cronbach α 의 범위	급내 상관계수의 범위
교수방법(12문항)	.89 - .99	.25 - .92
학습성취(5문항)	.88 - .96	.20 - .94
강의전반(4문항)	.77 - .94	.22 - .95

Cronbach α 는 일반적으로 .80이상이면 신뢰도가 높다고 해석되는데 2003년 1학기 분석결과 교수방법 .89-.99로 상당히 높았고, 학습성취와 강의전반 요인이 Cronbach α 도 높은편이다. 따라서 강의평가지에 사용된 질문들이 각 요인은 문항의 동질성이 매우 높다고 해석할 수 있다.

급내 상관계수(intraclass correlation coefficients)는 같은 문항에 대해 같은 수업을 수강한 학생들의 평가가 얼마나 일치하는가를 파악하는 지수로 일반적으로 평가인원이 많을수록 높은 상관이 예상된다.

Cashin(1995)의 연구에서는 응답자가 10명일 경우 .59, 20명인 경우 .89, 40명 이상일 경우 .91의 급내 상관계수를 보였으나 본 연구의 결과는 응답자의 수가 많아져도 급내상관계수가 높아지는 경향을 보이지는 않았다. 학생수가 117명인 경우에도 급내상관계수가 .47을 보인 과목도 있었고, 학생수가 12명인 경우에도 급내상관계수가 .92인 경우도 있었다. 수업에 따라서 다른 급내 상관계수를 보이고 있으나 일반적으로 학생수가 40명이하인 과목은 낮은 급내 상관계수를 보이고 있다. Cashin(1995)는 .70이상이면 문항에 대한 학생들의 의견이 일관성을 보이는 것으로 해석하고 있다. 본 연구에서 급내 상관계수가 .70이상인 수업은 전체 수업 5195개중 4523로 87.06%의 비율을 보였다.

(3) 반응의 성실성

설문 조사 결과 많은 교수들이 강의평가의 문제점으로 학생들의 반응의 성실도에 문제가 있다는 지적을 하였다. 현재까지 반응의 성실도를 분석한 연구는 없다. 그러나 연구자가 자료를 분석하면서 많은 학생들이 모든 문항에 같은 평정을 한 것을 보고 실제로 이런 응답이 얼마나 되는지를 분석하게 되었다. 21개의 문항에 대하여 2001년 1학기부터 2003년 1학기까지 모든 문항을 같은 평정을 학생들의 비율을 조사하였다. 조사 결과는 아래와 같다. 매학기 거의 50%의 학생이 21개 모두 같은 평정을 하였으므로 학생들의 반응의 성실성은 매우 낮다고 할 수 있다. 50%의 학생들 중에는 실제로 그렇게 생각하여 모든 문항에 같은 평정을 한 학생들도 포함될 가능성이 있으나, 약 50%의 학생들이 모든 문항에 같은 평정을 한다는 것은 강의평가 반응에 대한 성실성을 의심할만한 수치이다.

2001년 2학기 : 102,192 명 중 54,326 명 (53.1%)

2002년 1학기 : 109,313 명 중 53,461 명 (48.9%)

2002년 2학기 : 104,920 명 중 55,153 명 (52.5%)

2003년 1학기 : 119,522 명 중 60,227 명 (50.3%)

V. 결론 및 논의

1. 결론 및 강의평가 개선안

강의평가를 하는 것은 강의평가를 통해서 교수효과성에 대한 정보를 얻고, 그 정보를 교수에게 피드백해서 강의의 질을 개선하고 우수한 강자에게 적절한 보상을 주어 궁극적으로 강의를 개선하기 위함이다. 본 연구는 한대학의 사례만을 사용하고, 강의평가 자료도 한정되어서 일반화에는 한계를 갖는다. 그러나 이런 한계에도 불구하고 본 연구의 결과는 강의평가에 대하여 시사하는 바가 크다. 특히 낮은 강의평가의 타당도와 신뢰도와 같은 강의평가의 문제점은 다른 많은 대학들이 공통으로 갖고 있는 문제라고 할 수 있다. 김성열외(2001)는 전국의 4년제 대학 109개를 대상으로 강의평가에 대한 설문조사를 통하여 많은 대학들이 강의평가를 내실 있게 운영하기 보다는 형식적으로 운영하고 있고, 대부분 대학들이 평가 설문지를 다른 대학 것을 베껴 강의의 특성을 제대로 반영하지 못하고 있고, 이 때문에 학생들도 건성으로 답하고 있는 악순환이 계속되고 있다고 지적하고 있다. 본 연구의 결과 I 대학의 강의평가의 타당도와 신뢰도는 매우 낮다고 할 수 있다. 이런 강의평가 결과를 자료로 활용하는 것은 문제가 있고, 잘못된 의사결정을 하게 할 수도 있다.

강의평가에서 필요한 정보를 얻으려면 강의평가의 문항내용이 타당해야 한다. 부적절한 문항과 문항 기술이 적절하지 않은 문항은 유용한 정보를 제공하지 못한다. 많은 연구들이 강의평가가 교수효과성에 대한 유용한 정보를 제공하지만 강의평가의 내용이 타당할 것을 전제로 하고 있다. 따라서 교수효과성에 대한 구체적이고 명확한 정의와 함께 교수효과성의 하위요인을 조작적으로 정의하고 이것을 나타내는 지표가 무엇인가를 명확하게 하는 것이 필요할 것이다. 따라서 교수이론이나 학습이론에 근거하여 교수효과성에 대한 구인을 추출하고, 평가문항을 개발하고 타당화 과정을 거쳐야 한다. Marsh(1994)의 연구에 따르면 강의평가는 하나의 구인이 아닌 여러 구인들로 이루어져 있다. 따라서 강의 평가의 개념, 목적, 범위, 평가 항목, 결과의 활용 정도에 대해 관련 분야의 타당하고 신뢰로운 전문성 있는 연구결과가 축적되고 난 후에 신중하게 개발되어야 한다. 강의평가도구는 교수효과성의 하위 구인을 구체화하고, 하나하나의 구인에 내용타당도와 구인타당도를 검증해야 한다. 또한 심

도 있는 평가를 위해 강의평가 질문은 비교적 상세하고 구체적으로 기술해야 한다(김성열외 2001).

우리의 강의평가는 평가에 포함된 질문들도 학생들이 판단하기 어려운 강의내용이나 교수 자료에 대한 것이 많다. 평가문항도 교수에 대한 평가라기보다는 강의개선에 정보를 제공하는 측면에서 ‘자신이 노력한 만큼 성적이 공정하게 주어졌나’, ‘주어진 과제가 창의적으로 수행해야 하는 것이었나’ 등과 같이 구체적인 한 부분에 대하여 학생들이 판단이 가능한 부분에 초점이 맞춰져야 한다.

또한 강의평가는 여러 교수요인, 학생요인, 환경요인과 유의미한 상관을 보이고 있다 (Marsh, 1984). 학생들의 학습동기, 예상성적, 교양과 전공과목, 강의의 난이도 등은 강의평가에 영향을 미친다는 연구 결과가 있으므로, 이런 요인에 대한 실제적인 연구를 통하여 다양한 평가지를 활용하거나, IDEA처럼 가중치를 두는 방법에 대한 고려가 필요할 것이다.

본 연구에서 검증한 바에 따르면 강의평가의 신뢰도는 매우 낮다. 많은 연구들이 강의평가의 신뢰도는 내적일치도 계수인 Cronbach α 를 제시하고 강의평가 도구가 신뢰롭다는 주장을 하였으나, Marsh(1984)는 이런 내적 일치도 계수가 다른 학생들을 사용함에 의하여 분산이 커지는 오차로 인하여 과대 추정되는 경향이 있다고 지적하고, 강의평가의 신뢰도 지수로서 내적일치도 계수를 사용하지 말 것을 주장하고 있다. 본 연구에서도 내적일치도 계수는 비교적 높으나 다른 신뢰도 계수인 안정성계수는 매우 낮았다. 또한 많은 교수들이 우려하는 바와 같이 강의평가 반응의 성실성이 매우 낮았다. 학생들 자신이 설문조사에서 응답한 것도 실제로 강의평가에 문항을 잘 읽고 응답하는 것이 아니고, 강의에 대하여 생각하지 않고 그냥 응답한다는 비율이 높았다. 또한 학생들이 21문항 전부 같게 평정한 것이 매 학기 50% 정도 된다는 것은 강의평가의 신뢰도에 대하여 심각하게 고려해야 한다는 것을 보여주고 있다. 학생들이 실제로 그렇게 생각하여 모든 문항에 같은 응답을 할 수도 있으나 매 학기 약 50%의 학생들이 모든 문항에 같은 평정을 한다는 것은 강의평가 반응의 성실성을 의심할 정도로 높은 수치이다. 학생들이 강의평가에 대한 인식과 강의평가 방법에 대한 재고가 필요할 것이고 이런 낮은 신뢰도의 원인에 대한 조사와 학생들의 성실한 반응을 유도하기 위한 방법을 강구해야 할 것이다.

2. 연구의 제한점 및 제언

본 연구는 강의평가의 타당도와 신뢰도를 분석하고 강의평가 개선방안을 탐색하기 위하여 설문조사와 강의평가 결과분석을 하였다. 본 연구의 결과는 한 대학의 사례를 중심으로 이

루어졌고, 설문조사 또한 한 대학의 학생과 교수에 한정되어 이루어졌다. 그러므로 본 연구의 결과를 모든 대학의 강의평가에 일반화 하는 데는 한계가 있고, 자료의 해석에도 한계가 있을 것이다. 본 연구에서는 한정된 자료로 인하여 밝히지 못한 다음과 같은 주제에 대하여 체계적인 연구가 이루어져야 할 것이다.

강의평가와 학업성취도 또는 학습의 관계를 밝혀야 한다. 실제로 대학수업에서 학업성취도를 무엇으로 보고, 어떻게 측정할 것인가에 대한 논의는 미국의 연구에서도 다양하다. 일반적으로 대학의 평점과의 관계를 밝히려는 연구가 많으나, 일부 연구는 멀티섹션 강의(하나의 과목을 분반하여 여러 명의 강사가 같은 주제와 교재를 사용하는 가르치는 강의)에서 사전검사와 사후검사 사이의 획득점수를 학업성취도 지표로 하여 강의평가와 상관을 밝히려 하고 있다(Marsh, 1984). 강의평가의 목적이 수업의 개선을 통하여 학생들의 학업성취도를 제고 하고자 함이므로 실제로 강의평가를 통하여 학생들의 학업성취도가 증가하는지에 대한 경험적인 연구가 필요하다. 또한 강의평가와 교수의 자기평가, 동료평가, 졸업생의 평가, 행정가 평가와의 상관을 통하여 준거관련 타당도에 대한 연구도 이루어져야 할 것이다.

강의 이외에 강의평가에 영향을 주는 외적 요소들에 대한 체계적인 분석이 필요하다. 국내에서도 일부 연구(성균관대학교 대학개발센터, 2002; 김명화, 2001; 최상일의, 2001)는 교수요인, 학생요인, 환경요인에 대한 관련성을 입증하고 있으나 체계적인 연구는 미흡한 편이다.

강의평가에 대한 연구는 그 결과가 연구에 따라 다르다. 따라서 이런 연구를 아우르고, 강의평가의 실제에 시사를 줄 수 있는 메타연구가 필요하다. 외국에서는 강의평가에 대한 연구가 축적되어 이미 1980년대부터 이런 메타연구가 나오고 있어서 강의평가의 효과에 대한 긍정적인 결과를 얻었다. 또한 관련요인을 밝혀서 이런 부분에 대한 고려가 실제에 반영되고 있다. 우리나라는 1990년대부터 강의평가가 시작되었으나 강의평가에 대한 연구는 적고, 강의평가에 대한 체계화된 연구는 거의 없다.

참고문헌

- 김명화(2002). 강의평가 자료분석을 통한 강의평가지 개선방안. *교육문제연구*, 16, 81-101.
- 김성열, 박재완, 김종철, 강현석(2001). 대학 학사과정 강의평가제 실태분석을 통한 교육업적 평가모형 개발연구. *교육인적 자원부 정책연구2001-일-30*.
- 김성훈(2001). 수업개선을 위한 수업평가. 강의평가제의 개선을 위한 워크샵. 인하대학교 교수학습센터.
- 성균관대학교 대학교육개발센터(2002). 성균관대학교 강의평가 제도 연구. 성균관대학교 대학교육개발센터 대교 2002-2-1.
- 송미영, 지은림(1994). 강의평가 설문지에 관한 문항분석 연구. *교육평가연구*, 17(2), 263-283.
- 인하대학교 교수학습센터(1998). 강의평가 설문지 유형 개발에 관한 연구결과 보고서. 인하대학교.
- 조벽(1997). 강의평가. *공학 교육과 기술*, 4(3), 66-73.
- 최상일, 이은실(2001). 학생수업평가 도구의 양호도 분석. 포항공과대학교 대학교육개발센터 연구보고 01-2.
- 홍익대학교(2001). 강의평가 설문문항의 개발에 관한 연구(연구보고서). 서울: 홍익대학교
- 황윤환, 김인수(2001). 강의평가 제도 보안을 위한 정책연구. 광주교육대학교 정책연구보고서.
- Abrami, P.C., & d'Apollonia, S.(1991). Multidimensional students' evaluations of teaching effectiveness-generalizability of "N=1" research: comment on Marsh(1991) *Journal of educational psychology*, 83(3), 411-415.
- Abrami, P.C., Perry, R.P., & Leventhal, L.(1982). The relationship between student personality characteristics, teacher ratings, and student achievement. *Journal of educational psychology*, .74(1), 111-125.
- Armstrong, J.S.(1998). Are student ratings of instruction useful. *American Psychologist*, 53, 1223-1224.
- Braskamp, L.a., & Ory, J.C.(1994). *Assessing Faculty Work*. San Francisco: Jpssey-Bass Publishers.
- Cohen, P.A.(1981). Student ratings of instruction and student achievement: A meta-analysis of multi-section validity. *Review of educational research*, 51(3), 281-309.
- Cashins, W.E.(1988). Student ratings of teaching: A summary of the research Idea Paper, No.20. *Center for Faculty Evaluation and Development Research*, 41(5), 511-535.
- Cashins, W.E.(1995). Student ratings of teaching: The research revisited.. IDEA Technical Report No.32.

- Centra, J. A. 1973. Self-ratings of college teachers: A comparison with student ratings. *Journal of Educational Measurement* 10:287-95.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51, 281-309
- Costin, F., Greenough, W.T., & Menges, R.J.(1971). Student ratings of college teaching: reliability, validity, and usefulness. *Review of educational research*, 41(5), 511-535.
- Guthrie, E.R.(1954). *The Evaluation of Teaching: A Progress Report*. Seattle: University of Washington.
- Hoyt, D.P. & Cashin, W.E.(1977). Development of the IDEA system. IDEA Technical Report No.1.
- Levinson-Rose, J., & Menges, R.J.(1981). Improving college teaching: A critical review of research. *Review of educational research*, 51(3), 403-434.
- Marsh, H. W. (1982) SEEQ: A Reliable, Valid and Useful Instrument for Collecting Students' Evaluations of University Teaching. *British Journal of Educational Psychology*, 52, 77-95.
- Marsh, H.W.(1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of educational psychology*, 76(5), 707-754.
- Marsh, H.W.(1994). Weighting for the right criteria in the instructional development and effectiveness assessment(IDEA) system: global and specific ratings of teaching effectiveness and their relation to course objectives. *Journal of educational psychology*, .86(4), 631-648.
- McKeachine, W.J.(1979). Student ratings of faculty: A reprise. *Academe*, 65, 384-397.
- Morsh, J.E., Burgess, G.G., & Smith, P.N.(1956). Student achievement as a measure of instructional effectiveness. *Journal of educational psychology*, .47, 79-88.
- Office of Institutional Research Suffolk County Community College(1998). Student ratings of instruction - report and recommendation. [Http://instsrv.sunysuffolk.edu/strate.htm](http://instsrv.sunysuffolk.edu/strate.htm).
- Schienker, D.E., Mckinnon, N.C.(1994). Assessing faculty performance using the student evaluation of instruction. ED371 667.
- Soloman, D.(1966). Teacher behavior dimensions, course characteristics, and student evaluations of teachers. *American Educational Research Journal*, 3, 35-47.
- Ting, K.(2000). A multilevel perspective on student ratings of instruction: lessons from the Chinese experience. *Research in Higher Education*, 41(5), 637-661.
- Whitlock, G.H.(1966). The evaluation of teaching effectiveness. Proceedings, Sixth Annual Forum of the Association for Institutional Research.
- Whiltly, S.E. & O.Doyle, K.J.(1979). Validity and generalizability of student ratings from

between-classes and within-class data. *Journal of educational psychology*, 71(1), 117-124.

* 논문접수 2005년 5월 17일 / 1차 심사 2005년 6월 16일 / 2차 심사 2005년 7월 15일

* 김명화: 서울교육대학교 초등교육과를 졸업하고, The University of Iowa(Iowa City)에서 석사학위를 취득하였으며, 고려대학교 교육학과에서 교육측정 및 평가로 박사학위를 취득하였다. 현재 한국교육과정평가원 교수 학습개발본부 연구원으로 재직중이며, 주요 논문으로는 "대규모 검사에서 구성형 문항을 활용하기 위한 컴퓨터 채점의 타당화", "수학수행평가 도구의 타당화: 사고력, 창의성, 학습태도의 변화를 중심으로", "학생의 인지구조 및 변화의 진단방법으로서의 Bayesian 추론망 모델의 효과성" 등이 있다.

* e-mail: hwa@kice.re.kr

Abstract

Validity and Reliability of Lecture Evaluation

Myounghwa Kim*

This study analyzed validity and reliability of lecture evaluation scale, which used in I university and surveyed students and instructors who attended in I university about lecture evaluation. Although the research about lecture evaluation in United States of America are many and systematic, it is rare in Korea.

The validity and reliability of lecture evaluation of I university are very low. And so it needs to improve upon the system of lecture evaluation and lecture evaluation scale. The result of this research may reflect on the situation of other university as there are very similar to many university in Korea in the system of lecture evaluation according to Kim's research. Therefore, we need to improve on the lecture evaluation scale and to conduct to research about lecture evaluation.

Key words: lecture evaluation, validity, reliability, lecture evaluation scale

* Korea Institute of Curricula and Evaluation