

# 연구성과 측정을 위한 h-지수의 개량에 관한 연구

## Some Improvements on H-Index : Measuring Research Outputs by Citations

이재윤(Jae-Yun Lee)\*

### 초 록

Hirsch가 제안한 h-지수는 인용을 통해서 개인의 연구 성과를 측정하려는 시도이다. h-지수는 용이한 산출 방법과 함께 지수의 강건성 등이 인정되면서 발표 이후 다양한 분야에서 이를 적용하거나 보완하는 연구가 활발히 이루어지고 있다. 이 연구에서는 우선 h-지수와 이를 보완한 g-지수를 비롯한 변형 지수에 대해서 현재까지 전개되고 있는 논의를 살펴보았다. 그리고 h-지수와 g-지수의 단점을 보완할 수 있는 개량 지수를 몇 가지 제안하고 가상 자료와 실제 자료에 대해서 측정해보았다. 측정 결과 제안한 지수들이 h-지수와 g-지수를 대체할 가능성이 있음을 확인하였다.

### ABSTRACT

The h-index, also called as Hirsch-index, is a new tool for measuring research outputs by citations. This h-index is not only easy to calculate, but also robust enough to handle various citation data. After its suggestion by Hirsch in 2005, many researchers applied the h-index to their own areas, and some others tried to improve the weak points of the h-index such as low discriminating power. Firstly, several of these efforts are reviewed in the present article, and then novel indexes are suggested to measure research outputs by citations more fairly and reasonably. Calculating these indexes on both artificial data and real data showed that the newly suggested indexes in this article can replace the h-index and its variants.

키워드 : 인용, 인용지표, 연구성과측정, h-지수, g-지수

citations, citation measures, measuring research outputs, h-Index, g-index

---

\* 경기대학교 문헌정보학과 조교수 (memexlee@kgu.ac.kr)

■ 논문접수일자 : 2006년 8월 16일

■ 게재확정일자 : 2006년 9월 15일

## 1. 서 론

SCI를 비롯한 Thomson사의 인용색인 DB가 학술활동과 그 평가에 미치는 영향력에서 알 수 있듯이, 인용빈도가 학술활동에 대한 계량적인 척도처럼 인식되고 있는 것이 현실이다. 심지어 최근 국내 일부 대학에서는 논문이 게재된 학술지의 ISI 영향력 지수 수준에 따라서 교수의 연구업적 평가에 가산점을 차등적으로 부여하는 정책을 채택하는 것으로 나타났다(윤희운, 김신영 2005). 이와 같이 ISI 영향력 지수를 연구자 개인의 성과를 평가하는 용도로 사용하는 것에 대해서는 SCI DB를 제공하는 Thomson사 측에서도 무리한 적용이므로 주의할 것을 당부하 바 있다(Adam 2002).

그러나 현실적으로 연구성과에 대한 계량적인 측정 수단으로 손쉽게 활용할 수 것은 인용이 유일하기 때문에 이를 완전히 무시하는 것도 올바른 태도라고 말하기는 어렵다. Thomson사도 영향력 지수로 직접 개인을 평가하는 것에 대해서 언급하였을 뿐이며 개인 연구 성과에 대한 인용분석 자체를 비판한 것은 아니었다. Moed(2005)는 인용분석 자체가 평가라고 할 수는 없지만, 계량서지적 지표가 학술활동을 평가할 때 질적인 수준에 대한 통찰을 얻고 판단을 내리는 데 도움을 받는 하나의 연구 평가 도구는 될 수 있다고 언급하였다.

h-지수는 이와 같은 개인의 연구 성과 평가에 참조할 수 있는 계량서지적 지표로 2005년에 제안된 것이다. 캘리포니아 대학 샌디에이고 캠퍼스의 물리학자인 Jorge Hirsch는 h-지수를 제안하는 발표전 논문(Hirsch 2005)을 2005년 여름에 E-print archive에 올렸다.

일주일 후 Nature지의 소식란에 Ball(2005)이 Hirsch의 논문을 소개하면서 h-지수는 여러 학문공동체의 폭넓은 관심을 끌게 되었다.

h-지수는 공식의 간단명료함, 손쉬운 산출 방법, 지수의 강건성 등이 인정되면서 발표 이후 계량정보학 분야 뿐만 아니라 다양한 분야에서 활발한 후속 논의가 이례적으로 신속하게 이어지고 있다. 후속 연구는 주로 h-지수의 타당성 검토, 특정 학문 분야에의 h-지수 적용 시도, h-지수의 다른 용도 제안, h-지수의 보완 등에 대해서 전개되고 있다.

이 연구에서는 우선 h-지수와 이를 보완한 g-지수를 비롯한 변형 지수를 살펴본 다음 h-지수와 g-지수의 단점을 동시에 보완할 수 있는 새로운 지수를 제안하였다. 그리고 실제 자료를 대상으로 인용을 통한 성과 측정을 실험적으로 시도해보았다.

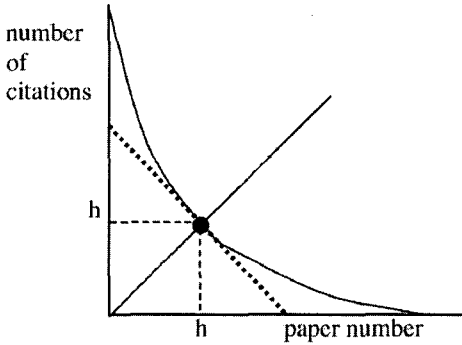
## 2. h-지수

### 2.1 h-지수의 정의 및 특성

h-지수의 정의는 다음과 같다(Hirsch 2005).

어떤 과학자의  $N_p$ 개 논문 중에서 h개의 논문은 최소 h번 이상씩 인용되었고, 나머지  $N_p - h$ 개 논문은 모두 개별 인용빈도가 h번 이하이면 그의 h-지수는 h이다.

이 정의에서 “h번 이하”라는 부분은 E-print 아카이브에 처음 올려진 발표전 논문의 1판에



〈그림 1〉 인용빈도 분포와 h-지수 (Hirsch 2005)

서 4판까지는 “h번 미만”으로 되어있었으나 오류임을 지적받고 2005년 9월에 올린 5판 원고에서 수정하였다. 이 부분은 Glnzel(2006)에 의해서 지적되었으며 Rousseau(2006a)는 다음과 같이 h-지수의 수정된 정의를 제시하였다.

연구자 S가 단독 혹은 공저로 발표한 문헌을 인용빈도 순으로 나열하고 동률을 무시하여 일련번호를 부여하였을 때, 상위 h개의 문헌이 최소 h회 인용되고 h+1번째 문헌의 인용빈도가 h+1회 미만이면 S의 h-지수는 h이다.

풀어서 설명하자면, 어떤 연구자의 논문을 인용빈도가 높은 순부터 나열하였을 때 논문의 인용빈도가 논문의 순위보다 크거나 같은 마지막 논문의 순위가 그 연구자의 h-지수가 된다. 예를 들어 h-지수가 20이라면, 최소 20회 이상 인용된 논문을 20편 발표했다는 뜻이다.

인용빈도의 분포가 〈그림 1〉의 굵은 선과 같다면 총 인용빈도는 굵은 선 아래의 면적이

〈표 1〉 연구자의 논문 인용빈도 가상 사례

인용 순위	연구자 A의 논문	연구자 B의 논문	연구자 C의 논문	연구자 D의 논문
1	10	4	3	20
2	7	4	3	3
3	5	4	1	2
4	4	4	1	1
5	3	3	1	1
6	2	2		
7	2	2		
8	2	2		
9	1	1		
10	1	1		
총인용	37	27	9	27
평균인용	3.7	2.7	1.8	5.4
h	4	4	2	2

다. 이 그림에서 h-지수는 원점에서 45도 각도로 출발한 직선과 마주치는 지점의 x좌표(순위)가 된다. 실제 인용빈도는 실수가 아닌 정수이므로 이 그림처럼 곡선형태가 아닌 계단식으로 감소한다. 〈표 1〉에 가상의 자료로 h-지수를 산출한 예를 소개하였다. Web of Science와 같은 인용색인 DB를 이용하면 이와 같은 인용 순위는 손쉽게 산출할 수 있다.

물론 h-지수 이외에 총 인용빈도 등 연구자를 평가하는 다른 지표도 여러 가지가 있다. Hirsch는 다른 지표에 대해서는 다음과 같이 비판하였다.

- 발표논문 총 수 : 생산성을 반영하지만 논문의 중요도나 영향력은 측정하지 못한다.
- 총 인용빈도 : 총 영향력을 나타내지만 완벽한 산출이 어려우며 공저자로 참여한 한 두 논문의 인용빈도가 예외적으로 높은 경

우에 과대평가될 위험이 있다.

- 논문당 평균 인용빈도 : 연구 경력이 긴 연구자와 짧은 연구자를 비교할 수 있지만, 완벽한 산출이 어려우며 생산성이 높은 연구자에게 불리한 단점이 있다.
- 일정 빈도 이상 인용된 논문의 수 : 앞선 지수들의 단점을 대부분 극복할 수 있으나, 인용빈도의 기준값을 정하는 것이 임의적이어서 누군가에게는 유리하게, 또 다른 누군가에게는 불리하게 작용할 수 있다.
- 인용빈도로 일정 순위 이상인 상위 논문의 인용빈도 집합 : 앞선 지수들의 단점을 대부분 극복할 수 있다. 그러나 하나가 아닌 여러 개의 값을 살펴야 하며, 일정 순위를 정하는 것이 임의적이어서 역시 누군가에게 유리하거나 불리한 상황이 발생한다.

이에 반해서 h-지수는 인용빈도가 높은 소수 논문에 좌우될 여지를 없애고 연구 생산성과 발표논문의 중요도를 함께 반영하는 척도라고 평가된다(Lehmann et al. 2006).

Hirsch(2005)는 h-지수를 제안한 같은 논문에서 h-지수가 연구 경력이 오래되었을수록 유리한 점을 감안하여 m이라는 척도도 제안하였다. 연구자의 최초 논문 발표 이후 경과한 햇수를 n이라고 하였을 때 m은 다음과 같이 산출된다.

$$m = \frac{h}{n}$$

즉, m은 h-지수로 측정되는 연구자의 인용 영향력이 어느 정도로 빠르게 달성되었는가를 반영한다. 연구자의 영향력을 물체의 이동에

비유해서 말하자면, h-지수는 총 이동 거리에 해당하고 m은 평균 이동 속도에 해당한다고 할 수 있다. m이 크면 클수록 연구자의 인용 영향력이 급속하게 향상되었음을 나타낸다.

## 2.2 h-지수를 이용한 연구자 성과 측정 사례

h-지수는 이용하는 인용정보원에 따라서 다른 결과가 나타날 수 있다. 많은 경우에 Thomson ISI Web of Science 데이터베이스를 이용하여 측정하지만, Lehmann(2006)과 같이 특정 분야 원문 데이터베이스에 인용 정보가 포함된 것을 이용하는 사례도 있다. 또한 Google Scholar를 이용한 사례도 있으며, 아예 연구자의 이름을 입력하면 Google Scholar를 대신 검색해서 h-지수를 산출해주는 사이트 <<http://www.brics.dk/~mis/hnumber.html>>도 등장했다. 물론 이런 자동 산출 결과는 동명이인의 문제를 해결하지 못하므로 그대로 믿을 바는 못된다. 많은 연구자들이 인용정보를 검색한 후 직접 상위 인용빈도 문헌을 확인하는 절차를 거쳐서 h-지수를 확정하였다.

처음 Hirsch(2005)가 h-지수를 제안하면서 시험적으로 측정한 대상은 물리학 분야와 생물학 및 생의학 분야의 유명 연구자였다. 물리학 분야에서는 E. Witten이 110으로 가장 높았고, 생물학 및 생의학 분야에서는 S. H. Snyder가 191로 가장 높게 측정되었다. Hirsch(2005)는 또한 개인이 아닌 집단을 대상으로 측정한 결과도 소개하였는데, 자신이 속한 UCSD 물리학과와 h-지수를 118로 보

고하였다.

Glänzel & Persson(2005)은 계량서지학 분야의 역대 Price 메달 수상자의 1986년 이후 발표논문들을 대상으로 측정하였다. 최고는 17로 나타났다. Bar-Ilan(2006)은 동일한 Price 메달 수상자에 대해서 Google Scholar를 이용해서 h-지수를 산출해본 결과 평균적으로는 Web of Science를 이용했을 때와 비슷한 결과를 얻었으며, 단행본 저술이 많거나 제목이 복잡한 논문을 쓴 연구자는 상이한 결과가 나타났다고 보고하였다.

Cronin & Meho(2006)는 미국의 생존 유력 정보학자 31명의 h-지수를 산출하되 자기 인용을 포함한 경우와 제외한 경우를 모두 구하였다. 이들은 DIALOG를 이용해서 SCI, SSCI, A&HCI 데이터베이스를 대상으로 31명의 정보학자가 제1저자로 발표한 논문이 인용된 경우를 모두 검색하였다. 이들의 조사에서는 N. J. Belkin이 20, T. Saracevic이 19로 최상위권을 차지하였다. 자기인용을 제외하였을 경우에는 Belkin과 Saracevic이 19로 공동 1위에 해당하였다. 자기인용을 포함한 경우의 순위와 제외한 경우의 순위는 한 두 명을 제외하고는 거의 유사하게 나타났다. 또한 자기인용을 제외한 피인용빈도 순위와 h-지수 사이에도 높은 상관관계가 나타났다(Spearman 순위상관으로 99% 수준에서 유의함). 그러나 이들은 h-지수가 일부 연구자에 대해서는 총 피인용빈도로는 알 수 없는 특성을 알려주는 장점이 있다고 언급하였다.

Bornmann & Daniel(2005)은 생의학 분야의 박사후연구비 장기 지원을 신청한 지원자를 대상으로 h-지수를 측정해본 결과 5년

간 승인된 연구자의 h-지수 평균이 3.84로 탈락된 연구자 350명의 h-지수 평균 2.72보다 높았다고 보고하였다.

Batista et al.(2005)은 Web of Science를 이용해서 주소지에 브라질이 포함된 논문의 물리학, 화학, 생물학/생의학, 수학 분야 연구자에 대해서 측정하였으며 최고값은 각각 37, 29, 24, 14로 나타났다.

Kelly & Jennions(2006)는 생태 및 진화 생물학 분야의 주요 7개 학술지 편집위원을 대상으로 측정하였다. 이들의 분석에서는 h-지수가 성별이나 수록학술지, 소속 국가에 따른 차이가 있는 것으로 나타났다.

Lehmann(2006)은 고에너지물리학 분야 데이터베이스인 SPIRES 데이터베이스를 이용하여 h-지수를 측정하였다.

Van Raan(2006)은 Hirsch(2005)가 시험적으로 해보았던 연구집단의 h-지수 산출을 체계적으로 시도하였다. 그는 네덜란드의 화학 및 화공학 분야 147개 연구집단이 1991년부터 1998년 사이에 발표한 문헌의 인용정보를 Web of Science를 이용해서 수집하였다. 인용빈도 산출은 논문이 발표된 연도를 포함해서 3년간의 인용만 고려하는 방식을 택했다. 이는 ISI의 영향력지수 산출에서 출간된 해의 2년 뒤까지의 인용만 고려하는 것과 유사하지만, 당해연도의 인용까지 포함하는 것이 다르다. 산출한 h-지수를 연구자들의 평균과 비교해본 결과 h-지수가 규모가 작은 연구집단에게는 불리하다고 결론지었다.

이외에도 h-지수를 이용해서 학술지 영향력 지수와 유사하게 학술지 단위의 인용성과를 측정하거나(Braun, Glänzel, & Schubert

2005 ; Rousseau 2006b), m 척도를 이용해서 물리학 분야 주요 연구주제의 활발한 정도를 측정 한 사례(Banks 2006)도 발표되는 등 h-지수의 응용 사례가 더욱 다양해지고 있는 실정이다.

심지어는 메타 h-지수에 대한 논의도 등장하였다. 이는 특정 학문분야에 최소한 h이상의 h-지수를 가진 연구자가 h명 있다는 뜻이다 (<http://11011110.livejournal.com/10507.html>). 이 메타 h-지수는 그 학문분야의 규모나 인용밀집도 등을 반영한다고 볼 수 있다.

### 3. h-지수에 대한 비판과 보완

#### 3.1 h-지수에 대한 비판

##### (1) 자기인용으로 인한 문제

Hirsch(2005)는 h-지수를 제안하면서 자기인용에 크게 좌우되지 않는다고 주장하였으나, Cronin & Meho(2006)가 미국 정보학자에 대해서 분석한 결과에 따르면 31명 분석 대상자 중에서 자기인용을 포함한 결과 순위가 4위 상승한 경우와 7위 하락한 경우가 있었다. 이는 자기인용의 포함 여부가 최소한 일부 연구자의 h-지수에는 영향을 끼친다는 것을 시사한다. 또한 Kelly & Jennions(2006)는 논문을 많이 발표하는 연구자일수록 자신의 논문을 인용할 기회가 많아지며, 공저자가 많은 논문도 역시 개별 공저자가 다시 인용할 가능성이 높아지므로 자기인용이 h-지수에 영향을 끼친다고 주장하였다. 결국 인용빈도값 자체만큼은 아니더라도 h-지수 역시 자기인용

에 어느 정도는 영향을 받음을 감안할 필요가 있다.

##### (2) 분야간 비교의 어려움

Hirsch(2005)도 언급하였지만 h-지수로는 상이한 분야의 연구자를 비교하기가 어렵다. 분야에 따라서 값의 차이가 크기 때문이다 (Batista et al. 2005 ; Cronin & Meho 2006 ; Kelly & Jennions 2006). 이는 특히 공저자 수에 영향을 많이 받는 것으로 나타나는데, 공저자가 많은 물리학 분야는 높은 경향이 있으며, 단독 저자 논문이 많은 수학 분야는 상대적으로 낮은 h-지수가 측정된다 (Batista et al. 2005).

##### (3) 동률값 빈발로 인한 낮은 변별력

h-지수는 동률 순위의 연구자가 많이 나타나므로 변별력이 약하다(Batista et al. 2005). 계량정보학자에 대한 Glanzel & Persson(2005)의 연구에서는 1위인 17을 기록한 연구자가 3명으로 나타났으며, Cronin & Meho(2006)가 조사한 미국 정보학자의 h-지수 순위는 공동 11위가 무려 8명에 달하였다. 상대적으로 지수의 수준이 높은 물리학, 화학, 생명과학 등의 분야에 속한 주요 연구자를 조사한 연구들에서는 동률값에 대한 지적이 없었다. 따라서 연구자간 인용이 활발하지 않고 규모가 작은 영역에 대해서 분석할 때에는 지수의 변별력이 떨어지는 것을 피할 수 없을 것이다. 비록 인용이 활발하고 규모가 큰 물리학이나 생명과학 분야에서라도 피인용빈도가 그리 높지 않은 연구자에 대해서 분석할 경우에는 낮은 변별력이 문제가 될 여지가 많다.

#### (4) 연구 생산성에 좌우됨

h-지수는 이론적인 최고값이 발표한 논문의 수이다. 예를 들어 발표한 논문마다 수 백회 이상 인용되더라도 총 다섯 편만 발표했다면 h-지수는 5이고, 단 한 편만 발표했다면 h-지수는 1에 불과하다. 또한 개인이나 연구집단은 생산성이 높을수록, 즉 발표한 논문이 많을수록 자기인용의 기회가 많아지므로 h-지수가 높아질 가능성이 있다. Van Raan(2006)은 화학 분야 연구팀에 대한 전문가들의 평판과 h-지수를 비교해본 결과 h-지수는 규모가 적은 연구팀에게 불리하게 나타났다고 보고하였다.

이는 학술지에 대해서 h-지수를 측정할 경우에도 마찬가지이다. 대표적인 SCI 등재 학술지의 h-지수를 측정한 연구에 따르면 ISI 영향력 지수 1위와 2위인 두 학술지는 수록 논문 수가 적은 리뷰지이기 때문에 h-지수로는 상위권에 포함되지 못하였다(Braun, Glänzel, & Schubert 2005). 따라서 학술지의 h-지수는 ISI 영향력지수와 정반대로 수록 논문 수가 적은 리뷰지에게 불리한 척도라고 할 수 있다. 이런 특징은 학술지의 규모를 감안하는 평가지수인 Euro-Factor(Hofbauer et al. 2002)와 유사한 것이다.

#### (5) 낮은 인용빈도 수준에서의 적용 문제

현실적인 측면에서 h-지수가 가지는 다른 한계점은 인용이 활발하게 이루어지는 분야 중에서도 인용빈도가 어느 정도 이상인 주요 연구자에 대해서 측정할 경우에만 유효성이 보장된다는 점이다. 인용이 활발하지 않은 분야에서는

동물이 많이 나타날뿐더러, 빈도가 낮은 연구자는 앞의 <그림 1>과 상당히 다른 인용빈도 분포를 보이는 경우도 흔하다. 예를 들어 다섯 개의 논문이 인용된 두 연구자가 한 사람은 다섯 논문 모두 인용빈도가 1이고, 다른 사람은 네 논문은 인용빈도 1인데 나머지 한 논문은 인용빈도 10이 되는 경우일지라도 h-지수로는 둘 다 1이라는 값을 가진다. h-지수의 장점이라고 할 수 있는 지수의 강건성이 낮은 인용빈도 수준에서는 오히려 문제가 되는 것이다.

### 3.2 h-지수의 보완에 관한 기존 연구

#### (1) $h_I$ 지수

Batista et al.(2005)은 h-지수의 수준이 분야마다 다를 수 있음을 지적하고, 그 원인이 상이한 공저 관례 때문이라고 주장하였다. 이들은 공저자 수를 고려하여 보정한  $h_I$  지수를 제안하였다. 브라질의 4개 기초과학분야에 대해서 측정해본 결과 h-지수와 달리  $h_I$  지수는 값의 분포가 분야마다 큰 차이 없이 거의 일치하는 것으로 나타났다.  $h_I$  지수 공식은 다음과 같다.

$$h_I = h / \langle N_a \rangle = h^2 / N_a^{(T)}$$

$\langle N_a \rangle$ 는 h-지수 산정에 포함된 h개 논문의 평균 공저자 수이며  $N_a^{(T)}$ 는 h개 논문의 총 저자수(중복 허용)를 뜻한다. 이 공식은 분야별 차이를 극복할 수 있을 뿐만 아니라, 단독저자로 참여한 경우보다 공저자로 참여한 논문의 인용 영향력을 낮춰주는 효과도 얻을 수 있다. 다만 각 논문의 공저자 수를 모두 파악하지 않으면 산출이 불가능한 것이 단점이다.

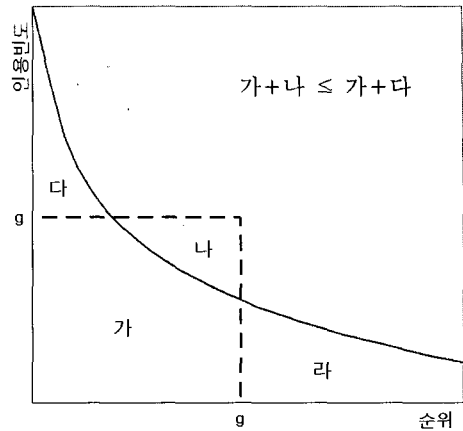
(2) g-지수

Egghe(2006a ; 2006b)는 일단 h-지수 산정에 포함된 인용빈도 상위 논문이 추가로 인용되더라도 지수의 향상에 전혀 기여하지 못하는 점을 비판하고 이를 보완하는 g-지수를 제안하였다.

주어진 논문집합에서 인용빈도 상위 g개 논문의 인용빈도 합이 g의 제곱 이상인 최하위 순위 g가 이 논문집합의 g-지수이다. 이 경우 상위 g+1개 논문의 인용빈도 합은 (g+1)의 제곱보다 작다.

풀어서 나타내면, 어떤 연구자의 논문을 인용빈도가 높은 순부터 나열하였을 때, 논문의 인용빈도 누적합계가 논문 순위의 제곱보다 크거나 같은 마지막 논문의 순위가 그 연구자의 g-지수가 된다. 위 정의에서 논문집합이라는 표현을 쓴 이유는 h-지수나 g-지수를 저자 단위가 아닌 다양한 대상에 대해서 측정하는 상황을 고려했기 때문이다.

g-지수 산출에 대한 이해를 돕기 위해 <그림 2>를 제시하였다. <그림 2>에서 보듯이 인용빈도 분포선과 가로, 세로 크기가 g인 정사각형을 그렸을 때 순위가 g 이상인 논문의 인용빈도 합은 '가' + '다' 영역이고, g의 제곱은 '가' + '나' 영역이다. 한 변의 길이가 g인 정사각형('가' + '나')의 넓이는 순위 g위 이상인 논문의 인용빈도 총합('가' + '다')에 대한 근사값이라고 할 수 있다. 결국 g-지수 산출이란 '나' 영역이 '다' 영역보다 작거나 같게 되는 가장 큰 정수 g를 구하는 문제이다.



<그림 2> g-지수 산출 개념도

예를 들어 g-지수가 20이라면, 인용빈도 상위 논문의 인용빈도 합이  $20 \times 20 = 400$ 회 이상이면서  $21 \times 21 = 441$ 회보다는 작다는 뜻이다. g-지수 산출과정에는 상위권 논문의 갯수만 아니라 개별 인용빈도가 그대로 반영되므로 이를 무시하는 h-지수에 비해서 직관과 더 일치하는 결과를 얻을 수 있다고 Egghe는 주장하였다.

<표 1>의 가상 자료에 대해 적용해보면 연구자 A, B, C, D의 h-지수는 각각 4, 4, 2, 2 이지만, g-지수는 5, 4, 2, 5가 된다. h-지수가 똑같이 4인 연구자 A와 B, 그리고 같은 2가 되는 연구자 C와 D중에서 상위 논문의 인용빈도가 더 높은 연구자 A와 연구자 D가 각각 더 큰 g-지수를 가진다. 따라서 g-지수로 평가한 결과가 직관에 더 가깝다는 Egghe의 주장은 설득력이 있다. 단점이라면, 인용빈도를 그대로 반영하므로 연구자 D와 같이 단 한 편의 논문 때문에 연구자 A와 대등한 높은 지수가 산출되는 경우가 발생할 수 있다는 점



이다. 반면에 h-지수의 장점은 한 두 편의 논문 때문에 지수가 크게 변화하는 것을 방지하는 특성이라는 점을 고려하면 g-지수와 h-지수는 상호 보완적이라고 볼 수 있다.

한편 Egghe(2006b)는 g-지수를 h-지수로 나눈 g/h도 흥미로운 척도라고 언급하였다. 이 값은 1 이상의 값을 가지며 값이 클수록 인용빈도가 최상위 논문에 집중됨을 뜻하는 척도가 된다.

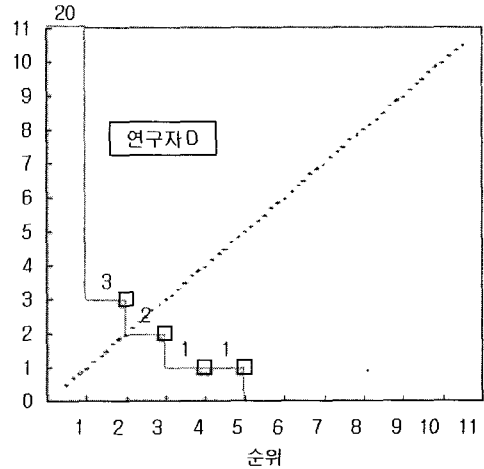
### (3) $h_r$ -지수와 $g_r$ -지수

h-지수뿐만 아니라 g-지수도 정수값으로 산출되며 지수가 10이내 정도 수준이면 동률 순위가 너무 흔해서 변별력이 약하다. 이를 다소 보완하기 위해서 소수점 이하까지 산출하는 방법도 가능하다.

Rousseau(2006a)는 순위가 h인 논문의 인용빈도가 h보다 클 경우에는 선형보간법을 이용해서 실수값인  $h_r$ 을 측정하는 방법을 언급하였다. 이때는 직선  $y=x$ 와 인용빈도 연결선과의 교점의 x좌표를 산출한다. 앞의 가상 데이터에서 연구자 D의 경우를 나타낸 <그림 3>을 보면 h-지수는 2인데 h위 문헌의 인용빈도는 3이다. 이때 점 (2, 3)과 점 (3, 2)를 이은 선과 45도 선과의 교점 (2.5, 2.5)의 x좌표(=y좌표)인 2.5가  $h_r$ -지수이다.

이와 비슷하게  $g_r$ -지수는 곡선  $y=x^2$ 과 각 순위에서의 인용빈도 누계 지점을 연결한 선과의 교차점의 좌표로 산출한다. 이 경우에도 교차점의 좌표는 가로축과 세로축이 같다.

실수값으로 측정하도록 g-지수를 보완하는 방법은 이보다 더 간단한 방법이 있다. 앞의



<그림 3> h위 문헌의 인용빈도가 h보다 큰 경우

<그림 2>에서 보았듯이 g-지수 산출은 한 번이 g인 정사각형의 면적으로 순위 g위 이상 논문의 총 인용빈도에 대한 근사값을 구하는 문제이다. 이때 근사값이 아닌 실제 g위 이상 논문의 인용빈도 총합의 제곱근을 산출하면 된다. 이 값은  $g=g_r$ 이 아닌 경우라면  $g_r$ 보다 아주 조금 작은 값이 되지만 그 차이는 무시해도 좋을 정도이며 우열을 가리기에는 충분하다.

$h_r$ -지수와  $g_r$ -지수는 각각 h-지수 및 g-지수보다 크거나 같게 되며 동률 순위를 다소 줄여주는 효과가 있지만 산출과정이 다소 복잡한 것이 문제이다. 또한 변별력을 다소 높여주는 것 이외에는 h-지수와 g-지수의 단점을 그대로 가진다.

## 4. h-지수와 g-지수의 개량 지수 제안

앞에서 살펴본 바와 같이 h-지수와  $h_r$ -지

수는 일단  $h$ 위 이상에 자리잡은 논문의 인용빈도가 이후에 아무리 높아지더라도 지수에 반영되지 않는 단점이 있다. 반면에  $g$ -지수와  $g_r$ -지수는 지나치게 최상위 논문의 인용빈도에 좌우된다는 약점이 있다. 이와 같이 상반된 두 가지 단점을 모두 극복할 수 있는 방안을 이하에 제시해보았다.

#### 4.1 $h_s$ -지수와 $g_s$ -지수

최상위 논문의 인용빈도에 전혀 영향받지 않는  $h$ -지수와, 반대로 최상위 논문의 인용빈도에 지나치게 영향받는  $g$ -지수의 절충방안으로,  $h$ 위 이내 논문의 인용빈도의 제곱근을 합하여 산출하는  $h_s$ -지수를 다음 공식과 같이 제안한다.

$$h_s = \sum_{r=1}^h \sqrt{C(r)}$$

여기서  $C(r)$ 은 순위가  $r$ 인 논문의 인용빈도이다.  $g$ -지수처럼 인용빈도를 그대로 합산하지 않고 제곱근을 취해서 합하는 이유는 네트워크의 빈익빈 부익부 원리(Barabasi 2002) 때문이다. 일단 인용빈도가 높아서 주목을 받은 논문은 그렇지 못한 논문에 비해서 추가 인용되기 쉽다. 실제로 Redner(1998)의 분석에 의하면 인용빈도별 논문의 수는 멱함수 법칙을 따르며 인용빈도가 2배 수준인 논문의 수는 대략  $2^3=8$ 배 정도 드문 것으로 나타났다. 따라서 인용빈도가 두 배라는 것이 곧 두 배의 영향력이라고 받아들이기는 어렵다.  $h$ 회 이상 인용된 논문의 인용빈도에 제곱근을 취하여 합산한  $h_s$ -지수는,  $h$ -지수와 달리 최상위

논문의 인용빈도 차이를 지수값에 반영하면서도  $g$ -지수처럼 한 두 논문에 의해 지수값이 좌우되는 상황은 방지하는 장점을 가진다.

기하학적으로 해석해보면  $h_s$ -지수와  $g$ -지수는 둘 다 인용빈도에 의한 면적을 정사각형으로 나타냈을 때 이 정사각형의 한 변의 길이를 측정하는 방식이다. 다만  $g$ -지수가 <그림 2>와 같이  $g$ 위 이상인 논문의 인용빈도를 모두 합한 면적의 정사각형을 다룬 반면에,  $h_s$ -지수는  $h$ 위 이상인 개별 논문의 인용빈도에 해당하는 면적을 가진 여러 정사각형을 각각 다루는 점이 다르다. 즉  $g$ -지수는 인용빈도를 합하여 제곱근을 취한 것에 가까운데,  $h_s$ -지수는 각 논문의 인용빈도에 제곱근을 먼저 취하여 합하는 점이 다른 것이다.

$h_s$ -지수는  $h$ -지수보다 항상 크거나 같으며  $h$ 가 1인 경우에만 같다. 최솟값은  $h$ 위 이내 논문의 인용빈도가 모두 동일하게  $h$ 인 경우에 가지게 되는  $h\sqrt{h}$ 이다. Egghe(2006b)가 제시한  $g/h$ 처럼 인용빈도가 최상위 논문에 집중되는 정도를 1부터 시작하는 값으로 나타내기 위해서는 다음 공식을 적용하면 된다.

$$\frac{h_s}{h\sqrt{h}}$$

<표 1>의 가상 데이터에 적용해보면 연구자 A, B, C, D의  $h_s$ -지수는 각각 10.0, 8.0, 3.5, 6.2이고, 최상위 집중도는 각각 1.13, 1.49, 1.00, 2.58이다.  $h$ -지수로는 동률이던 A와 B, 그리고 C와 D 사이의 우열이 드러나므로  $h_s$ -지수가 변별력이 높음을 알 수 있다. 또한 연구자 D의 경우에는 인용빈도가 최상위 논문에 집중된 덕분에  $g$ -지수로는 연구

자 A와 함께 공동 1위였지만,  $h_S$ -지수로는 네 명중 세 번째로 나타났다. 이와 같이  $h_S$ -지수는 최상위 논문의 인용빈도를 제한적으로 반영함으로써 h-지수와 g-지수의 상반된 단점을 동시에 극복함을 알 수 있다.

$h_S$ -지수와 동일한 방식으로 g-지수를 개량한  $g_S$ -지수를 측정할 수 있다.  $g_S$ -지수는 g-지수로 산출된 g보다 인용빈도 순위가 높거나 같은 논문의 인용빈도에 제곱근을 취하여 합하는 지수이다.  $h_S$ -지수보다 더 많은 논문의 인용빈도를 반영할 수 있다는 장점은 있으나, g 값 자체가 최상위 논문의 인용빈도에 크게 좌우되므로  $g_S$ -지수도 정도는 덜하지만 g-지수와 마찬가지로 문제점을 가진다.  $g_S$ -지수의 공식은 다음과 같다.

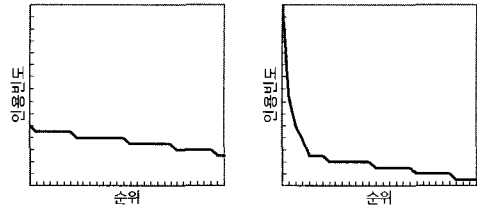
$$g_S = \sum_{r=1}^h \sqrt{C(r)}$$

## 4.2 L-지수

h-지수는 <그림 1>에서와 같이 인용빈도 분포선과 원점에서 출발한 45도 선과의 교차점의 x좌표에 해당한다. 그런데 연구자의 논문이 인용된 실제 빈도 분포는 곡선이 아닌 계단식으로 감소할뿐더러 <그림 1>과 같은 분포를 보이지 않고 <그림 4>와 같이 다양한 인용빈도 분포를 나타내는 경우도 적지 않다.

따라서 인용빈도가 최상위 논문에 집중된 분포를 보이면 상대적으로 45도 선과 교차하는 지점이 앞쪽으로 치우치게 되며, 상위 논문이 고르게 인용된 분포에서는 교차점이 뒤쪽으로 치우치게 된다.

이 연구에서는 h-지수에서 사용된 45도 선



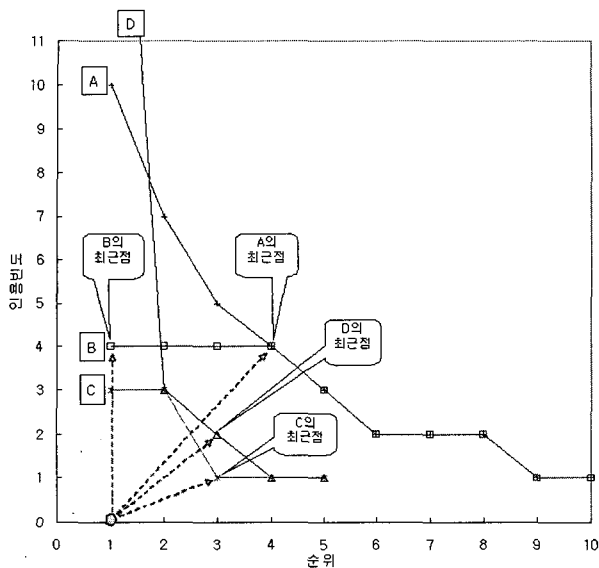
<그림 4> 다양한 인용빈도 분포 형태

을 버리고 각 연구자의 상이한 인용빈도 분포 형태를 반영할 수 있는 새로운 지수로 L-지수를 다음과 같이 제안한다.

특정 논문집합에 속한 논문을 인용빈도가 큰 순서로 정렬하여 순위가 r인 논문의 인용빈도를  $C(r)$ 이라고 하면, 이 논문집합의 L-지수는 각 점  $(r, C(r))$ 과 점  $(1, 0)$  사이의 거리  $L(r)$  중에서 최단거리이다. 거리는 유클리드 거리로서 다음 공식으로 산출한다.

$$L(r) = \sqrt{(r-1)^2 + C(r)^2}$$

풀어서 설명하자면, 각 논문의 인용빈도를 높은 것부터 그래프에 표시하였을 때 각 점과 좌표  $(1, 0)$ 과의 거리를 측정하여 가장 짧은 거리를 L-지수값으로 삼는다. 거리측정의 기준이 되는 원점으로  $(0, 0)$ 이 아닌  $(1, 0)$ 을 사용하는 이유는, 순위의 최저값은 1이고 인용빈도의 최저값은 0이기 때문이다. 만약 모든 논문이 인용빈도 0이라면 각 논문을 나타내는 점의 좌표는  $(1, 0), (2, 0), (3, 0)...$ 과 같이 되어 L-지수는 최단거리 0이 된다. 또한 모든 논문이 동일하게 n회 인용된 경우에는 L-지수는 최단거리 n이 된다.



〈그림 5〉 L-지수 산출 예

앞의 가상 사례에 대해서 최근점을 파악하여 L-지수를 산출하는 예를 〈그림 5〉에 제시하였다. 이 그림에서 연구자 A, B, C, D의 최근점은 각각 (4,4), (1,4), (3,1), (3,2)이고 L-지수는 5.0, 4.0, 2.2, 2.8이 된다. 연구자 B는 연구자 A와 h-지수로는 같은 4였지만, h 지점에서 상위로 올라가도 인용빈도가 증가하지 않아서 최근점이 원점 (1,0)과 더 가까워지는 결과가 되었다. 이와 같이 L-지수는 인용빈도의 분포 형태가 상이한 자료를 비교할 경우에 h-지수보다 공정한 산출이 가능하다는 것을 알 수 있다.

### 4.3 $h_A$ -지수

h-지수나 g-지수는 좌표축 공간에서 축에 내려 그은 연직선 상의 좌표를 이용하고, L-지수는 거리를 측정하여 지수값을 산출한다.

축 위의 좌표나 원점에서의 거리 이외에 점의 위치와 관련하여 흔히 측정되는 값으로 각도가 있다. 각도를 이용한 척도로 대표적인 것이 바로 코사인 계수이다. 이 연구에서는 이와 같은 각도와 코사인 계수를 이용한 지수인  $h_A$ -지수를 다음과 같이 제안한다.

특정 논문집합에 속한 논문을 인용빈도가 큰 순서로 정렬하여 순위가 r인 논문의 인용빈도를  $C(r)$ 이라고 하면 각 논문을 점  $(r, C(r))$ 로 나타낼 수 있다. h-지수로 산출된 순위 h 이상인 각 논문마다 원점 (0,0)으로부터 각 논문  $(r, C(r))$ 을 잇는 직선과 x축 사이의 각도에 대한 코사인 값을 구한 후 1에서 뺀 다음 합산한 값  $h_A$ 가 이 논문집합의  $h_A$ -지수이다. 이를 공식으로 나타내면 다음과 같다.

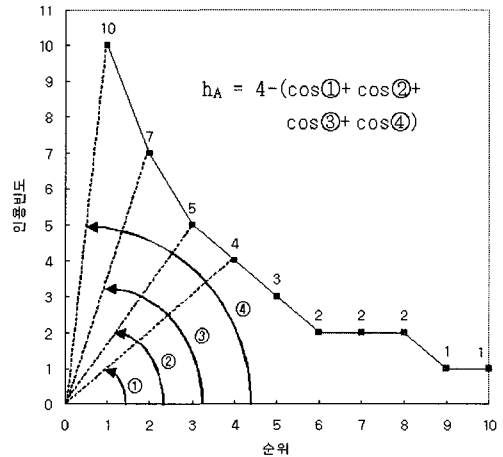
〈표 2〉 가상 자료에서  $h_A$ -지수 산출 사례

인용 순위	연구자 A의 논문		연구자 B의 논문		연구자 C의 논문		연구자 D의 논문	
1	10	0.900	4	0.757	3	0.684	20	0.950
2	7	0.725	4	0.553	3	0.445	3	0.445
3	5	0.486	4	0.400	1	0.051	2	0.168
4	4	0.293	4	0.293	1	0.030	1	0.030
5	3	0.143	3	0.143	1	0.019	1	0.019
6	2	0.051	2	0.051				
7	2	0.038	2	0.038				
8	2	0.030	2	0.030				
9	1	0.006	1	0.006				
10	1	0.005	1	0.005				
h	4		4		2		2	
$h_A$	2.40		2.00		1.13		1.40	

$$h_A = \sum_{r=1}^h \left(1 - \frac{r}{\sqrt{r^2 + C(r)^2}}\right)$$

$$= h - \sum_{r=1}^h \frac{r}{\sqrt{r^2 + C(r)^2}}$$

코사인 값을 1에서 빼주는 이유는 각도가 클수록 코사인 값은 작아지기 때문이다. 원점 (0,0)으로부터 각 논문 (r,C(r))을 잇는 직선과 x축 사이의 각도는 인용빈도가 높을수록 커지며 이론상 인용빈도가 무한대이면 90도가 된다. 이 각도에 대해서 코사인값을 구하면 최대 1(인용빈도 0일 때 각도 0도)에서 최소 0(인용빈도 무한대일 때 각도 90도) 사이의 값이 된다. 따라서 코사인 값을 1에서 빼면 각 논문은 인용빈도가 0일때 0부터 시작하여 인용빈도가 커질수록 1에 가까워지는 값을 가지게 된다.  $h_A$ -지수는 1 미만인 값을 h개 더하는 것이므로 동일한 대상에 대해서 산출하면 h-지수보다 항상 작다.



〈그림 6〉  $h_A$ -지수 산출 예

〈그림 6〉은 앞의 가상 자료에서 연구자 A의 경우를 나타낸 것이다. 연구자 A의 h-지수는 4이므로,  $h_A$ -지수는 4이내 논문으로 산출한다. 인용빈도 1위부터 4위까지 논문의 좌표는 각각 (10,1), (7,2), (5,3), (4,4)이고 원점에서 각 점을 이은 선과 x축과의 각도는 〈그림

6)의 ①, ②, ③, ④와 같다. 각각에 대해서 코 사인을 취한 후 1에서 빼면 0.90, 0.73, 0.49, 0.29가 되고 이를 합한 2.40이 연구자 A의  $h_A$ -지수이다. 같은 방법으로 <표 1>의 가상 자료의 연구자 B, C, D에 대해서  $h_A$ -지수를 산출하면 <표 2>와 같다.

가상자료에 대한 산출 결과에서 보듯이  $h_A$ -지수는  $h_S$ -지수와 마찬가지로 연구자 A와 B 사이의 우열과 연구자 C와 D 사이의 우열을 판별할 수 있으며, 연구자 D와 같이 한 논문 때문에 지나치게 높은 값을 가지는 것을 방지하는 특성이 있다.

## 5. 실제 자료에 대한 측정 결과

### 5.1 개인 연구자에 대한 측정

계량정보학 분야의 최고 권위인 Price 메달 수상자에 대해서는 Glnzel & Persson(2005)이 1986년 이후 인용 자료를 수집해서 분석한 바 있다. Egghe(2006b)는 1972년 이후의 인용 자료를 수집하여 더 완전한 분석을 수행하였는데, 여기에 포함된 14명에 대한 인용빈도 자료를 사용하여 앞에서 다룬 여러 지수를 산출해본 결과는 <표 3>과 같다. 각 지수의 특성을 비교하기 위해서 인용빈도 1위 논문의 인용빈도(max)와 연구자별 상위 16위 이내 논문의 인용빈도 합계(sum16)를 제시하였다. 16위 이내 합계를 산출한 이유는 Egghe(2006b)에 수록된 각 저자별 인용빈도 리스트 중에서 가장 짧은 경우가 16위까지였기 때문이다.

<표 3>에서 지수에 따라 순위가 가장 달라지는 저자는 White이다. White는  $h$ -지수는 12로 가장 낮은 값이지만,  $g$ -지수,  $h_S$ -지수,  $g_S$ -지수,  $h_A$ -지수로는 Leydesdorff, Egghe, Rousseau의 세 연구자보다 높게 나타났다. White의 상위 논문 인용빈도 합계(sum16)가 다른 세 연구자의 두 배 내외에 이르는 것을 감안할 때  $h$ -지수는 상위 논문의 개별 인용빈도를 무시하는 특성 때문에 White에 대해서 과소평가하였다고 판단된다. 정도는 덜하지만  $L$ -지수도  $h$ -지수와 비슷한 성향을 보였다.

<표 4>의 순위상관 분석 결과에서도 흥미로운 사실이 드러난다. 최댓값과의 상관을 보면 여섯 가지 지수 중에서  $g$ -지수와  $g_S$ -지수가 0.6대로 높았고, 그중에서도 특히  $g$ -지수는 유일하게 99% 수준에서 최댓값과 유의한 상관이 있는 것으로 나타났다(유의수준 99.2%). 앞에서 살펴본 바와 같이  $g$ -지수가 인용빈도를 그대로 합산하기 때문에 최상위 논문의 인용빈도에 크게 좌우된다는 점이 다시 확인되었다. 또한  $g_S$ -지수도 이런  $g$ -지수의 단점을 어느 정도는 공유하고 있음이 드러났다(98.4% 유의수준에서 최댓값과 상관 있음). 반면에 합산을 하지 않는  $h$ -지수와  $L$ -지수는 95% 유의수준에서도 최댓값과 상관이 없는 것으로 나타났다.

개인 연구자에 대한 지수 산출 결과를 검토해본 결과,  $h$ -지수는 상위 논문의 인용빈도가 높은 연구자에 대한 과소평가의 문제가 나타났고,  $g$ -지수는 최상위 논문의 인용빈도에 크게 좌우된다는 단점이 다시 확인되었다.

〈표 3〉 Price 메달 수상자 14인의 지수 산출 결과 (h-지수순 정렬)

연구자	max	sum16	h	g	h <sub>s</sub>	g <sub>s</sub>	L	h <sub>A</sub>
Garfield	625	2364	27	59	270.8	391.4	36.9	21.5
Narin	112	1063	27	40	190.5	243.9	35.6	18.2
Braun	125	913	25	38	165.7	227.3	33.6	16.5
van Raan	108	554	19	27	104.2	136.4	24.2	12.0
Small	305	1393	18	39	146.9	193.8	23.3	13.4
Schubert	124	714	18	30	111.1	158.9	24.8	12.2
Glänzel	124	555	18	27	99.8	136.9	25.5	11.9
Moed	108	614	18	27	104.7	136.7	24.2	12.0
Martin	156	634	16	27	95.5	133.7	22.6	11.1
Ingwerson	120	646	13	26	84.3	114.6	16.4	9.5
Egghe	47	345	13	19	61.5	82.5	17.0	8.4
Leydesdorff	79	351	13	19	61.7	81.6	17.0	8.4
Rousseau	25	238	13	15	51.0	57.9	15.3	7.6
White	128	601	12	25	74.9	111.5	17.0	8.6

〈표 4〉 개인 자료에 대한 지수간 스피어맨 순위상관

	h	g	h <sub>s</sub>	g <sub>s</sub>	L	h <sub>A</sub>
max	0.333	0.675	0.597	0.628	0.419	0.597
sum16	0.646	0.908	0.881	0.873	0.632	0.868

## 5.2 대학 단위의 측정

연구 성과의 측정 대상으로 자주 언급되는 단위가 바로 대학교이다. 교육인적자원부에서는 매년 SCI CD-ROM을 분석한 SCI급 논문 게재 실적을 국가별, 대학별로 발표하고 있다.

2003년 실적 발표 자료(교육인적자원부 2004)를 참고하여 게재 논문 수 상위 10개 대학을 대상으로 앞에서 다룬 여러 지수를 산출해보았다. 인용 정보는 출판년도가 2003년인

논문 중에서 저자 주소 항목에 각 대학의 이름이 포함된 경우를 Web of Science에서 검색하여 파악하였다. 모든 검색은 2006년 7월 13일과 14일에 걸쳐 수행하였고 산출 결과는 〈표 5〉에, 순위는 〈표 6〉에 제시하였다.

〈표 5〉와 〈표 6〉을 보면 1위와 2위는 max(최고 인용빈도)를 제외하면 서울대와 연세대로 모든 척도에서 동일하지만, 그 이하에서는 다소 차이가 있다. h-지수 기준으로는 KAIST와 포항공대가 연세대와 동률 2위로 나타나서 h-지수가 변별력이 낮음을 다시 확인할 수 있다. g-지수와 h<sub>s</sub>-지수, g<sub>s</sub>-지수로는 고려대가 3위로 나타났다. 고려대의 총 논문 수는 6위인데 반해서 최고인용빈도로는 2위임을 감안하면 이 세 지수는 최상위권 논문의 인용빈

〈표 5〉 2003년 국내 10개 대학 인용 지수

	논문 수	max	h	g	h <sub>s</sub>	g <sub>s</sub>	L	h <sub>A</sub>
서울대	3662	470	39	67	352.5	512.1	52.2	25.9
연세대	1741	259	30	52	245.0	357.0	41.0	21.0
KAIST	1474	70	30	38	190.7	233.5	39.8	18.9
성균관대	1214	197	23	35	145.3	200.1	29.7	14.9
한양대	1170	78	20	29	114.2	152.7	27.5	13.0
고려대	1147	313	27	47	210.1	304.5	37.5	18.8
포항공대	970	105	30	41	203.2	261.0	39.2	19.2
경북대	924	280	24	38	159.0	221.9	32.6	15.8
인하대	694	58	17	22	83.5	102.6	20.9	10.5
울산대	569	67	18	25	94.6	123.8	24.8	11.6

〈표 6〉 2003년 국내 10개 대학 인용 지수 순위

	논문 수	max	h	g	h <sub>s</sub>	g <sub>s</sub>	L	h <sub>A</sub>
서울대	1	1	1	1	1	1	1	1
연세대	2	4	2	2	2	2	2	2
KAIST	3	8	2	5	5	5	3	4
성균관대	4	5	7	7	7	7	7	7
한양대	5	7	8	8	8	8	8	8
고려대	6	2	5	3	3	3	5	5
포항공대	7	6	2	4	4	4	4	3
경북대	8	3	6	5	6	6	6	6
울산대	9	10	10	10	10	10	10	10
인하대	10	9	9	9	9	9	9	9

도에 영향받는다. L-지수와 h<sub>A</sub>-지수로는 포항공대가 3위로 나타나고 고려대의 순위는 5위로 낮아져서 두 지수가 최상위 논문의 인용빈도에 영향을 덜 받음을 알 수 있다.

전체적으로는 3위 이하에서 게재 논문 수에

따른 순위와 인용에 따른 순위는 다르게 나타났다. 성균관대와 한양대는 논문 수로는 고려대, 포항공대, 경북대보다 상위이지만, 인용 지수로는 이들보다 하위로 나타났다. 울산대와 인하대의 순위도 논문 수와 달리 인용 지수로는 맞바뀌었다.



〈표 7〉 국내 10개 대학 자료에 대한 지수간 스피어맨 순위상관

	h	g	h <sub>S</sub>	g <sub>S</sub>	L	h <sub>A</sub>
논문 수	0.742	0.693	0.721	0.721	0.794	0.745
max	0.644	0.839	0.806	0.806	0.661	0.685

각 지표간 스피어맨 순위상관을 구한 〈표 7〉을 보면 최댓값(max)과의 상관 측면에서 h-지수는 가장 낮고 g-지수는 가장 높다. 또한 h<sub>S</sub>-지수와 g<sub>S</sub>-지수도 최댓값과의 상관이 높게 나타난다. 최상위권 논문의 인용빈도를 지수에 반영하는 세 지수가 최댓값과의 상관이 높은 것은 당연한 결과이다. 이와 동일하게 h<sub>A</sub>-지수도 최상위권 논문의 빈도를 반영하지만 그 정도가 덜하다는 것이 분석결과에 나타났다. L-지수는 게재 논문 수와의 상관이 h-지수보다도 더 높아서 가장 높게 나타났고 최댓값과의 상관은 h-지수 다음으로 낮게 나타나서

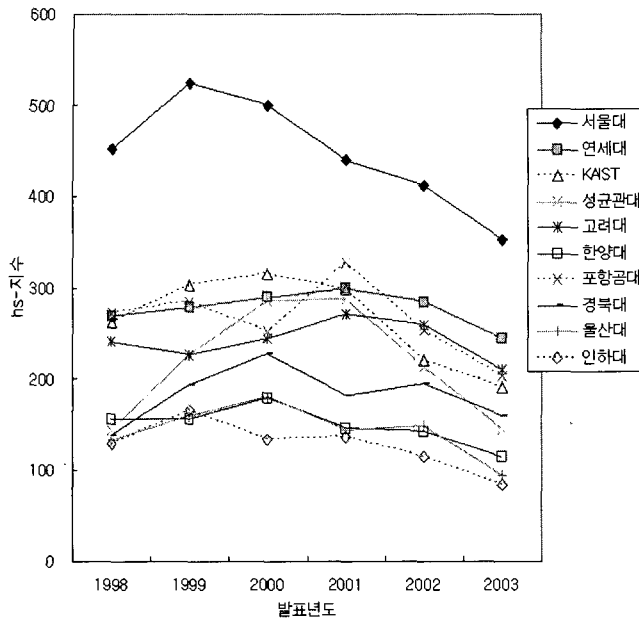
h-지수와 유사한 성향을 보임을 알 수 있다.

1998년부터 2003년까지 6년간 10개 대학의 국제학술지 발표논문 인용정보를 앞의 방법과 마찬가지로 Web of Science를 사용하여 검색하고 여러 지수 중에서 hS-지수를 산출한 결과를 〈표 8〉과 〈그림 7〉에 제시하였다.

최근에 발표된 논문일수록 인용될 기회가 적으므로 동일한 조건이라면 지수값은 최근 년간 감소하는 것이 정상이다. 서울대의 경우에 1999년에 발표한 논문들의 h<sub>S</sub>-지수가 가장 높고 이후 점차 감소하는 것은 이런 이유 때문이라고 해석된다. 그러나 KAIST, 한양대, 경북대, 울산대는 2000년 발표논문이 가장 높고 연세대, 성균관대, 포항공대, 고려대는 2001년 발표논문이 가장 높게 나타났다. 결국 2000년 이후 서울대와 상위권 사립대학의 h<sub>S</sub>-지수 격차가 다소 감소한 것이다. 또한 KAIST의 hS-지수 순위는 2000년의 2위를 고비로 2003년 발표논문 기준 5위까지 하락한 것으로 나타

〈표 8〉 1998-2003 hS-지수

	1998	1999	2000	2001	2002	2003
서울대	452.4	524.0	500.3	440.8	413.0	352.5
연세대	269.5	279.0	290.1	299.3	284.3	245.0
KAIST	262.9	303.2	315.7	298.4	220.0	190.7
성균관대	143.2	226.8	285.9	286.6	213.3	145.3
고려대	241.6	226.2	245.1	271.0	258.6	210.1
한양대	155.6	156.2	179.2	146.2	142.2	114.2
포항공대	273.0	284.9	251.9	327.6	252.6	203.2
경북대	139.1	193.7	227.4	181.2	194.9	159.0
울산대	132.5	159.5	179.6	143.9	148.3	94.6
인하대	129.1	165.8	134.5	136.0	114.0	83.5



〈그림 7〉 국내 10개 대학의 1998-2003 h<sub>s</sub>-지수

난다. 1단계 BK21 사업이 1999년부터 시작 되었음을 감안하면, BK21 사업을 통한 지원이 전통적으로 국가 지원이 집중되어 있던 서울대와 KAIST보다는 상위권 사립대학에 더 긍정적인 영향을 끼쳤음을 시사하는 결과이다.

## 6. 결론

인용을 통한 연구성과 측정 지표로 Hirsch가 제안한 h-지수와 이를 보완한 Egghe의 g-지수 및 기타 변형 지수를 살펴보고 이들의 단점을 보완할 수 있는 새로운 지수로 h<sub>S</sub>-지수, g<sub>S</sub>-지수, L-지수, h<sub>A</sub>-지수를 제안하였다.

가상 자료 및 실제 개인 연구자와 대학을 대상으로 지수를 산출해본 결과 파악된 각 지수

의 특징은 다음과 같다.

첫째, h-지수는 한 두 논문의 인용빈도에 좌우되지 않지만, 동률이 많이 나타나서 변별력이 낮으며 최상위 논문의 인용빈도 증가를 전혀 반영하지 못하는 단점이 있다.

둘째, g-지수는 h-지수와 달리 최상위 논문의 인용빈도를 반영하지만, 정도가 지나쳐서 한 두 논문의 인용빈도에 좌우될 여지가 크다.

셋째, 새롭게 제안한 지수는 모두 h-지수와 g-지수 사이의 절충 방안이 되어 최상위 논문의 인용빈도 차이를 제한적으로만 반영하는 특성을 가진다. 그중에서도 최상위 논문의 인용빈도 차이를 반영하는 정도는 h<sub>S</sub>-지수가 가장 크고 h<sub>A</sub>-지수가 그 다음, L-지수가 가장 적은 것으로 나타났다.

넷째, 가상 자료와 실제 자료에 대해서 적용

해본 결과 제안한 개량 지수가 h-지수 및 g-지수에 비해서 더 합리적인 결과를 산출하는 것으로 나타났다.

향후 h-지수를 비롯하여 이 연구에서 제안한 개량 지수를 활용하면 인용빈도를 이용한

성과 측정을 다방면으로 시도해볼 수 있을 것이다. 특히 국내 자료의 경우에는 외국에 비해서 인용빈도가 낮음을 고려해보면 h-지수보다는 변별력이 높은  $h_S$ -지수 등이 더 유용할 것으로 기대된다.

## 참 고 문 헌

- 교육인적자원부. 2004. 『03년도 SCI 과학기술논문 분석 내용』. 교육인적자원부.
- 윤희윤, 김신영. 2005. 학술지 영향계수와 연구업적 평가비중의 상관성 분석. 『정보관리연구』, 36(3) : 1-25.
- Adam, D. 2002. Citation analysis : The counting house. *Nature*, 415 : 726-729.
- Ball, P. 2005. Index aims for fair ranking of scientists. *Nature* 436 : 900. <<http://www.nature.com/nature/journal/v436/n7053/full/436900a.html>>.
- Banks, M. G. 2006. An extension of the Hirsch index : indexing scientific topics and compounds. Submitted to *Scientometrics*. <<http://arxiv.org/abs/physics/0604216>>.
- Barabasi, Albert-Laszlo. 2002. *Linked : The New Science of Networks*. Perseus Books Group.
- Bar-Ilan, J. 2006. H-index for Price medalists revisited. *ISSI Newsletter*, 2(1) : 3-5.
- Batista, P. D., M. G. Campiteli, O. Kinouchi, and A. S. Martinez. 2005. An index to quantify an individual's scientific research valid across discipline. Submitted to *Scientometrics*. <<http://arxiv.org/abs/physics/0509048>>.
- Bornmann, L. and H. D. Daniel. 2005. Does the h-index for ranking of scientists really work? *Scientometrics*, 65(3) : 391-392.
- Braun, T., W. Glanzel, and A. Schubert. 2005. A Hirsch-type index for journals. *The Scientist*, 19(22) : 8.
- Cronin, B., and L. Meho. 2006. Using the h-index to rank influential

- information scientists. *Journal of the American Society for Information Science and Technology*, 57(9) : 1275-1278.
- Egghe, L. 2006a. An improvement of the H-index : the G-index. *ISSI Newsletter*, 2(1), 8-9.
- Egghe, L. 2006b. Theory and practise of the g-index. To be published in *Scientometrics*.
- Glänzel, W. 2006. On the h-index : a mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, 67(2) : 315-321.
- Glänzel, W., and O. Persson. 2005. H-index for Price medalists. *ISSI Newsletter*, 1(4) : 15-18.
- Hirsch, J. E. 2005. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102 : 16569-16572.
- H-number. [online]. [cited 2006.7.10]. <<http://www.brics.dk/~mis/hnumber.html>>.
- Hofbauer, R., M. Frass, B. Gmeiner, and A. D. Kaye. 2002. *Euro-Factor<sup>TM</sup> : The New European Scientific Currency*. Vienna : VICER Publishing. [online]. <cited 2005.9.20>. <<http://www.vicer.org/b-eurofact.htm>>.
- Lehmann, S., A. D. Jackson, and B. E. Lautrup. 2006. Measures and mismeasures of scientific quality. [online]. [cited 2006.6.30]. <<http://arxiv.org/abs/physics/0512238>>.
- Moed, H. F. 2005. *Citation Analysis in Research Evaluation*. Springer.
- Redner, S. 1998. How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B*, 4 : 131-134.
- Rousseau, R. 2006a. A case study : evolution of JASIS' Hirsch index. [online]. [cited 2006.6.28]. <<http://eprints.rclis.org/archive/00005430/>>.
- Rousseau, R. 2006b. New developments related to the Hirsch index. [online]. [cited 2006.6.28]. <<http://eprints.rclis.org/archive/00006376/>>.