

# 국가연구개발사업 연구과제 선정방식 개선에 관한 연구 : 한국연구재단 지원사업을 중심으로

A Study on Evaluation and Screening Methods of Research Proposals for National  
Research Foundation of Korea Grants

이윤규(Youn-Kou Lee)\*, 손충근(Chung-Geun Son)\*\*

## 목 차

- |                     |                                 |
|---------------------|---------------------------------|
| I. 서론               | IV. 국가연구개발사업 연구과제 선정방식<br>및 문제점 |
| II. 주요 선행연구         | V. 사례분석 및 대안의 제시                |
| III. 연구과제의 선정과 동료평가 | VI. 결론                          |

## 국 문 요 약

국가연구개발사업의 투자효율성을 높이기 위해서는 무엇보다도 평가시스템이 잘 갖춰져야 한다. 왜냐하면 혹시라도 특정 평가자에 의해 평가결과가 왜곡 된다면 정부가 바라는 최적의 연구팀을 선정하지 못할 것이고, 이로 인해 바람직한 연구성과를 기대하기 어려울 것이기 때문이다. 현행 국가연구개발사업 평가과정에서는 특정 평가자의 편견(특이치)을 차단하기 위해 대부분 최고/최저점 제외방식을 사용하고 있다. 그러나 우리가 관례적으로 사용하고 있는 이 방식에 한계가 있을 수 있다. 따라서 본 연구에서는 실증분석을 통해 이 방식이 갖고 있는 한계에 대한 문제점을 도출하고, 이를 보완할 수 있는 현실적 대안을 제시함으로써 국가연구개발사업 평가제도의 신뢰성 제고 방안을 모색하였다. 이러한 결과들은 피 평가자들의 평가결과에 대한 수용도 제고 및 국가연구개발사업 효율성 증진에 기여할 수 있을 것이다. 또한 순위결정의 방법론 측면에서 다양한 분야에 도움을 줄 것으로 기대된다.

핵심어 : 국가연구개발사업, 연구과제 평가, 채점방식, 순위결정

※ 논문접수일: 2009.6.30, 1차수정일: 2009.8.20, 게재확정일: 2009.8.24

\* 경기대학교 회계세무학과 교수, lyk0149@hanmail.net, 031-249-9448

\*\* 한국연구재단 선도연구지원 팀장, chung@nrf.go.kr, 042-869-6822, 교신저자

## ABSTRACT

---

To enhance the efficiency of national R&D programs, it is the most important thing to establish the optimized evaluation system. This is because the distorted evaluation could be an obstacle to select the best suited research team and lead the desirable research result.

In current evaluation process of national R&D program, it is the most frequently used method for eliminating an evaluator's biased view to exclude both the maximum and minimum scores. However, there are several limitations in this stereotype method. Therefore, this study analyzes some real cases to derive the problems from the stereotype method and then recommends the complementary alternatives which can enhance the reliability of the evaluation system for national R&D programs. The results acquired from this study seem to be helpful to increase the agreement of the applicants on the evaluation results and ultimately to maximize the efficiency of national R&D program. In addition, they can be applied to various field with respect to the methodology for ranking decision.

Key Words : National R&D Program, Evaluation of research projects, scoring methods, ranking decision

---

## I. 서 론

국가경쟁력 제고를 위해 전 세계적으로 연구개발 투자규모가 증가되고 있으며, 우리나라의 국가연구개발사업 투자규모도 2007년 9조원을 넘어섰고, 2010년에는 13조원에 도달할 것으로 예상된다. 이처럼 국가연구개발사업 투자규모가 증가됨에 따라 자연스럽게 연구비 투자효율성 제고에 대한 관심이 커지고 있다. 국가연구개발사업의 투자효율성 증대를 위해서는 무엇보다도 최적의 평가시스템이 확립되어야 한다.

국가연구개발사업의 평가주체는 연구관리전문기관, 피 평가자 및 평가자로 구성되어 있다. 이들 평가주체 중에서 일반적으로 생각할 때 연구과제 선정의 결정권이 연구관리전문기관에 있는 것으로 생각하기 쉽다. 왜냐하면 평가방법을 정하고, 평가자를 선임하는 등 연구과제 선정과 관련한 모든 일을 관장하기 때문이다. 그러나 현실에서는 평가자들의 역할이 매우 중요하다. 연구관리전문기관에서 객관적인 평가시스템을 구축하고 공정하게 평가자를 선임한다고 해도 특정 평가자에 의해 평가결과가 왜곡 된다면 정부가 바라는 최적의 연구팀을 선정하지 못할 것이고, 이로 인해 바람직한 연구성과를 기대하기 어려울 것이기 때문이다. 그래서 다소 차이는 있지만 현행 국가연구개발사업 연구과제 선정과정에서는 평가결과의 공정성 확보를 위해 대부분 최고/최저점수 제외방식을 사용하고 있다. 이는 현행 국가연구개발사업 연구과제 선정방식이 대부분 동료평가(peer review)에 의존하고 있음으로 인해 발생할 수 있는 극단치를 제외시키겠다는 것이다.

현행 한국연구재단(National Research Foundation of Korea) 등 에서 적용하고 있는 우리나라 국가연구개발사업 연구과제 선정은 동료평가를 근간으로 하되, 피 평가자에 관련된 이해관계자를 사전에 배제시키는 상피제도를 적용하고 있다. 이 제도는 평가의 공정성을 높이기 위한 방안이지만, 이로 인해 세부 기술분야와 다소 상이한 전공자가 평가자로 참여 할 수도 있으며, 해당 기술분야의 최고 전문가들이 평가자 후보에서 제외될 수 있는 한계가 있다. 또한 현행 한국연구재단의 연구과제 선정방식은 평가자들의 평가점수로 우선순위를 정하고, 예산의 범위내에서 지원하는 방식이기 때문에 평가점수 절대치 보다는 순위정보가 더 중요할 수 있다. 다시 말해서 예산사정이나 기술분야의 특성에 따라 85점을 받은 과제가 탈락되는 경우가 있는가 하면 75점을 받은 과제가 선정되는 경우도 있을 수 있기 때문에, 몇 점을 받았느냐가 중요한 것이 아니라 신청한 과제 중에서 몇 등을 하였느냐가 더 중요할 수 있는 것이다. 다시 말해, 현행 한국연구재단에서 사용하고 있는 동료평가는 평가집단내의 평가자간 합의(consensus)를 도출하는 과정으로 볼 수 있으며, 그 합의 정도를 파악하기 위해서는 각 평가자가 부여한 평가점수의 절대치 비교 보다는 각 평가자가 인식하고 있는 순위 비교가 더 의미

있는 정보라는 것이다.

이상에서 살펴본 바와 같이 현행 한국연구재단의 연구과제 선정방식은 지나치게 공정성을 추구한 나머지 그 분야에서 최고의 전문가들이 평가위원에서 제외될 위험이 있고, 최고/최저 점수 제외로 인한 정보의 손실이 크며<sup>1)</sup>, 여전히 특정 평가자의 의도된 점수에 의해 선정결과가 좌우될 수 있는 현실적인 문제점을 안고 있다. 따라서 본 연구에서는 사례분석을 통해 현 한국연구재단의 국가연구개발사업 연구과제 선정방식이 안고 있는 한계에 대한 문제점을 도출하고, 이를 보완할 수 있는 현실적 대안을 제시함으로써 국가연구개발사업 평가제도의 신뢰성 제고 방안을 모색하고자 한다. 이러한 결과들은 피 평가자들의 평가결과 수용도 제고 및 국가연구개발사업 생산성 증대에 기여할 수 있을 것이다. 또한 향후 다양한 분야에서 순위결정의 방법론<sup>2)</sup> 측면에서 도움을 줄 것으로 기대된다.

## II. 주요 선행연구

국내·외 대부분의 연구기관에서는 연구과제 평가자의 평가점수를 산정하는 과정에서 최대/최소점수를 제외하고 평균하는 방식을 택하고 있지만 한국연구재단의 인문사회과학분야에서는 패널내 평가점수의 최대값과 최소값의 차가 40점 미만인 경우 산술평균을 내며, 40점 이상일 경우 극단치(표준화한 z값이 1.3 이상인 점수)를 배제한 후 평균을 내는 방식을 택하고 있다. 이러한 평가점수 산정방법에 관한 유사 연구사례를 소개하면 다음과 같다.

이종승외 2인(1998)의 “학술연구 지원과제의 심사 평가방식 개선방안 연구”에서는 한국학술진흥재단의 공모과제 심사 및 선정의 공정성과 형평성을 향상시킬 수 있는 방안을 연구하기 위해 심사결과에 있어서 개별 심사자의 ‘엄격성’ 차이로 인한 심사자 집단내 및 집단간 차이를 해결할 수 있는 수리적 대안을 모색하였다. 이들은 극단점수의 문제를 해결하고, 세부 학문영역간 과제선정의 불균형을 막기 위하여 가장 적절한 대안은 심사자의 수를 늘리고, 표준점수<sup>3)</sup>에 의하여 극단점수를 배제시키는 방안을 제시하였다. 그러나 이 방법은 정보의 손실이 너무 많고, 평가자가 인식하고 있는 순위정보를 반영하지 못하는 단점이 있다.

- 1) 대부분 평가위원회는 7인 내외로 구성되므로 최고/최저점을 제외하면 전체 평가정보의 71.4%만 사용하고 있으며, 이는 26이상을 극단치로 제외하는 경우(95.45%) 보다도 정보손실이 크다.
- 2) 기록경기가 아닌 채점방식의 스포츠분야(피겨스케이팅, 기계체조 등) 평가점수 산정이나 인사고과(다면평가) 등 다수의 심사위원(평가위원)이 참여하여 순위를 결정해야하는 경우 다양하게 응용이 가능 할 것이다.
- 3) 여기서 말하는 표준점수는 극단점수를 구별하기 위해 연구자가 제안한 것으로, 표준점수 1.3점 이상을 극단치로 간주 할 경우 80.6%의 자료를 활용하는 것이므로 약 20%의 평가정보 손실이 발생한다.

김동욱 외 1인(2003)의 “대학평가에서 평가자간 오차 크기의 검정 및 평가자 객관성의 제고 방안”에서는 평가자의 객관성을 검토하기 위하여 같은 팀에 속한 평가자들의 평가결과 차이를 분석하여 같은 평가팀 내의 평가자간 평가결과에 상당한 차이가 있음을 발견하고, 평가결과와 평가자 신뢰도, 즉 객관도의 향상을 위해서 평가자간 일치된 평가기준의 설정, 적절한 점수분포의 확보, 평가기간과 평가인원의 확대 그리고 평가위원 훈련 프로그램의 강화 등의 대안을 제시하였다. 그러나 이 연구에서는 평가자간 평가결과의 차이가 있음을 실제 자료를 바탕으로 제시한 의미는 있으나, 현실을 해결할 수 있는 구체적인 대안으로는 부족한 점이 있다.

조만형 외 2인(2004)의 “연구과제 심사평가시스템의 한미 비교연구”에서는 한국학술진흥재단과 미국과학재단의 심사/평가시스템을 비교분석하여 우리나라의 심사/평가시스템을 발전 시키는데 도움이 될 수 있는 문제점을 도출하고자 하였다. 이 연구에서는 한국학술진흥재단은 전문가들에 대한 신뢰도 보다는 심사/평가의 공정성에 더 많은 관심을 갖고 있다고 지적하고, 공정성은 제도에 의해서 확보되기도 하지만 그 보다는 전문가의 전문적 판단에 의해서 더 완벽하게 확보될 수 있다고 주장하고 있다. 즉 연구과제 평가의 주요쟁점 중에서 공정성과 전문성을 동시에 확보할 수 있는 평가제도 개선이 절실히 필요하다는 것이다.

양정모외 1인(2008)의 “연구지원사업 표준점수 산정방식 개선을 위한 연구”에서는 2007년 한국학술진흥재단의 온라인 심사 및 패널심사를 통해 수집된 5,190개 과제의 심사정보를 바탕으로 시뮬레이션 한 결과 전체 심사 패널의 95%는 최대점수와 최소점수가 만점의 30%를 벗어나지 않는 것으로 파악되었으며, 이를 바탕으로 최대점수와 최소점수의 차가 만점의 30% 이내에 들어오면 산술평균을 내고, 그렇지 못할 경우 T-분포의 95% 신뢰도 구간 추정범위 밖의 값은 극단치로 평균점수 산정에서 제외하는 방법을 제시하였다. 그러나 이 방식도 여전히 정보손실이 많고, 순위정보를 반영하지 못하는 한계가 있다.

위의 선행연구들과 분야는 다르지만 연구과제 선정에 의미 있는 문제점을 제시하고 있는 선행연구로는 최병일(2002)의 “댄스스포츠의 채점방식에 관한 비교연구”가 있다. 이 연구에서는 기존의 합산식(Total system) 채점방식의 문제점을 지적하고, 미국 등 선진국에서 적용하고 있는 스케이팅시스템(Skating system) 채점방식의 도입을 주장하고 있다. 스케이팅시스템 방식이란 각각의 심사위원이 부과한 순위의 과반수만을 적용하여 누적하는 방식으로써 특정 심사위원의 편파관정으로 인한 불이익을 줄이기 위한 장치이다. 이 방법은 프로야구 MVP 선정을 위한 기자단 투표에서도 사용되는 방식으로 무작위 인기투표와는 달리 전문 평가자집단 내에서 순위를 정하는 방법이라는 점에서 우리나라 국가연구개발사업 연구과제 선정방법에 시사하는 바가 있다.

### III. 연구과제의 선정과 동료평가

#### 1. 연구과제의 선정

연구과제 평가시스템에서 가장 중요한 과제는 전문성, 공정성, 효율성의 목표를 동시에 달성해야 한다는 것이다(Kostoff, 1997, NRC, 1997, Chubin & Hackett, 1990). 이러한 3가지 목표가 제대로 충족될 때 최적의 연구과제가 선정될 수 있다<sup>4)</sup>. 그런데 문제는 3가지 목표를 동시에 달성하기가 쉽지 않을 뿐만 아니라 목표들이 서로 상충될 수도 있다는 점이다(Foltz, 2000, Rennie, 1993). 가령 지나치게 공정성을 강조하다보면 전문성과 효율성이 떨어질 수도 있고, 반대로 지나치게 효율성만 강조하게 되면 전문성과 공정성이 떨어질 수 있다. 연구과제 평가시스템의 이슈는 어떻게 하면 전체적으로 이러한 목표들의 달성도를 동시에 높일 수 있는가에 초점이 맞추어져야 한다(U.S.DOE, 1998).

전문성의 문제는 연구계획서를 심사하는 심사위원들이 연구계획서의 내용을 평가할 만한 학문적 전문성을 갖추었는가의 문제이다(Chubin, 1994, NRC, 1996). 한국연구재단과 NSF에서는 연구과제 선정의 타당성을 높이기 위해 기본적으로 동료평가제도(peer review)를 활용하고 있다. 동료평가제도의 목적은 연구과제 선택 과정에서 여과과정을 만들자는 것이다(Royal Society, 1995, Chubin & Hackett, 1990).

평가의 공정성은 연구과제를 평가할 때 학문적 우수성 외의 다른 변수에 의하여 평가결과가 영향을 받지 않도록 하는 것이다. 연구지원 현장에서 연구사업을 수행하면서 부닥치는 문제는 심사의 공정성을 확보하기가 쉽지 않다는 점이다(Bozeman, 1993, Cicchetti, 1991). 대부분의 경우, 연구과제에 선정된 지원자들은 불평이 없지만 탈락된 지원자들은 공정성에 대해서 의문을 제기한다. 학자들 세계는 여러 가지 인연으로 얽혀있어서 익명성을 보장하기 힘든 상황이 발생한다(Chubin & Hackett 1990). 가령 특정 학교를 중심으로 동문들간에 봐주기식의 평가를 한다든가 사전에 심사대상자의 인적사항을 알고 비합리적인 기준으로 평가를 하는 등의 상황들이 발생할 여지가 많다(조만형, 홍형득, 김갑수, 2004). 서울소재 6개 대학, 지역소재 6개 대학 총 12개 대학을 선정한 후, 국가연구개발사업 수행경험이 있는 교수를 대상으로 전자우편을 통하여 설문조사한 결과에 의하면 심사결과에 대한 공정성은 대부분 보통에서 만족을 느꼈지만 불만족하다는 응답도 20%정도를 차지하였다(윤덕균, 2006). 공정성은 심사자들의 편견에서 비롯되는 경우가 많고 편견은 불공정한 평가결과를 초래하고, 결과적으로 부

4) 선행 연구자들이 제시한 3가지 목표에서 전문성과 공정성은 평가결과의 신뢰성을 지지하는 두 축이며, 연구과제 선정과정에서는 두 가지를 동시에 확보하기에는 어느 정도 한계가 있는 것으로 간주되고 있다.

적절한 연구과제가 최종적으로 선정되게 한다(조만형, 홍형득, 김갑수, 2004).

사회과학의 방법론에서 신뢰도(reliability)는 관찰자간 신뢰도(interobserver reliability), 측정도구의 신뢰도(the reliability of the instrument), 일반화 신뢰도(generalizability)로 구분될 수 있다(Mitchell, 1979, Conrad & Maul, 1981). 이 중에서 연구과제 선정에 관련되는 신뢰도는 관찰자간 신뢰도에 속하는 평가자간 신뢰도(interrater reliability) 및 평가자간 동의도(interrater agreement)이다. 평가자간 신뢰도는 평가자들이 평가한 값의 절대치와는 상관 없이, 평가치들의 상대적인 관계의 균형적인 정도, 평가치들간의 분산의 일관성에 대한 비율, 평가치들간의 상관관계의 정도, 평가대상에 대한 순서의 유사성을 말한다(Kozlowski & Hatrup, 1992, Tinsley & Weiss, 1975). 평가자간 동의도는 평가자들이 평가대상에 대하여 절대치가 같은 평가를 하는 정도를 나타내며, 평가대상에 대하여 점수로 평가하는 경우 평가자간 동의는 평가자가 서로 똑같은 수치를 부여하는 것을 의미한다(Bartko, 1976, Tinsley & Weiss, 1975).

이와 같이 평가자간 신뢰도 및 동의도 중에서 평가대상의 순위나 평가자간 판단의 일관성에 관심이 있을 경우에는 평가자간 신뢰도를, 평가대상의 절대적인 점수 또는 평가자간 판단의 기준치에 초점이 있을 경우에는 평가자간 동의도가 중요하다. 따라서 과제평가에서는 평가자간 동의도 보다는 평가자간 신뢰도가 더 중요하다(차종석, 김영배, 1994).

## 2. 동료평가(peer review)

과학기술 해당분야의 전문가를 활용하는 동료평가는 과학기술의 질적인 측면을 평가하는 가장 전통적이고 중요한 평가기법으로 인식되어 왔다. 이러한 동료평가는 이른바 "good science"에 대한 선정·지원을 결정하는데 있어 보편적으로 많은 강점을 갖고 있지만 반면에 역기능적인 단점도 있다. 대부분의 이러한 바람직하지 못한 역기능은 평가에 참여하는 이해관계자들의 지적활동에 따른 심리·인간행동학적 측면에서 나타난다. 동료평가는 인간의 지적활동(human intellectual activity)에 중심을 두고 있기 때문에 이러한 인간의 심리적인 측면에서 이해의 충돌로부터 발생하는 다양한 주관적인 편파성이 평가에 부정적인 영향을 미치는 것이며 따라서 동료평가는 이러한 인간의 주관적 판단이 주변의 여건에 의해 영향을 받아 보편타당한 객관성과 공정성을 일정 부분 상실하고 있다는 점에서 크게 비판을 받고 있다(Bozeman, 1993).

평가에 있어 동료평가방법은 대개의 경우 일부 동료평가자의 결정에 의존적이다. 따라서 이들 전문 평가자를 어떻게 선정하느냐가 평가방법의 신뢰도를 가름하는 결정적인 요인이 된다

(Luukkonen-Gronow, 1987). 평가자의 선정은 상당히 어려운 과정이며 평가자 패널 구성에는 분야에 대한 전문성이외에도 소속, 지역 및 인력의 분배 등의 요소들이 함께 고려된다(Chubin, 1994). 따라서 동료평가의 신뢰성을 개선하기 위한 제안들은 대부분 평가자 선정을 어떻게 할 것인가에 달려 있다고 결론을 내리고 있다. 구체적으로 평가자의 수를 확대하는 방안을 고려해 볼 수 있으나, 단순히 평가자 규모를 확대하는 것은 평가의견 불일치(discrepancy) 정도의 증폭과 평가 업무와 비용이 증가하는 등 여러 가지 단점이 있어 신중을 기해야 하며, 따라서 평가자 구성을 양적인 측면이 아닌 질적으로 향상시키기 위해 평가자의 구성에 있어 내부 및 외부평가자(대학이외의 정부출연연구소 및 산업계 전문인력을 포함하는) 뿐만 아니라 경우에 따라 외국의 전문가도 포함시키는 등 다양하게 할 것과 평가자 그룹의 집단이기주의를 회피하기 위하여 평가자를 선정함에 있어 연구과제 신청자의 의견도 고려하여 확정하는 방안이 제시되고 있다(Wessely & Wood, 1999).

#### IV. 국가연구개발사업 연구과제 선정방식 및 문제점

##### 1. 연구과제 선정방식

현재 한국연구재단 등에서 지원하는 대부분의 사업은 발표평가를 통해 지원과제를 선정하고 있으나 사업의 특성에 따라 서면평가와 발표평가를 병행하는 경우도 있다.<sup>5)</sup> 서면평가를 하거나 발표평가를 하거나 모두 평가자의 전공, 연구경력, 관련분야 논문실적, 지역, 성별, 연령, 피 평가자와의 이해관계 등 일정기준에 의거 선정한 평가자들이 평가를 하고, 평가한 점수를 합산하여 순위를 정한다. 이후에 교육과학기술부에 설치되어 있는 사업별 추진위원회의 검토를 거쳐 예산의 범위 안에서 우선순위에 따라 지원과제를 선정한다. 이러한 일련의 연구과제 선정과정에서 본 연구의 주제는 평가위원 선정방법과 평가점수 산정방식이다.

한국연구재단 지원사업의 평가위원 선정방법은 원천기술개발사업(구, 특정연구개발사업)과 기초연구사업등 사업에 따라 조금 차이가 있다. 기초연구사업은 피 평가자와 평가자간의 상관도를 산출하여 일정점수 이상인 평가자만 제외하는 방식인 반면, 원천기술개발사업은 상관점수와 관계없이 피 평가자와 이해관계로 보여지는 경우는 무조건 배제하는 방식을 적용하고 있

5) 중견연구사업(구 국가지정연구실사업)의 경우 1차 서면평가를 통해 일부 과제를 선별하고, 선별된 과제를 대상으로 2차 발표평가를 실시하여 연구과제를 선정하고 있으며, 예산규모에 비해 지원과제가 많은 경우에도 1차 서면평가를 통해 발표평가 대상과제를 선별하는 경우도 있다.



다. 그러나 대부분의 사업이 위와 같이 이해관계자를 사전에 배제하였음에도 불구하고 혹시 발생할 수도 있는 특정 평가자의 편견을 제거하기 위해서 평가점수 산정시 최고점수와 최저점수를 제외한 평균점수로 순위를 정하고 있다<sup>6)</sup>. 기초과학연구사업의 상관도 분석기준은 <표 1>과 같다.

<표 1> 기초연구사업의 상관도 분석 기준

구 분	기초연구사업의 상관도 분석 기준
피 평가자와 평가자 후보간의 관계에 따른 반영요소	· 사제지간의 경우(0.4점) · 동일한 지도교수 아래서 수학한 경우(0.3점) · 동일기관(대학)에서 박사학위를 취득한 경우(0.2점)·석사 학위를 취득한 경우(0.14점) · 학사 학위를 취득한 경우(0.06점)

\* 피 평가자와 평가자 후보간의 관련도가 일정 수준을 초과하는 경우 평가자 후보에서 제외함을 원칙으로 하되 불가피한 경우에는 예외로 함.

## 2. 현행방식의 문제점

연구과제 평가결과는 신뢰성이 가장 중요하므로 신뢰성을 확보하기 위해서 평가과정에 전문가들이 참여하여야 하고, 평가과정이 공정하여야 한다. 그러나 연구과제 평가를 동료평가에 의존하고 있는 현실에서는 종종 평가결과의 신뢰성에 대한 의문이 제기되고 있고, 이를 보완하기 위해 많은 노력들을 하고 있다. 한국연구재단에서도 연구과제 평가과정에서 혹시 발생할 수도 있는 정보비대칭의 문제<sup>7)</sup>를 최소화하기 위해 평가자 선정시 상피제도<sup>8)</sup>를 적용하고 평가점수 산정시 최고/최저점 제외방식을 채택하고 있다. 하지만 어떤 제도라도 완벽할 수는 없듯이 현행제도의 문제점은 전문성 손실과 평가정보 손실이라는 큰 비용을 지불함에도 불구하고 정보비대칭문제가 여전히 존재할 수 있다는 것이다.

6) 한국연구재단의 대부분사업(기초연구개발사업, 원천기술개발사업, 원자력연구개발사업 등)은 평가위원회를 7인 내외로 구성하고 최고/최저점수를 제외한 평균으로 순위를 정하고 있으며, 산업기술평가원의 산업기술개발사업도 동일하다. 그러나 정보통신연구진흥원과 농림기술관리센터에서 지원하는 일부 사업들은 전체평균으로 순위를 정하는 사업도 있다.

7) 현대 정보경제학(The Economics of Information)의 핵심적 토대를 마련한 정보비대칭(Information Asymmetry)이론은, 2001년도 노벨경제학상 수상자인 미국 캘리포니아대학교 Akerlof 교수에 의해 정립된 이론이다. 이 이론은 모든 경제주체가 동등한 수준의 정보를 지닌 상태에서 움직인다고 본 전통적 경제이론의 입장이 실제의 경제현실에는 들어맞지 않는다는 점으로부터 출발한다. 예를 들면 중고차시장에서 중고차를 사려는 사람에 비해 팔려는 사람이 차의 결함에 대해 훨씬 더 잘 알고 있다는 것이다.

8) 평가의 공정성 증대를 위해서 동일기관에 근무하는 경우, 동일기관에서 학위를 취득한 경우 등 피평가자와의 상관도를 계산하여 일정수준을 초과하는 경우 평가자 후보에서 제외시키는 제도를 말한다.

## 1) 상피제도로 인한 전문성의 손실

한국연구재단에서 지원하는 사업은 대부분 평가의 공정성 확보를 위해 상피제도를 적용하고 있고 이 과정에서 평가의 공정성을 지나치게 중시한 나머지 전문성 확보가 어려워지는 한계가 발생한다. 예를 들면 피 평가자와 평가자가 사제지간이거나 동일한 지도교수 아래서 수학한 경우 외에도 같은 대학교에 근무 한다거나, 같은 대학을 나왔다는 이유로 평가자 풀에서 제외되는 것이다. 심지어는 연령대가 달라 일면식도 없지만 출신 고등학교가 같다는 이유로 평가위원에서 배제시키는 사업도 있다. 이와 같이 공정성을 지나치게 중시함에 따라 그 분야 최고 전문가로 알려진 사람이 제외되거나, 피 평가과제 세부전문분야와 다소 상이한 평가자를 평가에 참여시키는 경우도 발생 할 수 있는 것이다. 대표적인 사업으로 국가지정연구실사업(National Research Laboratory Program)을 들 수 있다. 이 사업의 경우 엄격한 상피제도를 적용하다보니 평가자 선정에 많은 어려움이 발생하고, 이로 인해 연구관리의 효율성이 떨어지는 사례가 발생할 수 있다. 이 사업의 초기에는 특정과제와 상관도가 높은 경우 평가자로 위원회에 참여는 하되 상관도가 높은 과제만 평가하지 않으면 되도록 하였다. 그러나 경쟁이 치열해지고 평가의 공정성에 대한 관심이 높아짐에 따라 평가그룹내의 한 과제만이라도 상관도가 높은 사람은 원천적으로 평가자풀에서 제외시킴으로써 평가자 섭외에 더욱 어려움이 발생하였고, 해당 기술분야 최고 전문가들이 평가자풀에서 제외되는 경우가 발생하게 되었다. 이 주장은 비록 상관도가 높은 과제를 직접 평가하지는 않지만 평가에 참여하고 있는 다른 평가자들에게 영향을 줄 수 있다는 것이다. 이는 전문성보다는 공정성을 조금 더 확보하겠다는 의도이지만, 평가자풀이 작은 기술분야의 경우 가끔은 그 분야 최고 전문가(과제 제안자)의 과제를 상대적으로 비 전문가(평가자)가 평가했다는 비난의 대상이 되기도 하고, 전문성을 양보한 만큼 공정성이 더 확보되었다는 보장, 즉 정보비대칭의 문제가 해소되었다고 보기도 어려울 것이다. 결국 상피제도(이해관계자 사전배제방식)는 같은 교회를 다니거나, 같은 아파트에 거주하거나, 또는 같은 동호회 활동을 하고 있는 관계이거나 하는 등의 보이지 않지만 매우 우호적일 수 있는 인간관계까지 제거할 수는 없는 것이다. 이와 같은 이유로 국가연구개발사업의 수행에 있어서 정보비대칭의 문제는 한국연구재단과 평가자간에 발생하는 정보비대칭이 훨씬 더 중요하게 다루어야하는 주제일지도 모른다<sup>9)</sup>.

## 2) 최고/최저점수 제외로 인한 평가정보의 손실

### (1) 정규분포의 가정

국가연구개발사업 연구과제 평가과정에서 관례적으로 사용하고 있는 최고/최저점수 제외방

9) 정보비대칭에 대한 기존연구들은 대부분 연구비 지원부처인 정부와 연구비 수혜자인 연구책임자간의 정보 불균형으로 인한 도덕적 해이(moral hazard)를 다루고 있다.

식은 양 극단치를 제외하겠다는 것이다. 이는 자료가 너무 많아서가 아니라 특이치를 제외하고 정상치 자료만을 활용하겠다는 것으로 배경에는 평가점수가 정규분포를 이룰 것이라는 가정을 하고 있는 것이다. 그러나 현실적인 평가점수는 한쪽으로 치우치는 경우가 많이 있다. 그리고 또 정규분포가정에 우리가 따르는 것은 일반적으로 정규분포를 가정하려면 샘플크기가 20개는 넘어야 하지만 국가연구개발사업 평가과정의 샘플크기(과제별 평가점수)는 겨우 7개 내외로 매우 작다는 것이다. 결론적으로 정규분포를 이루는 자료에서 최고값과 최저값이 극단치 또는 특이치가 되지만, 가령 어떤 평가위원회에서 평가점수가 92, 94, 89, 91, 65점으로 구성되어 있다면, 65점(최저점)은 특이치가 분명하지만 94점(최고점)을 특이치로 보고 제외한다는 것은 무리가 있다. 반대로 78, 80, 79, 81, 98점으로 구성되어 있다면 이 경우도 마찬가지로 98점(최고점)은 특이치가 분명하지만 78점을 특이치로 보고 제외한다는 것은 무리가 있다. 또 하나의 경우로 86, 87, 84, 86, 85점으로 구성된 평가위원회가 있다면 단지 최고점, 최저점이라는 이유로 87점과 84점을 제외시켜야 하는가에 고민이 생기는 것이다. 왜냐하면 이 점수들 중에서 극단치로 불만한 점수는 없기 때문이다.

## (2) 평가정보의 손실

한국연구재단의 경우 대개 한 평가위원회별로 평가자를 7인 내외로 구성하는 것을 원칙으로 하고, 평가 대상과제 수에 따라 조정하고 있다. 여기에서 최고점과 최저점을 제외하면 전체 평가정보의 71.4%만 사용하고 있는 것이다. 이는 교육과학기술부에서 개선안으로 검토했던, 26이상을 극단치로 제외하는 경우(95.45%) 보다도 평가정보의 손실이 크다. 참고로 한국연구재단의 인문사회과학분야에서 주로 사용하고 있는 표준점수제 하에서도 표준점수 1.3점 이상을 극단치로 간주 할 경우 80.6%의 자료를 활용하는 것이므로 약 20%의 평가정보의 손실이 발생하고 있다. 물론 남은 자료의 정당성 확보를 위해서 어느 정도 자료의 손실은 감수해야 하지만 국가연구개발사업을 평가할 정도의 전문성을 갖추었거나 지명도가 있는 평가자가 많은 시간을 투자하여 생성한 정보라면 가능한 어떤 형태로든 사용하는 것이 비용효과적일 것이다. 결론적으로 최고/최저점수 제외 방식은 우리가 언제부터 사용했으며, 왜 사용하는지에 대한 큰 고민 없이 관례적으로 사용하고 있었던 것이다.

## V. 사례분석 및 대안의 제시

### 1. 사례분석

본 사례는 2005년도 NRL사업의 한 평가위원회 평가결과를 연구의 목적에 맞게 조금 각색

한 자료이다. 자료에 대해 간단히 설명하면, 5개 과제를 9명의 평가자가 평가하였으며, 평가자A와 평가자G는 상피제도에 저촉되어 과제1번을 제외한 나머지 과제만 평가한 경우로 흔히 있을 수 있는 평가결과 사례이다.

〈표 2〉 사례 : 과제별 평가점수

평가자	과제1	과제2	과제3	과제4	과제5
A	-	90	86	75	81
B	95	85	72	77	90
C	79	81	90	74	80
D	100	90	75	80	75
E	98	90	81	95	86
F	87	80	79	82	-
G	-	90	68	77	86
H	92	86	85	76	77
I	93	90	62	78	74
총 점	644	782	698	714	649
표준편차	7.1	4.1	9.1	6.4	5.8
단순평균	92.0	86.9	77.6	79.3	81.1
현행평균	93.0	87.3	78.0	77.9	80.8

주 : 점수란에 공란인 것은 평가자와 해당과제 제안자 간에 상관도가 높아서 평가하지 않은 것임.

이 사례에서 눈여겨 볼 평가점수는 과제3번과 과제4번이다. 총점으로는 과제4번이 과제3번에 비해 무려 16점이 높지만, 최고/최저점수를 제외하고 평균하는 방식의 현행방식으로 평균을 구하면 77.9점으로 과제3번에 비해 0.1점이 낮게 나온다. 만약 5개 과제 중에서 20%인 1과제가 탈락되는 단계평가로 가정한다면 0.1점으로 인해 과제4번은 탈락되는 것이다. 0.1점이라는 미미한 차이도 중요하지만 표준편차도 과제3번에 비해 적다는 점 때문에 더욱 아쉬움이 있다. 이 사례에서 시사하는 바는 최고/최저점수 제외로 인해 과제의 선정과 탈락에 큰 영향이 있을 수 있으므로 극단치(특이치)를 어떻게 볼 것이냐가 매우 중요하다는 것이다. 본 연구에서는 위의 사례를 대상으로 다양한 편차를 구해보고 이들 편차가 의미하는 바를 분석하고 궁극적으로는 현실적인 대안을 도출하고자 한다.

### 1) 평가자간의 편차

〈표 3〉은 평가자간의 평가점수 편차이다. 좀더 정확하게는, 각 과제별로 평가자들이 부여

한 점수에서 자기 점수를 제외한 나머지 평가자들의 평가점수를 평균한 값과의 차이를 나타낸 것이다, 즉 동일한 과제에 다른 평가자들에 비해 얼마나 높은 점수를 부여하였는지 또는 낮은 점수를 부여하였는지를 보여주는 자료이다. 평가자들의 가치부여기준(채점 성향)의 차이를 나타내는 숫자인 동시에 특정과제에 대한 영향력 정도를 나타낸다. 평가자 C는 대체로 다른 평가자에 비해 점수를 낮게 채점하는 경향(부호가 대체로 음수)이 있으며, 평가자 E는 대체로 다른 평가자들에 비해 점수를 높게 채점하는 경향(부호가 전부 양수)이 있음을 볼 수 있다. 또한 평가자C는 과제1번에 대하여 다른 평가자들에 비해 15점 이상 낮게 채점하였으며, 반대로 과제3번에 대하여는 다른 평가자들에 비해 14점을 높게 채점한 것으로 나타났다. 마찬가지로 평가자E는 과제4번에 대하여 약 18점을 높게, 평가자I는 과제3번에 대하여 약 18점을 낮게 채점한 것으로 나타나고 있다. 여기에서 우리는 평가자C, E, I가 다른 평가자들과 판이하게 다른 점수를 부여한 이유에 대하여 정성적으로 기술한 평가의견서를 확인할 필요성이 있음을 느끼게 된다.

〈표 3〉 평가자간의 편차

평가자	과제1	과제2	과제3	과제4	과제5
A	-	3.5	9.5	-4.9	-0.1
B	3.5	-2.1	-6.3	-2.6	10.1
C	-15.2	-6.6	14.0	-6.0	-1.3
D	9.3	3.5	-2.9	0.8	-7.0
E	7.0	3.5	3.9	17.6	5.6
F	-5.8	-7.8	1.6	3.0	-
G	-	3.5	-10.8	-2.6	5.6
H	0.0	-1.0	8.4	-3.8	-4.7
I	1.2	3.5	-17.5	-1.5	-8.1

## 2) 과제간의 편차

〈표 4〉는 과제들 간의 편차를 나타내는 자료이다. 한 평가자가 특정과제에 부여한 점수와 이 과제를 제외한 다른 과제에 부여한 점수들의 평균값과의 차이를 나타낸 자료이다. 즉 동일한 평가자가 특정과제에 다른 과제에 비해 얼마나 점수를 높게 부여하였는지 또는 얼마나 낮게 부여하였는지 보여주는 자료이다. 특정과제에 대한 평가자간의 상대적 가치인식의 차이인 동시에 특정과제에 대한 평가자의 영향력을 나타낸다. 평가자C는 다른 과제에 비해 과제3번에 대하여 매우 후한 점수를 부여한 반면 과제4번에 대하여 상대적으로 낮은 점수를 부여하고

있다. 반대로 평가자E는 과제4번에 대하여 다른 과제에 비해 후한 반면 과제3번에 대하여는 매우 낮게 평가하고 있다. 여기에서는 절대값의 크기보다는 부호가 주는 의미가 크다. 예를 들어 다른 평가자들은 나머지 4개 과제에 비해 과제1번을 높게 평가한 반면 평가자C는 낮게 평가하고 있으며, 평가자E는 나머지 4개 과제에 비해 과제4번을 높게 평가한 것으로 나타나고 있다. 특히 평가자C와 평가자E는 과제3번과 과제4번에 대해 판이하게 다른 방향으로 점수를 부여하고 있음을 볼 수 있다. 이것이 의미하는 바는 과제별 순위 편에서 고찰하겠지만 과제3번과 과제4번에 대한 평가자C와 평가자E의 점수는 분명히 특이치로 보여 진다는 것이다.

〈표 4〉 과제간의 편차

평가자	과제1	과제2	과제3	과제4	과제5
A	-	9.3	4.0	-10.7	-2.7
B	14.0	1.5	-14.8	-8.5	7.8
C	-2.3	0.3	11.5	-8.5	-1.0
D	20.0	7.5	-11.3	-5.0	-11.3
E	10.0	0.0	-11.3	6.3	-5.0
F	6.7	-2.7	-4.0	0.0	-
G	-	13.0	-16.3	-4.3	7.7
H	11.0	3.5	2.3	-9.0	-7.8
I	17.0	13.3	-21.8	-1.8	-6.8

### 3) 과제별 순위

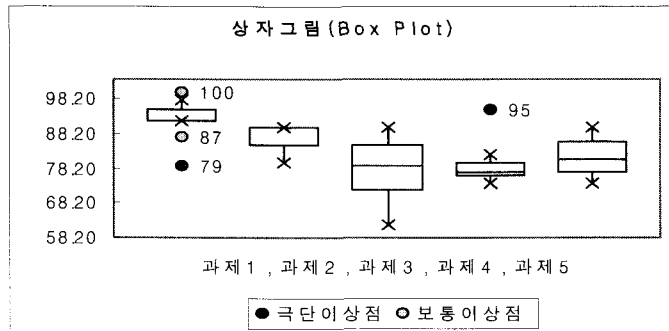
〈표 5〉는 위의 사례를 점수가 아닌 순위로 표시한 것이다. 평가자들이 우수하다고 인식하고 있는 과제별 순위이다. 즉 평가자별로 자기가 생각하고 있는 1등인 과제와 2등인 과제 등 과제간의 상대비교를 가능케 하는 자료이다. 예를 들어 평가자B는 자기가 평가한 과제 중에서 과제1번이 가장 우수하고 과제3번이 가장 부족한 것으로 평가하고 있으며, 평가자H는 자기가 평가한 과제 중에서 과제1번이 가장 우수하고 과제4번이 가장 부족한 것으로 평가하고 있다. 그러나 여기에서도 특이한 것은 평가자C이다. 대부분의 평가자들이 평가대상 과제 중에서 과제1번이 가장 우수하다고 평가한 반면 혼자만이 최하위에 가까운 4등을 부여하고 있고, 대부분의 평가자들이 과제3번에 대하여 최하위 등위를 부여한 반면 평가자C만이 1등을 부여하고 있다. 특히 경쟁과제인 과제4번을 최하위인 5등으로 평가하고 있다는 점이 특이하다. 이는 평가자C만이 다른 평가자들이 모르는 과제3번에 대한 긍정적인 정보가 있거나, 이해관계자로 오해할 수 있는 소지가 있는 것이다.

〈표 5〉 과제별 순위

평가자	과제1	과제2	과제3	과제4	과제5
A	-	1	2	4	3
B	1	3	5	4	2
C	4	2	1	5	3
D	1	2	4	3	4
E	1	3	5	2	4
F	1	3	4	2	-
G	-	1	4	3	2
H	1	2	3	5	4
I	1	2	5	3	4

#### 4) 극단치의 해석

특이치를 판단하는 방법중에 Tukey(1977)의 상자-수염그림(Box-whisker plot)이 있다. 그림 1은 위의 사례를 상자-수염그림(Box-whisker plot)으로 나타낸 것이다. 상자-수염그림은 두 개 이상의 집단분포를 비교하기 위하여 사용되는 그래프로, 최소값, 최대값, 중앙값, 제1사분위수, 제3사분위수를 상자(box)와 수염(whisker)형태로 표시하여 그리는 그래프이다. 이 그래프는 수치자료의 요약을 특이값과 함께 그래프로 표현한 것으로서 전체자료의 분포상태, 특이값의 유무 등 여러 집단의 수치자료를 비교 하고자 할 때 유용하게 이용된다. 상자-수염그림에서 보통이상치는 몸통(전체자료의 50%) 길이의 1.5배, 극단이상치는 몸통길이의 3배 이상으로 표현되고, 통계학자들에 의하면  $2\sigma$  와 Tukey의 이상치가 거의 일치한다는 것이 증명되었다. 결국 Tukey의 상자-수염그림에서는 전체자료의 95% 정도를 정상치로 포함하고 5%를 이상치로 간주 한 것이다. 그림 1에서는 과제1번의 최하점수(79점)와 과제4번의 최고점수(95점)가 극단값에 해당하고, 과제1번의 두 점수(100점, 87점)가 보통이상치로 나타났다. 다시 말하면 상피제도를 적용하여 사전에 이해관계자를 제외시키고, 또한 평가자 교육을 충분히 실시하였음에도 불구하고 엄연히 평가현장에서는 극단치가 존재할 수 있음을 보여주고 있다. 그리고 정보의 손실부분을 살펴보면, 현행방식의 경우 전체 자료중 10개의 자료가 제외된다. 그러나 그림 1에서 보여지는 것 같이 보통이상치와 극단이상치를 모두 제외한다고 해도 기존방식의 40%수준으로 줄일 수 있으며, 경우에 따라 극단이상치만 특이치로 해석할 경우 정보의 손실을 20% 수준으로 줄일 수도 있다.



(그림 1) 상자그림

## 5) 사례분석 결과

이상과 같은 간단한 사례 분석을 통해 다음과 같은 결과가 도출되었다. 첫째, 예상치 보다 편차가 크게 발생하고 있다는 것이다. 둘째, 평가결과가 다른 평가자들과 판이하게 다른 평가자가 있다는 것이다. 셋째, 최대/최소점수가 반드시 특이치가 아닐 수 있다는 것이다. 이와 같은 결과를 바탕으로 개선안을 도출한다면, 평가자간 편차가 10이상인 경우 제외, 과제간 편차에서 부호가 다른 경우 제외, 상자-수염그림에서 보통이상치와 극단이상치 제외 등 다양한 방법이 가능할 것으로 생각된다. 그리고 사례하나로 일반화하기에는 무리가 있으므로 지속적인 관찰이 필요하겠지만, 현행 방식으로는 특정평가자의 숨은 의도를 제거할 수 없기 때문에 새로운 방식으로 개선할 필요성이 절실하고, 또한 평가자별 평가이력 관리도 철저히 할 필요성이 있는 것으로 보여 진다.

## 2. 대안의 제시

현행 방식은 최고/최저점수를 극단치로 간주해서 제외하고 평균을 구하는 방식이다. 그러나 극단치가 아닐지라도 최고점수, 최저점수일 경우 무조건 제외되는 한계가 있었다. 그래서 최고와 최저에 관계없이 평균에서 가장 먼 값 또는 평균에서 일정값( $2\sigma$ ) 이상 먼 값만 제외하는 방법과 점수가 아닌 순위정보를 이용하는 방법, 점수와 순위를 동시에 고려하는 방법 등 다양한 대안이 있을 수 있다. 이러한 각각의 방법을 간략히 설명하면 다음과 같다.

### 1) 상자-수염 방식

상자-수염방식은 기존의 방식과 같이 점수를 기준으로 한다. 그러나 기존 방식과는 달리 무



조건 최고/최저점수를 제외시키는 것이 아니라 일정범위를 벗어나는 값만 제외시키는 것이다. 이 방식은 비록 최고점수 일지라도 대부분의 평가자가 높은 점수를 부여하였다면, 이 최고점수를 부여한 평가자도 정상적으로 평가를 수행했다고 보는 것이다. 또 경우에 따라서는 최고점수가 아닐지라도 일정범위를 벗어났다면 극단치로 간주하여 제외시키는 것이다. 그림 1에서 살펴보면 과제1번의 경우 이상치 두개와 극단이상치 1개, 과제4번의 경우 극단이상치 1개로 나타나고 있으므로 4개의 점수만 제외하지는 것이다. 즉 나머지 과제에서는 비록 최고/최저점수일지라도 다른 평가자들과 일정범위 내에 있는 경우 정상치로 간주하지는 것이다. 이 방식의 장점은 기존방식보다 정보의 손실을 획기적으로 줄일 수 있으며, 어느 정도는 평가그룹내의 평가자간에 동의하는 정도를 반영할 수 있다는 것이다. 다시 말해서 특정 평가자의 편견이나 의도된 행동에 의해 좌우되기 쉬운 현행제도의 단점을 보완하여, 대부분의 평가자들이 우수하다고 생각하는 과제를 선정할 수 있고, 대부분의 평가자가 부족하다고 생각하는 과제를 제외시킴으로써 연구과제 선별력이 좋아질 수 있다는 것이다. 그러나 무엇보다도 큰 장점은, Tukey(1977)의 상자-수염그림 방식은 통계학분야에서 수십년간 검증된 방법론이기 때문에 실무에서 바로 적용하기에 부담이 없다는 것이다.

## 2) 스케이팅 방식

스케이팅 방식이란 스포츠 분야에서 폭넓게 사용되고 있는 방식으로 각각의 심사위원이 부과한 순위의 과반수이상 만을 적용하여 누적하는 방식이며, 특정 심사위원의 편파관정으로 인한 불이익을 줄이기 위한 장치이다(최병일, 2002). 아래 <표 6>을 예로 쉽게 설명하면, 과제1번의 경우 이 과제가 1등이라고 인정한 평가자는 6명으로 이미 평가자 9명중 과반수이상인 과제1번이 1등이라는 점에 동의한 것이다. 과제2번의 경우 이 과제가 1등이라고 인정한 평가자는 2명으로 아직 평가위원 과반수를 획득하지 못했기 때문에 다음 등위인 2등이라고 인정한 평가자 수를 더해가는 것이다. 그래서 적어도 과제2번이 1등이거나 2등이라고 인정하는 평가자가 6명이 되었고, 과반수를 넘었기 때문에 이 과제는 대부분의 평가자들이 생각할 때 적어도 2등은 된다고 인정한 것이다. 그러나 과제4번의 경우는 과제2번의 경우와 조금 다르다. 과제2번과 같이 적어도 이 과제가 2등은 된다고 인정한 평가자가 과반수는 넘었지만 총 득표에서 과제2번 보다 적기 때문에 3등이 되는 것이다. 이와 같은 방법으로 순위를 정하면 과제3번이 가장 낮은 등위를 받게 되는 것이다. 부연 설명하면 과제3번의 경우 대부분의 평가자들이 4등이나 5등으로 인정하고 있다는 것이다. 이 방식은 점수가 아닌 순위를 중시한다는 점에서 국가연구개발사업 연구과제 선정에 적용이 가능하며, 무엇보다도 평가그룹내의 평가자간에 인식하는 수준의 정도를 반영할 수 있다는 점이 장점이다. 다시 말해서 특정 평가자의 편

견이나 의도된 행동에 의해 좌우되기 쉬운 점수방식의 단점을 보완하여, 대부분의 평가자들이 우수하다고 생각하는 과제를 선정할 수 있고, 대부분의 평가자가 부정적으로 생각하는 과제를 제외시킴으로써 연구과제 선별력이 좋아질 수 있다는 것이다.

〈표 6〉 과제별 득점 순위

순 위	과제1	과제2	과제3	과제4	과제5
1	6	2	1	2	2
2	-	4	1	3	2
3	-	3	1	2	4
4	1	-	3	2	-
5	-	-	3	-	-
결 과	1	2	5	3	4

### 3) 표준화점수 방식

표준화점수 방식은 평가자들이 부여한 점수와 평가자들이 인식하고 있는 순위를 동시에 고려하여 기존의 점수방식에서 야기되는 정보비대칭의 문제를 해소할 수 있는 방식이다. 상세히 설명하면 먼저 각 평가자가 각 등위에 부여한 점수의 평균을 구하여 각 등위별 표준화점수를 먼저 산정한 다음, 해당 과제별로 획득한 등위의 표준화평균<sup>10)</sup>을 구하는 것이다. 위의 사례를 예로 계산하면 1등에 대한 표준화점수는  $92.78((90+95+90+100+98+87+90+92+93)/9)$ 이 된다. 이 값은 각 평가자들이 1등이라고 생각한 과제에 부여한 점수의 평균이다. 동일한 방법으로 계산하면 표준화점수는 〈표 7〉과 같이 산출된다. 다음 단계로는 이렇게 산출된 각 등위별 표준화점수를 각 과제가 획득한 등위 득수를 곱하여 평균을 구하는 것이다. 과제1번의 경우를 예로 들면 평가자 6명이 이 과제를 1등으로 인정했고, 나머지 한명이 4등으로 인정했기 때문에  $90.5((92.78 \times 6 + 76.67 \times 1)/7)$ 가 되는 것이다. 동일한 방법으로 점수를 산출하면 〈표 7〉에서 보는바와 같이 과제2번은 86.7, 과제3번은 79.1, 과제4번은 80.0, 과제5번은 80.6이 된다.

〈표 7〉 등위별 표준화점수

순위	1	2	3	4	5
표준화점수	92.78	87.33	81.78	76.67	73.33

10) 표준화평균 =  $(\sum(\text{표준화점수} \times \text{등위득수}) / \text{총득수})$

이 방식은 평가점수의 절대치를 그대로 활용함으로써 과제별로 섬세한 차이를 인정할 수 있고 동시에 평가자별로 인정하고 있는 순위정보도 활용할 수 있기 때문에, 단 하나의 정보도 버리지 않고 활용하면서 정보비대칭을 최소화 할 수 있다는 장점이 있다. 다시 말해서 특정 평가자의 편견이나 의도된 행동에 의해 좌우될 수 있는 여지를 원천봉쇄 함으로써 정말로 우수한 과제가 선정될 수 있다는 것이다. 또한 평가그룹이 달라도 순위를 정하는데 어려움이 없으며, 평가자의 평가성향(평가 점수를 후하게 주거나 박하게 주는)을 그대로 반영할 수 있다는 장점이 있다. 무엇보다도 이 방식이 갖고 있는 큰 장점은 특정 평가자가 특정과제를 유리하게 도와주기 위한 행동이 오히려 해당 과제를 더욱 불리하게 만들 수 있고, 또한 특정과제를 도와주기 위해 이 과제와 경쟁에 있는 과제를 비정상적으로 평가했을 경우 그 결과로 인해 오히려 당초 도와주기로 의도했던 과제가 더욱 불리해지는 메카니즘을 갖추고 있다는 것이다. 예를 들어 위의 사례에서 평가자C는 과제3번을 1등으로 하여 높은 점수를 부여한 반면 경쟁관계에 있던 과제4번을 5등으로 하여 74점이라는 낮은 점수를 부여함으로써, 평가자C가 과제3번을 돕고자하는 의도가 있었다고 가정한다면, 평가자C의 의도된 행동은 결코 과제3번에게 유리한 방향으로 작용하지 않는다는 것이다. 평가자C가 5등에 74점을 부여한 이 행동은 과제3번이 5등이라고 인정하는 평가자들이 더 많기 때문에(3명) 과제4번보다 과제3번에 더 많은 영향을 미치고 결국은 과제3번의 평균점수를 낮추는데 기여하기 때문이다. 물론 나머지 평가자들이 객관적으로 평가했다는 전제가 필요하고, 절대값이 아니라 평균화 과정에서 다소 희석된다는 한계는 있다. 하지만 어쨌든 그 결과는 과제4번에는 2번밖에 영향이 없지만 과제3번에는 3번이나 영향을 주게 되는 것이다. 지금까지의 논의결과는 <표 8>과 같이 정리될 수 있다.

지금까지의 논의결과를 정리하면 다음과 같다. 첫째, 기존방식과 같이 점수를 중심으로 연구과제를 선정한다면 최고/최저점수 제외방식 보다는 통계학에서 이미 널리 사용되고 있는 특이치 제외방식인 상자-수염방식이 좀더 타당할 것이다. 둘째, 결국은 예산범위 내에서 연구비 지원의 우선순위를 정하는 평가라면 점수가 아닌 순위를 정하는 방식, 그것도 평가그룹 내에서 동의도를 반영한 스케이팅 방식이 적절할 것이다. 마지막으로 전문성과 공정성 두 마리 토끼를 잡고 싶다면, 현실적으로 적용이 간편하고 가장 합리적이면서 정보의 손실도 없고, 이해관계자라고 보여지는 특정 평가자의 편견으로부터도 어느 정도 자유로울 수 있는 표준화점수 방식이 적절할 것이다. 각 대안별 장단점을 요약하면 <표 9>와 같다.

〈표 8〉 과제별 점수, 순위 및 다양한 평균값

구 분	과제1		과제2		과제3		과제4		과제5	
	점수	순위	점수	순위	점수	순위	점수	순위	점수	순위
평가자 A	-	-	90	1	86	2	75	4	81	3
B	95	1	85	3	72	5	77	4	90	2
C	79	4	81	2	90	1	74	5	80	3
D	100	1	90	2	75	4	80	3	75	4
E	98	1	90	3	81	5	95	2	86	4
F	87	1	80	3	79	4	82	2	-	-
G	-	-	90	1	68	4	77	3	86	2
H	92	1	86	2	85	3	76	5	77	4
I	93	1	90	2	62	5	78	3	74	4
총 점	644		782		698		714		649	
표준편차	7.1	1.1	4.1	0.8	9.1	1.4	6.4	1.1	5.8	0.9
총평균	92.0	1.4	86.9	2.1	77.6	3.7	79.3	3.4	81.1	3.3
현행평균	93.0		87.3		78.0		77.9		80.8	
스케이팅		1		2		5		3		4
표준화 평균	90.5		86.7		79.1		80.0		80.6	

〈표 9〉 각 대안별 장단점

구분	장점	단점
상자-수업 방식	<ul style="list-style-type: none"> <li>통계적으로 검증된 방식이다.</li> <li>정보의 손실을 줄일 수 있다.</li> <li>평가자간 합의성(consensus)을 반영한다.</li> </ul>	<ul style="list-style-type: none"> <li>점수의 절대값에 여전히 영향을 받는다.</li> <li>순위정보가 반영되지 못한다.</li> </ul>
스케이팅 방식	<ul style="list-style-type: none"> <li>순위를 정하는 방법으로 많이 활용되고 있는 방식이다.</li> <li>전체정보를 활용하므로 정보의 손실이 없다.</li> <li>평가자간 합의성(consensus)을 반영한다.</li> </ul>	<ul style="list-style-type: none"> <li>순위만 반영하여 과제별로 미세한 점수차를 반영하지 못 한다.</li> <li>극단치 값이 여전히 영향을 준다.</li> </ul>
표준화점수 방식	<ul style="list-style-type: none"> <li>점수와 순위를 동시에 반영할 수 있다.</li> <li>전체정보를 활용하므로 정보의 손실이 없다.</li> <li>평가자간 합의성(consensus)을 잘 반영한다.</li> <li>과제별 미세한 점수 편차도 반영할 수 있다.</li> <li>특정 평가자의 의도된 행동을 방지 할 수 있다.</li> </ul>	<ul style="list-style-type: none"> <li>상당히 희석되지만 여전히 극단치 값이 영향을 준다.</li> <li>아직 검증되지 않은 방법이다.</li> </ul>

## VI. 결 론

오늘날 동료평가는 과학연구에서 옥석을 가려내는 “여과기”이자 과학 활동을 떠받치는 근간으로, 과학의 자율성을 지키는 상징과도 같은 존재로 널리 인정받고 있다. 그러나 이와 같은

동료심사의 높은 위상에도 불구하고, 1980년대부터 동료평가의 전제와 실행에 결부된 다양한 문제점들이 산발적으로 제기되기 시작하였다(김명진, 2007). 그중에 하나가 평가의 신뢰성 문제이다.

평가의 신뢰성에 대한 국제적 관심은 2003년 솔트레이크시티 동계올림픽 당시 동구권 심판들의 담합의혹으로 제기된 판정시비로 인하여 국제빙상연맹(ISU)이 채점방식을 심판 9명, 총 평균방식에서 심판 14명, 무작위 7명 평균방식으로 개선하였고, 영국 공학물리과학연구협회(EP SRC) 의장인 John O'Reilly는 2006년에 "Peer Review-It's Present and Future"에서 동료평가의 한계를 지적한바 있다. 또한 같은 해에 미국의 국립과학재단(NSF)은 "NSF 6개년(2006-2011) 전략계획" 발표를 통해 연구투자의 효율성 및 질적수준 향상을 위하여 다차원적 전문가 리뷰를 실시하고, 외부평가시 평가점수와 함께 심사평도 중요한 지표로 삼겠다고 선언함으로써 정량적인 평가점수뿐만 아니라 정성적인 전문가의 의견을 반영함으로써 채점방식의 한계를 보완하였다.

이러한 추세를 반영하여 본 연구에서는 국가연구개발사업 연구과제 선정과정에서 전문관리기관과 평가자간에 정보비대칭성이 존재할 수 있음을 확인하고, 이를 최소화 할 수 있는 방안을 제시코자 하였다. 이는 전문관리기관과 평가자간에 정보비대칭성을 최소화하는 것이 국가연구개발사업 연구비 투자효율성 증대에 기여한다는 전제이다. 즉 다시 말해서 연구과제 선정시 가능한 객관적으로 정확히 평가되어 연구능력이 탁월한 팀이 선정되어야 연구결과도 기대할 수 있을 것이라는 일반적인 인식에서 출발한 것이다. 특정 평가자에 의해 함량 미달의 연구팀이 선정되거나, 우수한 연구팀이 탈락한다면 정부가 바라는 최적의 연구팀을 선정하지 못할 것이고, 이로 인해 바람직한 연구성과를 기대하기 어려울 것이기 때문이다.<sup>11)</sup>

주요 연구성과는 다음과 같다. 첫째, 지금까지의 평가자 선정방식은 철저한 상피제도를 적용한 사전배제방식이었지만 사후검정 방식으로 전환함으로써 해당분야 최고 전문가의 활용이 가능하고, 평가자 섭외로 인한 과중한 업무로드를 경감할 수 있는 방안을 제시하였다. 둘째, 지금까지의 점수중심 연구과제 선정방식에서 점수와 순위를 동시에 고려하는 방식으로 개선하여, 공정성과 전문성을 동시에 확보하고, 연구관리전문기관과 평가자간의 정보비대칭 문제를 최소화 할 수 있는 방안을 제시하였다. 이러한 결과들은 실무적인 측면에서 도움이 될 뿐만 아니라 향후 다양한 분야에서 순위결정의 방법론 측면에서 도움을 줄 수 있을 것으로 기대된다.

11) 물론 다소 함량 미달인 연구팀이 열심히 연구한다면 더 좋은 연구성과를 이룰 수 있지만 선정되면 모두 열심히 연구한다는 것을 전제. 즉 연구팀별 도덕적 헤이 정도는 같은 것으로 간주. 또한 전문관리기관에서는 평가자의 공정성 담보를 위해 여러 가지 제척수단(동일기관 소속, 사제지간, 출신 학교 등)을 강구하고 있으나 현실적으로 한계가 있다.

본 연구의 한계는 다음과 같다. 첫째, 최고/최저점수를 제외하던 기존의 방식에서와 마찬가지로 극단치의 해석에 대한 한계가 여전히 남아 있다는 것이다. 양 극단에 있는 평가자들이 특정과제를 돕고자 하거나 방해 하고자 하는 의도가 있는 것이 아니라 그 분야에 최고 전문가일 수도 있다는 것이다. 둘째, 해당 기술분야에 영향력 있는 전문가 한사람이 평가자로 참여하게 되면, 그 전문가의 의견에 다른 평가자들이 영향을 받을 수 있다는 것이다. 이러한 경우는 현재 방식에도 존재하지만, 본 연구자가 제안하는 방식을 적용할 경우 자칫 이해관계자로 오해받을 수 있는 소지가 많아진다는 것이다. 셋째, 세 방식 모두 평가자가 의도적으로 약간 높거나 약간 낮은 점수를 부여함으로써 결과에 영향을 미치고자 한다면 이를 파악하기 어렵다는 것이다. 넷째, 하나의 사례분석을 통해 일반화하기에는 한계가 있다는 것이다. 즉 본 사례와 유사한 경우가 얼마나 빈번하게 일어나는지 다양한 사례를 제시하고, 제시된 대안의 효용성을 분석하지 못했다는 것이다. 향후 연구과제로는 다양한 실제자료에 근거한 실증분석을 통해 정보비대칭으로 인한 역선택의 문제를 밝히거나, 극단치들이 그 분야 최고 전문가인지, 아니면 비전문가인지, 또는 실수로 인한 것인지 등 극단치의 특성에 대한 분석적 연구가 필요할 것으로 생각된다. 또한 본 연구에서 제시한 세 가지 방식 모두 우선순위에 대한 결과가 상이하고, 어떤 방식이 더 신뢰성이 있는지에 대한 추가적인 연구가 필요할 것이다.

## 참고문헌

- 김동욱, 김현철 (2003), “대학평가에서 평가자간 오차 크기의 검정 및 평가자 객관성의 제고 방안”, 「통계연구」, vol 11.
- 김명진 (2007), “동료심사의 문제점과 대안에 대한 모색”, 「자연과학」, 제23호.
- 조만형, 홍형득, 김갑수, (2004), “연구과제 심사평가시스템의 한미 비교연구”, 「한국사회와 행정연구」, 제15권 제3호.
- 양정모, 한승환 (2008), 「연구지원사업 표준점수 산정방식 개선을 위한 연구」, 한국학술진흥재단.
- 윤덕균 (2006), 「국가연구개발사업의 평가결과 수용성 제고를 위한 방법론 연구」, 한국과학기술기획평가원.
- 이종승, 김성훈, 허숙 (1998), “학술연구 지원과제의 심사평가방식 개선방안 연구”, 한국학술진흥재단.
- 차종석, 김영배 (1994), “평가자간 신뢰도 및 동의도에 관한 분석적 고찰”, 「경영학연구」, 제

23권 특별호.

- 최병일 (2003), “댄스스포츠의 채점방식에 관한 비교연구 -합산식 채점방식과 스케이팅 채점 방식을 중심으로”, 연세대 교육대학원 석사학위논문.
- Bartko, J. J. (1976), “On Various Intraclass Correlation Reliability Coefficients”, *Psychological Bulletin*, Vol.83, No.5, pp.762-765.
- Bozeman, B. (1993), *Peer Review and Evaluation of R&D Impacts*, Chapter 5 in *Evaluating R&D Impacts: Methods and Practice*, Barry Bozeman and Julia Melkers (eds.), Boston: Kluwer Academic Publishers.
- Chubin, D.E. and Hackett, E.J. (1990), *Peerless science: Peer review and U.S. science policy*, New York: State University of New York Press.
- Cicchetti, Domenic V. (1991), “The reliability of peer review for manuscript and grant submissions. : a cross-disciplinary investigation”, *The behavioral and Brain sciences*, 14: 119-186.
- Conrad, E. and T. Maul, (1981), *Introduction to Experimental Psychology*, New York : John Wiley & sons.
- Foltz, Fran A. (2000), “The ups and downs of peer review: making funding choice for science”, *Bulletin of science, Technology & Society*, 20(6): 427-440.
- Kostoff, R.N. (1997). Research program peer review : principles, practices, protocols (<http://www.dtic.mil/dticlkostoff/index.html>)
- Kozlowski, S. W, and K. Hattrup (1992), “A Disagreement About Within Group Agreement : Disentangling Issues of Consistency Versus Consensus”, *Journal of Applied Psychology*, Vol.77, No.2, pp161-167
- Luukkonen-Gronow, T. (1987) “Scientific Research Evaluation: A review of methods and various contexts of their application”, *R&D Management* 17(3) p207-221.
- Mitchell, S. K. (1979), “Interobserver Agreement, Reliability, and Generalizability of Data Collected in Observational Studies”, *Psychological Bulletin*, Vol.86, No2, pp376-390
- NRC. (1996), *Environmental Management Technology-Development program at the Department of Energy*. Washington, D.C: National Academy press.
- Rennie, D. (1993), More peering into editorial peer review. *JAMA*, 20: 2856-2858.
- Royal Society. (1995), *peer review: An Assesment of Recent Development*. London:

The Royal Society.

Tinsley, H. E. and D. J. Weiss, (1975), "Research Methodology", *Journal of Counseling Psychology*, Vol.22, No.4, pp358-376

U.S.DOE. (1998), *Implementation Guidance for the Office of Science and Technology Technical peer Review Process*. Chicago: U.S Department of Center for risk Excellence.

#### 이윤규

---

경기대학교에서 경영학 박사학위를 취득하고 현재 경기대학교 회계세무학과 교수로 재직 중이다. 주요 저서는 회계원리, 원가회계, 환경 및 사회공헌회계론 역서로 진화하는 환경경영이 있고 관심분야는 기업 사회회계, 사회경제회계, 환경회 등이다.

#### 손충근

---

고려대학교에서 경영학 석사학위를 취득하고 현재 한국연구재단에서 책임연구원으로 근무 중이다. 관심 분야는 연구개발비 투자효율성, 무형자산 평가, 국가연구개발사업 기획, Open Innovation 등이다.