

강의 평가제가 정착되지는 않았지만 이에 대한 찬반론이 대학가에 뜨겁게 일고 있다. 현재로서는 강의 평가제에 대한 찬반론은 크게 둘로 나뉘고 있는데, 한 편에서는 강의 평가제가 대학 발전을 위해 꼭 필요하고, 다른 편에서는 강의 평가는 교수권 침해일 뿐 아니라 그 신빙성에도 의문의 여지가 많기 때문에 시일을 두고 검토해 보아야 한다는 회의적 보수론을 편다.

강의 평가



몇년 전부터 우리나라에도 본격적인 강의 평가제를 실시한다는 “개혁” 기치가 있어 왔고, 이미 몇몇 대학에서는 강의 평가제를 실시하고 있다. 아직 강의 평가제가 정착되지는 않았지만 이에 대한 찬반론이 대학가에 뜨겁게 일고 있다. 현재로서는 강의 평가제에

대한 찬반론은 크게 둘로 나뉘고 있는데, 한 편에서는 강의 평가제가 대학 발전을 위해 꼭 필요하고, 또 가능한 빠른 시일 안에 그 제도가 정착되어야 한다는 개혁적 급진론이 주장하고, 다른 편에서는 강의 평가는 교수권 침해일 뿐 아니라 그 신빙성에도 의문의 여지가 많기 때문에 시일을 두고 검토해 보아야 한다는 회의적 보수론을 편다.

그런데 이 논란에는 실용주의적 교육 철학에 대한 찬반론과는 달리, 일종의 “주도권과 자존심의 겨루기”가 개입된 인상을 풍긴다. 그 이유는 아마도 전자 쪽이 대개 교육부나 대학의 행정 담당 쪽이고, 후자는 대다수의 일선에서 강의하는 교수들이기 때문일 것이다. 그리고 필자가 지난 수 년 동안 한국의 20여 대학을 방문하면서 얻은 느낌은 많은 논란들이 강의 평가의 목적, 의미, 용도, 방법에 대한 구제성에 관한 것 보다는 새로운 체제를 일몰 단편적

으로 수용하느냐 마느냐의 감정 개입이 더 큰 듯하다. 따라서 이 논문에서는 주도권이 나 감정 문제가 아닌 순수 실용면에서, 그리고 개별적인 대학 안에서만이 아닌 국가적, 세계적 조류에 비추어 강의 평가제의 가치와 그 우려점, 그리고 향후 방향 등에 대한 객관적인 연구 결과를 소개하려고 한다.

논지를 펴기 전에 우선 강의 평가제는 단일 방식이 아니라는 것부터 소개를 해야 할 것 같다. 다시 말해 강의 평가를 하는 방법은 평가자와 평가 내용, 평가 방법이 매우 다양하다. 평가자만 보더라도 미국에서는 강의 평가를 학생들이 하는 것도 있고, 동료 교수들이 하는 것, 교육 전문인이 하는 것, 학생들이 졸업 후에 하는 것, 기업인이 하는 것, 또는 학과에 관련된 외부 인사가 하는 것, 등등 그 종류가 각양각색이다. 한 대학에서 이 중 몇 개의 평가제를 사용하기도 한다. 그리고 평가 내용을 보더라도 교수평가(teacher evaluation)에 중점을 두는 것도 있지만, 과목 자체에 대한 평가(course evaluation)에 비중을 더 크게 두는 것도 있다.¹⁾

그런데 한국에서는 이렇게 다양한 강의평가 방식 중에 단 한 가지 방식-즉 강의평가란 수업을 받는 학생들이 그 수업 담당 교수를 평가하는 것 - 만으로 인식되고 있다. 따라서 이후 이 글에서는 쓰이는 “강의 평가”는 잠정적으로 “학생들에 의한 교수/강의 평가”로 축소하여 사용하기로 한다. 거듭 강조하거니와 학생에 의한 교수 평가 외에도 강의 평가 방법은 많이 있을 수 있으며, 학교의 특성과 목표에 맞고, 또 대학, 교수, 학생 모두의 발전에 유익한 각기 고유한 평가 방법을 개발하는 것도 중요한 이슈가 되리라 믿는다.

이런 점을 염두에 두고 이 글에서는 크게 세 가지 이슈에 대해 논하고자 한다.

첫째, 외국 대학에서는 어떤 목적으로

강의 평가를 하는가? 둘째, 과연 강의 평가에 신빙성이 있는가? 셋째, 국내 대학에서도 강의 평가를 해야만 할까?

외국 대학에서는 강의평가를 왜 하는가?

강의 평가의 기원으로 말하자면 소크라테스 시대에까지 거슬러 올라갈 수 있다. 이처럼 강의 평가란 대학의 유래와 거의 같이 존속해 왔다고 해도 과언이 아니다. 그 후로 오랜 동안 강의 평가에 대해서는 양극화로 대립되어 왔는데, 대개 교수들은 반대 투쟁하고, 대학 행정 당국과 학생들은 옹호하는 쪽이었다.

미국의 경우 1924년도에 이미 강의 평가에 대한 체계적인 찬반론 연구들이 나와 있을 정도로 강의 평가제 뿐 아니라 그 찬반론에 대한 논란의 역사도 깊다. 그러나 대학별로, 또는 교수 재량에 맡겨오던 강의 평가를 1970년대부터는 거의 전국 대학 차원에서 실행해 오고 있다. 자율성을 생명으로 여기는 미국에서, 그것도 자기 존중감이 가장 높은 집단의 하나라는 교수들을 대상으로 한 강의 평가가 지역과 대학 수준에 상관없이 1970년부터 거국적으로 진행되었다는 것은 아주 특이한 현상이다.

미국에서 당연히 강의 평가제에 대해서 거국적인 논쟁이 활발하게 벌어졌었는데, 1970년부터 10여 년 넘게 들끓던 논쟁은 1980년 전후로 가라 앉고 이제는 완전 정착되었다. 아마 한 10여년 동안 강의 평가에 대한 방대한 연구결과도 누적되고, 또 찬반론의 “소용돌이” 와중의 “거품”도 가라 앉게 되어 감정보다는 “객관적 데이터”를 놓고 선택할 수 있었기 때문이 아닌가 한다. 아무튼 강의 평가에 대한 논쟁은 1980년 대 초로해서 종결되어 이제는 논란의 대상이 아니라 수강신청, 학점제 같이 대학의 당연한 절차가 되었다.

그러면 왜 강의 평가를 선택하고, 어떤 목적으로 사용하고 있는가? 미국 대학에서 강의 평가를 하는 데에는 크게 두 가지의 목적이 있다. 첫째는, 교수들 스스로가 강의의를 더 잘 할 수 있도록 학생들로부터 피드백을 얻기 위한 발전지향적 목적(formative evaluation)이고, 둘째는 대학 행정 쪽에서 교수의 능력과 기여도를 “점수로 환산하여” 승진이나 보수(급료) 책정과 같은 인사 결정을 할 때 반영하기 위한 평가지향적 목적(summative evaluation)이다.

우선 첫째 발전지향적 목표에 대한 논란과 연구 결과를 보자. 교수들 스스로가 학생들로부터 피드백을 받아 강의의 질을 발전시키는 것이 목표라면, 강의 평가를 받을수록 교수는 점점 더 훌륭한 강의를 하게 되지 않을까? 만일 그렇다면 대학교 전체의 강의 평가 평균 점수는 매 학기 조금씩 올라가야 할 것이다. 과연 그럴까?

미시간 공대의 경우는 학교 전체의 강의 평가 평균 점수가 놀랍게도 지난 10년간 조금도 달라지지 않았다. 이런 현상에 대해 원인을 알아보려고 했으나 유감스럽게도 결론을 단정지을 만한 연구가 되어 있지 않았다.²⁾ 그런데 그 원인을 규명하는 것은 쉽지가 않다. 원인이 복합적일 수 있기 때문이다. 예를 들어 강의 평가의 도구(나중에 자세히 언급하겠음)가 부실한 것도 한 가지 요인일 수가 있을 것이다. 그런데 나의 견해로는 지난 10년 동안 강의 평가 점수에 큰 변동이 없는 이유는 아마도 “제로섬” 현상으로 플러스와 마이너스가 양극화하여 평균치 만으로는 변동이 없는 것처럼 나타난 데 있지 않을까 한다.

다시 말해 높은 평가를 받은 교수는 강의 평가를 신뢰하고, 학생들의 반응을 존중하며, 다음 평가를 더 높이 기대하기 때문에 아마 평가 점수가 향상되었을 것이다.

그 반대로 강의 평가 점수가 좋지 못한 교수는(정말 피드백이 꼭 필요한 교수들은) 평가를 보는 순간 자존심도 상하고, 원인을 자신의 교수법 이외의 다른 요인으로 돌리려고 하며, 학생들의 평가를 무시하거나 받아들이지 않기 때문에 평가 점수는 낮은 채로 변동이 없거나 더 낮아지게 되었을 것이다. 점점 더 향상하는 교수와 점점 더 퇴보하는 교수의 점수가 합산되면 표면상으로는 “변동없음”으로 나타난다는 해석이다.

따라서 이 양극화 현상을 막고 교수가 강의 평가를 통해서 더욱 분발, 향상한다는 첫번째 목표를 달성하려면, 강의 평가의 두 번째 목표, 즉 대학 행정 쪽에서 강의 평가 점수로써 교수의 능력과 기여도를 인정하는 평가지향적 목표가 수반되어야 하는 것이다. 그래서인지 이제까지의 관행에 비추어 보면, 많은 대학에서 두 번째 목표에 더 비중을 두는 경향이 있다. 그러나 인사 결정을 할 때 강의 평가 점수가 높은 교수들은 당연시하고, 낮은 교수에게는 ‘별점을 가하는’ 식으로 운영하여 평가에 대해 상당히 부정적인 인식을 가지게 하고 있다. 따라서 두 번째 목표는 첫 번째 목표와 상반되기도 한다. 이 점에 대해서는 차후에 더 논의하기로 한다.

다음은 강의 평가의 신빙성 여부에 대한 논란과 그 신화에 대한 검토이다.

강의 평가는 과연 신빙성이 있는가?

어느 도구의 신빙성을 따지자고 하면 최소한 두 가지를 만족시켜야 한다. 신뢰성(reliability)과 타당성(validity)이 있느냐 하는 문제이다.

우선 신뢰성에 대한 논쟁을 살펴보자. 아마도 강의 평가를 한다면 가장 첫 질문이 과연 학생들의 판단을 신뢰할 만한 것이냐 하는 문제일 것이다. 아직도 미성숙한 20대 전후의 대학생들이 과연 자신들보다 연

장자인 교수들의 평가를 올바르게 하겠느냐, 감정이나 선입견에 좌지우지하지 않겠느냐 하는 질문은 미국에서도 오랜동안 강의 평가의 신뢰성을 따질 때에 가장 자주 거론되던 문제였다. 지난 70년 동안 강의 평가에 대한 연구들을 종합 분석해 본 결과로는 확정적으로, “학생들의 강의 평가는 신뢰성이 아주 높다”는 것으로 판명되었다.³⁾ 일 예로 카신이 강의 평가의 신빙성에 대하여 행해진 1,600 연구를 모두 검토해 본 결과 학생들의 평가 점수의 상관도는 .90에서 .94로 상당히 높으며,⁴⁾ 이후의 많은 후속 연구에서도 대략 비슷하게 높은 상관도가 나타난다.⁵⁾⁶⁾⁷⁾ 즉 일관성이 있다는 것이다.

또 알리아모니와 헉스너(Aleamoni & Hexner, 1980)가 학생들의 평가 점수와 1)동료교수, 2)전문교육가 3)졸업생들의 평가 점수와의 상관 관계를 검토해 본 결과 이들 간의 평가 점수에 상당히 높은 일치율이 있다는 결론을 내렸다.

다음은 강의 평가의 신뢰성과 타당성에 대한 “신화”와 그 “검증”의 요약이다.

신화 1. 강의 평가는 교수의 인기도와 직결된다.

학생들은 진실한 교육자·보다는 약장사 같이 유창한 말재주로 강의를 재미있게 이끌어 가는 교수에게 후한 평가를 준다는 믿음이 팽배하다. 또 실력은 없어도 학생들에게 인기를 끌기 위해 점수를 후하게 주는 교수들이 높은 평가 점수를 받는다는 “설”이 있다. 그러나 학생들은 “말 잘하는 것”과, “강의 내용의 수준과 질”을 구분하고, “학생에게 친절하다”와 “전문 지식과 실력이 있다”와의 구분을 명확히 한다고 한다.

이것은 학생들의 평가 내용을 항목별로 분석해 본 결과이다. 어느 한 항목이 높다고 해서 나머지 모든 항목이 일률적으로 높

지가 않다는 것이다.

신화 2. 교육의 효과는 즉각 나타나는 것이 아니기 때문에 수업 직후의 평가는 의미가 없다.

그러나 연구 결과로 보면 학기 말에 한 평가와 5년이나 10년 후에 내린 평가와 일치함을 알 수 있다. 다시 말해 진실이 깃든 훌륭한 강의는 지금 당장도 인정받지만 나중에 회고해 보아도 높은 평가가 지속된다는 것이다.

신화 3. 강의 평가는 강의생 숫자가 크면 불리하다.

대강의를 하는 교수와 소강의를 하는 교수가 일률적인 평가 점수로 비교된다는 것은 매우 불공평하다는 인상을 준다. 그러나 강의 평가 점수와 수강생의 크기와 상관도를 비교한 결과 여기에는 아무 상관관계가 없음이 나타났다. 단, 대학원생 강의와 15명 이하의 강의에서만 강의 평가 점수가 약간 높게 나올 뿐, 그 이상은 수십명부터 수 백명까지도 별 차이가 없다는 결론이다.

신화 4. 강의 시간이 평가 점수에 불리할 때가 있다.

예를 들어 이른 오전이라든지, 늦은 오후, 혹은 점심 시간 전 후 등은 출석율도 낮고 학생들도 강의에 집중을 안하여 강의 평가 점수가 낮아질 것이라는 우려가 있지만 이들 사이에는 전혀 상관 관계가 없는 것으로 나타났다. 심지어 주말이나 주초, 주중의 강의 요일과도 무관한 것으로 나타났다.

이 같이 누적된 연구를 통하여 그릇된 신화도 밝혀졌지만, 약간의 상관관계가 입증된 요인들도 있다. 예를 들어 “필수” 과목은, 일반적으로 “선택” 과목들보다 평가 점수가 좀 낮게 나오는 경향이 있다고 한

다. 그러나 평가 점수와 가장 상관관계 지수가 높고, 직접적인 영향을 미치는 요인은 “학생들이 얼마나 배웠는가” 라는 항목이라는 것이다. 여기에는 논란의 여지가 없을 정도로 정확하고 일관적인 평가 점수가 반영된다는 것이다.

학생들은 왜 평가를 정확히 할까? 그 답은 간단하다. 강의의 수준과 내용이 자신들의 이익과 직결되기 때문이다. 또 어느 교수를 한 학기 동안 지속적으로 직접적으로 관찰하고 경험하는 사람들이 바로 학생들이기 때문이기도 하다. 다른 외부 전문가들은 한, 두 번 참석하고 평가를 할 지 모르지만, 학생들은 가장 오랫동안, 가장 가까이서 교수의 일거일동을 경험하는 것이다. 그러기에 “인상”이나 “성급한 편견”이 크게 좌우하지 않는다는 것이다.

이상으로 강의 평가의 신뢰성과 타당성에 대한 검토를 해 보았다. 결론은 학생들의 평가는 신뢰할 만 하고, 그 타당성도 입증되어 있다는 말이다. 그러나 아직도 풀지 못한 문제는 많다. 그 가장 큰 숙제가 바로 앞서 언급했듯이 왜 강의 평가를 수십년 해 왔어도 강의의 전반적인 향상에는 효과가 나타나지 않는다는 문제이다.

여기서 이 문제를 짚고 넘어가야 할 듯하다. 그래야만, 앞으로 우리나라 대학에서 강의 평가제를 도입하더라도 소위 “선진국”의 과오를 되풀이 하지 않을 수 있기 때문이다.

강의 평가제가 강의 수준 향상과 무관한 이유

강의 평가는 앞서 “평가지향적 목표”에서 언급했듯이 교수평가의 일부로서 쓰이기도 하지만 아직까지 그 실효성에 긍정적 결과보다는 부정적 인식을 확산하는 데에 더 기여한 듯한 인상을 준다. 그 가장 큰 원인은 “보상체계”에 있다고 본다.

미국 대학에서 교수평가는 일반적으로 크게 연구, 강의, 봉사 등 세 분야로 나뉜다. 이 중 연구와 강의가 중점을 이루는데, 연구 업적은 쉽게 수량화(연구비와 논문수)하여 양적 비교를 할 수 있지만 교수의 교육 활동을 평가 점수로 수량화하는 것은 논란의 여지가 있다. 더우기 “자본주의” 원리에 입각하여 대학을 “운영” 해 온 미국에서는 “좋은 논문을 내야 연구비를 많이 지원 받는다” 라는 원칙에 입각하여 이제까지 “대학을 살리는 데 연구비 따오는 것을 최상책”으로 여겨왔다.

그 결과 논문은 연구비를 끌어오기 위해서 하고, 연구비는 교수평가에 직접 반영이 되는 연쇄 고리를 만들게 되었다. 예를 들어 논문 200점이 커트라인인데 400점을 달성한 교수는 그만큼 연구비 지원도 많이 받았을 확률이 높고, 그만큼 대내외로 “우수하고 유능한” 교수라는 인정을 받는다. (실제로 대부분의 대학에서 논문수와 연구비는 교수 연봉과 승진 속도에 정비례로 반영된다.)

그러나 강의 평가는 잘 해도 “아무 소용 없는 일”로 간주된다. 학생들에게 열성적으로 강의하느라고 연구비를 못 따온다면 “헛고생” 하는 것으로 여겨진다는 것이다. 그러나 평균치보다 못하면 교수평가 때에 감점을 받는다. 체제가 이러니 연구에 스트레스 받는 교수들이 “강의는 대충 중간 기준치만 하면 되고, 연구는 남보다 더 잘 할 수록 이익이다” 라는 생각을 하지 않을 수가 없게 되는 것이다. 다시 말해 연구는 “상” 받는 개념이라 최대한을 추구하지만, 강의는 “매”를 피하는 개념이라 최소한으로 향한다는 말이다.

강의 평가가 제구실을 하기 위해서는 강의 평가도 연구 평가에 걸맞는 대우를 해 줘야한다. 미국에는 연구비를 많이 따오면 강의 부담을 줄여 준다. 그러나 강의를 잘

하는 교수에게 연구 부담을 줄여 준다는 말은 들어 본 적이 없다. 나 자신 실제로 경험해 보았는데, 몇 년 전에 내가 큰 액수의 프로젝트를 “따 왔다”는 소식과 미국 공학교육학회의 우수 교수(교육)상을 수여하게 되었다는 소식을 동시에 받았다.

과장, 학장, 총장, 이사장으로 부터 각각 축하의 글과 메시지를 받았는데 모두 연구 프로젝트에 대한 언급만 있었지 교육상에 대해서는 말이 없었다. 이런 행정적 “편견”이 사라지지 않고는 학내에 강의 평가가 제대로 목적을 이루지 못할 것이다.

물론 미국내에서도 1990년 대에 들어서 “교수들이 연구에 급급하느라고 교육 수혜자인 학생들에게 너무 소홀하다”라는 자성의 소리가 크고, 스탠포드, 프린스턴과 같은 명문 대학에서는 앞 다투어 강의 수준 향상에 역점을 두고 보상 제도를 강화하는 정책을 쓰고 있다. 그러나 어떤 문화라도 일단 정착이 되고나면 타성과 관성이 생겨서 쉽게 바뀌어지지 않는다. 그래서 오히려 “아직” 시작하지 않은 우리나라 대학에 더 큰 희망을 걸어볼 만한 것이다.

국내 대학에서도 강의 평가를 해야만 할까?

우리나라 대학에 강의 평가를 도입하는데 있어 큰 걸림돌이 있다. 외국 대학의 경우처럼 오랜 논란 과정도 없이, 일시적으로 또 “하향식”으로 강의 평가를 시행하려고 하기 때문에 다른 무엇보다도 감정적인 거부 반응이 가장 장애가 되고 있는 듯 하다.

우선 교수의 입장에서 강의 평가제는 불쾌하다. 왜냐하면 본래 교수란 의무감, 책임감 등이 강하고 대단히 자발적인 (self-



motivated) 사람들이기 때문에 “하향식”, “일방적” 평가제란 교수의 속성에 대단히 위배된다. 또 교육 철학에 입각해서 보더라도 진정한 교육은 믿음을 토대로 한 교수와 학생들 사이의 튼튼한 유대관계에서만 가능하다고 믿는데 “불신”에서 비롯된 듯한 인상을 주는 평가제가 교수들의 사고방식이나 “체질”에 맞을 수가 없을 것이다. 그런데 이보다 더 중요한 점은 이 모든 평가제가 실용주의 (utilitarian) 철학에 관한 논란으로 이어지고 있다는 것이다. 대부분의 교육자들은 교육의 결과는 학생들이 졸업한 이후 평생 동안 알게 모르게 작용하는 것이며, 따라서 일시적이고 실증적인 측정 (empirical measurement)에 의하여 가늠할 수가 없다고 믿는다. 그러니 이득지출의 삭막한 숫자놀이에 의하여 이루어지는 평가제는 교수들의 이념적 반발을 유발시키지 않을 수가 없는 것이다.

그러면 교육 수혜자인 학생들과 그들의 교육비를 부담하는 부모, 그리고 대학 교육의 배출자들을 고용하는 일반 기업인들의 입장에서는 교수 평가제가 어떤 의의가 있

을까?

현재 우리나라에서는 대학에 대한 일반 대중의 관심이 양극화 되어 있다. 대학 입시까지의 투자와 관심은 세계 최고이지만, 일단 입학한 후의 대학 “교육 과정”에 대한 관심은 거의 전무하다 해도 과언이 아니다. 이런 현상은 바로 대학입학이 “지배계급”이 되는 최저 문턱이라는 대중의 인식 때문이다.

그러나 사회가 성숙해지고, 대학 졸업자들 사이의 경쟁력이 높아지면서 부터 학력(學歷) 보다는 학력(學力)이 중요하다는 실속을 차리게 된다. 직업을 얻기 위해 4년제 대학을 졸업한 후 다시 전문대학에 다니는 일이 최근 늘고 있는데 이런 어처구니 없는 일이 바로 대학 교육에 대한 일반 대중의 인식 변화를 재촉한다.

결과적으로 일반 대중은 지속적으로 하락하는 대학 교육의 사회·경제적 가치를 검토/비판하게 된다.

국가 경제적 측면에서도 대학교육을 비판할 수 밖에 없는데, 그 이유는 세계를 대상으로 한 우리나라의 경쟁력이 이미 저임금으로부터 고(高) 기술로 변해야 하는 시점에 와 있기 때문이다. 우리나라가 지난 30여 년간 꾸준히 눈부신 발전을 할 수 있었던 이유는 가히 세계 최고라 할 수 있는 대한민국 고졸자들의 막강한 실력을 믿고 저임금 경쟁을 치열하게 해왔기 때문이라고 한다. 그러나 첨단 과학 기술 경쟁은 대졸자들의 창의력과 전문 기술에 의존하는 경쟁이다. 바로 여기에 문제가 있다. 과연 우리나라 기업들이 우리나라 대학 졸업자들의 실력을 믿고 첨단 과학 기술 경쟁에서 버텨낼 수 있겠느냐 하는 의문이다.

이 때문에 기업을 포함한 국민 대중들이 대학 교육에 대한 인식을 크게 변화하고 있다. 현상태의 우리나라 대학 교육이 개인적으로 보아도 신분 상승(개인 경쟁력)에 별 도움이 안 될 뿐 아니라, 국가적으로도 우

리 나라가 선진국으로의 상향 이동(국제 경쟁력 증진)을 하려는 데에 아무 도움이 안 되고 있다는 요지이다. 이런 대중의 인식 변화는 대학 교육 제도에 대한 불신을 초래하고, 이어서 투자 가치에 대한 반발을 촉구한다. 이런 사회 분위기에서 대학 교육의 질을 높이기 위한 교수 평가제는 필수로 여겨지게 된다. 그리고 점점 “열린 사회”로 바뀌어 가고 있는 시대 흐름에 부합하여 평가의 결과도 공개되기를 요구할 것이다.

맺음말

아무리 평가제가 교수들의 교육 철학과 감정에 거부감을 준다 하더라도, 대학 교육에 대한 일반 대중의 불신과 회의에 대한 검토방안의 도구로 교수 평가제가 요구되어지는 만큼, 앞으로 어쩔 수 없이 보편화 되리라고 본다. 대중의 인식을 철학적 논쟁으로 바꾸기는 쉽지 않을 것이기 때문이다. 또한 교수들이 계속해서 대중의 요구를 기피한다 하더라도 학생들에 의한 평가 외에 평가제가 외부 기관(행정 당국이나, 기업, 혹은 졸업생)에 의해 행해질 수도 있을 것이다.

따라서 문제는 평가제를 시도하는 방법에 있어서 교수들이 사회의 비판에 이끌림을 당할 것인지, 아니면 교수들이 능동적으로 평가제를 대학교육 개선의 방안으로서 적극 이끌어 나갈 것인지의 차이로 압축될 것이다. 다시 말해 문제는 평가를 하느냐 마느냐에 있지 않고, “어떻게” 하느냐만 남은 것이다.

참고문헌

1. Handbook of Teacher Evaluation, ed. Jason Millman, Sage Publications: Beverly Hills, 1981.
2. McKeachie, W.J. (1979) “Student ratings of faculty: a reprise.”

- Academe, 65: 384-397.
3. Gillmore, G.M.(1973) "Estimate of reliability coefficients for items and subscales of the Illinois Course Evaluation Questionnaire" (Research Report No. 341). Urbana: University of Illinois.
 4. Cashin, W.E.(1990) "Assessing Teaching Effectiveness." In Seldin, Peter and Associates, How Administrators Can Improve Teaching, San Francisco: Jossey-Bass.
 5. Costin F., W.T.Greenough, and R.J.Menges(1971) "Student ratings of college teaching: reliability, validity, and usefulness." Review of Educational Research, 41: 511-535.
 6. Hogan, T.P.(1973) "Similarity of student ratings across instructors, course and time." Research in Higher Education, 1: 149-154.
 7. Aleamoni, L.M. and M. Yimer(1973) "An investigation of the relationship between colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness." Journal of Educational Psychology, 64: 274-277.
 8. Aleamoni, L.M. and P.Z.Hexner(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, 9: 67-84.
 9. Ware J.E. and R.G. William(1977) "Discriminate analysis of student ratings as a means of identifying lecturers who differ in enthusiasm or information giving." Educational and Psychological Measurement, 37: 627-639.
 10. Reardon, M. and L.K. Waters(1979) "Leniency and halo in student ratings of college instructors." Educational and Psychological Measurement, 39: 159-162.
 11. Marsh, H.W. and J.U. Overall(1979) "Long term stability of student's evaluations." Research in Higher Education, 10: 139-147.
 12. Wood, K., A.S. Linsky and M.A. Strauss(1974) "Class size and student evaluations of faculty." Journal of Higher Education, 45: 524-534.
 13. Yongkittikul, C., G.M. Gillmore, and D.C. Brandenburg(1974) "Does the time of course meeting affect course ratings by students?" Research Report No. 346. Urbana: University of Illinois.
 14. Pohlmann, J.T.(1975) "A multivariate analysis of selected class characteristics and student ratings of instruction." Multivariate Behavioral Research, 10(1): 81-91.