

2025 Vol. 01

차세대리포트

연구 성과, 어떻게 평가할 것인가?



펴낸곳

한국과학기술한림원
031)726-7900

펴낸이

정진호

발행연월

2025년 3월

홈페이지

www.kast.or.kr

기획·편집

한국과학기술한림원 정책연구팀

디자인·인쇄

경성문화사
02)786-2999

이 보고서는 복권기금 및 과학기술진흥기금의 지원을 통해 제작되었으며,
모든 저작권은 한국과학기술한림원에 있습니다.

차세대리포트

발 간 사

오늘의 날씨와 주요 뉴스, 그리고 나의 일정을 알려주는 인공지능 비서로 하루를 시작하고, 자율주행 자동차를 타고 이동하며, 디지털 치료기기로 건강을 관리하는 일상. 이처럼 지금 시대를 살아가는 우리는 매 순간을 과학기술과 함께하고 있다. 앞으로 과학기술은 우리의 삶을 보다 빠르게, 더욱 혁신적으로 변화시키게 될 것이다. 그리고 이 변화의 중심에는 과학기술 최일선에서 새로운 지식과 기술을 발견하기 위해 노력하는 과학기술인들이 있다.

우수한 젊은 과학기술인 그룹인 ‘한국차세대과학기술한림원(Young Korean Academy of Science and Technology, Y-KAST)’ 회원들과 연구 현장 최 일선에서 활약하고 있는 최고의 젊은 과학자들은 차세대리포트를 통해 과학기술 분야의 최신 기술과 연구 동향, 이슈를 대중들이 이해하기 쉽게 소개하고 있다. 특히 우리 사회가 직면한 도전 과제를 과학기술로 해결하기 위한 방향을 제시하고, 연구자들이 더욱 연구에 전념할 수 있는 환경 조성과 정책적 지원에 대한 목소리를 담아내어 과학기술인들의 노력을 통해 대한민국이 과학기술 강국으로 성장할 수 있도록 하고자 차세대리포트를 발간해 오고 있다.

한국과학기술한림원은 차세대리포트를 통해 다양한 이슈에 대해 새로운 시각과 신선한 의견을 전하고자 노력해 왔으며, ‘과학기술 분야 최신 동향’과 ‘사회적 이슈 및 현안’이라는 측면을 함께 다루기 위해 주제의 선정에서부터 발간에 이르기까지의 모든 과정을 치열한 고민을 통해 진행하고 있다.

이 같은 고민을 통해 선정된 2025년도 차세대리포트의 첫 번째 주제는 바로 ‘연구 성과, 어떻게 평가할 것인가?’이다. 우리나라가 과학기술 역량을 꾸준히 높여가기 위해서는 연구의 질적 가치와 성과를 공정하고 합리적으로 평가할 수 있는 체계를 확립함으로써 우수한 연구 성과를 만들어 갈 수 있는 환경을 조성하는 것이 필수적이기 때문이다.

그러나 공정하고 합리적이며, 효율적인 연구 성과평가 체계 구축을 위해서는 학문 분야별 특성을 비롯하여 다양한 요소에 대한 세심한 고려와 공감대 형성을 통한 접근이 필요하다. 이번 차세대리포트는 국내외 연구 성과의 평가 방법에 대한 동향을 살펴보고, 보다 선진화된 연구 성과평가 제도 구축을 위한 아이디어를 제시하는 데 초점을 두었다. 이를 통해 연구자들이 혁신적인 연구에 몰입할 수 있는 환경이 조성되고 연구 성과를 평가하는 과정에서 반드시 필요한 공정성과 효율성은 한층 강화될 수 있는 계기가 마련되기를 기대해 본다.

2025년 3월
한국과학기술한림원 원장
정진호

참여자 소개



김상우 | 연세대학교 의과대학 교수

의생명시스템정보학 분야를 대표하는 과학자로 생물정보학을 바탕으로 유전체를 연구하고 그 결과물을 실제 환자의 진단과 치료에 활용하는데 집중하고 있다. 실제 환자데이터를 이용하여 질병의 유전적 원인과 메커니즘을 규명하고, 정확하고 새로운 분석을 해낼 수 있는 알고리즘과 소프트웨어를 개발하는데 주력하고 있다.



박주영 | 서울대학교 건설환경공학부 교수

지속가능한 삶을 위한 효율적인 자원 사용에 대해 고민하며, 플라스틱, 산업 클러스터, 수자원 인프라, 건물 등을 중심으로 자원효율성을 평가하고 순환경제 전략을 탐구하는 융합분야 연구를 수행하고 있다.



이상규 | 기초과학연구원 인지 및 사회성 연구단 연구위원

뇌과학과 합성생물학의 융합분야를 대표하는 연구자로 분자 광유전학 기술 및 합성단백질 제작을 통해 뇌 기능을 분자 수준에서 관찰하고 조절하는 연구를 수행하고 있다. 최근 뇌 기능을 시냅스 수준에서 규명하는 연구와 뇌 질환의 새로운 치료방법을 제시하는 연구를 수행 중이다.



이성주 | 서울대학교 산업공학과 교수

서울대학교 기술인텔리전스 연구실에서 미래 예측과 기술전략 수립 관련 연구를 수행하고 있다. 특히 특허와 논문을 포함한 과학기술 분야의 텍스트 데이터를 활용해 유망기술을 발굴하고 과학기술 변화를 감지하려는 연구를 수행하고 있다.



이학연 | 서울과학기술대학교 산업공학과 교수

기술과 비즈니스, 인공지능의 접점을 개척하는 융합 연구자로, 현재 BK21 4단계 데이터사이언스와 비즈니스 포텐셜 교육연구단 단장을 맡고 있다. 데이터로부터 유망 기술 및 비즈니스 기회를 발굴하고, 인공지능을 활용하여 혁신 기회의 유망성을 평가하는 연구를 선도하고 있다.

목차

| | |
|-------------------------------------|----|
| 발간사 | 01 |
| 참여자 소개 | 02 |
| 들어가기 | 04 |
| 연구 성과평가 제도 관련 현황 및 제언 | |
| I. 연구 성과평가 방법론 | 06 |
| 가. 연구 성과평가의 기준과 방식 | 06 |
| 나. 인용기반 지표 개요 | 08 |
| 다. 새로운 대안으로 부상하고 있는 알트메트릭과 AI | 13 |
| II. 연구 성과평가 제도 국내동향 | 16 |
| 가. 우리나라 연구 성과평가 방법의 변천 과정 | 17 |
| 나. 분야별 연구 성과평가 방법의 변화 | 18 |
| 다. 연구자의 경력 단계에 따른 학문적 우수성 평가 | 20 |
| 라. 현재 연구 성과평가 기준 및 한계 | 21 |
| III. 연구 성과평가 제도에 대한 젊은 과학자들의 생각과 제언 | 22 |
| IV. 연구 성과평가 제도 해외동향 | 29 |
| 가. 연구평가 개혁의 배경과 대표적인 글로벌 이니셔티브 | 30 |
| 나. 연구 성과평가 개혁을 위한 RRA의 원칙 및 약속 | 31 |
| V. 정책 제언 | 36 |
| 맺음말 | 41 |

오늘날의
학계 환경에서는
아마도

내가
찾고자 하는 입자를
발견하지 못했을 것이다

이 말은 현대 물리학의 기념비적 성과로 여겨지고 있는 힉스입자^{Higgs boson}를 발견한 공로로 지난 2013년 노벨물리학상을 수상한 ‘피터 힉스^{Peter Higgs}’ 박사가 한 인터뷰에서 밝힌 소감이다. 처음 힉스입자의 존재를 예측한 1960년대의 연구 환경과 현재 상황이 크게 다르다는 점을 강조하기 위해 그런 소감을 말했다는 것이 후일담으로 전해지고 있다.

실제로 힉스 박사가 이 같은 말을 한 이유에 대해 당시 소감을 들은 대다수의 과학자들은 충분히 공감한다는 의견을 내놓은 바 있다. 힉스입자를 예측한 것은 1960년대이지만, 입자의 존재가 확인되기까지는 그로부터 수십 년이 걸렸기 때문이다. 이렇게 오랜 시간 동안 진행된 연구와 발견의 과정이 현재의 연구 환경이라면 지속하기가 어려웠을 것이라고 생각한 것이다.

그렇다면 힉스 박사가 언급한 현재의 연구 환경은 과연 어떤 상황일까? 과거에 비해 연구 성과와 관련된 압박이 커지면서, 장기적인 연구가 어려워지고 있다는 점은 많은 과학자들이 인지하고 있는 사실이다. 또한 연구과제를 선정하는 과정에서 행정적 절차가 많아졌으며, 창의적인 연구가 위축되고 있는 점도 분명하다. 경쟁적인 학계 분위기는 또 어떤가? 연구자 간 경쟁이 심화되면서, 협력보다는 개인적인 성과에 집중하는 경향이 강해지고 있는 중이다.

연구 환경이 이렇게 변화된 데에는 여러 원인이 있겠지만, 무엇보다 우수한 연구 성과를 내야 한다는 압박감이 가장 클 것이다. 과학자와 공학자, 인문학자 모두가 각자의 자리에서 우수한 연구 성과를 내기 위해 오늘도 끊임없이 노력하고 있는 것이다.

이처럼 모두가 우수한 연구 성과를 내기 위해 노력하다 보니, 연구 성과를 평가하는 방법에 대해서도 많은 연구가 진행되고 있다. 연구 성과의 평가는 주로 정량적 지표를 중심으로 이루어져 왔으며, 논문 수와 인용 횟수, 그리고 논문의 영향력 지수 등이 연구의 가치를 가늠하는 대표적인 기준으로 사용되었다. 그러나 이러한 정량적 평가 방식은 연구의 질적 가치를 제대로 반영하지 못한다는 한계를 보여 온 것도 사실이다.

이에 따라 연구 성과에 대한 평가 기준도 시대적 흐름에 따라 변화하면서, 과거 정량적 평가 방식만으로 평가될 수 없다는 인식이 확산되고 있다. 획기적인 연구 성과임에도 단기간 내에 높은 논문 인용 수치를 기록하지 못하는 경우도 있으며, 파급효과가 큰 연구일지라도 기존의 평가지표만으로는 그 가치를 제대로 평가하기 어려운 경우가 많았기 때문이다.

이런 모순이 발생하는 이유에 대해 전문가들은 ‘연구 성과를 평가하는 것은 단순히 숫자만으로 환원될 수 없는 복잡한 과정’이라는 점을 지적하고 있다. 예를 들어, 어떤 연구자가 10년 동안 한 가지 주제에 몰입하여 획기적인 발견을 이루어냈다고 가정했을 때, 이 연구자의 성과는 단순히 논문의 수나 인용 횟수로 평가하는 것이 아니라 해당 연구가 사회에 미친 영향, 기술의 상용화 가능성, 그리고 미래 연구에 대한 기여도 등을 종합적으로 고려해야만 한다. 그러나 현재의 평가 체계는 이러한 요소들을 충분히 반영하지 못하고 있다.

이 같은 문제를 해결하기 위해, 최근 들어 국내는 물론 해외에서도 연구 성과를 다각적으로 평가하려는 시도가 활발히 이루어지고 있다. 연구자의 성과를 평가할 때 논문의 양적 지표뿐만 아니라, 연구의 사회적 영향력, 산업계와의 협력 성과, 그리고 연구의 혁신성 등을 종합적으로 평가하는 ‘다차원 평가 체계’ 등이 그것이다.

이는 연구의 양적 성과를 중시하는 기존의 평가 방식에서 벗어나, 연구의 질적 가치를 더욱 합리적으로 평가할 수 있는 새로운 패러다임이라고 할 수 있다. 연구 성과평가의 새로운 패러다임과 관련한 사례로는 영국의 ‘연구우수성평가^{REF}’ 제도나 네덜란드의 ‘전략평가프로토콜^{SEP}’ 시스템 등이 꼽힌다.

이 외에도 연구 성과의 공정한 평가 체계를 구축하기 위해서는 연구 윤리 준수를 독려하고, 장기 연구 프로젝트를 지원하며, 연구자의 경력 개발을 도와줄 수 있는 다양한 정책적 지원이 고려되어야 한다.

따라서 이번 차세대리포트에서는 연구 성과평가 체계와 관련된 국내의 동향을 면밀하게 분석하고, 이를 바탕으로 새로운 시각의 정책적 개선 방안을 제안하고자 한다. 이와 함께 연구 성과평가의 기준이 어떻게 변화하고 있으며, 선진국들은 어떤 방식으로 연구자를 지원하고 있는지 등도 살펴볼 것이다. 또한, 국내 연구 환경의 현실을 반영하여 보다 공정하고 효율적인 평가 체계를 구축하기 위한 정책 아이디어도 제시하고자 한다.

연구 성과평가 체계를 둘러싼 논의는 단순히 연구자 개인의 문제를 넘어 국가 연구 경쟁력과 직결되는 중요한 과제다. 그러다 보니 공정하고 체계적인 시스템을 마련함으로써 연구자들이 창의적이고 혁신적인 연구를 수행할 수 있도록 지원하는 것 또한 필수적인 제도로 정착하고 있다. 이번 차세대리포트가 연구 성과평가 체계를 개선하기 위한 의미 있는 출발점이 되기를 기대하며, 앞으로도 지속적인 논의와 발전이 이루어지기를 기원한다.



연구 성과평가 방법론



가 연구 성과평가의 기준과 방식

연구 성과평가는 다양한 목적을 위해 이루어진다. 연구자 채용, 승진, 업적 평가 및 포상, 연구 과제 선정 등에서 연구자의 역량을 평가하는 중요한 과정이다. 평가의 관점은 목적과 맥락에 따라 달라질 수 있지만, 대체로 연구의 질적 우수성을 평가하는 것에 초점을 두고 있다. 여기서 중요한 것은 바로 ‘우수한’ 연구의 기준이 명확하게 무엇인가 하는 것이다. 일반적으로 연구의 우수성은 △엄밀성rigor △독창성originality △영향력impact 등 크게 세 가지 기준으로 평가된다.

엄밀성은 연구가 얼마나 철저하고 정확한 방법을 통해 수행되었는지에 대한 기준을 의미한다. 연구 과정에서 발생하는 데이터의 수집을 비롯하여, 실험 및 분석 등에서 학문적 기준을 엄격히 준수해야 한다. 또한 동일한 조건에서 같은 방법을 사용했을 때 유사한 결과를 얻어야 비로소 과학적 연구로 인정받을 수 있다.

또한 독창성은 기존 연구와의 차별성을 의미하는 것으로서, 새로운 이론이나 방법론을 제시하거나 기존 연구에서 나타난 한계점을 해결할 수 있어야 한다.

마지막 기준인 영향력은 연구 결과가 학술적으로나 사회적으로 미치는 파급 효과를 의미한다. 학술적 영향력은 후속 연구에서의 활용도를, 그리고 사회적 영향력은 정책, 산업, 기술 개발 등 현실 문제의 해결에 기여하는 정도를 포함하는 것이다.

그렇다면 엄밀성과 독창성, 영향력은 어떻게 평가할 수 있을까? 일반적으로 연구 성과평가는 두 가지 방식으로 이루어지는데, 첫 번째는 ‘동료 평가^{peer review}’이고, 두 번째는 계량서지 지표^{bibliometrics}이다.

동료 평가는 같은 분야의 전문가들이 연구의 우수성을 평가하는 정성적 평가 방식이고, 계량서지 지표는 서지 데이터를 통해 연구 성과를 측정하는 정량적 평가 방식이라고 할 수 있다. 정성적 평가 방식인 동료 평가는 전문가들의 질적 평가를 통해 연구의 전반적 우수성을 판단할 수 있는 장점이 있지만, 평가자의 주관에 따라 결과가 달라질 수 있으며 무엇보다 평가 과정에서 시간이 오래 소요되고 비용이 많이 발생하는 단점을 갖고 있다.

반면에 정량적 평가는 연구 성과를 숫자로 표현하여 객관적으로 비교하기 쉽게 해주는 장점이 있지만, 피인용수 같은 지표만으로는 연구의 질적 우수성을 충분히 대변하지 못한다는 한계가 있다. 따라서 이 두 가지 방식은 각각의 평가 기준과 목적에 맞게 보완적으로 활용되어야만 한다.

엄밀성과 독창성은 해당 분야에 대한 깊은 전문 지식이 있어야만 평가가 가능하기 때문에 객관적으로 수치화하는 것은 매우 어려운 작업이다. 이러한 이유로 논문 투고 및 심사 과정에서 동료 평가를 통해 평가되는 경우가 대부분이다.

한편, 영향력은 논문이 출판된 이후에 실질적으로 발생하며, 주로 정량적 지표로 측정된다. 사회적 영향력은 범위 설정과 객관적인 측정이 어려워서 대부분의 평가에서 반영되지 않고 있지만, 학술적 영향력은 후속 논문에서 얼마나 인용되었는지를 기준으로 비교적 쉽게 측정할 수 있다.

나 인용기반 지표 개요

(1) 논문 평가지표

논문의 가치를 측정하는 데 있어 ‘피인용수’는 중요한 역할을 하는 지표다. 이는 한 논문이 다른 연구자들에 의해 얼마나 많이 인용되었는지를 나타내는 수치이기 때문이다. 연구자들이 자신의 논문에서 다른 논문을 인용한다는 것은 그 논문의 내용이 의미 있고 가치 있다고 판단했다는 의미로 해석할 수 있다.

이러한 피인용수 정보는 최근 학술 데이터베이스의 발전으로 누구나 쉽게 확인할 수 있다. 이는 연구자들뿐만 아니라 학생들이나 일반인들도 특정 논문의 중요성을 간단히 파악할 수 있게 해주는 계기가 되었다. 따라서 피인용수는 논문의 영향력을 나타내는 중요한 지표로서, 연구 결과가 학계에서 얼마나 활발히 공유되고 후속 연구에 활용되는지를 보여준다.

하지만 피인용수가 연구의 질적 우수성을 대변할 수 있는가에 대한 의문이 제기되는 것도 사실이다. 실제로 많은 실증 연구에서 동일 논문에 대한 동료 평가 점수와 피인용수의 상관관계가 예상보다 낮았으며, 일부 분야에서는 통계적으로 유의미한 관계조차 발견되지 않은 사례도 있다. 이는 논문의 인용이 연구의 우수성뿐만 아니라, 여러 사회적 요인의 영향을 받을 수도 있다는 점을 가리키는 것이다.

다시 말해 연구자가 작성한 논문에서 참고문헌을 인용하는 행위 자체가 사회적 행위이기 때문에, 지역, 성별, 언어에 따라 인용이 편향될 수 있다는 뜻이다. 연구자 본인이나 소속 연구 그룹의 논문, 또는 투고 대상 저널의 논문을 우선적으로 인용하는 자기 인용^{self-citation} 현상이 나타나기도 한다.

또한 논문 자체의 특성도 인용 가능성에 영향을 미친다. 예를 들어, 최신 유행을 다루는 논문은 일반 연구 논문보다 주목받을 가능성이 크며, 특히 리뷰 논문은 그 특성상 더 많은 인용을 받는다. 논문의 길이가 길수록 다양한 내용을 포함할 수 있어 인용될 확률이 높아지고, 오픈 액세스 저널에 실린 논문은 접근성이 높아 연구자들이 더 자주 참고하는 경향이 있다.

가장 큰 문제는 학문 분야별로 인용 패턴의 차이가 크다는 것이다. 예를 들어 의학이나 생명과학 분야에서는 연구 결과가 빠르게 축적되고 후속 연구도 신속하게 진행된다. 이 때문에 의학이나 생명과학 분야에서의 논문 출판과 인용은 매우 활발하다. 반면에 수학이나 사회과학 등의 분야는 상대적으로 피인용수가 낮게 나타난다.

각 학문 분야에서 논문에 포함되는 참고문헌의 수 역시 서로 다른 양상을 보인다. 이론 연구 비중이 큰 수학 논문의 참고문헌 수는 상대적으로 적은 반면, 기존 연구를 폭넓게 검토하여 논의의 기초로 삼는 사회과학 분야의 논문은 훨씬 많은 수의 참고문헌을 포함하므로 당연히 논문당 평균 피인용수도 높아질 수밖에 없는 것이다.

이런 문제점들 때문에 서로 다른 학문 분야에 속한 연구 성과나 연구자들의 업적을 비교할 때 피인용수를 기준으로 삼는 것은 부적절하다는 것이 학계의 의견이다. 하지만 대학의 교수 업적 평가와 같은 현실에서는 서로 다른 학문 분야 간의 비교를 피하기 어려운 상황이 존재한다. 학문 분류 체계상 동일 분야 내에서도 세부 분야마다 인용 패턴이 다르며, 여러 분야가 혼합된 융합 학문의 경우에는 평가가 더 복잡해질 수밖에 없다.

이러한 문제를 해결하기 위해 최근 들어 학문 분야별 인용 특성을 반영할 수 있는 정규화된 지표들이 선을 보이고 있다. 대표적으로 엘스비어^{Elsvier}의 SCOPUS 데이터 베이스를 활용한 FWCI^{field-weighted citation impact}가 있는데, FWCI는 분야별 논문들의 평균 피인용수 대비 특정 논문의 피인용수 비율로 계산된다.

FWCI 값이 1보다 작으면 해당 분야의 평균보다 낮고, 1보다 크면 평균보다 높음을 의미한다. 하지만, 이 정규화 지표 역시 여전히 피인용수를 기준으로 하므로 한계가 있다. 또한 학문 분야의 구분 방법, 특정 저널의 분류 방식 등에 따라 지표 값이 달라질 수 있다는 것도 문제점으로 지적되고 있다.

(2) 저널 평가지표

피인용수가 연구 성과평가지표로서 가지는 가장 큰 현실적 한계는 시점의 문제다. 논문이 게재된 이후 오랜 시간에 걸쳐 피인용이 이루어지므로, 피인용수를 통해 실제 논문의 영향력을 측정하기까지 상당한 시간이 필요한 것이 현실이다.

그러나 대학이나 연구소의 연구자 업적 평가는 보통 논문이 게재된 당해 말부터 곧바로 그 성과를 반영하며, 연구 제안서 심사에서도 연구자의 역량을 평가할 때 일반적으로 최근 3년 또는 5년 등 단기간의 실적만을 고려한다. 이러한 이유로 대부분의 경우 논문의 진정한 영향력이 드러나기 전에 평가가 진행된다. 그래서 피인용수를 지표로써 사용하기가 어렵고, 사용하더라도 한계가 있을 수밖에 없다.

이에 대한 대안으로 최근 들어 개별 논문의 피인용수 대신 해당 논문이 게재된 저널의 평균 피인용수를 논문의 우수성을 측정하는 지표로 활용하고 있다. 바로 ‘학술지 영향력 지수^{IF} Impact Factor’인데, IF는 해당 저널이 발행한 논문들의 평균 인용 횟수, 즉 인용률을 의미한다. IF는 특정한 2년의 기간 중 발행된 논문들이 그 다음 해에 받은 총 인용 횟수를 해당 2년간 발행된 전체 논문 수로 나누어 산출된다. 예를 들어 IF 2024는 2022년과 2023년에 발행된 논문들이 2024년에 인용된 총 횟수를 2022년과 2023년에 발행된 논문 수로 나눈 값이며, 2025년 중순에 발표된다.

IF의 본래 목적은 연구자들의 저널 선택이나 기관의 저널 구독 결정을 돕기 위한 것이었다. 하지만 2000년대 이후 국내 연구 성과평가의 초점이 양적 평가에서 질적 평가로 전환되면서, IF는 연구의 질적 우수성을 측정하기 위한 대용 지표로 본격적으로 활용되기 시작했다. IF는 계산 방법이 간단하고 직관적이어서 저널의 영향력을 쉽게 파악할 수 있는 장점이 있지만, 이를 연구 논문의 우수성 평가지표로 사용하는 것의 적절성 여부에 대해서는 다음과 같은 문제점들이 제기되고 있다.

첫 번째 문제는 IF가 저널 단위의 평균적 우수성을 나타내는 지표이므로, 개별 논문의 우수성을 대변하지 못한다는 것이다. 실제 연구에 따르면, 가장 많이 인용된 15%의 논문이 전체 인용의 절반을 차지하고, 상위 50%의 논문이 전체 인용의 90%를 차지하는 것으로 나타났다. 이는 일부 높은 피인용 논문들이 저널 전체의 IF를 크게 올릴 수 있음을 의미한다.

두 번째 문제는 IF 산출에 사용되는 DB의 범위가 제한적이라는 점이다. Web of Science DB는 약 23,000개의 저널을 수록하고 있는데, 이는 전 세계 저널의 10% 수준에 불과하다. 특히 영어권인 미국 저널에 편중되어 있고, 학술적으로 중요한 도서나 학술대회 논문 등이 제외된다는 문제를 안고 있다.

이어서 세 번째는 IF의 계산 방식에서 여러 가지 문제가 발생하고 있다는 점이다. IF 산출 공식에서 분모에는 일반 논문과 리뷰 논문 등 인용 가능한 항목만 포함되지만, 분자에는 사설, 서신 등 저널에서 발간하는 모든 유형의 출판물에 대한 인용이 포함되어 있는 까닭에 왜곡의 여지가 존재한다는 사실이다. 또한 분자의 피인용 횟수 산출 시, 자기인용, 즉 동일 저널 내에서 이루어지는 인용을 포함시키는 것에 대한 논란도 존재한다. 이 외에도 2년이라는 제한된 인용 기간은 중요한 문제점으로 지적되고 있는데, 2년이라는 기간은 한 논문이 출판된 후 다른 연구자들이 이를 검토해서 새로운 논문에 반영하기에는 매우 부족한 시간이다.

마지막으로 학문 분야별 IF 값의 차이가 크다는 점도 문제로 지적되고 있다. 예를 들어, 수학 분야 상위 IF는 3~4점대에 머무는 반면, 생명과학 분야의 상위 저널들은 30점을 초과하는 경우가 많다. 이러한 학문 분야별 IF 격차는 이전에 언급한 논문 인용 패턴의 차이에서 기인한 것이다. 따라서 IF 값만으로 서로 다른 학문 분야의 연구 성과를 비교하는 것은 부적절하다. 그럼에도 불구하고 연구 성과평가 과정에서는 서로 다른 학문 분야 간 비교가 불가피한 상황이 빈번하게 발생하게 되는데, 이러한 경우 IF 백분위^{percentile}와 같은 상대적 지표를 활용하는 것이 보다 합리적인 대안이 될 수 있다.

2021년부터 도입된 JCI^{Journal Citation Indicator} 또한 대안으로 고려할 수 있다. JCI는 개별 논문이 속한 학문 분야와 출판 연도를 고려하여 기대 인용 횟수를 산출한 후, 실제 발생한 인용 횟수를 이 기댓값으로 나눈 것이다. 이는 앞서 언급한 개별 논문의 FWCI와 유사한 개념으로, 특정 저널의 JCI가 1보다 크면 평균을 초과하는 수준이고, 1보다 작으면 평균 이하의 영향을 나타낸다. 그러나 이 지표들 역시 저널이 어느 학문 분야로 분류되는지에 따라 지표 값이 크게 달라질 수 있다는 문제를 갖고 있다.

(3) 연구자 평가지표

‘연구 논문’이 아닌 ‘연구자’의 우수성이나 연구 역량을 평가하는 문제는 더욱 복잡한 양상을 띤다. 예를 들어, 단 한 편의 매우 영향력 있는 논문을 발표한 후 더 이상의 연구 성과를 내지 않는 연구자와, 개별 논문의 영향력은 상대적으로 낮지만 꾸준히 다수의 논문을 발표하는 연구자 중 누가 더 우수한 연구자일까? 이 질문에 쉽게 답하기 어려운 이유는 연구자를 평가할 때 생산성과 영향력이라는 두 가지 차원을 동시에 고려해야 하기 때문이다.

이러한 문제의식에서 출발하여 개발된 연구자 평가지표 중 하나가 바로 h-index이다. h-index는 특정 연구자가 발표한 논문 중 h번 이상 인용된 논문이 h편 이상 존재함을 나타내는 지표이다. 예를 들어, 어떤 연구자의 h-index가 10이라면 이는 해당 연구자가 최소 10회 이상 인용된 논문을 10편 이상 보유하고 있다는 의미다. h-index는 연구자의 생산성과 영향력을 단 하나의 숫자로 표현할 수 있다는 점에서 분명하면서도 실용적인 지표로 인식되고 있지만, 워낙 단순하다 보니 지표로서 몇 가지 한계점을 보이는 것도 사실이다.

우선 h-index는 시간이 지남에 따라 누적되는 특성이 존재한다. 이로 인해 오랜 기간 연구를 수행한 연구자들은 자연스럽게 높은 h-index를 보유하게 되는 반면에 우수한 연구 성과를 내더라도 경력이 짧은 신진 연구자들은 상대적으로 낮은 h-index를 기록하게 되는 구조적 불균형이 발생한다. 또한 h-index는 논문 저자의 역할이나 기여도를 구분하지 않는 문제점이 있다. 주저자나 교신저자로서 핵심적인 역할을 수행한 논문이나 단순 공동 저자로 참여한 논문을 동일하게 취급하기 때문에, 실질적인 연구 기여도를 정확히 반영하기 어려운 측면이 있는 것이다. 이 밖에도 h-index는 피인용수를 기반으로 산출되는 지표이므로, 학문 분야 간 직접적인 비교도 어렵다.

(4) 정량지표의 부작용

여러 가지 정량적 평가지표들은 연구 성과를 직관적이고 효율적으로 측정할 수 있다는 점에서 분명 유용한 평가 도구로 평가받아 온 것이 사실이다. 하지만 정량지표에 대한 의존도가 지나치게 높아져서 나타나는 여러 가지 부작용들이 있다. 특히 연구 성과를 평가하는 데 있어 IF가 무분별하게 활용되면서, 학문 생태계 전반에서 우려스러운 부정적 현상들이 나타나고 있다.

가령 IF가 높은 저널에 논문을 게재하는 것이 연구 성과를 평가하는 절대적인 기준으로 인식되면서, 저널 간에 치열한 IF 경쟁이 시작되었다. 이로 인해 ‘IF 게이밍^{gaming}’이라는 왜곡된 관행이 확산되고 있다. IF 게이밍은 IF를 인위적으로 부풀리기 위해 편집자나 출판사, 때로는 저자와 심사자가 취하는 부적절한 행위를 의미한다. 일부 저널에서는 논문 게재의 조건으로 해당 저널의 최근 논문들을 일정 수 이상 인용할 것을 요구하는 경우도 있다.

또한 단일 저널 내에서 자기 인용이 쉽게 포착될 수 있으므로, 여러 저널이 서로의 논문을 상호 인용하는 일종의 ‘인용 카르텔’이 형성되기도 한다. 일부 저널들은 출판 시점을 전략적으로 조정하여 IF를 부풀리기도 하며, 2년이라는 기간 내 피인용 횟수를 최대한 확보하기 위해 논문을 조기에 온라인으로 공개하고 공식 출판은 지연시키는 방식도 사용된다.

이 뿐만이 아니다. 출판물의 유형 또한 IF를 조작하는 수단이 된다. IF 계산식에서 논문 수분모에는 포함되지 않지만 피인용수분자에는 포함되는 사설^{editorial}이나 서신^{letter}을 다수 발행하거나 인용 가능성이 높은 리뷰 논문의 비중을 늘려가는 방식이다. 또한 단기간에 주목받을 수 있는 주제를 특집호 형태로 빠르게 출간하는 경향도 보인다.

따라서 IF를 높여야 하는 부담을 지고 있는 편집자들은 ‘학술적으로 우수한 논문’ 보다는 ‘빠른 시일 내에 많은 인용이 기대되는 논문’을 우선적으로 선택하게 된다. 이로 인해 연구자들은 게재 가능성이 높은 혹은 단기적으로 인용이 보장되는 유행중인 연구 주제에 집중하게 된다. 이는 결과적으로 연구 주제의 다양성을 저해하고 학문의 장기적 발전 기반을 약화시키는 요소 중 하나가 되고 있다.

다 새로운 대안으로 부상하고 있는 알트메트릭과 SI

(1) 알트메트릭

학술적으로 평가되는 영향력은 주로 인용기반 지표를 통해 측정되어 왔으나, 사회적 영향력은 대부분 연구 성과평가에서 간과되어 왔다. 이는 사회적 영향력이 중요하지 않아서가 아니라, 측정이 어렵기 때문이다. 그렇지만 디지털 환경에서 연구 성과물이 빠르게 공개되고 쉽게 공유되는 시대가 열리며, 연구의 사회적 영향력을 수치화하려는 새로운 시도가 등장했다. 바로 알트메트릭^{altmetrics}이다.

알트메트릭은 대안적인^{alternative} 지표^{metrics}를 의미하는데, 전통적인 인용기반 지표와는 달리 온라인에서 발생하는 다양한 언급, 공유, 다운로드 등 활동을 종합적으로 측정하여 연구의 사회적 영향력을 평가하려고 한다. 예를 들어, 블로그나 위키백과, 뉴스 등에서의 언급, 트위터 및 페이스북, 유튜브와 같은 소셜 플랫폼에서의 공유 및 ‘좋아요’ 수 등이 해당된다. 또한 멘델레이^{mendeley}와 조테로^{zotero} 같은 참고문헌 관리 도구에서 다른 연구자들에 의해 얼마나 많이 저장되었는지가 주요 측정 대상이 된다. 이 외에도 기존 학술 데이터베이스에 포함되지 않는 정책 보고서나 특허와 같은 다른 문헌에서의 인용도 측정된다.

물론 단점도 있다. 알트메트릭은 연구가 발표되는 즉시 반응을 실시간으로 확인할 수 있는 장점이 있지만, 연구 성과평가에 직접적으로 활용하기에는 근본적인 한계가 존재한다.

단순히 소셜 미디어에서 널리 회자된다고 해서 그 연구의 영향력이 높다고 볼 수 없기 때문이다. 대중의 흥미를 끌기 쉬운 주제는 과대평가될 위험이 있으며, 때로는 부정적인 이유로 많은 주목을 받은 경우에도 높은 점수를 받을 수 있다. 따라서 알트메트릭은 기존의 전통적 지표를 대체하기보다는 보완적인 도구로 사용되어야 하며, 공식적인 평가지표로 활용되기보다는 연구자나 기관이 연구 성과의 다양한 영향력을 강조하고 홍보하는 수단으로서 그 가치를 찾을 수 있다.

(2) 인공지능 기반 평가

인공지능^{AI} 또한 연구 성과 측정을 위한 새로운 대안으로 떠오르고 있다. AI 기술의 발전은 연구 성과평가에 새로운 가능성을 제시하고 있는데, 2010년대 후반부터 기계 학습과 딥러닝의 비약적인 발전으로 연구 성과평가에도 AI를 접목하려는 시도가 활발해지고 있다. 초기 AI 활용은 주로 예측 모델 개발에 집중되어 있었다. 피인용수는 유용한 성과지표지만, 인용이 축적되기까지 시간이 걸린다는 한계가 있다. 하지만 만약 출판 시점의 논문 정보만으로 미래의 피인용수를 예측할 수 있다면 어떨까? 이러한 아이디어에서 출발한 연구들은 논문의 초록, 본문 텍스트, 저자의 피인용수, 참고문헌 정보를 바탕으로 논문의 피인용 횟수나 등급을 예측하는 AI 모델을 개발했다.

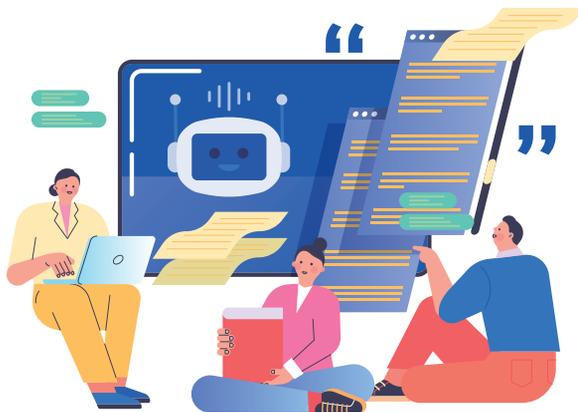
그러나 AI라고 해서 완벽한 것은 아니다. 이러한 AI 모델들의 예측 정확도가 대부분 70%를 겨우 넘는 수준에 그쳤기 때문이다. 영국의 한 연구에서는 2014년부터 2018년 사이에 발표된 수만 건의 논문들을 대상으로, 세 등급으로 구분된 동료 평가 결과를 인공지능 모델이 예측할 수 있는지에 대한 실험이 이루어졌다. 논문의 서지 정보, 저널의 영향력, 논문 텍스트 정보를 총체적으로 활용했음에도, 가장 높은 성능을 보인 화학 분야에서조차 정확도는 72% 수준에 머물렀다. 따라서 이러한 피인용수에 대한 예측 결과는 연구 성과평가에 활용하기에 한계가 분명하다.

따라서 AI 예측 모델은 유망 분야 탐색, 역량 진단, 계획 수립 등 기관의 전략적 연구 개발 의사결정에 참고 자료로 활용 가능하지만, 연구자 업적 평가나 연구비 배정 같은 중요한 의사결정에는 적합하지 않다. 특히 대부분의 AI 모델이 가진 블랙박스 특성으로 인해 예측 결과에 대한 명확한 설명이 어렵다. 이로 인해 합리적이고 공정한 프로세스가 요구되는 연구 성과평가에 직접 활용하기에는 현재로서는 무리다.

최근 생성형 인공지능 기술의 발전으로 ‘AI 에이전트^{agent}’가 연구평가 분야에서 새로운 가능성을 보여주고 있다. AI 에이전트는 인간 대신 자율적으로 의사결정을 할 수 있는 AI 시스템으로, 대규모 언어 모델^{LLM}을 기반으로 외부 데이터 연결과 추가 학습을 통해 개발된다. 이 기술은 이미 소프트웨어 개발이나 특허 명세서 작성 등 전문가 영역에서 큰 성과를 내고 있다.

연구평가 분야에서는 학술지 논문이나 연구 제안서의 우수성을 평가하는 데 AI 에이전트가 활용될 수 있다. 특히, 여러 역할을 가진 복수의 AI 에이전트가 서로 토론하며 평가를 진행하고, 그 과정과 결과가 기록되어 평가의 근거와 이유를 투명하게 확인할 수 있다는 점에서 기존의 블랙박스 AI 모델보다 유용성이 높다고 평가된다.

하지만 현재 AI 에이전트의 평가 결과는 인간 전문가 수준에 미치지 못하며, 편향성과 윤리적 문제도 존재해 동료 평가를 완전히 대체하기는 어렵다. 다만, 특정 분야의 과거 논문 데이터를 추가로 학습하거나 실제 리뷰 결과를 매칭하여 특화된 모델을 개발한다면 AI 에이전트의 평가 수준을 크게 높일 수 있다. 이를 통해 동료 평가의 주관성, 높은 비용, 긴 소요 시간 등의 문제를 해결할 수 있을 것으로 기대된다. 특히, 전 세계적으로 논문 투고량이 급증하면서 심사자 선정과 심사 프로세스에 어려움을 겪는 학술지와 학회에서 AI 에이전트를 초기 스크리닝 도구로 활용하는 방안이 주목받고 있다.





연구 성과평가 제도 국내동향



연구 성과 · 어떻게 평가할 것인가?

연구 성과를 평가하는 작업은 단순히 숫자로 나열할 수 없는 복잡한 과정이다. 그러다 보니 과학기술이 발전하면 발전할수록 연구의 성과를 어떻게 측정할 것인가에 대한 고민은 과거에 비해 훨씬 더 커졌다고 할 수 있다. 일반적으로 연구 성과를 평가하는 방법은 접근 방식에 따라 △프로젝트 단위 평가 △결과물 중심 평가 △기여도 평가 등 크게 세 가지로 구분한다.

프로젝트 단위 평가는 개별 연구 과제의 목표 달성도와 수행 과정의 충실성을 면밀히 분석하는 방식이고, 결과물 중심 평가는 논문이나 특허, 또는 기술 실용화 등 구체적 산출물이 학계와 산업계에 미친 파급효과를 심층적으로 살펴보는 방식이다. 그리고 기여도 평가는 연구자가 축적해 온 학문적 성과를 전체적인 관점에서, 해당 연구자의 학술적 역량과 학문 발전에 대한 기여도를 종합적으로 판단하는 것이다.

가 우리나라 연구 성과평가 방법의 변천 과정

우리나라의 연구 성과평가 시스템은 시대적 요구와 학문적 발전에 발맞추어 끊임없이 변화해왔는데, 그 변화의 흐름은 시기에 따라 크게 세 단계로 나눌 수 있다.

(1) 정량적 평가 중심

1990년대부터 2000년대 초반까지는 연구의 양적 성과에 무게를 둔 평가가 대부분이었다. 예를 들어 SCI(E) 등재 논문 수나 특허 출원 및 등록 실적 등이 핵심 지표로 자리를 잡았다. 그러나 이러한 양적 지표 중심의 평가는 학문의 질적 발전보다 양적 확장에 매달리도록 만드는 부작용을 낳았다. 그 결과 논문 수를 늘리기 위해 질보다 양을 우선시하는 경우도 있었으며, 이는 연구의 본질적 가치를 훼손하는 문제를 발생시켰다.

(2) 정성적 평가 확대

2000년대 중반에 접어들면서 학계에서는 정량적 평가 방식의 문제점이 지적되었고, 그에 대한 해결 방안으로 연구의 실질적 가치를 측정하려는 움직임이 본격화되기 시작했다. 피인용 횟수 같은 영향력 지표의 도입, 연구자가 자신의 대표 업적을 선별하여 제시하는 방식의 채택 등이 그것이다. 이 같은 변화들은 연구의 내실을 다지는 계기가 되었으나, 평가의 객관성 확보라는 새로운 과제를 남겼다. 정성적 평가는 연구의 질적 측면을 강조하지만, 평가자의 주관이 개입될 여지가 크다는 한계를 가지고 있기 때문이다.

(3) 사회적 영향 평가 강조

정량적 평가나 정성적 평가와 달리 최근의 평가 체계는 연구 성과의 사회적 파급효과를 핵심 가치로 삼고 있는 것이 특징이다. 연구가 실제로 사회 문제 해결과 정책 수립에 어떤 기여를 했는지, 또는 기술 사업화를 통해 산업현장에서는 어떤 실질적 변화가 일어났는지가 주요 평가 대상이 되고 있기 때문이다. 2006년부터 과학기술정보통신부와 한국과학기술기획평가원이 주관하는 ‘국가연구개발 우수성과 100선’은 이런 동향을 보여주는 대표적 사례라고 할 수 있다.

그림 1 국가연구개발 우수성과



출처: 한국과학기술기획평가원

표 1 연구 성과평가 방법의 변천과정

| 시기 구분 | 주요 평가 방식 | 세부 평가 지표 | 특징 및 한계점 |
|--------------------------------|-----------------------------------|--------------------------------------|------------------------------------------|
| 정량적 평가 중심 (1999년대~2000년대 초) | - 논문 실적 - 특허 실적 | - SCI(E) 등재 논문 수 - 특허 출원/등록 건수 | - 양적 성장 중심 - 외형적 성과 치중 - 질적 평가 미흡 |
| 정성적 평가 확대 (2000년대 중반 이후) | - 논문 영향력 - 연구 우수성 - 대표업적 평가 | - 피인용 횟수 - h-index - 대표 논문 평가 | - 질적 평가 도입 - 연구 독창성 증시 - 객관성 확보 과제 |
| 사회적 영향 평가 (최근) | - 사회문제 해결 - 정책 기여도 - 기술 사업화 | - 문제해결 효과성 - 정책 반영도 - 경제적 파급효과 | - 실용성 강조 - 사회적 가치 증시 - 종합적 평가 |

출처: 기초과학연구원 이상규

나 분야별 연구 성과평가 방법의 변화

연구 성과평가는 분야별 특성에 따라 다른 접근법이 필요하다. 기초과학은 물론 응용과학과 공학, 그리고 의생명과학 및 인문사회과학처럼 서로 다른 분야의 연구 성과평가 방법의 특징과 변화과정을 정확하게 진단하는 것이 평가 시스템 발전에 도움이 될 수 있다.

(1) 기초과학 분야

과거 기초과학 분야의 연구 성과평가는 논문 수 및 인용 횟수와 같은 정량적 지표에 크게 의존했다. 이러한 접근법은 연구자의 생산성과 영향력을 객관적으로 측정할 수 있다는 장점이

있었으나, 연구의 질적 가치와 혁신성을 온전히 담아내지 못한다는 비판이 있었다. 이를 극복하기 위해 연구의 혁신성, 학문적 기여도, 사회적 영향력 등을 아우르는 정성적 평가 체계가 도입되었다. 최근에는 여기서 한 발 더 나아가 학제 간 융합과 사회적 가치 창출이라는 새로운 평가 기준이 주목을 받고 있다.

(2) 응용과학 및 공학 분야

응용과학 및 공학 분야는 처음부터 실용성에 중점을 두어, 기술 개발 실적이나 특허 관련 지표(인용 분석, 지역적 확장성, 유지 기간) 등을 핵심적인 평가 기준으로 삼았다. 이러한 평가 시스템은 시간이 흐르면서 기술 이전과 상용화 성과라는, 보다 실질적인 지표들을 포함하고 기술의 혁신성을 정성적으로 평가하는 방향으로 진화했다. 연구 결과의 산업적 활용도와 경제적 파급효과를 중시하는 흐름을 반영한 것인데, 최근 들어서는 이런 산업적 활용도와 파급효과에 더해 지속 가능성과 사회적 책임이라는 시대적 요구가 새로운 평가 기준으로 대두되고 있다. 예를 들어 친환경 기술이나 사회 문제 해결형 연구 등이 더욱 각광받고 있는 것이다.

(3) 의생명과학 분야

의생명과학 분야 역시 과거에는 논문의 양적 수준 외에 임상 연구의 결과물을 토대로 연구 성과를 평가했다. 그러나 시간이 지나면서 연구의 실제 임상 적용 가능성과 새로운 치료법 개발이라는 실용적 측면이 더욱 부각되면서, 의료 현장에서의 실질적 기여도가 핵심적인 평가 요소로 자리를 잡았다. 최근 들어서는 의료 패러다임이 환자 중심으로 전환되고 예방 의학의 중요성이 강조되면서, 환자의 삶의 질 향상과 질병 예방에 초점을 맞춘 연구들이 높은 평가를 받고 있다.

(4) 인문사회과학 분야

인문사회과학 분야의 연구 성과평가는 저서와 학술 논문의 출판이라는 전통적 기준에서 출발했다. 이러한 평가 체계는 점차 연구의 학문적 파급력을 측정하는 인용 지표와 학술상 수상 실적 등을 포함하는 보다 포괄적인 형태로 발전했다. 그러나 최근들어 연구의 사회적 영향력과 공공의 참여도가 새로운 평가 기준으로 부상하고 있으며, 특히 정책 수립과 사회 문제 해결에 기여하는 실용적 연구의 가치가 새롭게 조명받고 있다.

표 2 연구 성과평가 방법의 분야별 특성

| 학문 분야 | 초기 | 중기 | 최근 |
|---------|--------------------------------------|--------------------------------------|--------------------------------------|
| 기초과학 | - Impact Factor - 논문 수 - 인용 횟수 | - 연구 혁신성 - 학문적(수) 기여도 - 독창적 평가 | - 학제간 융합 - 사회적 가치 - 지속 가능성 |
| 응용과학/공학 | - 기술개발 실적 - 특허 지표 | - 기술이전 실적 - 상용화 성과 - 산업 활용도 | - 지속가능 기술 - 사회적 책임 - 친환경성 |
| 의생명과학 | - 임상연구 결과 - 논문 실적 | - 임상 적용성 - 새로운 치료법 - 의료현장 기여 | - 환자 중심 연구 - 삶의 질 향상 - 예방의학 중시 |
| 인문사회과학 | - 저서 출판 - 학술 논문 | - 학술적 파급력 - 연구 독창성 - 학술상 수상 | - 정책 기여도 - 사회문제 해결 - 공공 소통성 |

출처: 기초과학연구원 이상규

다 연구자의 경력 단계에 따른 학문적 우수성 평가

연구자의 경력 단계별 평가는 각 성장 국면의 특성을 반영한 차별화된 접근을 필요로 한다. 국내 학계는 이를 △신진 △중견 △시니어 연구자 등으로 구분하여, 각 단계에 적합한 평가 체계를 운영하고 있다.

(1) 신진 연구자

신진 연구자에 대한 평가는 현재의 성과보다 미래의 성장 잠재력에 초점을 맞추고 있다. 연구 주제의 독창성과 혁신적 접근 방식, 그리고 기초 연구 능력의 숙달도 및 학술 네트워크 형성 능력 등이 주요 평가 대상이 된다. 이 단계의 평가는 주로 지도 교수나 선임 연구자의 멘토링 및 피드백, 또는 연구 계획서의 타당성 검토나 학술 발표에서의 수행 능력 등을 통해 파악할 수 있다.

(2) 중견 연구자

중견 연구자 단계에서는 실질적 연구 성과와 학술적 영향력이 평가의 중심축을 이룬다. 주요 학술지 논문 실적과 인용도, 그리고 연구비 수주 실적, 프로젝트 운영 능력, 후학 양성 기여도, 학회 활동 참여도 등이 종합적으로 고려된다. 이러한 평가는 정기적인 연구 업적

심사와 동료 연구자들의 전문적 평가, 연구 프로젝트 수행 이력 분석 등을 통해 진단할 수 있다.

(3) 시니어 연구자

시니어 연구자에 대한 평가는 학문적 리더십과 사회적 영향력이 판단의 기준이 된다. 해당 분야에서의 학술적 권위는 물론, 연구소나 학과 운영을 통한 기관 발전 기여도, 정부 및 산업계 자문 실적, 국제 학술 네트워크 구축 성과 등이 핵심 평가 요소라 할 수 있다. 이들의 공헌도는 국내외 학술상 수상 경력, 국가 정책 수립 참여 실적, 국제 학술지 편집 위원 활동이나 학회 임원 경력 등을 통해 평가를 받는다.

라 현재 연구 성과평가 기준 및 한계

(1) 현재의 평가 기준

오늘날 우리나라의 연구 성과평가는 정량적 지표와 정성적 지표가 유기적으로 결합된 복합적 체계를 이루고 있다. 저널 영향력 지수와 논문 실적 및 피인용 지수, 그리고 특허 성과, 기술료 수입 등의 객관적 지표가 대표적이다. 그리고 이와 함께, 연구의 혁신성과 사회적 기여도, 정책 활용도, 기술 사업화 실적 등 질적 측면을 아우르는 다면적 평가가 이루어지고 있다.

(2) 평가 체계의 한계

현행 평가 체계는 과거에 비해 많이 개선되었음에도 불구하고 여전히 몇 가지 근본적 한계를 안고 있는 것이 사실이다. 정량적 지표는 연구의 실질적 가치와 영향력을 온전히 담아내지 못하는 태생적 한계를 지니고 있고, 정성적 평가는 평가자의 주관이 개입될 여지를 완전히 배제하기 어렵다는 단점을 갖고 있다. 또한 연구 성과의 사회적 가치를 객관적으로 측정하고 계량화하는 문제 역시 여전히 해결해야 할 과제로 남아있다.



연구 성과평가 제도에 대한 젊은 과학자들의 생각과 제안



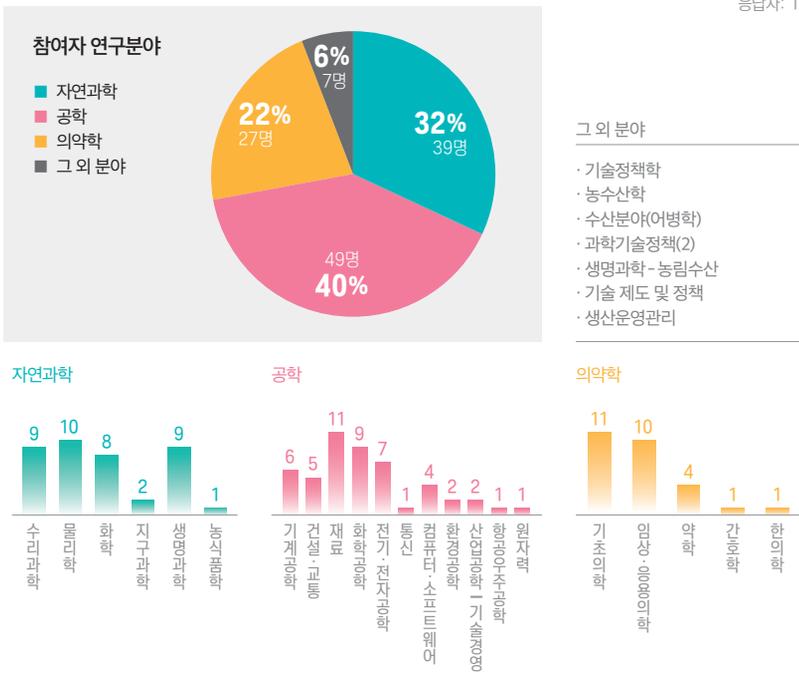
연구 성과 · 어떻게 평가할 것인가?

현재의 연구 성과평가 제도는 많은 노력에도 불구하고 풀기 어려운 점들이 존재하고 때로는 불합리한 문제가 발생하기도 한다. 또한 같은 논점에 대해서도 분야와 상황에 따른 시각의 차이가 나타날 가능성도 충분히 있다. 이에, 본 차세대리포트에서는 실제 연구 현장에서 우수한 연구를 수행하고 있는 젊은 과학자들의 생각을 알기 위해 다음과 같이 설문조사를 수행했다.

설문 대상은 한국차세대과학기술한림원 회원 및 동문회원이었으며, 총 122명이 설문에 응하였다. 첫 번째 질문은 ‘귀하의 연구 분야는 어느 쪽에 가장 가깝습니까?’였다. 참여자들은 크게 공학(49명, 40%), 자연과학(39명, 32%), 의약학(27명, 22%) 및 그 외(기술정책, 농림, 수산분야 등 7명, 6%) 분야에 종사하고 있으며, 이를 통해 다양한 분야의 입장을 확인할 수 있었다.

설문 1 **귀하의 연구분야는 어느 쪽에 가장 가깝습니까?**

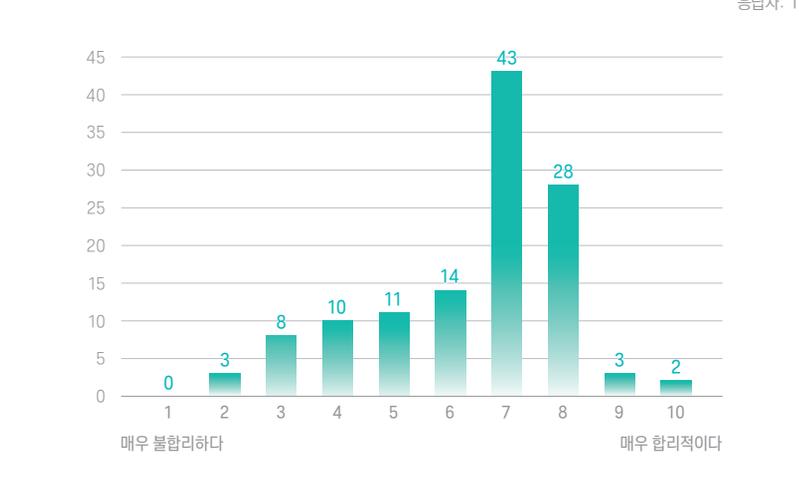
응답자: 122명



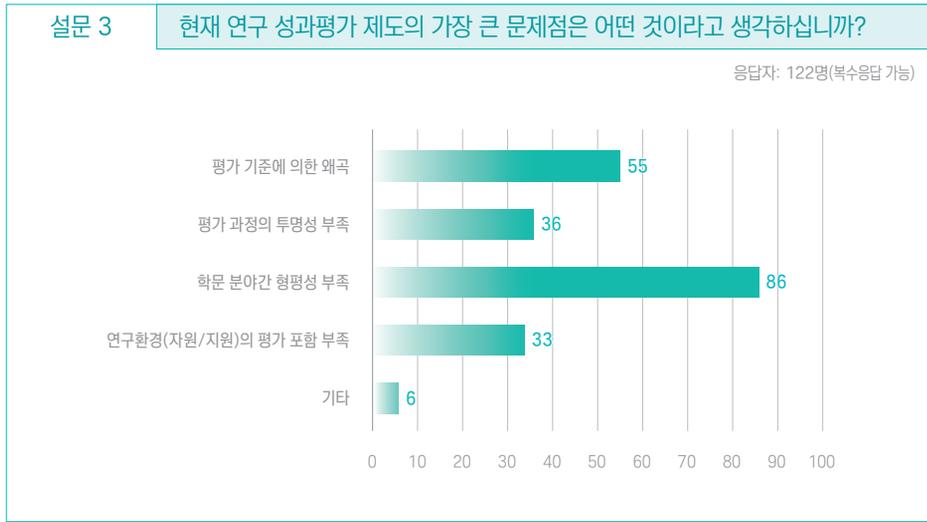
이어진 설문조사는 총 7문항으로 구성되었으며, 전반적인 연구평가에 대한 신뢰도, 문제점, 그리고 이를 해결하기 위한 구체적인 방안을 조사했다.

설문 2 **현재 우리나라에서 연구 성과의 우수성을 평가하는 방법과 체계는 어느 정도로 합리적이라고 생각하십니까?**

응답자: 122명



응답자들은 국내 연구 성과평가 제도에 대한 신뢰 점수를 10점 만점 기준 약 7점 정도에 해당한다고 느끼고 있었다. 다만 불합리성에 대해서는 매우 불합리하다(2점)가 3명에 그쳤지만 보통 정도(5-6점)에도 다수의 응답이 있었기 때문에 응답자별로 느끼는 체감의 정도는 차이가 있는 것으로 나타났다.

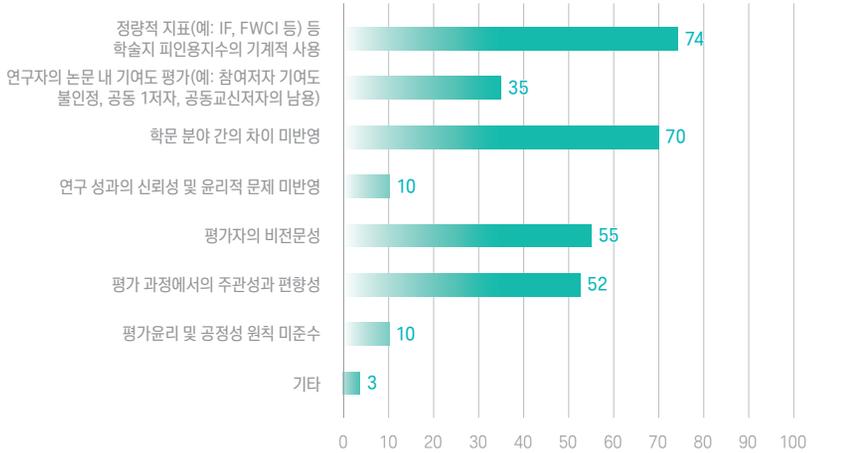


현재 연구 성과평가 제도의 현재 가장 큰 문제점을 묻는 설문에서 응답자들은 학문 분야간 형평성 부족을 가장 중요한 이슈로 꼽았다. 이는 앞서 살펴본 것처럼 연구의 우수성을 정성적 지표와 정량적 지표로 평가하는 방식이 동일 분야 내에서는 어느 정도 조정과 논의가 가능하지만, 연구 분야가 달라지면 문제가 커진다는 점과 일치한다. 또한 평가지표 자체가 연구평가를 왜곡하는 문제도 크게 체감하고 있었다. 반면에 평가 과정의 투명성에 대한 문제 인식은 상대적으로 낮았다.

설문 4

연구논문의 우수성을 평가하는 데에 가장 시급히 개선되어야 할 부분은 무엇이라고 생각하십니까?

응답자: 122명(복수응답 가능)

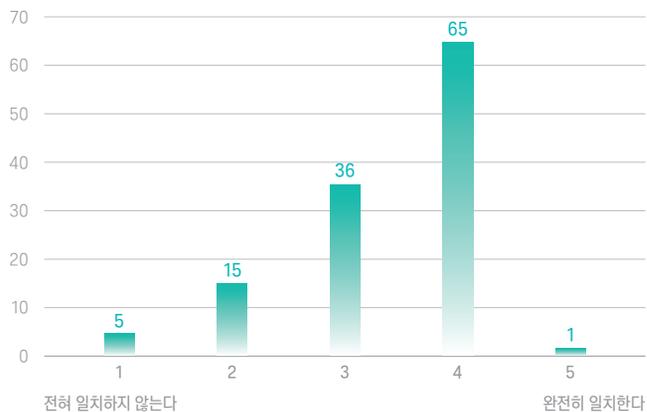


논문의 우수성을 평가하는 데 가장 큰 걸림돌은 IF, FWCI 등 정량적 지표를 기계적으로 적용하는 것이라는 의견이 지배적이었다. 또한, 이러한 지표가 학문 분야별 차이를 반영하지 못한다는 점도 주요한 문제로 지적되었는데, 평가자의 비전문성과 편향성은 개별 논문 평가보다는 연구자 우수성 평가에서 더 큰 문제로 나타났다. 반면, 연구 결과의 신뢰성이나 연구 윤리성에 대한 문제의식은 비교적 낮은 수준인 것으로 드러났다.

설문 5

학술 커뮤니티에서 저널의 권위는 해당 저널의 IF와 얼마나 일치한다고 생각하십니까?

응답자: 122명

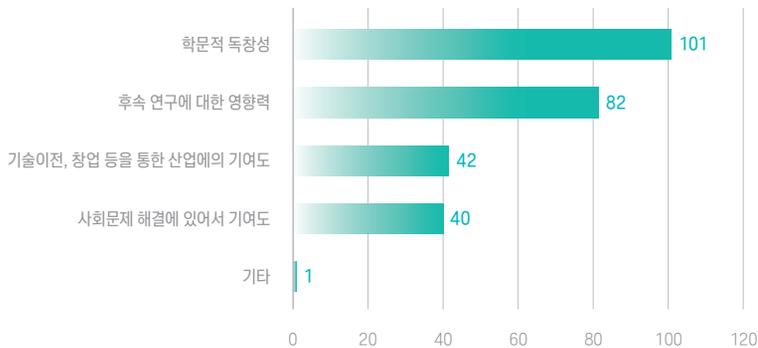


저널의 ‘영향력 지수¹⁾’와 저널의 연구 성과의 상관관계를 묻는 질문에서는 예상외로 저널의 IF를 저널의 우수성을 반영하는 지표로 인식하고 있었다. 이는 비록 저널의 IF가 특정한 원인으로 높아지더라도, 높은 IF에 게재하려는 쏠림현상에 의하여 경쟁이 치열해지고, 그 결과 보다 높은 수준의 논문이 게재되는 현상이 어느 정도 반영된 것으로 보인다. 그럼에도 불구하고 완전히 일치한다는 답변은 1명에 불과하여, 다양한 보조지표 개발 혹은 정성적 평가가 동반되어야 함을 재확인할 수 있었다.

설문 6

연구 성과의 우수성 평가에 있어서 가장 중요한 요소가 무엇이라고 생각하십니까?

응답자: 122명(복수응답 가능)



어떠한 연구가 우수한 연구임을 묻는 설문에서는 독창성과 영향력, 그리고 산업 및 사회에 대한 기여도 순으로 응답했다. 앞에서 서술된 바와 같이 연구의 우수성은 게재 시점에 서 판단될 수 있는 것과 게재된 이후에 중장기적으로 평가될 수 있는 부분으로 구분되는데, 게재 시점에서 바로 판단될 수 있는 독창성이 연구의 우수성을 즉시 평가하는 데에 유용하기 때문인 것으로 보인다.

설문 7

연구 성과의 평가 과정에서 실제로 겪거나 목격한 불합리한 과정이 있으면 적어주십시오.

응답자: 58명

1. 평가 기준 및 지표의 문제점

- 저널 Impact Factor(IF)에 의존한 기계적인 평가 방식의 부작용 (특정 시기에 특정 분야가 과대평가됨)
- 연구자의 기여도 평가에서 공동 주저자(1저자, 교신저자)가 남용되는 사례
- 논문 수(IF, FWCI 등)만을 기준으로 한 정량적 평가의 부적절성
- 마이너한 연구 분야가 상대적으로 저평가되는 문제
- 연구비 대비 성과평가보다는 단순한 산출물(논문 수)만 강조하는 경향

2. 평가 과정에서의 문제점

- 평가자의 전문성이 부족하여 잘못된 평가가 이루어지는 사례
- 평가위원 선정 과정에서의 불공정성과 인맥 및 정치적 개입
- 연구자의 소속이나 특정 학문 분야에 따른 배척 현상
- 평가자의 주관적 견해가 평가 결과에 영향을 미치는 문제
- 특정 연구실 출신 평가위원이 같은 연구실 연구자를 평가하는 사례
- 평가 과정에서 목소리가 큰 평가자의 의견이 지나치게 반영되는 문제

3. 연구 환경 및 제도적 문제점

- 수도권 대학에 대한 역차별 문제
- 연구비 삭감이 평가 없이 일방적으로 이루어지는 문제
- 융합 연구 분야에서 평가자의 전문성이 부족하여 부적절한 평가가 이루어지는 경우
- 협력 연구를 장려하기보다 평가 기준이 이를 저해하는 방향으로 작용하는 문제

연구자들이 실제로 체감한 연구 성과평가의 불합리한 사례에 대해서 묻는 질문에는 응답자의 약 절반인 58명이 답변을 제시하였다. 매우 다양한 사례들이 수집되었는데, 요약하면 크게 ① 평가 기준 및 지표의 문제점 ② 평가 과정에서의 문제점 ③ 연구 환경 및 제도에 대한 문제점으로 나눌 수 있었다.

그중에서도 가장 많은 사례는 분야별 정량지표의 차이와 연구에서의 기여도에 집중되었다. 예를 들면 특정 분야, 및 특정 시기에 급격한 상승을 보인 저널의 정량지표에 의한 성과의 과대 평가가 불합리한 문제로 지목되었고, 유행을 타지 않는 분야에서 정량적 지표만을 적용하는 까닭으로 피해를 입었다는 내용도 다수 응답되었다. 또한 연구에 기여했음에도 불구하고 주저자(1저자 및 교신저자)가 아니라는 이유로 그 성과가 전혀 인정받지 못하는 현실에 대해서도 여러 응답자가 지적하였다.

설문 8

답변에 기반한 귀하가 생각하는 평가 제도 개선안을 부탁드립니다.

응답자: 71명

1. 평가 기준의 다양화 및 정성 평가 강화

- 저널 IF뿐만 아니라 h-index, Total Citation, FWCI 등 다양한 지표를 고려
- 연구의 독창성, 사회적 파급력, 장기적인 피인용 지수를 반영한 평가 체계 구축
- 각 학문 분야의 특성을 반영한 평가 기준 마련 (예: 순수수학, 컴퓨터 과학, 인문학 등)
- 논문 게재 후 최소 3년 이상의 피인용 지수를 반영하는 장기적 평가 체계 도입
- 학회 발표 논문(특히 컴퓨터 과학 분야)의 중요성을 인정하는 방향으로 제도 개선

2. 공정한 평가 프로세스 및 심사위원 개선

- 심사위원의 전문성 강화 및 피드백 평가 제도 도입
- 평가위원 선정의 공정성 확보, 특정 연구실 및 인맥 영향력 축소
- 블라인드 평가 시스템 도입(저자, 소속 정보 제외)
- 해외 전문가 풀을 활용한 객관적인 평가 보완
- 심사위원의 평가에 대한 후속 피드백을 제공하여 지속적인 질적 향상 유도
- 평가 책임제 도입: 대형 과제외의 경우 동일한 심사위원장이 지속적으로 평가

3. 연구 환경 개선 및 도전적 연구 장려

- 좋은 연구가 실패하더라도 기획 및 평가가 가능하도록 지원
- 연구 성과의 평가에서 연구비 대비 성과를 반영하는 체계적 접근 필요
- 연구자가 IF가 낮더라도 우수한 저널에 게재한 경우 해명할 수 있는 항목 추가
- 지역 및 기관별 연구 평가 기준의 차이를 줄이기 위한 공식적 전문가 합의 도출
- 연구자가 평가를 위한 충분한 시간을 가질 수 있도록 평가 수당 증액 및 평가 책임제 도입
- 학문 분야별 전문가 집단을 활용하여 우수 연구를 공정하게 평가

4. 학문 분야별 맞춤 평가 체계 구축

- 분야별 특성을 반영한 맞춤형 평가 기준 마련
- 기초과학: 논문의 창의성 및 독창성 평가
- 산업화 연구: 실용성 및 산업화 가능성 반영
- 융합 연구: 기존 평가 방식의 한계를 고려하여 유연한 평가 적용
- 약탈적 저널이 IF가 높다고 단순 정량평가에서 유리하지 않도록 정성적 평가 강화
- 전공 분야별 차등 평가를 적용하고, 각 연구 분야 커뮤니티의 의견을 존중
- 사회적 기여도 평가지표 추가 (지역사회 발전, 성평등, 지속가능성 등)

본 설문 의 가장 큰 목적이라고 생각할 수 있는 개선안에 대한 제언에는 주관식 답변임에도 불구하고 약 60%에 달하는 71명이 답변하여 연구평가 제도 개선에 대한 강한 필요성을 확인할 수 있었다. 이를 요약하면 △평가 기준의 다양화 및 정성평가 강화 △공정한 평가 프로세스 및 심사위원 개선 △연구환경 개선 및 도전적 연구 장려 △학문 분야별 맞춤 평가 체제 구축으로 나누어졌다.

이는 그동안 알아본 다양한 관점에서의 결론과 많은 부분이 일맥상통했다. 특히 연구 자체의 우수성에 대해서 결국 새로운 지표 개발은 물론, 정성적 평가 및 시차를 두고 중장기적으로 평가하는 부분과 그 영향력을 평가하는 것이 필요하다는 사실을 많은 연구자들도 느끼고 있음을 알 수 있었다. 또한 앞선 설문에서 가장 큰 문제로 언급되었던 학문 분야간 형평성에 대해서도 해당 학문 분야에서 중요하게 여겨지는 부분을 고려하는 것이 시급하다고 응답하였다.

연구 성과평가 제도 해외동향



최근 연구 환경과 문화가 변화하면서, 연구자들은 연구의 질을 높이고 더 나은 연구 환경을 만들기 위해 연구 평가 방식이 개혁될 필요가 있다고 인식하고 있다. 특히 ‘논문을 내지 않으면 도태된다(publish or perish)’는 문화 속에서 논문 수나 인용지수 같은 지표가 지나치게 강조되면서, 연구의 다양한 가치를 제대로 반영하지 못한다는 문제가 제기되고 있다.

예를 들어, 단순히 논문의 영향력을 수치화하는 지표만으로는 해당 연구가 지역사회에 미친 긍정적인 영향이나, 연구 데이터의 공개를 통해 학문 발전에 기여한 정도를 평가하기 어렵다. 이처럼 출판물 중심의 평가 방식은 연구자들이 논문의 질보다 양을 중시하도록 만들고, 신뢰할 수 없는 학술지나 학술대회의 난립을 초래할 수 있다.

또한, 혁신적인 연구보다는 안전한 연구를 선호하게 만들고, 부정적인 연구 결과는 보고되지 않으며, 과도한 경쟁 속에서 연구자 간 협업이 줄어드는 등의 부작용도 발생할 수 있다. 이런 문제들을 해결하기 위해서는 연구평가 방식이 보다 균형 잡힌 시각을 가질 필요가 있다.

가 연구평가 개혁의 배경과 대표적인 글로벌 이니셔티브

이러한 문제를 해결하기 위해 다양한 국제적 이니셔티브(global initiative)가 등장했는데, 그 중 하나가 ‘샌프란시스코 연구평가 선언(San Francisco Declaration on Research Assessment, DORA)’이다. DORA는 연구 평가 관행을 개선하기 위해 2013년 첫 번째로 마련된 지침으로서, 연구 성과를 평가할 때 정량적 지표에만 의존하지 말고 연구의 질적 성과와 사회적 기여를 종합적으로 평가해야 한다는 내용이 담겨있다.

이 외에도 2015년에는 ‘라이덴 선언(Leiden Manifesto)’이 발표되었는데, 이 선언은 연구지표의 책임 있는 사용을 위한 10가지 원칙을 제시하면서 정량적 평가는 전문가의 정성적 평가를 대체하는 것이 아니라 보완해야 한다는 점을 강조하고 있다. 또한 라이덴 선언은 연구 지표의 투명성, 다양성, 맥락화의 중요성을 강조하고 있으며, 연구자들이 보다 책임감 있게 연구 성과를 평가받을 수 있도록 환경 조성에 기여한다는 목표를 추구하고 있다.

그로부터 4년 뒤인 2019년에는 제6회 세계 연구윤리 회의에서 ‘홍콩 원칙(Hong Kong Principles)’이 공식화되었다. 이 원칙은 다양한 연구 성과를 인정하고 이를 연구평가에 반영할 것을 강조하고 있다. 홍콩 원칙은 연구자들이 연구 결과의 질을 높이기 위해 지켜야 할 여러 가지 규칙과 원칙을 제시하며, 연구자들의 윤리적 행동을 촉진하고자 만들어졌다.

이러한 국제적 지침들은 연구자에게 보다 나은 연구 환경을 제공하고, 연구의 질적 향상을 도모하고자 하는 의지를 담고 있다. 그러나 이러한 노력들은 국가와 조직에 따라 파편적으로 이루어져 온 것이 사실이다. 또한 연구 성과평가에 관심을 가지고 있는 나라들은 자국의 연구 환경에 맞춰 연구 성과평가 시스템을 발전시켜 나가고 있지만, 전체적인 방향성과 통일성이 부족하다. 따라서 성과를 평가하는 관행을 체계적으로 개혁하기 위한 공동의 노력이 필요하다.

이러한 상황에서 ‘연구 성과평가 시스템의 개혁을 위한 협약(Agreement on Reforming Research Assessment, RRA)’이 발표되어 주목을 끌었다. 40여 개국 350개 이상의 조직이 동의한 RRA의 최종 협약은 지난 2022년에 발표되었는데, RRA에 서명한 국가와 조직은 협약 체결 후 1년 이내에 정해진 실행 계획에 따라 연구평가 개혁을 이행하고, 5년 이내에 이행 상황을 공유해야만 한다. 이를 통해 협의된 시간 내에 가시적인 변화를 실행하고, 모범사례와 실행 함의점을 공유함으로써 연구 성과에 대한 평가 시스템의 개혁을 촉진하게 된다.

그림 2 연구평가 관행 개혁을 위한 글로벌 이니셔티브



나 연구 성과평가 개혁을 위한 RRA의 원칙 및 약속

RRA는 연구, 연구자, 연구 기관의 평가 방식을 개선하여 연구의 본질적 가치를 존중하고, 그 영향력을 극대화하는 것을 목표로 한다. 이를 위해 RRA는 동료 평가를 중심으로 하되, 정량지표를 책임 있게 활용하는 정성적 평가를 기본 원칙으로 삼는다.

RRA의 평가 대상은 연구 기관, 연구 과제, 개별 연구자 및 연구팀 등으로 광범위하다. 이러한 원칙은 연구비 배분, 과제 선정, 포상, 채용 및 승진 등 다양한 결정 과정에 활용될 수 있다. 물론, 국가 수준의 연구 성과평가나 연구 이외의 활동에 대한 기관 평가는 RRA의 직접적인 대상은 아니다. 그러나 RRA의 원칙은 이러한 영역에도 확장 적용될 수 있다.

또한 RRA는 연구 평가의 공정성과 투명성을 높이기 위해 4가지 기본 대원칙을 제시하고 있는데 △윤리 및 청렴의 최우선 △과학적 연구의 자유 보장 △연구기관 자율성 존중 △데이터와 인프라의 독립성과 투명성 보장이 그것이다.

첫 번째 원칙은 연구 성과에 대한 평가 전후로 최고 수준의 윤리 기준을 충족하는지 확인하며, 편향된 평가를 방지하기 위해 방법론의 엄격성과 도덕적 행동을 강조하는 것을 말한다. 또한, 연구 데이터와 결과의 조기 공유 및 외부 검증을 통해 과학적 완결성을 추구하는 것도 포함된다.

두 번째 원칙은 연구자가 선택한 주제, 이론, 방법론에 제한을 두지 않고, 평가 결과가 연구자와 기관에 실질적으로 활용하도록 하는 것이고, 세 번째 원칙은 평가의 독립성을 유지하고, 연구자 이동 등으로 인한 평가의 파편화를 방지하는 것을 뜻한다.

마지막으로 네 번째 원칙은 데이터 수집과 평가 기준이 명확하고 공개적으로 이루어지며, 피평가자가 평가 데이터와 분석 결과에 접근할 수 있도록 하는 것을 가리킨다. 이 원칙들은 연구평가의 신뢰성을 높이고, 연구자와 기관의 자율성을 보호하며, 과학적 자유를 존중하는데 기여한다. 이를 통해 RRA는 연구의 질과 영향력을 극대화하고, 연구 생태계의 건강한 발전을 도모하는데 적용된다.

4개의 기본 대원칙 외에 평가 기준 및 절차에 대한 원칙은 6개로 구성되는데, 연구의 질적 성과 및 영향력, 그리고 다양성과 포용성을 중심으로 하는 것이다. 첫 번째 원칙은 연구 성과에 대한 평가의 경우 아이디어의 독창성, 연구 수행의 전문성, 최신 기술을 뛰어넘는 결과와 같은 연구의 질적 성과를 기본적으로 고려해야 한다는 점이다. 특히 연구평가는 동료 심사가 중심이 되는 정성적 평가에 기반해야 하며, 정량적 지표는 필요한 경우 보조적으로 활용되어야 한다는 내용을 담고 있다.

두 번째 원칙은 지식 발전에 대한 기여와 연구 결과의 잠재적 영향을 고려하는 것이다. 연구 결과의 영향력은 과학적, 기술적, 경제적, 사회적인 측면에서의 장기적인 영향과 단기적 영향을 포함하는데, 이는 학문 분야와 연구 유형에 따라 다를 수밖에 없다.

이어서 세 번째 원칙은 연구 활동, 관행, 결과물의 다양성을 존중하고, 조기 공유와 개방적 협업을 장려하는 것을 의미한다. 동료 심사, 교육, 박사과정 학생 멘토링, 리더십 역할 등 다양한 활동을 평가에 포함하며, 필요에 따라 과학 커뮤니케이션, 홍보, 기업이 정신, 지식 가치화, 산학협력 등도 고려한다. 또한, 과학 출판물, 데이터, 소프트웨어, 모델, 방법론, 이론, 알고리즘, 프로토콜, 워크플로우, 전시물, 정책 기여 등 모든 연구 결과물을 평가 대상으로 삼는다. RRA는 연구자가 모든 분야에서 탁월할 수 없다는 점을 인정하며, 개별 연구자의

목표와 특성에 맞춰 평가할 수 있는 유연한 프레임워크를 제공한다. 그리고 이를 통해 연구의 다양성과 창의성을 극대화하고, 개방적이고 협력적인 연구 문화를 조성하는 것도 포함한다.

또한 네 번째 원칙은 다양한 학문 분야, 연구 유형, 연구자 경력 단계를 고려한 평가 기준과 절차를 적용하는 것이다. 기초 연구, 프론티어 연구, 응용 연구 등 다양한 연구 유형을 포괄하며, 주니어와 시니어 연구자의 경력 단계에 맞춰 차별화된 평가를 진행한다. 또한, 학제 간이나 분야 간 융합 연구와 같은 접근 방식도 평가에 반영한다. 연구평가는 과학 분야의 특성, 연구 목표, 과학적 노력의 다양성을 고려하여 유연하게 설계되고 이를 통해 각 분야와 연구자의 고유한 가치를 존중하며, 융합적이고 혁신적인 연구를 촉진하고, 공정하며 균형 잡힌 평가 체계를 구축하는 것을 제시한다.

다음으로 다섯 번째 원칙은 연구 이외의 활동을 포함하여 연구자의 역할과 경력의 다양성을 인정하고 존중하는 것이다. 오픈 사이언스 기술을 포함한 개별 연구자 기술, 역량 및 강점, 그리고 협업을 고려해야만 가능하다.

마지막 여섯 번째 원칙은 성평등 및 공정성, 그리고 포용성을 보장하는 것을 말한다. 성별 균형, 젠더, 인종, 민족, 성적 지향, 사회경제적 위치, 장애 등 광범위한 측면에서의 다양성을 고려하는 것이다.

한편, 약속(commitments)의 경우도 앞에서 소개한 원칙들처럼 여러 개의 약속들로 구성되어 있는데, 4가지 핵심 약속(다양성 관련 2가지, 부적절한 지표 활용 관련 2가지) 및 6가지 부가적인 보조 약속(새로운 연구 평가 기준, 도구, 절차에 관련된 3가지, 상호 학습, 모니터링, 증거 기반 관련 3가지) 등 총 10가지 약속이 그것이다.

첫 번째 핵심 약속은 연구의 목적과 특성에 따라 연구 기여와 연구자 경력의 다양성을 인정하는 것이다. 학술 논문 외에도 동료 평가, 팀워크, 협업 활동, 연구의 견고성, 개방성, 투명성, 포용성에 기여하는 다양한 활동을 평가하도록 설계되었다. 또한 언어와 상관없이 과학적·사회적 이익을 창출하는 다양한 결과물도 고려한다. 이 외에도 데이터 관리자, 소프트웨어 엔지니어, 데이터 과학자, 기술자, 홍보, 과학 자문, 커뮤니케이션 등 연구 과정에서의 다양한 역할과 경력을 인정하고 평가할 수 있다.

두 번째 핵심 약속은 동료 심사를 중심으로 하는 정성적 평가를 기반으로 하고, 정량적 지표는 보조적으로 활용하는 것이다. 동료 심사는 투명성, 공정성, 적절성, 기밀성, 무결성, 성평등, 다양성 등의 윤리적 원칙을 충족하도록 설계된 대표적인 정성적 평가 방법이다. 편견 등 동료 평가의 문제를 줄이기 위해 평가 관행을 정기적으로 점검하고 개선해야 하는데, 동료 심사에 기반한 정성적 평가는 연구자의 추가 노력을 필요로 할 수 있으며, 이는 승진 등에서 평가 대상으로 인정되어야 한다. 또한, 동료 평가와 함께 질적 평가를 위한 새로운 기준, 도구, 절차를 개발하거나 개선하여 활용할 수 있다.

세 번째 핵심 약속은 학술지 및 출판물 기반 지표의 부적절한 사용과 과도한 의존을 방지하는 것이다. 대표적 지표로는 학술지 영향력 지수^{IF}나 Article Influence Score^{AFIS}, 또는 h-index 등이 있다. 논문 수와 특허 수, 그리고 인용 수 등 연구자 지표에만 의존한 평가나 출판물 형식 및 언어와 관련된 지표에 의존한 평가, 그리고 기타 연구의 질적 성과와 영향력을 제대로 반영하지 못하는 지표에 의존하는 평가가 여기에 해당된다.

네 번째 핵심 약속은 연구 성과를 평가하는 데 있어서 글로벌 랭킹과 같은 순위 지표를 사용하지 않는 것이다. 현재 연구 기관 평가에 사용되는 글로벌 순위 평가는 공정하고 책임 있는 방식으로 이루어지고 있지 않기 때문에, 이러한 순위에 대한 홍보는 연구의 질적 성과를 순위와 동일시하는 잘못된 인식을 퍼뜨릴 수 있다. 따라서 가장 좋은 평가는 외부 기준과 방법론에 의존하지 않고 해당 연구기관이 어떠한 데이터와 순위 평가 방법론을 사용할지를 자율적으로 결정하여 평가하는 것이다. 다만 국가 또는 기관의 벤치마킹과 같은 순위 평가 결과의 활용이 불가피한 경우, 방법론의 한계를 인식하고 활용해야만 약속을 지키는 데 도움이 될 수 있다.

다섯 번째 약속부터는 보조 약속인데, 합의된 기간 내 연구평가 관행을 개선할 수 있도록 예산이나 인력 등의 필요한 자원을 투입하는 것을 의미한다. 이는 연구평가 관행을 개혁하기 위한 필수 조건으로, 여기에는 모니터링, 인식 제고, 교육, 데이터 수집 및 처리 도구와 같은 인프라 지원에 필요한 자원이 포함되고, 모든 경력 단계의 연구자 지원이 이루어져야 한다.

여섯 번째 약속은 연구평가 기준, 도구 및 절차를 검토하고 개발하는 것이다. 이 기준은 크게 대학, 연구센터, 연구 인프라를 대상으로 하는 기관의 평가 기준과 연구자 및 연구 프로젝트를 위한 기준으로 나뉜다. 두 기준 모두 학문 분야별, 경력 단계별 연구자, 피평가자와

평가자 모두가 참여하여 개발해야 하며, 연구의 질적 성과에 기여하는 다양한 활동을 고려해야 한다.

일곱 번째 약속은 연구평가 개혁에 대한 인식을 제고하고, 평가 결과 및 절차, 활용에 대한 투명한 정보, 지침 및 교육을 제공하는 것이다. 모든 경력 단계의 연구자와 이해관계자의 인식 제고를 위해 관련 조직 및 기관은 평가 기준, 도구, 및 절차에 대한 정보를 투명하게 제공하고, 지침을 마련하여 공개해야 하며, 활용 방안에 대해서도 교육해야 한다. 또한 피평가자에게는 평가 기준과 평가에 활용된 정보를 기밀이 유지되는 범위 내에서 접근할 수 있도록 허용해야 한다.

여덟 번째 약속은 협약에 서명한 연합 회원 내외부의 상호 학습을 위해 경험과 정보를 공유하는 것이다. 개혁에 대한 조직의 자율성을 보장하면서도, 경험 공유와 연구자 교류를 통해 서로 배우고 조직 간 평가 관행의 일관성을 제고해야 한다.

아홉 번째 약속은 원칙 준수 및 약속 이행에 대한 진행 상황을 공유하는 것이다. 조직 내 및 조직 간, 그리고 조직과 연구 커뮤니티 간 성공 사례 및 어려움 등의 진행 상황을 서로 공유함으로써 자가 점검을 촉진하고 이행 상황을 상호 모니터링할 수 있다.

마지막 열 번째 약속은 근거 기반 평가를 위해 명확한 근거와 최신 기법을 바탕으로 연구 관행, 기준, 도구를 평가하고, 데이터를 공개적으로 제공해 근거 마련과 연구를 제공하는 것이다. 기존 학술지 및 출판물 지표 평가의 한계가 드러나면서, 새로운 평가지표와 프레임워크를 활용한 정성적 평가의 시범 도입이 확대되고 있다. 따라서 증거와 엄격한 방법에 기반한 연구평가가 더욱 중요해졌는데, 조직은 데이터를 소유하고 통제된 상태에서 자체 평가를 진행하며, 연구평가 개혁이 목표를 얼마나 달성했는지 증거를 통해 확인해야 한다. 또한, 연구 데이터는 개인정보를 보호하면서 개방형 인프라를 통해 공개하고, 연구에 대한 리뷰 연구도 활발히 이루어져야 한다.



정책 제언



연구성과 · 어떻게 평가할 것인가?

젊은 과학기술인들의 동기부여를 위해 연구 성과평가 방법 관련 정책을 어떻게 고도화할 수 있는지 제시해보고자 한다.

연구 성과평가는 개인, 과제, 연구자, 기관 등 다양한 객체를 대상으로 실시하고 있다. 이는 승진, 포상, 과제 수주 등을 심사하는데 있어 중요한 기준으로 활용되고 있으므로, 우수한 연구를 지속하게 만드는 핵심 동력 중 하나라고 할 수 있다. 따라서 객관성, 투명성, 다양성, 윤리성에 기반한 공정한 성과평가 시스템 구축이 필수적이다.

하지만 연구 성과는 단일 업적물이 아닌 다양한 업적의 종합적 평가 대상이다. 또한 성과 창출 시점과 가치 실현 시점이 다르고, 평가 기준도 목적에 따라 다르게 적용되고 있으므로 복잡성으로 인해 균형 잡힌 평가 실현이 쉽지 않은 것이 현실이다. 따라서 이러한 특성을 고려한 섬세한 평가 체계 마련이 요구된다.

(1) 질적 평가 강화

우선, 연구 자체가 지닌 혁신성과 창의성, 산업 및 사회 문제 해결 가능성, 새로운 이론과 방법론 확립 같은 학술적 기여 등을 중점적으로 평가해야 한다. 연구의 중장기적인 파급력을 평가하기 위해 논문 게재 후 일정 시간이 지난 시점에 재평가를 진행할 수도 있을 것이다. 이는 양적 지표에 매몰되지 않고, 연구가 지닌 내적 동기를 복돋우는 방향으로 평가를 전환하려는 시도이다.

다만 질적 평가에 있어서는 명확한 평가 기준의 설정이 중요하다. 예를 들어, 파급력이 매우 높은 논문 1편의 연구 성과와 파급력이 어느 정도 높은 논문 3편의 연구 성과를 비교하거나, 혁신성이 매우 우수한 연구 성과와 사회 문제 해결에 크게 기여한 연구 성과를 비교하여 상대적인 우수성을 평가해야 하는 경우가 종종 발생한다. 이때 평가자별로 각기 다른 기준을 적용하지 않도록 충분한 사전 논의를 통해 평가 목적에 부합하는 명확한 평가 기준을 설정하는 게 필요하다. 명확한 평가 기준의 설정은 젊은 과학자들이 가장 중요하게 생각하는 평가 과정의 투명성을 높이는 데 있어서도 크게 기여할 것이다.

한편, 해당 분야의 중요하고 파급력 있는 저널을 IF 등의 정량지표가 아닌 전문가들의 정성적 평가에 의해 도출할 수도 있다. 각 분야별로 우수 연구 성과가 발표되는 저널에 대한 공감대가 어느 정도 형성되어 있는 만큼 해당 저널의 리스트를 만들어 활용하는 것이다.

또한 평가 결과가 연구개발 사업 기획이나 지원정책 수립, 연구자의 연구 방향 개선으로 연결되도록 환류 체계를 강화해야 한다. 평가가 그저 결과를 판단하는 데 그친다면 연구자들이 느끼는 피드백은 한계가 분명하기 때문에 정교한 반복 작업을 통해 연구 성과평가가 미래 연구 방향 설정과 과제 기획 등에 유용하게 활용될 때, 연구 생태계가 건강한 방향으로 발전할 수 있다.

(2) 평가자 관리 체계 개선

평가자의 수가 적을수록 편견과 주관성이 개입될 가능성이 높아진다. 따라서 평가자의 폭을 국내외로 확대하고, 이에 대한 정보를 지속적으로 제공하며, 우수한 평가자에게는 충분한 혜택을 제공하는 것이 필요하다. 이를 위해 산학연 글로벌 네트워크를 구축하여

국내의 연구자들이 연구 성과를 공유하고 검토할 수 있는 환경을 조성해야 하는데, 특히 해외 연구자들의 참여를 지원하는 것이 필요하다. 국내 평가자들만으로 이루어진 집단이 가진 한계로 인해 해외 연구자를 포함하려는 시도가 있었지만, 언어, 보안, 시간과 비용 등의 문제로 아직은 충분히 활용되지 못하고 있는 것이 현실이다.

그러나 글로벌 네트워크 내에서 연구 성과의 가치를 평가하는 것은 선도적인 연구를 수행하고 전문성 있는 평가자를 확보한다는 측면에서 반드시 필요한 만큼 이를 위한 행정적, 재정적 지원이 필요하다. 해외에서 활동하는 국내 연구자들의 참여를 독려하는 것도 한 가지 방안이다. 또한 평가자 집단을 산업체와 연구소로 확대하여 연구 성과의 가치를 다양한 관점에서 평가할 수 있도록 지원해야 한다.

현재 활동하고 있는 평가자의 활용도를 높이는 것도 중요하다. 이를 위해 평가자 정보를 주기적으로 현행화해야 한다. 특히 평가자 추천 알고리즘을 지속적으로 고도화하고 그 신뢰도를 모니터링하여 최적의 평가자를 추천해야 한다. 이 과정에서 소규모 연구 분야가 저평가되지 않도록 주의할 필요가 있다. 일반적으로 평가위원은 분야별로 구성되므로 소규모 분야의 경우 평가위원의 수가 제한적이라 평가의 영향력이 낮아질 수 있으며 심지어 서로 다른 분야가 통합되어 평가가 이루어지기도 한다.

한편, 연구 성과평가에 참여하는 것은 학문 발전에 중요한 기여 활동이므로, 이에 대한 충분한 보상을 제공해야 한다. 동시에 성과평가에 대한 평가자의 책임성을 확보해야 한다. 특히 우수한 평가자에 대해서는 수상이나 인센티브를 확대하여 평가자들의 동기를 부여하고 평가의 질을 높이기 위해 노력해야 한다.

또한 평가자들을 대상으로 연구 성과평가의 중요성과 공정한 평가를 위해 필요한 고려사항 관련 교육을 실시하여 평가의 전문성을 제고해야 한다. 결국 이러한 교육을 통해 연구 성과평가 참여 경험이 많지 않은 주니어 연구자들의 참여를 확대하는 데도 기여할 수 있기 때문이다.

(3) 평가 지원 도구 개발

동료 연구자들이 주관하는 질적 평가와 함께 AI 기반 데이터 분석을 활용하여 질적 평가를 지원하는 다양한 도구를 개발하는 것이 필요하다. 연구 성과에 대한 평가가 기존의 양적

평가에서 변화하고 있는 것은 사실이지만, 여전히 제한된 관점에서 평가가 이루어지고 있는 상황이다. 저널의 IF나 FWCI가 연구 업적의 우수성을 어느 정도 평가할 수 있다고 판단되지만, FWCI는 후속 연구에 미치는 파급효과만을 측정하며, 연구 성과가 창출된 후 충분한 시간이 지나야 정확한 평가가 가능하다는 한계가 있다.

따라서 AI 기반의 논문 및 특허 분석 시스템을 통해 혁신적인 연구 성과를 조기에 탐색할 수 있어야 한다. 연구공유 플랫폼에서의 기여, 소셜미디어에서의 논의, 보도 자료 등 다양한 측면에서 연구 성과의 영향력을 평가할 수 있는 보조지표를 개발하여 질적 평가를 지원할 필요가 있는 것이다. 또한 연구의 혁신성과 독창성, 연구 커뮤니티에서의 기여도, 그리고 해당 연구가 미치는 파급효과 등을 데이터 기반으로 정교하게 분석해 전문가들의 의사결정을 지원하는 도구를 개발해야 한다.

이 외에도 우수한 연구 성과들의 데이터베이스^{DB}를 구축하고 이를 분석하여 우수한 연구 성과들이 보이는 기본 특성을 추출함으로써, 우수한 연구 성과를 조기에 발굴하고 전문가들의 평가 부담을 완화하는 평가 지원 도구를 개발해야 한다. 실제 AI Scientist 등 인간을 지원하는 AI Agent가 연구 수행과 평가 과정에 어떻게 활용될 수 있을지 본격적으로 논의되고 있는 시점이다.

(4) 차별화된 평가 시스템 구축

평가가 이루어지는 상황이나 배경에 따라 차별화된 평가 시스템을 구축할 수 있도록 충분한 자원을 투입하고, 평가 시스템의 운영 결과를 지속적으로 모니터링하여 개선하는 것이 필요하다. 평가 목적, 학문 분야, 경력단계 등 평가가 이루어지는 상황에 따라 차별화된 시스템이 구축될 수 있도록 자원을 집중하는 것이 필요한 이유는, 이를 통해 도출된 우수한 연구 성과를 추적하고, 평가 시스템의 적절성을 지속적으로 점검하고 개선할 수 있기 때문이다.

특히 학문 분야별 차별화된 평가 시스템 구축은 가장 시급한 개선이 요구되는 분야이다. 학문 분야별 특성을 반영하여 기존의 논문 외에도 데이터셋, 소프트웨어^{SW}, 정책 제안, 특허, 기술이전, 창업, 산학협력, 국제공동연구 등 다양한 성과를 평가 대상으로 포함해야 한다. 또한 평가 목적과 경력단계에 따라 학제 간 연구, 산업적 및 사회적 파급력 등을 충분히 고려한 평가 시스템을 구현하는 것이 중요하다. 이 외에도 연구자의 장기적이고

지속적인 성과 창출 여부, 협력 네트워크 형성 및 후학 양성에의 기여도 등 여러 평가 기준을 반영하여 포괄적인 평가가 이루어져야 한다. 이를 통해 각자의 방식으로 과학기술 발전에 기여하는 연구자들을 격려할 필요가 있다.

또한 선도형 혁신시스템을 실현하기 위해 우수한 연구 성과를 발굴하고 이를 공정하게 평가할 수 있는, 연구자들이 신뢰할 수 있는 평가 시스템을 구축해야 한다. 이를 통해 지속 가능한 연구 생태계를 조성하고, 혁신적 성과의 창출과 확산을 촉진할 수 있다.

특히 평가는 단순히 성과를 측정하는 도구로만 여겨져서는 안되므로 과학기술인들에게 동기를 부여하고 도전·혁신적 연구를 추구하는 방향으로 설계되어야 한다. 그리고 이러한 시스템 구축을 위한 투자 확대는 필수적이다.

과학기술 분야의 연구 성과가 점점 다양해짐에 따라, 연구 성과평가에 대한 국제적인 관심도 높아지고 있다. 따라서 이러한 흐름에 발맞춰 우리나라가 국제적으로 통용될 수 있는 평가 가이드라인을 개발하고 공표함으로써, 과학기술 분야에서 글로벌 리더십을 확보하는 데 기여할 수 있을 것이다. 연구평가 방식에 대한 국제 선언이 지속된다는 것은 이상적인 연구평가 시스템에 대한 연구자들의 바람이 그만큼 간절함을 의미하며 이에 대한 정기적이고 지속적인 논의가 요구된다.

이번 차세대리포트를 통해 우수한 연구 성과의 정의와 평가에 대한 미래의 방향성을 조망해 보았다. 그 결과 공정하고 효율적인 평가 체계 구축은 연구 생태계의 건강한 발전을 위해 필수적 요소라는 점을 확인했다. 그리고 이를 위해서는 우수한 연구 성과를 어떻게 정의하고, 이를 어떻게 평가할지에 대한 방안을 심도있게 논의했다.

다양한 논의 속에서 우수한 연구 성과의 정의는 단순히 논문의 수나 인용 횟수와 같은 양적 지표에 국한되어서는 안된다는 점과 연구의 혁신성, 사회적 영향력, 학문적 기여도 등 질적인 측면을 포함하는 다면적 접근이 요구된다는 점 등이 제시되었다.

그리고 다양한 학문 분야의 특성을 고려한 맞춤형 평가 기준을 수립하고, 장기적인 연구 성과와 단기적 연구 성과를 조화로우면서도 균형 있게 평가할 수 있는 시스템을 구축해야 한다는 의견도 제시되었다.

평가 방식에 있어서도 기존의 방식을 보완할 수 있는 새로운 방법론 도입이 필요하다는 의견이 많이 모아졌는데, 대표적으로는 요즘 한창 주목을 받고 있는 인공지능^{AI}을 활용한 객관적 평가 시스템을 꼽을 수 있다. 이 외에도 연구 과정의 투명성과 재현성을 고려한 평가나 교육, 멘토링, 사회공헌 등 연구자의 다양한 활동을 종합적으로 고려하는 평가 방식 등도 함께 논의되었다.

사례 검토에 대한 조사도 관심이 가는 대목이다. 연구 성과평가 시스템과 관련된 국내외 동향을 살펴본 결과, 많은 국가와 기관에서 이미 이러한 문제의식을 가지고 다양한 시도를 하고 있음을 알 수 있었다.

이러한 동향을 바탕으로, 보다 효과적인 연구 평가 및 보상체계를 마련하고자 한다면 △연구 포트폴리오 시스템 도입 △분야별 맞춤형 평가지표 개발 △장기 연구과제의 의무적 지원 프로그램 △연구자 경력개발 지원 같은 새로운 제도 지원이 뒷받침되어야 할 것으로 보인다.

‘연구 포트폴리오 시스템 도입’은 연구자의 다양한 활동과 성과를 종합적으로 기록하고 평가할 수 있는 시스템을 구축함으로써, 단편적인 평가를 지양하고 연구자의 전체적인 기여도를 파악할 수 있는 장점이 있다.

그리고 ‘분야별 맞춤형 평가지표 개발’은 각 학문 분야의 특성을 반영한 평가지표를 개발하여, 획일화된 평가로 인한 부작용을 최소화하고 각 분야의 발전을 촉진하고자 하는 것이며, ‘장기 연구과제의 의무적 지원 프로그램’은 단기적 성과에 치중하지 않고 혁신적이고 도전적인 연구를 장려하기 위해 장기 연구 지원 프로그램을 확대하고 이에 대한 평가와 보상 체계를 별도로 마련하는 것이다.

마지막으로 ‘연구자 경력개발 지원’은 연구 성과에 대한 보상의 일환으로, 연구자의 경력 개발을 위한 다양한 프로그램, 예를 들면 국제 교류나 산학협력 등의 기회를 제공하고 이를 연구 성과평가에 반영하는 제도이다.

우수한 연구 성과의 평가 및 보상체계는 연구 생태계의 근간을 이루는 중요한 요소이다. 이를 개선하고 발전시키는 것은 단순히 연구자 개인의 만족도를 높이는 것을 넘어, 국가의 연구 경쟁력을 높이고 궁극적으로는 인류의 지식 발전에 기여하는 열쇠가 되는 것이다.

우리가 상상하는 미래의 연구 생태계는 더이상 단순한 양적 지표에 얽매이지 않고, 진정으로 연구 성과의 혁신성을 인정받는 곳이 될 것이다. 연구자들은 자신의 능력을 최대한 발휘하며, 사회는 이를 통해 얻어진 지식의 혜택을 누리게 될 것이다.

이러한 미래를 향해 나아가기 위해, 우리는 지금 여기에서부터 변화를 시작해야 한다. 공정하고 효율적인 평가 및 보상체계 구축을 위한 우리의 노력이 미래 연구 생태계의 건강한 발전으로 이어지기를 기대해 본다.

참고문헌

과학기술정보통신부·한국과학기술기획평가원(2023). 국가연구개발 우수성과 100선 추적조사.

과학기술정보통신부·한국과학기술기획평가원(2024). 2024 국가연구개발 우수성과.

이공래(2023). 대한민국 반세기 연구개발정책의 변천. 기술혁신연구, 31(1), 1-28. 기술경영경제학회.

한국과학기술기획평가원. (2018). 국가연구개발사업 평가제도의 성과분석 및 고도화 방안 연구.

4단계 BK21사업, 숫자 너머 질적 혁신을 선도한다.https://webzine.nrf.re.kr/magazine/2407/sub_1_02.php

Barbour, G. and Hazlett, H.(2023). From declaration to global initiative: a decade of DORA, The Declaration on Research Assessment (DORA) webpage.<https://sfдора.org/2023/05/18/from-declaration-to-global-initiative-a-decade-of-dora/>

Coalition for Advancing Research Assessment(2022). Agreement on Reforming Research Assessment.<https://coara.eu/agreement/the-agreement-full-text/>

Krüger, A. K., & Petersohn, S.(2022). From Research Evaluation to Research Analytics. The digitization of academic performance measurement. Valuation Studies, 9(1), 11-46.

Lee, H., & Shin, J.(2014). Measuring journal performance for multidisciplinary research: An efficiency perspective. Journal of Informetrics, 8(1), 77-88.

Leiden Manifesto for Research Metrics webpage<https://www.leidenmanifesto.org/index.html>

Reed, M. S., Ferré, M., Martin-Ortega, J., Blanche, R., Lawford-Rolfe, R., Dallimer, M., & Holden, J.(2021). Evaluating impact from research: A methodological framework. Research Policy, 50(4), 104147.

Thelwall, M.(2024). Quantitative Methods in Research Evaluation Citation Indicators, Altmetrics, and Artificial Intelligence. arXiv preprint arXiv:2407.00135.

Thelwall et al.(2023). Predicting article quality scores with machine learning: The UK Research Excellence Framework. Quantitative Science Studies, 4(2), 547-573.

Tyseretal.(2024). AI-Driven review systems: evaluating LLMs in scalable and bias-aware academic reviews. arXivpreprint arXiv:2408.10365.

World Conferences on Research Integrity Hong Kong Principles.<https://www.wcrif.org/guidance/hong-kong-principles>

차세대리포트(최근 3개년)

2022 우주 개척, 어떻게 해야 할까?

유전체 교정 작물, 식량안보의 대안이 될 수 있을까?

코로나19 엔데믹 전환과 롱코비드 문제 어떻게 대응할 것인가?

책임성 있는 AI를 위한 조건은?

2023 한계 돌파형 차세대 탠덤 태양전지 기술: K-Solar

새로운 의료서비스 혁명: 디지털 치료제

이론 연구와 실험 연구의 양극화 진정한 협력을 이루려면?

인공지능 언어모델의 기술 변천사와 미래 가능성

2024 국가 과학기술로 바라보는 RE100의 전략적 접근

뇌와 세상의 소통

진료지원인력과 미래 간호인력 활용 방안

한국과학기술한림원은,

대한민국 과학기술분야를 대표하는 석학단체로서 1994년 설립되었습니다. 1,000여 명의 과학기술분야 석학들이 한국과학기술한림원의 회원이며, 각 회원의 지식과 역량을 결집하여 과학기술 발전에 기여하고자 노력해오고 있습니다. 그 일환으로 기초과학연구의 진흥기반 조성, 우수한 과학기술인의 발굴 및 활용 그리고 정책자문 관련 사업과 활동을 펼쳐오고 있습니다.

한림석학정책연구는,

우리나라의 중장기적 과학기술정책 및 과학기술분야 주요 현안에 대한 정책자문 사업으로 한국과학기술한림원 회원들이 직접 참여함으로써 과학기술분야 및 관련분야 전문가들의 식견을 담고 있습니다. 한림연구보고서, 차세대리포트 등 다양한 형태로 이루어지고 있으며 국회, 정부 등 정책 수요자와 국민들에게 필요한 정보와 지식을 전달하기 위하여 꾸준히 노력하고 있습니다.

한국과학기술한림원 더 알아보기

 홈페이지 www.kast.or.kr

 블로그 kast.tistory.com

 포스트 post.naver.com/kast1994

 페이스북 www.facebook.com/kastnews





KAST 한국과학기술원
The Korean Academy of Science and Technology

(13630) 경기도 성남시 분당구 돌마로 42

Tel 031-726-7900 **Fax** 031-726-7909 **E-mail** kast@kast.or.kr



9 772635 716007
ISSN 2635-716X