

## 저자 순위를 이용한 개인별 연구 성과 지수 산출에 관한 연구

### A Study on Evaluation of Ideces an Individual's Scientific Productivity Using Author Rank

김완중, 노경란, 서진이, 이혜진  
한국과학기술정보연구원

Kim wan-jong, Noh kyung-ran, Seo jinny, Lee hye-jin  
Korea Institute of Science and Technology Information  
(KISTI)

#### 요약

연구자 개인의 연구 성과를 측정하기 위한 연구는 2005년 Hirsch가 h-지수(index)를 제안하면서 이와 관련된 후속 논의가 매우 활발하게 진행되고 있다. h-지수를 변형시킨 g-지수도 각광을 받고 있다. 하지만 이들 지수의 경우 연구에 기여한 비율에 상관없이 해당 논문의 인용 빈도만을 고려하여 지수가 정해진다는데 그 한계가 있다. 본 연구에서는 이 두 가지 지수의 단점을 보완하기 위해 각 지수에 저자의 기여도를 접목시킨 K-지수를 제안하고자 한다.

#### Abstract

There are many measuring tools to evaluate individual's scientific productivity using scientific articles based on bibliometrics. Especially succeeding discussions are accomplished at an trial for individual's scientific productivity after Hirsch's 2005 proposal using h-index. And there are many succeeding discussions and applications to evaluate journals, articles and qualitative level of researchers using h-index, g-index. But these indices have a same indices without regard to first author or other co-author if a article has a same number of citations. This research aims to propose K-index integrated between these indices and author rank.

## I. 서론

많은 사람들은 연구 생산성을 측정하기 위해 총 논문의 수, SCI(Science Citation Index) 혹은 SCIE(SCI Expanded)등 재 학술지 발표 논문의 수, 그리고 영향력 지수(IF : Impact Factor) 등을 사용한다. 하지만 연구 결과를 측정하기 위해 사용하는 이러한 기준들이 많은 약점들을 지니고 있다.

2005년 University of California(San Diego)의 물리학자인 Hirsch는 인용을 통해서 과학적 연구 성과를 측정하려는 시도를 하였다. h-지수가 단순명료하고, 편향되지 않으며, 부정확한 방법으로 조작하기 힘들다는 J. E. Hirsch의 발표가 있는 후, 계량정보학 분야와 다른 학술 커뮤니케이션 분야에서 h-지수의 유효성 평가, 특정 연구 영역에서 h-지수를 이용한 개인의 연구 생산성 측정, 다른 영역으로의 적용, 그리고 h-지수에 대한 보완 지수의 제안 등에 대한 후속 논의가 활발히 이루어지고 있다. 이후, Egghe는 h-지수의 단점을 보완하는 g-지수를 제안하였다. 또한 Batista et al.은 연구논문의 피인용 횟수는 공저자의 수가 많을수록, 자기인용이 많아지기 때문에 공저자의 수에 영향을 받는다고 하였다.

하지만 이들 지수는 연구자들이 작성한 논문에 얼마만큼의 기여를 했는지에 상관없이 동일한 평가를 받는 단점이 있다.

연구논문에서의 저자 순서는 논문의 기여도에 따라 정해지는 것이 일반적이다. 따라서 본 연구에서는 h-지수, g-지수에 각 논문에서의 저자 순위, 즉 논문의 기여도를 적용한 K-지수를 새롭게 제안하고자 한다.

## II. 선행 연구

### 1. h-지수

Hirsch는 Thomson Scientific의 데이터베이스인 Web of Science SCIE를 사용하여 h-지수를 제안하였다[1]. 개인의 출판기록 및 인용기록은 유용한 정보를 포함하고 있는 명확한 데이터이다. 이 정보는 출판된 논문의 수( $N_p$ ), 각 논문(j)에 대한 피인용 횟수( $N_c$ ), 그리고 논문이 출판된 학술지와 그 영향력을 나타내는 파라미터 등을 포함한다.

h-지수는 "어떤 과학자가  $N_p$  개의 논문 중에서 최소 h번 이상 인용된 논문이 h개이며, 나머지  $N_p-h$ 개 논문의 각 개별 인용 빈도가 h번 이하일 때, 그 연구자의 h-지수는 h"로 정의된다.

이 지수는 (i) 생산성과 영향력을 동시에 파악할 수 있고, (ii) 필요한 데이터는 Thomson Scientific의 Web of Science

데이터베이스를 통해 쉽게 접근 할 수 있으며, (iii) 극값에 민감하지 않으며, (iv) 수치를 과장하기 어렵고, (v) 인용에 관하여 가장 연관된 논문들에 대한 표본을 자동적으로 추출하는 장점을 가지고 있다[2]. 하지만 h 지수는 정수로 표시되기 때문에 많은 연구자들이 동일한 값을 가질 수 있으며, 그로 인해 연구자들을 구분하거나 나열하는 것이 매우 임의적이다. 따라서 또 다른 기준을 필요로 하는 단점이 있다.

### 2. h<sub>f</sub>-지수

최근의 연구에서, 논문의 피인용 건수는 저자의 수에 영향을 받는다는 것이 밝혀졌다[3]. 공저자수가 많은 논문을 통해 높은 h 값을 가진 과학자는 대부분 그의 실제 h 값보다 과장된다는 것이다. Batista et al.은 공저자의 수가 많을수록 미래의 자기 인용이 많아질 것이기 때문에 통계적 편향이 발생할 수 있다는 것을 전제로 공저자들에 의한 영향을 고려하여 이를 수정한 h<sub>f</sub> 지수를 제안하였다.

공저자들의 영향을 측정하기 위해서 Batista et al.은 h 개의 논문에서 연구자들의 공저자들의 평균을 h로 나누었다.

$$\langle N_a \rangle = N_a^{(T)} / h \tag{1}$$

$\langle N_a \rangle$ 는 h-지수를 산정하는데 포함된 h개 논문의 평균 공저자 수이며,  $N_a^{(T)}$ 는 h 개의 논문에 나타난 저자들의 총 수(중복허용)이다.

하지만 h<sub>f</sub>-지수는 각기 다른 공저자들의 순위를 고려하지 않았다. 만일 10명의 연구자가 하나의 논문을 작성했다면 이 10명의 공저자들은 논문에 대하여 각기 다른 기여를 했을 것이며, 그로 인해 각 저자들의 순위가 결정된다고 할 수 있다. 그렇기 때문에 하나의 논문을 공동으로 작성한 공저자들이 동일한 h<sub>f</sub>-지수를 갖는 것은 연구자들에 대한 평가를 하기 어려운 단점이 있다.

### 3. g-지수

Egghe[4]는 h-지수를 산정하는데 포함된 피인용빈도가 높은 상위 논문의 피인용건수가 늘어나더라도 지수값의 향상에는 영향을 미치지 않는다는 것을 지적하고 이를 보완하기 위한 g-지수를 제안하였다.

g-지수는 논문별 피인용횟수의 순위를 제공해서 산출하는 지수로, h지수를 보완하기 위해 나온 것이다. 지수를 산출하는 방식은 h지수와 거의 동일하다 .

$$g = \left( \frac{\alpha - 1}{\alpha - 2} \right)^{\frac{\alpha - 1}{\alpha}} T^{\frac{1}{\alpha}} \tag{2}$$

(T = 총 논문의 수)

g-지수는 연구자들의 논문을 피인용건수가 높은 순서로 나열했을 때, 논문의 인용빈도의 누적합계가 논문 순위의 제곱 ( $g^2$ )이상의 마지막 논문 순위가 그 연구자의 g-지수가 되며, g-지수는 h-지수보다 항상 크거나 같은 값을 가진다.

하지만 한 연구자가 피인용건수가 많은 논문이 한 편만 있으며, 다른 논문들의 피인용건수가 낮을 경우, g-지수는 그 연구자의 연구 생산성을 과장할 수 있으며, h-지수와 같이 해당 논문에서 본인의 기여도를 반영하지 않았기 때문에, 저자 순위에 상관없이 높은 지수를 가질 수 있다는 단점이 있다.

## III. K 지수

### 1 K<sub>h</sub>-지수

h-지수는 저자 순위를 고려하지 않은 상태로 해당 저자의 h-지수를 평가했으며, h<sub>f</sub>-지수는 평균 공저자수 만을 이용하여 지수를 산출했기 때문에 평가하고자 하는 연구자가 제1저자이든 제10저자 이든 상관없이 해당 논문의 피인용횟수를 반영하였다.

하지만 본 연구에서 제안하고자 하는 K<sub>h</sub>-지수는 다음과 같은 기준을 바탕으로 한다. 논문을 작성한 공동저자들은 본인의 저자 순위에 따라 가중치를 둔 저자점수를 갖는다. 각 순위별 저자점수는 어떤 한 논문의 각 저자들이 갖는 저자 점수는 상수 1에 논문의 공저자수를 더한 후 저자순위를 뺀 값을 다시 공저자수로 나누어 산출한다. 각 논문의 저자점수의 합은 단일 저자를 1점으로 하여 공저자가 1명씩 늘어날 때 마다 0.5점씩 추가 되도록 하였다.

$$A_S = \frac{(1 + N_{CA} - A_R)}{N_{CA}} \tag{3}$$

(A<sub>S</sub> : 저자 점수, N<sub>CA</sub> : 공저자수, A<sub>R</sub> : 저자순위)

[표 1] 공저자수 및 저자순위별 저자점수 예

공저자수	저자 순위별 점수					점수 합계
	1저자	2저자	3저자	4저자	5저자	
1명	1.00					1.00
2명	1.00	0.50				1.50
3명	1.00	0.67	0.33			2.00
4명	1.00	0.75	0.50	0.25		2.50
5명	1.00	0.80	0.60	0.40	0.20	3.00

[표 2] 논문 분석 결과

	Baltimore	Chambon	Dinarello	Evans	Gallo	Kishimoto	Moncada	Snyder	Ulrich	Vogelstein
논문 수	360	732	822	888	683	1,490	486	577	540	403
총피인용수	63,992	73,859	62,366	15,515	35,620	72,935	80,983	69,303	59,650	121,729
평균피인용수	177.76	100.90	75.87	17.47	52.15	48.95	166.63	120.11	110.46	302.06
최소저자수	1	1	1	1	1	1	1	1	1	1
최대저자수	74	40	29	484	25	74	15	122	21	45
평균저자수	4.34	6.38	5.58	6.29	6.55	7.89	4.48	5.25	6.06	7.80
최대저자순위	1	1	1	1	1	1	1	1	1	1
최소저자순위	12	17	29	223	24	74	15	87	19	35
평균저자순위	3.46	5.42	4.46	3.66	5.74	5.53	3.95	4.75	4.86	6.68

$$K_h = TC \times A_g \quad (4)$$

각 논문의 피인용횟수를 저자점수로 곱한 값을 내림차순 정렬한 후, 각 저자별 논문점수가 정렬된 논문순위보다 크거나 같은 값을  $K_h$ -지수라 한다.

### 2. $K_g$ -지수

Egghe et al.은  $g \geq 2$  이상의 누적 피인용횟수를 가진 논문의 최하위 순위가  $g$ -지수가 된다고 정의하였다. 하지만  $g$ -지수 역시  $h$ -지수와 마찬가지로 저자 순위를 고려하지 않았다. 본 연구에서는 이러한 단점을 보완한  $Kg$ -지수를 제안한다.

공저자 점수의 합 =

$$\sum \left( \frac{1 + N_{CA} - A_R}{N_{CA}} \right) = 1 + \frac{(N_{CA} - 1)}{2} \quad (5)$$

## IV. 분석 결과

$h$ -지수,  $h_r$ -지수,  $g$ -지수는 각 연구자들이 발표한 논문에서 본인의 연구 기여도, 즉 저자 순위를 고려하지 않았기 때문에 각 지수 값을 과장하는 단점을 지니고 있다. 이에 본 연구에서는 Hirsch가 그의 논문에서 실제 데이터를 측정하기 위해 사용했던 생물학 및 생의학 분야에서 가장 많이 인용된 10인의 연구자들을 대상으로 1985년부터 2005년까지의 문헌데이터를 기초로 하여 각 지수별로 분석하였다(표 2).

### 1. 논문 분석 결과

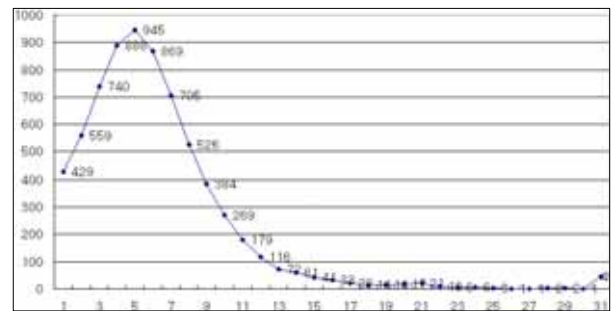
본 연구에 사용된 대상 문헌은 총 문헌수 6,981건 중 단일저자 문헌 429건 공저자문헌 6,552건으로 구분되었다. 6,552건의 공저자 문헌은 또 다시 2인 공저 559건, 3인 공저 740건, 4인 공저 887건, 5인 공저 945건, 6인 공저 868건, 7인 공저 705건, 8인 공저 526건, 9인 이상 공저 1,322건으로 나타났다. 6,981편의 논문에 나타난 저자는 총 44,371명(중복 포함)으로 나타났다

으며, 무려 484명의 공동저자가 작성한 논문도 1편이 발견되었다. 또한 평균 6.35명의 연구자와 공동연구를 수행하는 것으로 분석되었으며, 8인 이하의 공저비율이 81.1%를 기록했다(표 3).

[표 3] 공동저자수별 논문 현황

구분	논문편수	비율(%)	누적비율(%)
1인 단독	429	6.2	6.2
2인 공저	559	8.0	14.2
3인 공저	740	10.6	24.8
4인 공저	887	12.7	37.5
5인 공저	945	13.6	51.1
6인 공저	868	12.4	63.5
7인 공저	705	10.1	73.6
8인 공저	526	7.5	81.1
9인 이상	1,322	18.9	100.0
합계	6,981	100.0	-

1960년대 Price, Beaver 등[5]에 의해 수행되었던 공저관련 연구는 과학기술분야 학술 커뮤니케이션 흐름이 주로 공동 연구된 문헌의 형식을 통해 표출되고 있음을 입증해 주는 것이었다. 이후 공저자문헌의 질적 수준을 평가하는 방법과 공동연구의 본질 및 기능을 파악해 보려는 연구들로 계속되고 있다.

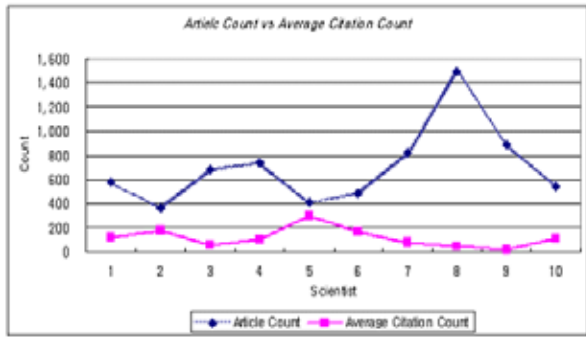


▶▶ 그림 1. 공저자 논문 수 현황

그림 1은 공저자에 관한 분석결과를 볼 수 있다. 주제분야별, 기간별로 공저율의 차이가 있겠지만 5인 공저 논문이 가장 많

은 945편이었다. 산업의 발달과 타 산업, 기술, 과학과의 융합이 자연스럽게 이루어지는 최근에 이러한 현상은 더 두드러지게 나타난다고 이해할 수 있다. 학제간 연구를 통해 공저율이 높은 주제 분야와 접목될 때 공저연구가 활발해지고 있으며, 자연과학 분야에서의 공동연구는 이미 일반화된 현상이라고 보고 있다. 그러므로 연구자들의 성과 분석시 저자순위를 고려하여야 보다 객관적인 평가가 가능하다.

연구자의 생산성 평가는 발표한 연구논문과 이로 인한 피인용건수로 정량적인 평가를 할 수 있다. 데이터 분석 결과, 연구자 S. H. Snyder, D. Baltimore, B. Vogelstein, S. Moncada, A. Ullrich 등은 타 연구자들에 비해 연구논문의 수는 적었으나 인용정도는 높은 부류로 분류되었다(그림 2).



▶▶ 그림 2. 연구자별 논문수 대비 평균 피인용건수 현황

## 2. 지수별 분석 결과

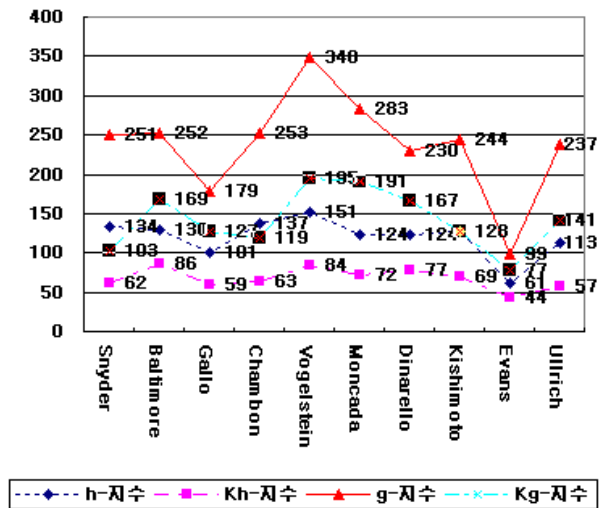
대상 데이터를 기초로 h-지수, Kh-지수, g-지수, Kg-지수 지수를 연구자별로 산출하여 분석 하였다. 10명의 연구자들에 대한 평균은 h-지수 120, Kh-지수 67, g-지수 238, Kg-지수 142로 나타났다. 각 연구자별 h-지수와 g-지수값을 비교했을 때 g-지수가 h-지수에 비해 모두 높게 나타났으며, 각 연구자들의 평균값은 g-지수가 약 두 배 정도 높은 수치를 나타냈다. 이는 g-지수가 h-지수보다 항상 크거나 같게 나타난다는 Egghe의 주장을 뒷받침하는 근거가 되었다.

[표 4] 연구자별 h, Kh, g, Kg 지수 및 순위 현황

구분	h-지수		Kh-지수		g-지수		Kg-지수	
	지수	순위	지수	순위	지수	순위	지수	순위
Baltimore	130	4	86	1	252	4	169	3
Chambon	137	2	63	6	253	3	119	8
Dinarello	124	6	77	3	230	8	167	4
Evans	61	10	44	10	99	10	77	10
Gallo	101	9	59	8	179	9	127	7
Kishimoto	128	5	69	5	244	6	128	6
Moncada	124	6	72	4	283	2	191	2
Snyder	134	3	62	7	251	5	103	9
Ullrich	113	8	57	9	237	7	141	5
Vogelstein	151	1	84	2	348	1	195	1
평균	120	-	67	-	238	-	142	-

또한 본 연구가 제안하는 연구자의 공저시 저자순위가 고려된 K-지수를 적용하여 보면, h-지수와 g-지수의 평균값에 비해 Kh-지수와 Kg-지수가 전반적으로 유사한 양상을 보여 약 56%, 60% 수준으로 감소하는 것으로 파악되었다. 이는 각 연구자들이 발표한 논문에서의 기여도, 즉 저자순위에 따라 1 보다 작거나 같은 각기 다른 저자 점수를 받기 때문이다.

연구자들을 각 지수별로 순위를 분석 했을 때, Vogelstein은 h-지수, g-지수, Kg-지수에서 모두 1위를 차지하고, 84의 Kh-지수를 가졌으나 h-지수가 4위였던 Baltimore의 86에 뒤져 2위를 기록했다(표 4). 연구자별로 분석된 결과를 보면 Baltimore, Dinarello는 K-지수 적용 후 좀 더 높은 평가치를 보였으며, Chambon, Snyder 등은 K-지수 적용 후 좀 더 낮은 평가치를 보였다.



▶▶ 그림 3. 연구자별 지수 현황

이를 좀 더 상세히 분석하기 위하여 각 연구자들의 평균 피인용건수와 각 지수간의 상관관계를 분석하였다. 각 지수별 상관계수는 h-지수 0.69, Kh-지수 0.72, g-지수 0.85, Kg-지수 0.75로 나타나 각 논문별 평균 피인용건수가 높을수록 각 지수도 높게 나타난다는 결과를 얻었다(표 5).

[표 5] 평균피인용건수와 각 지수간의 상관계수

구분	h-지수	Kh-지수	g-지수	Kg-지수
상관계수	0.693412	0.719944	0.851307	0.755173

## V. 결론

본 연구에서 제안한 K-지수는 연구자들의 성과를 평가하기 위한 기존의 h-지수, g-지수를 좀 더 보완한 척도라 할 수 있

으며 저자들의 기여도를 저자점수로 환산한 새로운 평가척도를 제안하여 객관성을 부여하는데 의의가 있다 할 수 있다. 하지만 향후 각 저자순위에 따른 가중치 적용에 대한 지속적인 보정과 각 논문발표건수, 총 피인용건수, 평균 피인용건수 등과 K-지수와와의 상관관계를 좀 더 면밀히 분석할 필요가 있다. 나아가 전 분야를 대상으로 이러한 현황분석이 되었으면 하는 바람이다. 또한 신뢰할 만한 데이터를 근간으로 하여 국내연구자에 대한 평가가 용이하도록 지속적으로 인프라 구축이 절실하고, R&D 연구성과나 연구자 평가를 위한 지속적인 척도 개발이 요구된다.

#### ■ 참고 문헌 ■

- [1] Hirsch, J. E., "An index to quantify an individual's scientific research output." *Proceedings of the National Academy of Sciences of the United States of America*, 102, pp. 16,569-16,572, 2005.
- [2] Batista, P. D., M. G. Campiteli, O. Konouchi, and A.S. Martinex., "Is it possible to compare researchers with different scientific interest?". *Scientometrics*, Vol. 68, No. 1, pp. 179-189, 2005.
- [3] Glänzel, W. and B. Thus., "Does co-authorship inflate the share of self-citations?". *Scientometrics*, vol. 61, No. 3, 2004, pp. 395-404.
- [4] Egghe L., "Theory and practice of the g-index", *Scientometrics*, Vol. 69, No. 1, pp. 131-152, 2006.
- [5] Price D. J. and Beaver D. D., "Collaboration in an Invisible College", *American Psychologist*, Vol. 21, 1011-1018, 1966.