

공공기관 경영평가에서 상대평가의 오류가능성 분석*

이 창 길**, 최 성 락***

< 목 차 >

- I. 서 론
- II. 공공기관 경영평가와 상대평가에 관한 선행연구 검토
- III. 상대평가의 오류가능성
- IV. 연구대상 및 변수
- V. 분석결과
- VI. 결론 및 정책적 시사점

< 요 약 >

본 연구는 2007년도 정부산하기관 경영평가결과를 대상으로 평가과정에서 발생하는 상대평가의 오류가능성을 분석하였다. 자료는 75개 정부산하기관에 대하여 총 12개 지표에 각각 9등급으로 부여된 평가점수를 활용하였다. 상대평가의 오류가능성은 지표별 기관 간 오류가능성과 기관별 지표 간 오류가능성으로 구분하여, 특정기관 특정지표의 평균평점이 상대평가 요인, 즉 다른 기관들 또는 다른 지표들 평점의 분산 값으로부터 영향을 받는지 여부를 분석하였다. 전체 기관을 대상으로 한 분석결과, 지표별 기관간의 평가에 있어서 다른 기관들 평점의 분산정도와 실제 평가점수가 유의미한 관계가 있는 반면, 기관별 지표간 평가의 경우 다른 지표들 평점 분산값과 실제 평가점수 간 관계가 유의미하지 않았다. 그리고 이러한 경향은 부정적 평점에서보다 긍정적 평점에서 더 크게 발생하고 있다. 또한 각 기관 유형별로 살펴본 분석 결과도 이와 유사하게 나타났다. 본 논문은 상대평가의 오류가능성을 최소화하기 위하여 절대평가를 가미한 상대평가의 운영, 기관 순위 중심에서 평가지표 순위 중심으로 전환, 평가 훈련 프로그램 시행 등의 정책적 시사점을 제공한다.

【주제어: 공공기관 경영평가, 상대평가, 평가오류, 평가 신뢰성】

* 이 논문은 2007년도 정부지원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2007-327-B00739).

** 세종대학교 행정학과 부교수(cklee@sejong.ac.kr)

*** 동양미래대학 경영학부 조교수(haihaba@chol.com)

논문접수일(2009.9.10), 수정일(2009.11.29), 게재확정일(2009.12.1)

I. 서론

공공부문에서 성과평가가 확산됨에 따라 이에 대한 이론적·제도적 연구도 증가하고 있다. 성과평가의 궁극적인 목적은 보상적 인센티브를 통하여 동기를 부여하고 궁극적으로 조직성과를 향상시키는 데 있다. 이러한 의미에서 성과평가는 평가자체로서도 긍정적 효과가 존재하긴 하지만, 피평가자에 평가결과를 환류(feedback)하는 과정을 통하여 궁극적으로 자기목적을 달성할 수 있다(노화준, 2002; 김종희, 2007). 따라서 이러한 성과평가가 성공할 수 있는 핵심적인 요소는 평가내용이 조직성과 향상에 실질적으로 도움이 되어야 한다는 타당성이라 할 수 있다. 평가의 타당성 확보의 전제요건은 측정도구로서 평가의 신뢰도이다(유민봉·임도빈, 2007). 평가의 객관성과 신뢰성이 없는 타당도는 평가제도 자체에 대한 회의를 낳고 조직성과를 오히려 저하시키는 요인이 되기도 한다. 즉 평가의 신뢰성은 피평가자들의 평가결과에 대한 수용성과 직접적으로 연계되어 있다(신동일·장소영, 2002; 김종희, 2007). 우선, 피평가자들은 대부분 평가의 발전적(developmental) 측면보다는 평가의 평가적(evaluative) 측면에 보다 높은 관심을 가지고 있다. 미래의 조직성과 향상보다는 현재의 평가결과, 즉 상대적 순위에 보다 많은 관심을 가지고 있는 경우가 많다. 또한 피평가자는 평가결과의 원인을 자기 자신의 요인보다는 평가자 또는 평가도구 요인에서 찾는 경우도 많이 존재한다. 피평가자는 특히 좋지 않은 평가결과에 대하여 피평가자 자신의 업무성과 책임으로 보기보다는 평가자 또는 평가도구에 대한 신뢰성 문제를 제기하는 경우가 많다(Boreman, 1978). 그러므로 평가자 또는 평가도구의 신뢰성이 평가결과에 대한 피평가자 수용 여부에서 중요한 요인으로 작용한다.

지금까지 측정도구로서 평가지표에 대한 연구와 논의는 많이 이루어졌으나(이석원, 2005; 이환범 외, 2005), 측정도구로서 평가자 측면의 오류가능성에 대한 연구는 거의 없는 것이 현실이다(Facteau and Craig, 2001; Nankervis and Compton, 2006). 특히 비계량 정성평가에 있어서 이러한 평가자의 오류는 평가의 신뢰도에 많은 영향을 미친다(박영범, 1998; 박명호·이성욱, 2000). 상대평가 제도는 이러한 비계량 평가지표의 평가과정에서 평가자들의 관대화 경향을 방지하고 평가결과의 실효성을 높이기 위하여 도입된 제도이다. 이러한 상대평가를 통하여 피평가자에 대한 상대적 순위를 부여하기 쉽고, 보상을 위한 명확하고 객관적인 기준을 제공할 수도 있다(Bernardin, 1978; Jenter and Kanaan, 2006). 그러나 상대평가는 강제배분 방식으로 피평가자로 하여금 평가내용보다는 평가순위에 대한 관심을 높게 할 뿐만 아니라 평가자로 하여금 평가의 발전적 측면보다 평가적 측면을 중시하게 하는 경향이 있다(Stanton, 1991; Gibbons and Murphy, 1990). 따라서 상대평가는 평가결과에 대한 기관별 또는 지표별로

강제 배분하는 평가방식으로서 상대적 평점과정에서 심각한 오류가 발생할 수 있고, 이는 곧 피평가자의 수용성을 해치는 결정적 요인으로 작용할 수 있다.

본 연구는 2007년도에 이루어진 정부산하기관 경영평가를 대상으로 피평가기관별 또는 평가지표별 상대평가 과정에서 발생할 수 있는 평가자 오류가능성을 분석하고자 한다. 2007년도 공공기관 경영평가 중 비계량 정성평가지표를 대상으로 하여 기관별·지표별 상대평가 과정에서 발생할 수 있는 평가오류의 존재 여부를 실증적으로 분석하고자 하는바, 궁극적으로 공공기관 경영평가의 수용성과 신뢰성을 보다 높일 수 있는 방안을 제시하고자 하는데 본 연구의 의의가 있다 하겠다.

II. 공공기관 경영평가와 상대평가에 관한 선행연구 검토

1. 경영평가 제도와 평가지표의 신뢰성

공공기관 경영평가의 기존연구를 살펴보면, 평가제도 자체에 대한 성공요건 및 제도개선에 중점을 둔 연구와 평가제도의 효과성에 관한 연구로 구분된다. 전자는 경영평가제도의 도입필요성과 함께 평가방법이나 절차, 평가방식, 그리고 평가주기 등 제도 전반에 대한 정책적 개선에 중점을 둔 반면(오연천, 1998; 김준기, 2001; 송대회, 2003; 광채기, 2008), 후자는 평가제도가 경영성과에 얼마나 영향을 미쳤는지 그 효과성에 대한 분석에 집중한다(광채기, 2003; 박석희, 2006; 광병환, 2005; 최성락 외, 2009). 따라서 전자는 사전적 의미가 강하며, 후자는 사후적 의미가 강하다.

평가제도 자체의 효과성에 대한 기존연구에서는 주로 평가의 타당성에 대한 연구가 많이 이루어졌다. 광채기(2003)는 지난 20여 년 동안의 공기업 경영실적을 바탕으로 공기업 평가제도가 공기업경영 개선효과에 영향을 미쳤는지 살펴 보았다. 박석희(2006)는 1965년부터 2004년까지 13개 정부투자기관에 대한 패널 데이터를 활용하여 공기업들의 관리업무 비율, 인건비 비율, 금융비용율, 인원 증감률 등을 중심으로 경영평가제도 도입이 조직역량 증대에 미치는 영향을 분석하였다. 윤기웅·공동성(2008)은 재정사업 자율평가와 관련해서, 경영평가 결과가 예산 절감에 어떤 영향을 미쳤는지 분석하였다.

그리고 최근에는 공기업 평가자체의 신뢰성과 일관성에 대한 연구가 증가하고 있다. 즉 평가의 효과성이나 평가제도적 문제점보다는 평가과정에서 발생하는 오류를 파악하고자 하는 연구들이다. 최성락 외(2009)는 2001년부터 2006년까지 공기업 평가결과를 바탕으로 평가지표의 개선효과를 분석하였고, 이창길·최성락(2009)은 2007년도 경영평가실적을 토대로 평가과정에서 평가자들의 평가

일관성 문제에 집중하여 후광효과의 존재가능성을 분석한 바 있다. 이는 공공기관에 대한 연구라기보다는 평가(evaluation)자체에 대한 연구로 귀착된다. 따라서 이러한 평가연구는 공공기관 뿐만 아니라 정부기관 및 민간기관에도 적용 가능하며, 경영평가는 물론, 기금평가, 혁신평가, 부패평가, 정부업무평가 등 다양한 분야의 평가에도 적용될 수 있다. 즉 평가의 타당성과는 달리 평가의 신뢰성문제는 평가대상과 분야에 상관없이 평가연구의 핵심적인 과제라 할 수 있다.

2. 비계량 정성평가와 상대평가

공공부문의 평가에 있어서 가장 문제가 되는 것 중 하나는 지표선정과 측정(measurement)의 문제이다(이환범 외, 2005; 이석원, 2005). 즉 공공의 가치는 계량적으로 측정하기 곤란하기 때문에 이들 지표의 경우에도 계량적 정량평가보다는 비계량적 정성평가에 의존하는 경우가 많다. 계량적 지표에 대한 평가결과는 피평가자의 수용성이 상대적으로 높은 반면 비계량적 지표에 대한 평가결과는 피평가자의 수용성이 상대적으로 낮을 것이다(Hedge and Teachout, 2000; 김종희, 2007). 이와 같이 비계량적 정성평가에 피평가자의 수용성과 평가과정의 신뢰성을 중심으로 비판적 의견이 증가하면서(Gabris and Ihrke, 2000), 평가지표 자체를 측정 가능한 계량지표만으로 구성하는 한편, 비계량지표에 대한 평가 신뢰성 확보를 위하여 평가방식과 과정을 합리적으로 통제하고자 하는 다양한 노력이 이루어져 왔다(박영범, 1998; 박명호·이성욱, 2000).

이러한 노력의 일환으로 평가관리자가 선택할 수 있는 가장 쉽고 유용한 대안은 역시 상대평가이다(Steers, 1975; Jenter and Kanaan, 2006). 상대평가는 관리자의 입장에서 보면 평가결과가 가장 객관적으로 명확하게 나타날 수 있고, 성과경쟁을 유도할 수 있는 중요한 수단이 된다(Dye, 1986; Gibbons and Murphy, 1990; Jenter and Kanaan, 2006). 계량평가의 경우 상대평가가 아닌 절대평가를 통해서도 명확하게 평가의 순위를 측정할 수 있는 반면, 비계량적 지표에 대한 평가의 경우 평가의 신뢰도를 해치지 않으면서 성과의 차이를 가장 명확하게 하기 위해서는 절대평가보다는 상대평가를 선호하게 된다. 즉 비계량 지표에 대한 평가의 경우 상대평가는 계량평가 만큼 피평가자 상호간의 차이를 명확히 할 수 있는 방법이 된다(박명호·이성욱 2000; 노화준, 2002).

하지만 상대평가는 피평가자 뿐만 아니라 평가자에게도 심리적 통제를 가함으로써 평가의 오류 또는 왜곡을 낳을 수 있다(최연희, 2002; Mark and Henry, 2004). 피평가자들은 상대평가를 통하여 협력보다는 경쟁을 지향하게 된다. 좋은 평가결과를 얻기 위한 중요한 전략은 동료와의 비교와 경쟁을 통해서 가능하다. 피평가자 상호간의 협력은 평가결과에 긍정적 영향을 주기보다는 부정적 영향을 주기 쉽다. 따라서 피평가자는 평가과정에서 평가자체의 실질적인 내용

보다는 다른 피평가자와의 비교 하에 평가자의 평가 신뢰성 문제와 평가의 상대적 순위에 높은 관심을 보이기 쉽다. 이러한 심리적 압력은 평가자에게도 유사한 평가왜곡 현상을 낳는다. 평가자들은 평가의 실질적인 내용의 상대적 차이를 명확히 하고 이를 평가에 반영하는 경우에는 문제될 것이 없다. 하지만, 평가자의 평가와 실제적인 성과와의 격차(decoupling)가 발생할 우려가 있고, 평가 기준별 또는 피평가자별 상대적 격차를 임의로 규정함으로써 피평가자 성과에 대한 과대평가와 과소평가 현상이 나타나기 쉽다(Gibbons and Murphy, 1990; Jenter and Kanaan, 2006). 이러한 왜곡 또는 평가오류 가능성은 피평가자 상호간 보다는 평가자 상호간에 보다 심각한 문제를 야기한다. 특히 평가의 공정성과 신뢰성 차원에서 피평가자의 오류 책임보다는 평가자의 오류 책임이 평가의 효과성을 저하시키는 주된 요인이 된다(이영식, 1998).

3. 정부산하기관 경영평가에서의 상대평가¹⁾

정부산하기관의 2006년도 실적을 평가하는 2007년도 평가단은 경영실적 평가의 합리성·객관성·공정성을 확보하기 위하여 대학교수, 공인회계사, 연구원 및 관련 분야 민간 전문가로 구성·운영되었다. 75개의 정부산하기관을 각 기관의 기능, 성격, 기존 평가체제의 존재 여부 등을 감안하여 금융·수익, 연수·교육훈련, 학술·연구지원 유형 등 8개 유형으로 분류하고, 각 유형에 대하여 평가위원들이 배정되었다. 즉 평가위원들은 8개 산하기관 유형 중 하나에 속하여, 자기에게 주어진 평가지표에 대하여 해당 유형에 포함되는 모든 산하기관을 평가하는 방식으로 이루어졌다. 평가 지표는 주요 평가 부문으로 종합경영, 주요 사업, 경영관리 3 부문으로 구분되며, 각 주요 부문마다 3-5개 정도의 세부 평가 부문으로 나뉘어진다(기획예산처, 2007).

이와 같이 정부산하단체의 경영평가 중 정성평가는 상대평가를 지향한다. 평가기준에 따라 기관별 상대적 평가가 이루어진다. 기관 내 기준별 상대적 배점이라기보다는 기준내 기관별 상대적 배점이 주어진다(기획예산처정부산하기관 경영평가단, 2007). 점수 등급은 기본적으로 각 평가 항목을 5개 등급으로 구분하고, 각각의 기본등급에 대해 보다 우수한 성과를 낸 경우 +점수를 부여하여 9등급으로 평가하도록 되어 있다. 종합경영, 주요사업 및 경영관리부문 3개 부문 100점 만점으로 평가한다. 점수 부여 방식은 유형내 기관 간 점수등급은 팀장주

1) 2007년 이전에는 정부투자기관, 정부산하기관 등으로 구분되어 별도로 경영평가가 이루어졌지만, 2007년 4월 1일 공공기관의 운영에 관한 법률이 시행되면서 현재는 정부투자기관과 정부산하기관 등이 통합되어 평가되고 있다. 정부산하기관이란 명칭도 현재는 소멸되었지만, 본 연구에서는 공공기관의 운영에 관한 법률이 본격적으로 시행되기 전에 수행된 정부산하기관 평가를 연구의 직접적인 대상으로 하였는바, 따라서 여기서는 정부산하기관의 경영평가 과정을 중심으로 논의하고자한다.

도의 팀원 합의를 토대로 C0를 중심으로 정규분포와 유사하게 상대적으로 부여 하도록 하고, 유형 내 기관수가 적은 3개 유형도 정규분포와 유사하게 점수를 배치하도록 되어 있다. A+, A0, B+, B0, C0, D+, D0, E+, E0 등 각각의 등급에 대하여 일정비율을 강제 배분하는 방식이다.²⁾ 특히, 점수가 전년도에 비하여 3 단계 이상 급변하는 경우 연도 간 일관성 확보 차원에서 그 이유를 정확하게 확인하여 논리적으로 서술하도록 하고 있다(이창길·최성락, 2009).

앞에서 지적한 바와 같이, 정부 산하기관에 대한 정성평가는 각 평가기준마다 평가위원이 배정되고, 각 평가기준별 평가위원은 자신이 속한 유형에 포함되는 모든 기관에 대한 평가를 한다. <표 1>은 유형별 상대평가의 방식이자 기본적인 구조를 나타내고 있다. 1유형의 평가기준 a에 대한 평가자는 1유형의 A기관, B기관, C기관의 평가기준 a에 대한 평가를 담당한다. 즉 평가점수 Aa, 평가점수 Ba, 평가점수 Ca가 한 평가자의 평가에 의해서 도출된다. 마찬가지로 2유형의 평가기준 b에 대한 평가자는 2유형 D기관, E기관, F기관에 대한 평가기준 b에 대한 평가를 하게 되어, 평가점수 Db, 평가점수 Eb, 평가점수 Fb가 한 팀의 평가자에 의해서 점수가 매겨진다. 즉 공공기관 경영평가에 있어서 상대평가는 기본적으로 동일한 평가기준 내 피평가기관 상호간의 상대적인 평점을 부여한다.

<표 1> 공공기관 경영평가에 있어서 상대적 평가방식

유형	평가 기관	평가기준 a	평가 기준 b	평가기준 c	평가기준 d
1유형	A 기관	평가점수 Aa	평가점수 Ab	평가점수 Ac	평가점수 Ad
	B 기관	평가점수 Ba	평가점수 Bb	평가점수 Bc	평가점수 Bd
	C 기관	평가점수 Ca	평가점수 Cb	평가점수 Cc	평가점수 Cd
2유형	D 기관	평가점수 Da	평가점수 Db	평가점수 Dc	평가점수 Dd
	E 기관	평가점수 Ea	평가점수 Eb	평가점수 Ec	평가점수 Ed
	F 기관	평가점수 Fa	평가점수 Fb	평가점수 Fc	평가점수 Fd

Ⅲ. 상대평가의 오류가능성

1. 심리적 과정으로서 상대평가

2) 본 평가 방법은 (구)정부산하기관에서의 평가 방법으로서 2009년 현재 공공기관 평가와는 다소 차이가 있다. 현재는 6단계 평가로 이루어지고 있으며, 공식적으로 절대적 평가 방식을 채택하고 있다.

상대평가의 오류 가능성은 평점의 강제적 배분에 있으면 배분과정에서 평가자의 심리적 조정과정이 개입됨으로써 발생한다. 예를 들면, 하나의 평가기준에 대하여 상대평가를 전체로 5개 기관에 대하여 평점 A, B, C, D, E를 각각 20%씩 부여한다고 하자. 이러한 상대평가는 두 가지 측면에서 오류 발생가능성이 있다. 첫째, 평가자가 4개 기관 평점의 평균이나 분산이 마지막 순서의 피평가기관의 평점에 영향을 미칠 수 있다는 점이며, 둘째, 평가자는 5개 기관을 모두 평가한 이후에 종합과정에서 자의적인 심리적 조정을 통하여 평점이 달라질 수 있다는 점이다. 이는 실제적인 성과와 상관없이 평가자의 자의적 조정과정에서 평점이 부여될 수 있음을 말한다. 이러한 현상은 평가기준이 다수이고 다양할 경우 보다 더 심각한 오류가능성을 내포한다. 즉 피평가기관과 평가기준 상호간의 복잡한 비교과정을 통하여 실제의 성과 내용과 상관없는 자의적인 평점이 이루어질 개연성이 높아진다. 만약 어느 한 피평가기관의 평점이 나머지 피평가기관들의 평점과 관련성이 높은 경우 전반적으로 평가의 오류가 발생했다고 추론할 수 있다. 이러한 경향은 엄격한 상대평가의 경우보다는 느슨한 상대평가, 즉 평가자의 자율성이 어느 정도 보장된 경우에 평가자의 왜곡 가능성이 더 크게 나타날 수 있다.

2. 기관별 평가와 지표별 평가의 상대성

상대평가로 인한 평점의 오류가능성은 지표별 평가의 상대성과 기관별 평가의 상대성으로 구분될 수 있다. 지표별 평가의 상대성은 하나의 지표 내에서 기관 상호간의 상대평가 오류가능성을 말하며, 기관별 평가의 상대성은 하나의 기관 내에서 지표 상호간의 상대평가 오류가능성을 말한다.

1) 지표별 기관 간 상대평가의 오류가능성

상대평가가 이루어질 경우 첫 번째 오류가능성은 평가기준 내에서 각 기관 간 평가점수를 심리적으로 조정하는 과정에서 발생할 수 있다. 특히 공공기관 평가의 경우 8개의 기관유형으로 구분되어 평가기준별 평가위원이 해당유형에 속하는 기관을 모두 평가하기 때문에 평가기준별로 기관유형 내 개별기관 간 상대평가를 하게 된다. <표 2>에서 보는 바와 같이, 평가자는 특정한 평가기준에 대하여 기관 간 수직적으로 상대평가를 하게 된다. 이와 같은 평가구조에서 평가기준 a에 대하여 평가점수 Aa, 평가점수 Ba, 평가점수 Ca 상호 간의 수직적 상대평가가 이루어진다. 이 과정에서 동일한 기관유형 내에서 개별기관에 대한 평점은 자기를 제외한 타 기관 평점에 의하여 심리적 영향을 받을 수 있다.

<가설 1> 평가지표별 기관 상호간의 평가점수 부여는 상대평가요인, 즉 다른 기관들의 평점과의 관련성이 높다.

<표 2> 지표별 기관 간 상대평가의 오류가능성 (수직적 상대평가)

오류가능성	평가 기관	평가기준 a	평가 기준 b	평가기준 c	평가기준 d
지표별 기관간	A 기관	평가점수 Aa	평가점수 Ab	평가점수 Ac	평가점수 Ad
	B 기관	평가점수 Ba	평가점수 Bb	평가점수 Bc	평가점수 Bd
	C 기관	평가점수 Ca	평가점수 Cb	평가점수 Cc	평가점수 Cd

2) 기관별 지표 간 상대평가의 오류가능성

상대평가가 이루어질 경우 두 번째 오류가능성은 평가기관 내에서 각 평가지표 간 평가점수를 심리적으로 조정하는 과정에서 발생할 수 있다. 특히 공공기관 경영평가에 있어서는 평가지표 상호간 상대평가보다는 평가지표 내 피평가 기관 상호간의 상대평가를 권장하고 있기 때문에 기관별 지표 간 발생하는 평가왜곡 가능성은 더욱 심화될 우려가 있다. 즉 기관별 지표 간 평점의 상대적 배분은 평가지침에 포함되어 있지 않다. 이와 같은 평가구조에서 A기관에 대한 평가점수 Aa, 평가점수 Ab, 평가점수 Ac, 평가점수 Ad 상호 간의 수평적 상대평가가 이루어진다. 따라서 <표 3>에서와 같이 평가주관기관 또는 평가팀에서 기관별 지표별 평가점수를 종합하는 과정에서 오류 가능성이 존재하게 된다. 이 과정에서 동일한 기관유형 내에서 개별기관에 대한 평점은 해당지표를 제외한 타 지표 평점에 의하여 심리적인 영향을 받을 수 있다.

<가설 2> 평가기관별 지표 상호간의 평가점수 부여는 상대평가요인, 즉 다른 지표들의 평점과의 관련성이 높다.

<표 3> 기관별 지표간 상대평가의 오류가능성 (수평적 상대평가)

오류가능성	평가 기관	평가기준 a	평가 기준 b	평가기준 c	평가기준 d
기관별 지표간	A 기관	평가점수 Aa	평가점수 Ab	평가점수 Ac	평가점수 Ad
	B 기관	평가점수 Ba	평가점수 Bb	평가점수 Bc	평가점수 Bd
	C 기관	평가점수 Ca	평가점수 Cb	평가점수 Cc	평가점수 Cd

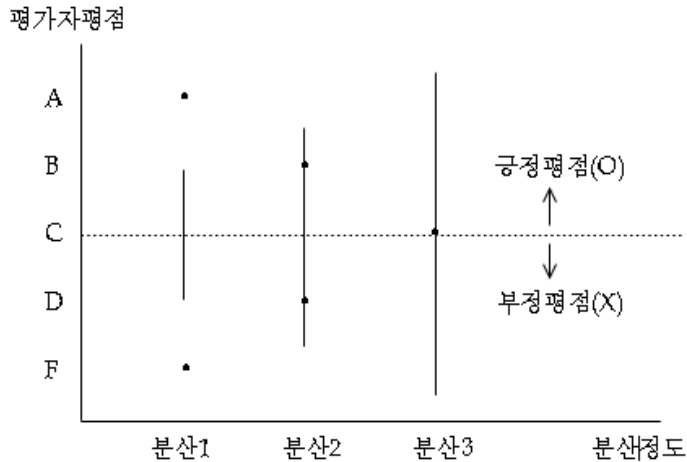
2. 긍정적 평가와 부정적 평가의 차별화

이러한 상대평가의 오류가능성은 실질적으로 긍정적 평가결과와 부정적 평가결과에 따라 차이가 있을 것으로 본다. 평가자의 긍정적 평점은 상대평가 요인

에 의하여 영향을 받기 쉬운 반면, 부정적 평점은 상대평가요인에 의하여 영향을 받지 않을 것으로 추론된다. 먼저 지표별 기관 간 상대평가의 경우 상대평가요인과 평점과의 관계는 높을 것이다. 즉 하나의 평가지표에 대하여 한 기관의 평점은 나머지 기관들의 평점과 연관성이 높을 것이다. 예를 들면 타 기관들의 평균 평점이 낮을 경우에는 해당기관의 평점은 높아진 반면, 타 기관들의 평균 평점이 높을 경우에는 해당기관의 평점은 낮아질 것이다. 따라서 긍정적 평점은 평가자의 입장에서는 다른 기관들에 대한 평점의 영향을 많이 받을 것이다. 즉 평점이 중앙에 집중되어 있을 경우 해당기관에 대한 평점은 상대평가로 인하여 긍정적인 평점에는 영향을 주지만 부정적 평점에는 영향을 주지 않을 것으로 판단된다. 평가자의 관대화, 중심화 경향에 따라 긍정적 평점의 경우 상대평가요인이 영향을 미칠 수 있는 반면, 부정적 평점의 경우에는 상대평가 요인에 의하여 영향을 받지 않을 것이다.(부정적 평점의 경우 긍정적 평가는 평균평점 이상의 평점이고 부정적 평점의 경우 평균평점 이하의 평점을 말한다.) 이 가설은 지표별 기관 간 상대평가와 기관별 지표 간 상대평가에도 동일하게 적용될 수 있을 것이다.

<가설3 > 긍정적 평점은 상대평가로 인하여 영향을 받지만 부정적 평점은 상대평가로 인하여 영향을 받지 않을 것이다.

<그림 1> 평가자 평점과 분산의 관계모형



IV. 연구대상 및 변수

1. 연구대상 및 방법

연구 자료는 상대평가 오류의 존재가능성을 분석하기 위하여 2007년도 정부 산하단체의 지표별 기관별 경영평가 결과를 활용하였다. 따라서 분석대상은 정부 산하단체에 대한 기관별 지표별 평점이 된다. 정부산하기관 경영평가는 (구) 정부산하기관관리기본법에 의해서 이루어졌으며, 2005년도 이후 매년 70여개 이상의 정부산하기관에 대하여 외부전문가에 의해서 이루어진 경영평가제도이다. 본 연구에서는 정부산하기관 경영평가 중 2006년도 실적을 대상으로 2007년도에 시행된 경영평가 결과를 직접적인 연구 대상으로 하고자 한다. 2007년도의 평가는 총 75개 정부산하기관에 대하여 총 12개 지표에 각각 9등급의 점수가 부여되었다.

2. 종속변수와 독립변수

1) 종속변수

본 연구에서 사용하는 종속변수는 특정지표 특정기관의 평가자 평점과 평균 평점 차이의 절대 값이다. 즉 [평가자 평가점수-전체 평균평점]의 절대 값인 표준점수를 활용한다. 단순한 절대평점을 종속변수로 사용하지 못하는 이유는 상대평가 요인에 의하여 평점이 어떻게 달라지는지 보기 위해서는 평균평점에 의해 표준화된 점수를 사용할 필요가 있기 때문이다. 본 연구에서는 상대평가 요인, 즉 다른 점수들의 분산도 차이에 의해서 평균 점수에 대한 점수 조정이 발생하는가를 살펴봄으로써 상대평가의 오류가능성을 확인하고자 하는바, 상대평가요인의 효과가 평균평점보다 높은지 낮은지 그 자체는 상관성이 없기 때문에 절대값을 사용하였다. 이러한 표준점수와 상대평가 요인과의 관련성에 의하여 상대평가에 의한 오류가능성을 추론하고자 한다.

2) 독립변수

독립변수는 상대평가 요인으로 자기평점을 제외한 타 기관 또는 타 지표 평점의 분산값이다. A 내지 E 평점에서 분산의 정도는 바로 상대평가의 정도를 나타낸다. 이러한 상대평가의 정도는 표준점수에 영향을 줄 수 있다는 가설에 따른 것이다. 평가 점수를 심리적으로 또는 종합평가과정에서 조정하는 경향이 존재하는가를 통계적으로 살펴볼 수 있는 방법은 다음과 같다. 중간 값이 C라고 할 때 기준에 부여한 점수가 A, E라고 하자. 그러면 위와 같은 점수 조정이 이루어질 경우 C 정도의 중간 값을 부여하게 될 것이다. 기준에 부여한 점수가

A, B 라고 하자. 그러면 점수 조정이 이루어질 경우 남은 점수에 C나 D를 부여한다. 기존에 부여한 점수가 C, C라고 하면, 남은 점수는 B나 D 정도의 점수를 부여한다. 즉 기존에 부여한 점수가 중간 값에서 멀리 떨어져 있을수록 남은 점수는 중간 값에 가까운 값을 부여하려는 경향이 강해진다. 또한 기존에 부여한 점수가 중간 값에 몰려 있을 때에는 남은 점수는 중간 값에서 멀리 떨어진 점수를 부여할 가능성이 높아진다. 즉 부여할 수 있는 점수의 중심 값이 결정되어 있을 경우(정부산하기관 평가에서는 C가 중간 값으로 규정되었었다.) 이미 부여한 점수의 분산 값이 크면 남은 점수는 평균과 비슷한 수준에서 부여할 가능성이 크고, 이미 부여한 점수의 분산 값이 적으면 남은 점수는 평균과 차이가 나는 수준에서 부여할 가능성이 높아진다. 즉 다음과 같은 관계가 성립한다. 자기 점수를 제외한 다른 점수들의 분산 값이 증가할수록 (X-m)의 표준점수인 절대값은 작아질 것이고, 자기 점수를 제외한 다른 점수들의 분산 값이 감소할수록 (X-m)의 표준점수인 절대 값은 커질 것이다³⁾.

본 연구의 독립 변수와 종속변수들 간의 상관관계를 살펴보면 다음의 <표 4>와 같다.

<표 4> 변수 간 상관계수

변수구분	평점분산 1 (지표별 기관간)	평점분산 2 (기관별 지표간)	평가점수-평균점수 1 (지표별 기관간)	평가점수-평균점수 2 (기관별 지표간)
평점분산 1 (지표별 기관간)	1.000			
평점분산 2 (기관별 지표간)	0.003	1.000		
평가점수-평균점수 1 (지표별 기관간)	-0.023	-0.023	1.000	
평가점수-평균점수 2 (기관별 지표간)	-0.049	-0.049	0.372	1.000

V. 분석결과

1. 지표별 기관 간 상대평가

1) 전체 및 유형별 회귀분석 결과

본 연구는 상대평가 과정에서 평가자 평점이 다른 평점으로부터 영향을 받는 지 여부를 분석하였다. 지표별 기관 간 상대평가는 특정지표 평점이 자기평점을

3) X는 자기 점수값, m은 평균값. 이하 (X-m)의 절대값을 ‘평균 차이 점수’로 부르고자 함.

제외한 다른 기관들 평점으로부터 영향을 받는지 여부이다. 본 연구는 이를 자기점수를 제외한 다른 기관들 점수의 분산을 통하여 분석하였다. 이와 같은 지표별 기관 간 상대평가의 분석결과는 <표 5>에서 보는 바와 같다.⁴⁾ 즉 <표 5>는 자기 점수를 제외한 다른 기관 점수에 대한 분산정도와 평균 점수와 자기 점수와의 절대 값 차이간의 관계에 대한 회귀 분석결과이다.

<표 5> 회귀분석 결과 (종속변수 : (평가 점수-평균 점수)의 절대값)

변수/유형	전체	1유형	2유형	3유형	4유형	5유형	6유형	7유형	8유형
절편값	0.965***	1.462***	0.945***	1.076***	0.759***	1.358***	0.953***	1.004***	0.004
평점분산 계수값	-0.074**	-0.483***	-0.083	-0.164	0.110	-0.223**	-0.092	-0.019	0.995***
R 제곱	0.004	0.076	0.008	0.014	0.008	0.086	0.012	0.000	0.204
F 유의값	0.028	0.000	0.352	0.137	0.218	0.011	0.309	0.822	0.000
빈도수	1078	165	105	150	195	75	90	119	179

※ 유의수준 * : 0.05-0.10, ** : 0.01-0.05, *** : 0.01이하

<가설 1>에서 설명한 바에 의하면, 자기제의 분산 값이 크다면 해당 기관의 점수를 평균치에 가깝게 점수를 부여할 것이고, 자기제의 분산값이 적다면 해당 기관의 점수를 평균에서 차이가 큰 점수를 부여할 가능성이 높다. 따라서 자기제의 분산값이 해당 기관의 점수에 영향을 미친다면, 그 계수값은 -로 산출된다.

<표 5>에서 모든 표본을 대상으로 하는 전체 회귀분석 결과 가설을 지지한 것으로 나타났다. 즉 종속변수인 자기평점과 평균평점과의 차이와 독립변수인 자기제의 분산값의 회귀계수가 -0.074로 상호 유의미한 관계를 보이고 있다. 즉 지표별 기관간 상대평가의 과정에서 해당지표에 대한 평점에서 다른 기관들의 평점이 특정기관의 평점에 영향을 준 것으로 나타났다. 특히 이러한 관계는 부(-)의 관계를 보이는 것으로 다른 기관들 평점의 분산이 높아질수록 해당기관의 평점과 평균평점과의 차이는 낮아지게 된다.

<표 5>에서 자기제의 분산값이 평균차이 점수에 미친 영향을 유형별로 보면, 8개 유형 중 6개 유형에서 계수가 부(-)의 값을 가진다. 이는 전체적으로 분석한 결과와 유사한 결과로 해석된다. 특히 유형별로 <1유형>, <5유형>, <8유형>에서 다른 기관들 평점의 분산값이 유의미한 영향을 미치는 것으로 나타났다. 이와 같이 자기점수 제외 분산값과 평균 차이 점수 간 관계를 회귀분석으로 살펴본

4) 구체적으로 평가점수 1의 자기점수 제외 분산은, 한 팀에 의해서 평가가 이루어지는 평가점수 1, 평가점수 2, 평가점수 3의 점수 중에서 평가점수 1을 제외한 평가점수 2, 평가점수 3의 분산이다. 평균차이 점수는 [평가점수 1-평가점수 1, 2, 3의 평균] 값의 절대값이다. 이 두 수치를 산정하여 자기점수 제외 분산값이 평균차이 점수에 영향을 미치는가를 분석하였다.

결과에 의하면, 자기점수 제외 분산 값은 어느 정도 해당 기관의 점수 부여에 영향을 미치는 것으로 판단된다. 평가 유형마다 영향이 미치는 경우도 있고 영향을 미치지 않는 경우도 존재하지만, 위 분석 결과는 기존에 부여한 점수의 분산 차이에 의해서 나머지 점수가 조정되는 경향이 어느 정도 존재할 수 있다는 것을 보여준다.5)

2) 긍정적 평가와 부정적 평가에 대한 회귀분석 결과

상대평가의 가능성을 <가설 3>에서 보는 바와 같이 평점이 긍정적인 경우와 부정적인 경우를 분리하여 분석한 결과는 <표 6> 및 <표 7>과 같다.

<표 6> (평가점수-평균평점) ≥ 0 인 경우

변수/유형	전체	1유형	2유형	3유형	4유형	5유형	6유형	7유형	8유형
절편값	1.043***	1.666***	0.979***	1.214***	0.559***	1.656***	0.991***	0.927***	0.257
평점분산값	-0.144***	-0.637***	-0.169	-0.324**	0.254*	-0.371***	-0.141	0.018	0.729***
빈도수	538	80	57	82	101	46	45	59	79
R제곱	0.014	0.108	0.026	0.047	0.030	0.206	0.033	0.000	0.170
F 유의값	0.006	0.003	0.231	0.049	0.084	0.006	0.235	0.896	0.000

※ 유의수준 * : 0.05-0.10, ** : 0.01-0.05, *** : 0.01이하

<표 7> (평가점수-평균평점) < 0 인 경우 (절대값)

변수/유형	전체	1유형	2유형	3유형	4유형	5유형	6유형	7유형	8유형
절편값	0.891***	1.245***	0.921***	0.908***	1.011***	1.039***	0.873***	1.078***	-0.340*
평점분산값	-0.009	-0.324*	-0.002	0.033	-0.052	-0.067	-0.005	-0.050	1.419***
R 제곱	0.000	0.044	0.000	0.001	0.003	0.009	0.000	0.004	0.255
F 유의값	0.842	0.053	0.982	0.811	0.620	0.553	0.972	0.642	0.000
빈도수	540	85	48	68	94	29	45	60	100

※ 유의수준 * : 0.05-0.10, ** : 0.01-0.05, *** : 0.01이하

먼저 <표 6>에서 보는 바와 같이 지표별 기관 간 평가에 있어서 평균평점 이상 긍정적으로 평가한 경우에는 다른 기관들 평점분산과 종속변수인 평균차이점과의 평점차이 상호간에 유의미한 부(-)의 관계를 보이고 있다. 즉 다른 기

5) <표 5>의 분석 결과 및 이하의 분석 결과에서 R 제곱 값이 매우 낮은 수치임에 유의할 필요가 있다. 이는 기관별 평가점수 차이는 다른 많은 독립변수에 의하여 설명될 수 있으며, 아주 적은 부분만이 평가 오류에 의해서 설명될 수 있다는 것을 의미한다. 즉 실제 평가 점수 차이 중 평가오류와 관련된 부분은 유의한 결과의 경우에도 그 정도는 매우 적은 값인 것으로 해석된다.

관들 평점의 분산값이 클 경우에는 평균평점과의 차이가 작아지는 경향을 보인다. 다른 기관들의 평점에서 어느 정도 분산되어 있는 경우 평균평점에 가깝게 평가하는 경향을 보인 반면, 다른 기관들 평점이 집중되어 있는 경우 평균평점에 멀게 평가하는 경향이 존재한다. 긍정적인 평점의 경우에는 상대평가에 의하여 다른 기관에 대한 평점이 해당기관의 평점에 영향을 미치는 것으로 해석된다.

<표 7>의 경우는 부정적인 평가의 경우이다. 부정적인 평가의 경우 긍정적인 평가와는 달리 다른 기관들의 평점 분산값과 평균평점과 평가점수와의 차이가 유의미한 관계를 보여주지 못하고 있다. 즉 다른 기관들의 평점 분산값이 크더라도 평균평점에 집중하는 경향은 보이지 않으며, 다른 기관들의 평점 분산값이 작더라도 평균평점에 멀게 평가하지는 않는 것으로 나타났다.

전체에 대한 회귀분석 결과와 마찬가지로 유형별 분석결과에 있어서도 유사한 결과가 도출되었다. 긍정적인 평점의 경우 <1유형>, <3유형>, <5유형>, <8유형>에서는 다른 기관들의 평점 분산값이 유의미한 관계를 보인 반면, 부정적인 평점의 경우에는 <유형 8>의 경우를 제외하고는 유의미한 관계를 보이고 있지 않다. 그리고 <유형 8>의 경우도 계수 값이 + 로서 가설에 적합하지 않다. 이는 부정적인 평가, 즉 평균평점 이하의 평가를 할 경우에는 다른 기관들의 평점 분산값에 따른 경향성을 보이지 않고 있어 상대평가보다는 관대화 경향이 강하게 작용하고 있는 것으로 해석된다.

2. 기관별 지표 간 상대평가

1) 전체 및 유형별 회귀분석 결과

앞에서 살펴본 바와 같이, 상대평가는 특정지표에 대한 기관간의 상대평가와 특정기관에 대한 지표간의 상대평가로 구분된다. 지표별 기관간의 상대평가와 별도로 기관별 지표간의 상대평가 또한 평점부여의 중요한 결정요인이다. 기관별 지표 간 상대평가는 특정기관 평점이 해당지표 평점을 제외한 다른 지표들의 평점으로부터 영향을 받는지 여부이다. <가설 2>에서 보는 바와 같이 자기점수를 제외한 다른 지표들 평점 분산값이 크다면 해당 지표의 점수를 평균치에 가깝게 점수를 부여할 것이고, 작다면 해당 해당지표의 점수를 평균에서 차이가 큰 점수를 부여할 가능성이 높다는 것이다. 따라서 다른 지표들 평점 분산값이 해당 지표의 평점에 영향을 미친다면, 그 계수값은 부(-)로 산출된다. 이러한 기관별 지표간 회귀분석결과는 <표 8>과 같다. 한 기관에 있어서 자기점수를 제외한 다른 지표 점수에 대한 분산정도와 평균점수와 자기점수와의 차이간의 관계에 대한 회귀 분석결과이다.

<표 8> 회귀분석 결과 (종속변수 : (평가 점수-평균 점수)의 절대값)

변수/유형	전체	1유형	2유형	3유형	4유형	5유형	6유형	7유형	8유형
절편값	0.065	0.291	0.039	0.028	0.260	0.337	-0.404	0.214	0.086
평점분산값	-0.095	-0.296	-0.036	-0.036	-0.268	-0.546	0.296	-0.253	-0.144
R 제곱	0.002	0.011	0.000	0.000	0.011	0.012	0.018	0.008	0.003
F 유의값	0.189	0.179	0.875	0.826	0.142	0.344	0.208	0.346	0.435
빈도수	1078	165	105	150	195	75	90	119	179

※ 유의수준 * : 0.05-0.10, ** : 0.01-0.05, *** : 0.01이하

<표 8>에서 모든 표본을 대상으로 하는 전체 회귀분석 결과 <가설 2>를 지지하지 않는 것으로 나타났다. 종속변수인 해당지표의 자기평점과 평균평점과의 차이와 독립변수인 다른 지표들 평점분산값의 회귀계수가 통계적으로 유의미한 관계를 보이지 않고 있다. 즉 기관별 지표간의 상대평가의 과정에서 한 기관 내 해당지표에 대한 평점에서 다른 지표들의 평점이 특정지표의 평점에 영향을 주지 않는 것으로 나타났다. 각 기관 유형별로 살펴볼 경우에도 통계적으로 유의미한 결과가 도출되지 아니하였다.

2) 긍정적 평가와 부정적 평가에 대한 회귀분석 결과

지표별 기관간 상대평가의 경우와 마찬가지로 기관별 지표 간 상대평가의 가능성도 <가설 3>에서 보는 바와 같이 평점이 긍정적인 경우와 부정적인 경우로 구분될 수 있다. 분석결과는 <표 9> 및 <표 10>과 같다.

<표 9> (평가 점수-평균 점수) ≥ 0 인 경우

변수/유형	전체	1유형	2유형	3유형	4유형	5유형	6유형	7유형	8유형
절편값	0.762***	1.352***	0.519**	0.524***	0.932***	1.680***	0.464**	0.940***	0.700***
평점분산값	-0.014	-0.526**	0.202	0.297*	-0.289*	-1.618***	0.329*	-0.174	0.156
빈도수	522	75	58	68	110	34	41	58	78
R 제곱	0.000	0.067	0.019	0.052	0.026	0.222	0.073	0.008	0.010
F 유의값	0.839	0.025	0.305	0.062	0.094	0.005	0.087	0.515	0.392

※ 유의수준 * : 0.05-0.10, ** : 0.01-0.05, *** : 0.01이하

<표 10> (평가 점수-평균 점수) < 0 인 경우 (절대값)

변수/유형	전체	1유형	2유형	3유형	4유형	5유형	6유형	7유형	8유형
절편값	-0.672***	-0.572***	-1.179***	-0.379***	-0.755***	-0.431	-1.038***	-0.793**	-0.799***
평점분산값	-0.077	-0.126	0.249	-0.328***	-0.095	-0.198	0.168	-0.017	0.121
R 제곱	0.003	0.008	0.036	0.094	0.006	0.006	0.009	0.000	0.005
F 유의값	0.227	0.401	0.200	0.005	0.482	0.637	0.520	0.950	0.471
빈도수	556	90	47	82	85	41	49	61	101

※ 유의수준 * : 0.05-0.10, ** : 0.01-0.05, *** : 0.01이하

분석 결과 전체 표본을 대상으로 할 경우 긍정적 평가와 부정적 평가 모두에서 다른 지표들 평점 분산값이 평균평점과 평가점수와의 차이에 통계적으로 유의미한 영향을 미치지 않는 것으로 나타났다.

다만, 유형별로 분석할 경우에는 유의미한 관계가 다소 나타났다. <표 9>에서 보는 바와 같이 긍정적인 평가의 경우 8개 유형 중 2개 유형에서 다른 지표들 평점 분산값과 평균평점-평가점수의 차이가 부(-)의 유의미한 관계를 보이고 있다. 즉 다른 지표들 평점 분산의 정도가 클수록 해당지표의 평균평점에 가깝게 평가하는 경향이 있으면, 다른 지표들 평점 분산의 정도가 작을수록 해당지표의 평균평점에 멀게 평가하는 경향을 보이고 있다. 또한 <표 10>의 부정적 평가의 경우에도 1개 유형에서 유의미한 부(-)의 관계를 보이고 있다.

3. 결과 종합 및 토의

기관 전체의 결과를 중심으로 지금까지의 분석결과를 종합하면, <표 11>에서 보는 바와 같이 <가설 1>의 지표별 기관 간 평가는 유의미한 반면 <가설 2>의 기관별 지표간 평가는 유의미하지 않는 것으로 분석되었다.

<표 11> 상대평가로 인한 기관별·지표별 평점의 오류가능성 통계분석결과(전체)

구분	지표별 기관간	기관별 지표간
전체	○ (높은 오류가능성)	× (낮은 오류가능성)
평균이상 평점	○ (높은 오류가능성)	× (낮은 오류가능성)
평균미만 평점	× (낮은 오류가능성)	× (낮은 오류가능성)

상대평가의 오류가능성, 즉 다른 기관들 또는 다른 지표들의 평점 분산값이 평균평점과 평가점수간의 차이에 영향을 주는 것은 지표별 기관간 평가에서 강하게 나타난 반면, 기관별 지표 간 평가에서는 거의 나타나지 않았다. 즉 이는 평가자들이 상대평가 과정에서 특정기관에 대한 지표간의 상대평가보다는 특정

지표에 대한 기관간의 상대평가가 보다 강하게 작용하고 있는 것으로 판단된다. 이는 총점 중심으로 기관 상호간의 상대적 순위를 정함으로써 특정지표를 평가하는 평가자는 기관 상호간의 상대평가에 중점을 두게 되는 경향이 있다는 것을 의미하며, 기관별 지표간의 평가에 중점을 두지 않는 것은 지표 간 상대적 우열을 중요시하지 않는 평가방침에서 기인한 것으로 해석된다.

또한 전체 기관을 대상으로 한 지표별 기관간의 긍정적 평점과 부정적 평점을 비교할 경우, <표 11>에서 보는 바와 같이 긍정적 평점에서는 상대평가의 오류가능성이 유의미한 것으로 나타난 반면, 부정적 평점의 경우 상대평가의 오류가능성이 유의미하지 않는 것으로 나타났다. 또한 전체 기관의 기관별 지표간 평가에서는 둘 다 유의미한 결과를 보이지 않고 있다. 즉 지표별 기관 간 평가에서 긍정적 평점을 할 경우 평가자들은 다른 기관들 평점 분산값에 민감하며 이에 따라 평점이 평균평점으로 중심화하는 경향을 보이며, 부정적 평점의 경우에는 다른 기관들 평점 분산과 상관없이 다른 요인에 의하여 관대화하는 경향이 있는 것으로 추정된다.

이러한 전체적인 결과는 기관유형별로 구분하여 분석할 경우 보다 구체적이고 실질적인 분석이 가능하다. 앞에서 언급한 바와 같이, 실제 정부산하기관 경영평가는 기관유형별로 구분하여 이루어지고, 평가자들도 기관유형별로 구성되었다. 또한 실제 기관에 대한 현장방문도 평가자 그룹별로 진행되기 때문에 기관유형별 상대평가의 오류가능성 분석이 중요한 의미를 가진다. <표 12>은 기관유형별로 구분하여 유의미한 관계와 그 방향성을 정리한 표이다.

<표 12> 상대평가로 인한 지표별·기관별 평점의 오류가능성 통계분석결과(유형별)

구분	지표별 기관간	기관별 지표간
전체	3개 기관유형(+1, -2)	0개 기관유형(+0, -0)
평균이상 평점	4개 기관유형(+1, -3)	2개 기관유형(+0, -2)
평균미만 평점	1개 기관유형(+1, -0)	1개 기관유형(+0, -1)

* ()의 숫자는 유의미하게 나온 기관유형의 수를 표시함. 다만, '+'는 정의 계수를 가진 기관유형의 수를, '-'는 부의 계수로 가진 기관유형의 수를 나타냄.

** $p < 0.05$ 의 유의수준을 기준으로 하였음.

<표 12>에서 보는 바와 같이, 전체적으로 48개의 분석모델⁶⁾ 중 총 11개의 분석모델에서 유의미한 결과가 나왔으며, 이중 8개 분석모델에서 다른 기관들 또는 다른 지표들의 평점 분산값과 평균평점-평가점수의 차이와 부(-)의 유의미한 관계를 보인 반면, 정(+)의 관계는 3개 모델에 불과하였다. 즉 기관유형별로 구

6) 8개 기관유형 × 3개 분석대상(전체, 평균이상, 평균미만) × 2개 분석기준(지표별 기관간, 기관별 지표간)으로 계산된 것임.

분하여 분석할 경우 상대평가의 오류가능성은 적지 않은 것으로 나타나며 다양한 기관유형에 넓게 퍼져 있는 것으로 해석된다.

VI. 결론 및 정책적 시사점

본 연구는 2007년도 정부산하기관 경영평가 결과를 대상으로 평가과정에서 발생하는 상대평가의 오류가능성을 분석하고자 하였다. 상대평가의 오류가능성을 지표별 기관 간 상대평가 오류 가능성과 기관별 지표 간 상대평가 오류 가능성으로 구분하였고, 다른 기관들 및 다른 지표들 평점의 분산값과 평균평점-평가점수의 차이와의 관계를 통하여 이를 실증하고자 하였다. 전체 기관을 대상으로 한 분석결과, 지표별 기관간의 평가에 있어서 상대평가 요인, 즉 다른 기관들 평점의 분산정도와 평균평점-평가점수의 차이 간에 유의미한 관계가 있는 반면, 기관별 지표 간 평가의 경우 다른 지표들 평점 분산값과 평균평점-평가점수의 차이 간에 유의미한 관계가 발견되지 않았다. 즉 지표별 기관 간 평가의 경우 다른 기관들 평점의 분산값이 클수록 평균평점에 가깝게 평점을 부여하는 경향을 보였다. 또한, 지표별 기관 간 평점에 있어서 긍정적인 평점의 경우 다른 기관들 평점의 분산정도가 평균평점-평가점수의 차이에 영향을 미치는 반면, 부정적인 경우에는 유의미한 결과가 도출되지 않았다. 그리고 기관유형별로 구분하여 분석한 결과 상대평가의 오류가능성을 좀 더 상세히 확인할 수 있었다. 많은 기관유형별 평가에서 다른 기관들 또는 다른 지표들 평점의 분산값과 평균평점-평가점수의 차이와의 관계가 유의미하게 나타났다.

이와 같은 분석결과는 다음 몇 가지 정책적 시사점을 제공한다. 첫째, 그동안 평가 과정에서의 오류로는 관대화 경향, 엄격화 경향, 평가 지표 이외 요소의 영향을 받는 후광효과 등이 주로 논의되어 왔다. 이에 본 결과는 이와같이 일반적으로 인정되는 평가 오류 외에 상대 평가로 인한 평가 오류가 존재할 가능성이 있다는 점을 제시한다. 평가 과정에서 발생하는 평가 오류를 감소시키기 위하여 평가 제도를 변경하는 등 많은 평가 노력이 이루어지고 있는바, 상대평가 오류도 하나의 오류로 인정하여 이를 최소화할 수 있는 방안도 정책적 평가 과제의 하나로 수용되어야 할 필요가 있다.

둘째, 상대 평가 오류를 감소시킬 수 있는 방안으로는 우선적으로 절대평가를 가미하는 방식의 도입을 고려할 수 있다. 상대 평가 오류는 평가가 상대평가로 이루어질때 발생하는 것으로 상대평가 시스템 하에서는 어느 정도 필연적인 측면이 존재한다. 상대평가와 절대평가 시스템 상에서 어떠한 평가 시스템을 선택할 것이냐의 문제에서는 절대평가 하에서 모두 점수가 높게 나오는 관대화

경향의 문제, 상대평가 하에서 높은 성과를 달성하였음에도 낮은 점수가 나오는 문제 등이 고려된다. 이러한 고려 사항 중에서 상대평가 하에서는 상대평가 오류가 존재한다는 것을 추가적으로 고려하여 평가 시스템을 선정할 필요가 있다. 상대평가 하에서 추가적인 오류 가능성이 존재한다면, 기존의 상대평가에 절대평가를 가미한 형태로 평가 제도를 운영하여 상대평가의 오류가능성을 최소화할 필요가 있다. 평가등급체계를 축소하거나 운영과정에서 절대평가의 기준을 명확히 하는 등이 절대평가를 가미하는 하나의 방안이 될 수 있다⁷⁾.

셋째, 전체적인 분석결과 상대평가의 오류가능성이 기관별 지표 간 평가보다는 지표별 기관간의 평가에서 주로 발생하고 있다. 이러한 결과는 경영평가에 있어서 기관별 평가 순위가 민감한 영향을 미치는 현실을 반영하고 있는 것으로 추정된다. 기관별 평가 순위가 평가자 및 피평가자에게 큰 영향을 미치기 때문에 기관별 평가에 있어서는 낮은 점수를 부여하지 않으려는 상대평가 오류 경향이 나타나고 있는 것으로 보인다. 이러한 문제를 보완하기 위해서는 평가지표별 기관순위에 대한 평가자의 심리적 부담을 줄이기 위하여 평가지표 간 상대적 우열을 중요시하는 방향으로 평가방침을 조정할 필요가 있다(Mark and Henry, 2004). 즉 기관 종합 평가 순위를 중시하는 현재의 평가 시스템 분위기에서, 종합순위를 제외하고 각 지표별 평가순위를 중시하고 각 지표별 평가 순위로 기관 간 차별화를 도모하는 방향으로 나아가는 것이 바람직하다.

넷째, 관대화 효과, 후광효과 등 평가자의 오류는 평가에 대한 사전 교육에 의해 많은 치유가 가능하다. 이러한 평가 오류는 평가자의 무의식이 많이 작용하기 때문에 그러한 오류 발생 가능성을 의식하고 평가한다면 많은 평가 오류를 사전에 치유할 수 있다. 상대평가 오류도 평가자의 무의식적 측면에 영향을 많이 받는 오류로서 사전에 그에 대해 평가자가 인식한다면 사전에 치유가 가능한 오류라 할 수 있다. 현재 평가자들에 대해서 특별한 평가 교육이 이루어지지 않고 있는바, 평가자들을 대상으로 사전에 다양한 평가 오류를 예방할 수 있는 평가자훈련프로그램을 시행하는 것이 요구된다(Brown, 1995; 김종희, 2007).

본 연구는 현재까지의 공공기관 경영평가의 연구에서 상대적으로 관심이 미약했던 평가방법의 신뢰성에 대한 실증적 분석을 시도하였다. 공공기관 평가에 있어서의 상대평가로 인한 오류 가능성을 분석하고, 상대평가 제도와 그 운영과정에 제도적 시사점을 제공하고자 하였다는데 그 연구의 의의가 있다. 정부 및 공공기관에서 평가와 보상이 시대적 흐름처럼 인식됨에 따라 높은 신뢰성과 객관성을 통하여 생산적인 평가제도의 정착과 발전에 기여할 수 있기를 기대하였

7) 앞에서 지적한 바와 같이, 본 논문은 2007년도 분석자료를 토대로 서술하였으나, 2009년 현재 이미 절대평가를 가미한 형태로 운영되고 있고, 평가등급도 9등급에서 6등급으로 축소는 등 개선이 일부 이루어졌다(기획재정부, 2009).

다. 본 연구 결과는 향후 평가의 신뢰성과 피평가자의 수용성을 제고할 수 있는 제도개선의 방향을 제시하고자 하는 기초적 연구로서 공공기관은 물론 정부기관에서 확산되고 있는 다양한 평가에도 적용 가능할 것으로 생각된다.

참고문헌

- 곽병한. (2005). 정부투자기관 경영평가지표의 인과관계에 관한 실증분석. 한국유통과 학회 2005년 동계 정기학술대회
- 곽채기. (2003). 정부투자기관의 비효율성 통제를 위한 경영평가제도의 역할 및 운영 성과. 『공기업논총』 15(1).
- 곽채기. (2008). 공기업 선진화방안 - 공기업 개혁을 위한 법·제도적 기반 구축 방안을 중심으로. 『한국행정학회 추계학술대회 발표논문집(6)』 pp. 81-100.
- 김종희. (2007). 지방공기업 경영평가의 수용도 제고방안, 『지방행정연구』 21(1).
- 김준기. (2001). 공기업 경영평가제도의 성과와 향후 과제 『한국행정연구』 10(1).
- 노화준. (2002). 공공부문 성과측정의 이슈와 정책평가제도의 개혁방향. 『한국정책분석 평가학회보』 10(2).
- 박명호·이성욱. (2000). 정부투자기관 경영평가를 위한 비계량 평가지표의 개선방향. 『공기업 논총』 12(1).
- 박석희. (2006). 공공부문의 성과관리와 조직역량: 차단된 시계열설계를 통한 13개 공기업의 실증분석. 『한국행정정보』 40(3). pp. 219-244.
- 박영범. (1998). 경영환경변화의 수용을 위한 정부투자기관 경영평가제도의 비계량지표의 개선방안. 『공기업논총』 10(1).
- 송대회. (2003). 공기업 경영평가제도의 회고와 발전 방향. 『공기업논총』, 15(1). pp. 1-10.
- 신동일. (2001). 채점 경향 분석을 위한 Rasch 측정 모형 적용 연구. 『Foreign Languages Education』 8(1). pp. 249-272.
- 신동일·장소영. (2002). 후광효과의 이해를 통한 채점 오류 분석 연구. 『Foreign Languages Education』 9(4). pp. 215-232.
- 오연천. (1998). 공기업 구조조정과 경영평가제도의 개선 방향. 『공기업논총』 10(1). pp. 3-10.
- 유민봉·임도빈. (2007). 인사행정론, 박영사.
- 윤기웅·공동성. (2008). 재정사업 자율평가제도의 평가. 『한국행정연구』 17(2). pp. 89-113.
- 이석원. (2005). 조정성과지표를 사용한 공공기관의 성과측정 : 정부투자기관경영평가

- 에의 적용. 『한국행정학보』 39(4). pp. 219-244.
- 이창길·최성락. (2009). 정부산하기관 경영평가에서의 후광효과(Halo Effect)에 관한 연구. 『한국행정학보』 43(3). pp. 151-172.
- 이영식. (1998). 영어 작문 채점에 대한 평가 신뢰도에 대한 분석, 『영어교육』, 52(1). pp. 221 -242.
- 이환범·송건섭·김병문. (2005). 지방공기업의 경영성과관리와 평가지표개발 : 자료포락 분석(DEA)에 의한 효율성 측정. 『한국사회와 행정연구』 15(4).
- 기획예산처. (2007). 『2007년도 정부산하기관 경영평가단 워크숍 자료집』.
- 기획재정부. (2009). 『2009년도 공기업·준정부기관 경영평가단 워크숍 자료집』.
- 최연희. (2002). 채점자 훈련이 영어 작문 채점에 미치는 효과에 대한 연구: FACETS 분석을 통한 신뢰도 변화 분석. 『응용언어학』 18(1). pp. 257-292.
- 최성락·박민정. (2009). 공기업 경영평가 제도의 성과분석. 『행정논총』 47(1). pp. 183-208.
- Bernardin, H. J. (1978). Effects of rater training on leniency and halo errors in student ratings of instructors. *Journal of Applied Psychology*, 63(3), 301-308
- Borman, W. C (1978) Exploring upper limits of reliability and validity in job performance ratings. *Journal of Applied Psychology*, 63(2), 135-144
- Brown, A (1995) The effect of rater variables in the development of an occupation-specific language performance test. *Languages Testing*, 12(1), 1-15
- Dye, Ronald A. (1992). Relative performance evaluation and project selection, *Journal of Accounting Research*, Vol. 30 No. 1, Spring, p. 27-52.
- Facteau,, J.D., and Craig, S.B. (2001). Are performance appraisal ratings from different rating sourced comparable? *Journal of Applied Psychology*, 86, 215-227
- Gabris, G., and K. Ihrke. 2000. Improving employee acceptance toward performance appraisal and merit systems: The role of credible leadership. *Review of Public Personnel Administration* 20(1): 41 - 53.
- Gibbons, Robert, and Kevin J. Murphy. (1990). Relative performance evaluation for chief executive officers, *Industrial and Labor Relations Review*, 43(3), 30-51.
- Hedge, J., and M. Teachout. (2000). Exploring the concept of acceptability as a criterion for evaluating performance. *Group and Organisation Management* 25(1): 22 - 44

- Jenter, Dirk and Fadi Kanaan. (2006). *CEO Turnover and Relative Performance Evaluation*, Working paper, Stanford University.
- King, L.M., Hunter J.E., & Schmidt.F.L (1980). Halo in a multidimensional forced choice performance evaluation scale. *Journal of Applied Psychology*, 65, 507-516
- Mark, M. M & Henry, G T. (2004) The mechanism and outcomes of evaluation influence, *Evaluation*, 10(1), 35-57
- Nankervis, Alan R. and Robert-Leigh Compton, (2006). Performance management: Theory in practice?, *Asia Pacific Journal of Human Resources* 2006; 44; 83-101
- Stanton, M (1991) Increasing the accountability of government sponsored enterprises: Next Steps, *Public Administration Review*, 51(6), 572-575
- Steers, R. M (1975) Problems in the measurement of organizational effectiveness . *Administration Science Quarterly*, 20: 546-558

Abstract

An Analysis on Rating Errors by Relative Performance Evaluation for Public Enterprises

Chang Kil Lee · Seong Rak Choi

This paper analyzes the possible existence of rating errors by relative performance evaluation for Korean public enterprises. Data was collected from the 2007 performance evaluation results for 75 Korean public enterprises. The results of the analysis show that rater errors are significantly involved in the process of relative performance rating among enterprises by individual indicators rather than among indicators by individual enterprises. In particular, such errors are found more significant in the ratings with lower scores rather than that with higher scores. The results are also found in the analysis by functional groups. These results suggest the complementary reconciliation of absolute rating by the level of performance and relative rating by forced distribution.

【Key words: relative performance evaluation, forced distribution, rating error, public enterprise】