

공저자 수를 고려한 h-지수 산출*

Calculating the h-index and Its Variants Considering the Number of Authors in a Paper

이재윤 (Jae Yun Lee)**

초 록

연구자 성과 평가를 위해 널리 사용되는 h-지수는 일관성 부족 문제와 공저자 수를 고려하지 않는다는 문제를 가지고 있다. 이를 극복하기 위해 h-지수와 g-지수, 그리고 공저 보정 방안을 검토하고 2004년부터 2013년 사이의 실제 KCI 데이터를 대상으로 분석해본 결과는 다음과 같다. 첫째, 일관성 결여 문제를 해소하기 위해서는 g-지수를 사용하는 것이 더 바람직하다고 판단된다. 둘째, 연구 성과의 양적인 측면과 질적인 측면을 한꺼번에 반영하는 복합 지수라는 h-지수와 g-지수의 고유한 특성을 유지하기 위해서는 반드시 공저를 보정하여 지수를 측정해야 한다. 셋째, 공저자 수로 나눈 인용빈도를 사용하는 h_C -지수와 g_C -지수를 적용하면 단독 저술 비중이 높은 인문학 분야 연구자도 공정하게 평가할 수 있고, 특정 분야나 특정 기관에 속한 연구자가 상위 순위를 과점하는 현상을 방지할 수 있다.

ABSTRACT

The h-index is a popular bibliometric indicator for evaluating individual researchers. However, it has been criticized for its inconsistency with reflecting increased number of citations and disregarding the number of co-authors in a paper. In order to overcome these problems, we examined the g-index and other Hirsch-type indices considering the number of co-authors. Test data collection was extracted from Korean Citation Index database published from 2004 to 2013. The results of this study are as follows: First, g-index is more reliable indicator than h-index with consistency. Second, number of co-authors must be considered to maintain the h-index as a complex indicator applying the quality and the quantity of research performance. Finally, h_C -index and g_C -index, with fractionalised counting of the papers, can fairly measure the research performance of humanities researchers, and successfully prevent specific disciplines or institutions occupying majority of top rankings.

키워드: 연구성과, h-지수, 공저자, 공동연구, g-지수, 연구 평가
research performance, h-index, co-authors, research collaborations, g-index,
research evaluation

* 본 논문에는 한국연구재단에서 구축하여 제공하는 한국학술지인용색인(KCI) DB 정보를 이용하였음.

** 명지대학교 문헌정보학과 부교수(memexlee@mju.ac.kr)

■ 논문접수일자: 2016년 5월 31일 ■ 최초심사일자: 2016년 9월 11일 ■ 게재확정일자: 2016년 9월 11일
■ 정보관리학회지, 33(3), 7-29, 2016. [http://dx.doi.org/10.3743/KOSIM.2016.33.3.007]

1. 서론

물리학자인 Hirsch(2005)가 제안한 연구자의 성과 평가 지수인 h-지수는 계량정보학 분야에 큰 파란을 일으켰다. 2001년부터 2010년 사이의 10년 간 계량정보학 분야 논문에서 인용된 논문을 분석한 연구(Lee & Choi, 2011)에서는, 모든 시대에 발표된 논문을 통틀어서 2005년에 발표된 Hirsch의 h-지수 제안 논문이 21세기 초반 10년 동안 가장 많이 인용된 것으로 나타났다. Hirsch(2005)가 다소 복잡하게 제시한 정의를 다른 연구자들이 쉬운 말로 다듬었는데, Rousseau(2006)는 “연구자 S가 단독 혹은 공저로 발표한 문헌을 인용빈도순으로 나열하고 동률을 무시하여 일련번호를 부여하였을 때, 상위 h개의 문헌이 최소 h회 인용되고 h+1번째 문헌의 인용빈도가 h+1회 미만이면 S의 h-지수는 h이다”라고 정의했으며, 이재운(2006)은 “어떤 연구자의 논문을 인용빈도가 높은 순부터 나열하였을 때 논문의 인용빈도가 논문의 순위보다 크거나 같은 마지막 논문의 순위가 그 연구자의 h-지수가 된다”라고 정의했다.

h-지수는 폭넓게 사용되고 있는 만큼 여러 상황을 고려한 다양한 단점이 지적되어왔다. 그중에서도 h-지수를 제안한 Hirsch 자신부터 최근까지 꾸준히 지적되어온 사항은 공저를 많이 하는 연구자에게 유리하다는 문제이다. 이는 h-지수를 실제 연구자 평가에 적용할 때 공정성을 가장 심각하게 저해하는 단점이다. 단독으로 발표한 연구이든, 10명이 공저한 연구이든 간에 구분하지 않고 h-지수 산출에 반영하기 때문이다. 학문 분야별로도 단독 연구를

많이 하는 분야의 연구자가 공동 연구를 많이 하는 분야의 연구자보다 불리하다는 문제가 발생하기도 한다. 또한 비교적 최근에는 Waltman과 Van Eck(2012)이 h-지수의 일관성 결여 문제를 지적하여 h-지수의 무비판적인 사용에 경종을 울린 바 있다.

이 논문에서는 이와 같이 h-지수에 대해서 지적되어온 주요 한계점을 공저 보정 문제와 함께 살펴본 후, 이에 대한 대안을 검토해보고 국내 연구자의 연구 성과 측정에 적용해서 결과를 검토해보았다. 이를 통해 공저를 보정하여 국내 연구자의 연구 성과를 측정하는 방안을 제시하고자 하였다.

2. 이론적 배경

2.1 h-지수와 g-지수

Hirsch(2005)가 제시한 h-지수의 정의는 문장으로 표현되어 있는데, 이를 공식으로 다시 표현해보면 인용빈도 순위 j번째 논문의 인용빈도를 cit_j 라고 할 때 다음과 같다.

$$h = \max_j \{j: cit_j \geq j\}$$

예를 들어 h-지수가 5라면, 최소 5회 이상 인용된 논문을 5편 발표했다는 뜻이다. h-지수 산출에 반영되는 h위 이내의 논문들을 별도로 Hirsch-core 또는 h-core라고 부른다. h-지수는 연구자 개인을 평가하기 위해 개발된 지수이지만 개인 이외에도 논문 평가(이재운, 최상

회, 2015; Schubert, 2009), 연구기관별 성과 평가(이재윤, 2006; 박지연, 김정은, 민윤경, 2010; van Raan, 2006), 저널 평가(고영만, 조수련, 박지영, 2013; 김판준, 이재윤, 2010; Braun, Glänzel, & Schubert, 2005), 연구주제 평가(Banks, 2006; Bar-Ilan, 2008) 등을 비롯한 다양한 영역에 응용되고 있다.

h-지수는 손쉽게 계산할 수 있으며 인용 데이터가 일부 불완전하더라도 산출에 큰 문제가 없다(Vanclay, 2007). 그러나 다음과 같은 단점이 지적되고 있다. 첫째, 동률값이 흔하여 변별력이 낮다(이재윤, 2006; Batista, Campiteli, Kinouchi, & Martinez, 2006). 둘째, 자기인용에 좌우되는 경향이 다소 있다(Cronin & Meho, 2006; Kelly & Jennions, 2006). 셋째, 분야 간 비교가 어렵다(Batista et al., 2006; Cronin & Meho, 2006). 넷째, 연구의 양에 많이 영향 받는다(이재윤, 2006; van Raan, 2006). 다섯째, 인용빈도가 낮은 연구자를 평가할 경우 효용성이 떨어진다(이재윤, 2006). 여섯째, 공저를 많이 하는 연구자에게 유리하다(Batista et al., 2006; Hirsch, 2005). 이런 단점을 개선하기 위한 여러 연구가 계속되고 있으며 h-지수를 개량하거나 보완한 사례가 다수 발표되었다.

h-지수를 보완하여 제시된 지수 중에서 가장 대표적인 것은 Egghe(2006)가 제안한 g-지수이다. h-지수 산출과 마찬가지로 특정 논문집합에 대하여 인용빈도순으로 정렬한 리스트를 생성한 다음, 상위 g개 논문의 인용빈도 합이 g의 제곱 이상인 최하위 순위 g가 이 논문집합의 g-지수 값이 된다. 최근에는 g-지수를 달리 해석하여 g순위까지 각 논문의 인용빈도 평균이 g회 이상인 경우로 보기도 한다(Schreiber, 2010).

이 해석에 따라 g-지수를 공식으로 나타내보면, 인용빈도 순위 j번째 논문의 인용빈도를 cit_j 라고 할 때, 다음 공식을 만족하는 가장 큰 g값이 g-지수이다.

$$g \geq \frac{1}{g} \sum_{j=1}^g cit_j$$

이와 같은 g-지수는 최상위 인용논문의 인용빈도를 반영하지 않는 h-지수의 단점을 보완한 지수로서 h-지수보다 변별력이 높다는 장점이 있다. 반면에 g-지수는 최상위 논문의 인용빈도에 크게 좌우된다는 단점이 있다(이재윤, 2006). 그러나 국내의 연구자 집단 규모가 해외 전체에 비해서 훨씬 작으므로 인용빈도도 낮게 나타난다는 차이점 때문에 이런 단점이 희석될 수 있다. 국내 인용데이터베이스인 KCI는 인용빈도 1위 논문의 인용빈도가 2016년 6월 14일 현재 341회이고, 10위 논문의 인용빈도가 208회에 머무르고 있으므로 수 천회 이상인 해외 인용데이터베이스에 비해서 g-지수를 적용할 때 고인용 논문으로 인한 문제가 덜 발생한다. 따라서 국내 인용데이터베이스를 이용한 연구자 평가를 위해서는 g-지수의 적용을 적극적으로 검토해볼만 하다.

2.2 평가 지수로서 h-지수의 일관성 문제

최근에 지적된 h-지수의 중요 단점은 개인이나 연구기관의 연구성과 평가에서 연구가 추가되거나 연구자가 추가될 경우에 순위가 직관에 반하여 뒤바뀌는 성향이 있을 수 있다는 점이다. Waltman과 Van Eck(2012)은 규모 의

존 지수(size-dependent indicators)와 규모 무관 지수(size-independent indicators)를 구분하면서 h-지수의 일관성 결여 문제를 지적하였다. 규모 의존 지수는 주로 누적 수치로 산출하므로 연구 성과가 추가될 때 지수 값이 감소되는 경우가 없는 지수이다. 규모 의존 지수의 대표는 논문 수, 총 인용빈도 등이며 고인용 논문 수(인용빈도 1% 상위, 5% 상위 등), h-지수도 이에 해당한다. 반면에 규모 무관 지수는 누적 수치를 연구성과 총량으로 정규화하는 지수이다. 대표적으로 논문 당 평균 인용빈도(CPP), 인용빈도의 중앙값 등이 있으며 고인용 논문 비율도 이에 해당한다. 동일한 지수 구분을 유소영과 이재윤(2008)은 합산 지향 지수와 평균 지향 지수로 구분하였는데, 합산 지향 지수 혹은 규모 의존 지수는 인용빈도나 논문 수만 고려한 것으로서 일반적으로 연구성과의 총 영향력을 측정하는 것으로 해석되며, 평균 지향 지수 혹은 규모 무관 지수는 논문 수를 분모로 하여 평균적인 영향력을 측정하는 것으로 해석된다.

Waltman과 Van Eck(2012)은 h-지수로 연구자의 성과를 비교할 경우에 발생하는 문제점을 다음과 같이 세 가지로 나누어 제기하였다.

첫째, 두 연구자가 각자 이전과 비교해서 동일한 비율의 성과를 달성하더라도 두 연구자의 h-지수가 역전될 경우가 있다. 이는 h-지수가 인용빈도와 논문수 증가를 선형적으로 반영하지 않기 때문에 나타날 수 있는 현상이다. 예를 들어 <표 1>의 왼쪽과 같이 연구자 A는 12편의 논문을 발표하면서 그중 9편은 12회, 나머지 3편은 4회의 인용을 받았고, 동료 연구자 B는 10편을 발표한 가운데 그중 7편은 15회, 나머지 3편은 5회의 인용을 받았다면 각자의 h-지수는

9와 7이 된다. 그런데 5년 후 <표 1>의 오른쪽과 같이 두 연구자가 이전 5년과 동일한 연구성과(같은 수의 추가 논문이 동일한 인용빈도를 가질 경우)를 산출할 경우에는 두 연구자의 h-지수가 10과 12로 역전되어 연구자 B가 더 높은 지수값을 가지게 된다. 반면에 g-지수를 기준으로 할 경우 5년 전의 h-지수는 연구자 A가 10, 연구자 B가 12이고 동일한 연구성과를 거둔 5년 후의 h-지수는 연구자 A가 12, 연구자 B가 14가 되어 일관되게 연구자 B가 높은 지수값을 가지게 된다.

둘째, 두 연구자가 동일한 수준의 추가 연구 성과를 얻었음에도 불구하고 h-지수 순위가 역전될 수 있다. <표 2>에서와 같이 연구자 C와 D가 각각 7편의 연구성과를 가지고 있는데, C는 5회 인용 5편, 2회 인용 2편이고, D는 6회 인용 4편, 3회 인용 3편이라고 하였을 때, 두 연구자의 h-지수는 각각 5와 4가 된다. 그런데 두 연구자가 공저로 두 편의 논문을 발표하였는데 두 편 모두 8회 인용된 경우 연구자 C의 h-지수는 변하지 않지만 D의 h-지수는 6이 되어 C보다 더 높은 h-지수를 가지게 된다. 반면에 공동연구 이전의 연구자 C와 D의 g-지수는 5로 동일하고, 공동연구 이후에도 g-지수는 6으로 두 명이 동일하게 상승하게 된다.

셋째, h-지수가 더 낮은 수준의 연구자 2명이 소속된 기관이, 더 높은 수준의 연구자 2명이 소속된 기관보다 더 높은 기관 h-지수를 가질 수 있다. 앞의 첫 번째 사례에서 5년 전과 5년 후를 동일한 연구자의 성과로 간주하지 않고 같은 수준의 성과를 얻은 두 연구자라고 간주할 경우에, h-지수가 9인 연구자 2명이 소속된 기관은 h-지수가 12가 되고 h-지수가 7인

〈표 1〉 두 연구자 A와 B의 5년 전/후의 각 논문의 인용빈도와 연구성과
(Waltman & Van Eck, 2012의 표에 g-지수를 추가하여 일부 수정함)

논문 순서	5년 전 인용빈도		5년 후 인용빈도	
	연구자 A	연구자 B	연구자 A	연구자 B
1	12	15	12	15
2	12	15	12	15
3	12	15	12	15
4	12	15	12	15
5	12	15	12	15
6	12	15	12	15
7	12	15	12	15
8	12	5	12	15
9	12	5	12	15
10	4	5	12	15
11	4		12	15
12	4		12	15
13			12	15
14			12	15
15			12	5
16			12	5
17			12	5
18			12	5
19			4	5
20			4	5
21		4		
22		4		
23	4			
24	4			
합계	120	120	240	240
평균	10	12	10	12
h-지수	9	7	12	14
g-지수	10	12	12	14

〈표 2〉 두 연구자 C와 D의 공동연구 이전/이후의 각 논문의 인용빈도와 연구성과
(Waltman & Van Eck, 2012의 표에 g-지수를 추가하여 일부 수정함)

논문 순서	공동연구 이전 인용빈도		공동연구 이후 인용빈도	
	연구자 C	연구자 D	연구자 C	연구자 D
1	5	6	8	8
2	5	6	8	8
3	5	6	5	6
4	5	6	5	6
5	5	3	5	6
6	2	3	5	6
7	2	3	5	3
8			2	3
9			2	3
합계	29	33	45	49
평균	4.14	4.71	5	5.44
h-지수	5	4	5	6
g-지수	5	5	6	6

연구자 2명이 소속된 기관은 h-지수가 14가 되는 경우가 발생할 수 있다. 이 경우에도 g-지수를 기준으로 평가하면 두 기관의 연구성과 평가는 높은 수준의 성과를 얻은 두 연구자가 소속된 기관이 낮은 수준의 성과를 얻은 두 연구자가 소속된 기관보다 일관되게 높게 유지된다.

이처럼 Waltman과 Van Eck(2012)이 지적한 h-지수의 일관성 결여 문제는 g-지수를 평가기준으로 삼을 경우에 대부분 해결됨을 알 수 있다. 현재 KCI에서는 h-지수만 제시하고 있는데, 평가지수를 적용할 때의 일관성 문제와 국내의 낮은 인용빈도 수준을 감안하여 g-지수의 채택을 깊이 고려해야 한다.

3. 공저를 보정한 h-지수와 g-지수 산출법

h-지수를 제안한 Hirsch(2005)는 공저를 많이 하는 저자에게 유리하게 지수가 측정되는 문제가 있으므로 평균 공저자 수로 h-지수를 나누는 방식을 고려해볼만하다고 언급한 바 있다. 이후에 여러 연구에서 공저자 수를 감안하여 지수를 산출할 수 있도록 h-지수를 변형하는 시도가 있었다. Hu, Rousseau, Chen(2010)은 공저자를 고려해서 h-지수를 산출하는 방식을 다음과 같이 네 가지로 구분하였다.

- (1) h-지수를 저자 수 평균으로 나누어주기 (Batista et al., 2006), 또는 고에너지물리학과 같이 극단적으로 저자가 많은 경우를 대비해서 저자 수 중앙값으로 나누어주기.

- (2) h-지수를 h-core에 있는 저자 수의 제곱근으로 나누어주기(Wan, Hua, & Rousseau, 2007), 또는 모든 논문에 대해서 저자 수의 제곱근으로 나누어주기(Chai, Hua, Rousseau, & Wan, 2008)
- (3) 인용 빈도나 논문 순위를 분수 계상(fractional counting)하여 h-지수 산출하기 (Egghe, 2007; Schreiber, 2008a, 2008b).
- (4) 저자 기입순서와 역할을 반영하기(Hagen, 2008; Zhang, 2009).

각 방식에 대해서 대표적인 변형 지수를 중심으로 아래에서 다루어본다.

3.1 h₁-지수

Batista 등(2006)은 h-지수의 수준이 분야마다 다를 것을 지적하고, 그 원인이 상이한 공저 관례 때문이라고 주장하였다. 이들은 공저자 수를 고려하여 보정한 h₁-지수를 제안하였다. 브라질의 4개 기초과학분야에 대해서 측정해본 결과 h-지수와 달리 h₁-지수는 값의 분포가 분야마다 큰 차이 없이 거의 일치하는 것으로 나타났다. h₁-지수 공식은 다음과 같다.

$$h_1 = \frac{h}{\langle N_a \rangle} = \frac{h^2}{N_a^{(T)}}$$

$\langle N_a \rangle$ 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 있다고 Batista 등(2006)은 주장했다. 이들이 공저자 수를 보정하기 전에는 물리학, 화학, 생물학, 수학 분야의 h-지수의 분포가 매우 다르게 나타났다. 그러나 공저자 수를 보정한 이후의 h_i -지수에서는 지수값을 평균으로 나누고 순위를 최댓값으로 나누어서 정규화해보면 네 분야의 지수 분포가 거의 겹치게 나타나므로 분야별 차이가 극복된 것으로 판단하였다.

공저자 수를 보정한 h_i -지수는 h-core 내 논문의 평균 저자 수만 사용하여 보정하기 때문에 h보다 낮은 순위에 있는 단독 저술 논문이 고려되지 않는 문제가 있다. 예를 들어 h가 10인 연구자의 h-core 내에 있는 논문이 2명이 공저한 12회 인용된 논문인 경우에는 $12/2=6$ 이고, 8회 인용된 단독 저술 논문이 있으면, 이 논문은 h-core에 속한 논문보다 저자 1인당 피인용 횟수가 많음에도 불구하고 지수 산출에 반영되지 않는다. 따라서 h-지수를 산출한 후에 공저자 수를 보정하는 h_i -지수는 불합리하다.

또한 h_i -지수는 공저자 수가 예외적으로 많은 논문이 h-core에 포함되어 있을 경우에 공저자 수 평균을 지나치게 높여서 불합리한 결과를 산출한다. 예를 들어 <표 4>의 가상 사례에서 연구자 A의 h-지수는 6이다. 그런데, 6위 이내에 포함된 논문의 공저자 수 평균은 2가 된다. h-core에 포함된 연구자 A의 논문은 대부분 단독 저술이지만 예외적으로 D3번 논문의 공저자가 6명으로 많기 때문에 커진 값이다. 이에 따라 연구자 A의 h_i -지수는 $6/2=3$ 이 된다. 그런데, 만약 연구자 A가 3번 논문에 참여하지 않았다고 가정할 경우, 즉 3번 논문을 발표하지 않았다고 가정하고 제외하게 되면, <표 5>와 같이 h-지수는 6이고 6위 이내 논문의 평균 저자 수는 1.17이어서 h_i -지수는 5.14가 된다. 이처럼 저자 수가 많은 특정 논문을 제외했을 때 오히려 h_i -지수가 커지는 현상이 발생할 수도 있는 것은 h-지수를 먼저 산출한 후에 공저자 수로 보정하는 방식의 치명적인 문제점이다.

<표 3> 연구자 A의 h_i -지수 산출

인용빈도 순위	논문번호	인용빈도 C	저자 수 Na	C / Na
1	D1	20	1	20
2	D2	17	1	17
3	D3	12	6	2
4	D4	10	2	5
5	D5	9	1	9
6	D6	8	1	8
7	D7	6	1	6
8	D8	6	2	3
9	D9	5	1	5
10	D10	5	1	5
h-core 평균		12.67	2.00	10.17
h_i -지수		$h / \langle Na \rangle = 6 / 2.00 = 3.00$		

〈표 4〉 연구자 A의 공저자가 많은 3번 논문을 제외하고 h_I -지수 산출

인용빈도 순위	논문번호	인용빈도 C	저자 수 Na	C / Na
1	D1	20	1	20
2	D2	17	1	17
3	D4	10	2	5
4	D5	9	1	9
5	D6	8	1	8
6	D7	6	1	6
7	D8	6	2	3
8	D9	5	1	5
9	D10	5	1	5
h-core 평균		11.67	1.17	10.83
h_I -지수		$h / \langle Na \rangle = 6 / 1.17 = 5.14$		

〈표 5〉 예외적으로 공저자가 많은 논문을 포함할 경우와 제외할 경우의 h-지수와 h_I -지수 산출 결과

	3번 논문 포함할 경우	3번 논문 제외할 경우
h	6	6
h-core의 평균 저자 수	2.00	1.17
h_I -지수	$6 / 2 = 3.00$	$6 / 1.17 = 5.14$

3.2 Pure h-지수(h_P -지수)와 Adapted Pure h-지수(h_{ap} -지수)

Wan, Hua, 그리고 Rousseau(2007)는 공저자 수를 고려한 지수로 Pure h-지수를 제안하였다. 이들은 Batista 등(2006)의 h_I -지수와 유사하게, h-core에 포함된 논문들의 평균 저자수의 제곱근으로 h를 나누는 방법을 사용하였다. 이들은 평균 저자수를 그대로 분모로 사용하지 않고 제곱근을 취해 사용함으로써 Batista 등(2006)의 h_I -지수와 달리 예외적으로 많은 공저자가 있는 상위 논문 때문에 불이익을 받는 것을 피할 수 있다고 주장하였다. 그러나 앞의 가상 사례에 대해서 〈표 6〉과 같이 pure h-지

수를 산출해보면, 3번 논문을 포함하면 4.24인데 이를 제외하면 오히려 5.55로 높은 값이 산출되어 정도의 차이만 있을 뿐 동일한 문제를 가진 것으로 나타난다. 이처럼 h-지수를 먼저 산출한 후 h-core에 포함된 논문의 공저자 수를 반영하는 방식은 모두 치명적인 단점을 가지고 있다.

Wan, Hua, 그리고 Rousseau(2008)는 pure h-지수를 다시 개량하여 h-core를 먼저 확정하지 않고, 각 논문마다 저자수의 제곱근으로 인용빈도를 나눈 값을 해당 논문의 등가 인용빈도(equivalent number of citations)로 삼은 후, 선형 보간법을 응용한 공식으로 다음과 같은 절차를 거쳐 adapted pure h-지수(h_{ap} -지

〈표 6〉 예외적으로 공저자가 많은 논문을 포함할 경우와 제외할 경우의 h-지수, h_f-지수, h_p-지수 산출 결과

	3번 논문 포함할 경우	3번 논문 제외할 경우
h-지수	6	6
h _f -지수	3.00	5.14
h _p -지수	4.24	5.55

수)를 산출하는 방법을 추가로 제안하였다.

여 다음 공식을 적용한다.

- 단계 1: 연구자 A가 공저한 각 논문의 공저자 수를 파악한다.
- 단계 2: 각 논문의 실제 인용빈도를 공저자 수의 제곱근으로 나눈 값으로 등가 인용빈도(equivalent number of citations)를 산출한다.
- 단계 3: 등가 인용빈도 내림차순으로 A의 각 논문의 순위를 도출한다.
- 단계 4: 등가 인용빈도가 정수로 떨어지지 않고 실수이므로 adapted pure h-지수(h_{ap}-지수)도 다음과 같이 실수로 산출한다. r위 논문의 등가 인용빈도를 C_E(r)이라고 하고, 선형 보간값을 C_E(x)라고 하

$$h_{ap} = \frac{(h_e + 1) \times C_E(h_e) - h_e \times C_E(h_e + 1)}{C_E(h_e) - C_E(h_e + 1) + 1}$$

등가 인용빈도 내림차순으로 산출한 h-지수를 h_e라고 할 때 h_{ap}는 h_e와 h_e+1 사이의 값이 되도록 선형 보간법으로 구한다. 〈표 7〉의 경우에 h_e는 3이고 C_E(3)은 4이며 C_E(4)는 3.78이므로, 선형 보간법을 적용하면 adapted pure h-지수(h_{ap}-지수)는 다음과 같이 3.82가 된다.

$$\begin{aligned} h_{ap} &= \frac{(h_e + 1) \times C_E(h_e) - h_e \times C_E(h_e + 1)}{C_E(h_e) - C_E(h_e + 1) + 1} \\ &= \frac{(3+1) \times 4 - 3 \times 3.78}{4 - 3.78 + 1} \\ &= 3.82 \end{aligned}$$

〈표 7〉 h_{ap}-지수 산출 사례

인용빈도 순위	논문 번호	인용빈도 C(r)	저자 수	등가 인용빈도 C _E (r)	등가 인용빈도 순위
1	D1	10	7	3.78	4
2	D2	10	7	3.78	5
3	D3	8	5	3.58	6
4	D4	8	5	3.58	7
5	D5	6	4	3	8
6	D6	6	4	3	9
7	D7	5	1	5	1
8	D8	5	1	5	2
9	D9	4	1	4	3
10	D10	4	2	2.83	10
11	D11	4	2	2.83	11

앞서의 h_I -지수 산출 사례에 적용해보면, h_e 는 6이고 $C_E(6)$ 은 6이며 $C_E(7)$ 은 5이므로, 선형 보간법을 적용하면 adapted pure h-지수 (h_{ap} -지수)는 다음과 같이 6이 된다.

$$h_{ap} = \frac{(h_e + 1) \times C_E(h_e) - h_e \times C_E(h_e + 1)}{C_E(h_e) - C_E(h_e + 1) + 1}$$

$$= \frac{(6 + 1) \times 6 - 6 \times 5}{6 - 5 + 1}$$

$$= 6$$

그러나 선형 보간법으로 지수 값을 실수로 산출하는 방법은, h-core의 경계에 있는 두 값만 활용하므로 h-core에 속한 전체 논문의 인용빈도를 반영하지 못하기 때문에 불합리한 면이 있다. 예를 들어 h_e 가 3이면 3위 논문의 등가 인용빈도와 4위 논문의 등가 인용빈도에 의해서 지수가 좌우되고, 1위 논문의 인용빈도는 아무리 크더라도 지수 값의 차이가 없게 된다.

3.3 h_m -지수

Schreiber(2008a, 2008b)가 제안한 h_m -지수는 순위를 분수 계상(fractional counting)하여

실질 순위(effective rank)를 산출하는 것이 특징이다. 즉, 순위가 1씩 증가하는 것이 아니라, 1/공저자수 만큼 증가하는 방식을 사용한다. 따라서 가장 많이 인용된 논문이 공저자 수가 2명이면 실질 순위 r_{eff} 는 1위가 아니라 $1/2 = 0.25$ 위가 된다. 두 번째로 많이 인용된 논문의 공저자 수가 2명이라면 이 논문의 실질 순위는 $1/2 + 1/2 = 0.5$ 위가 된다. 이렇게 인용빈도와 실질 순위를 동시에 분수 계상하여 산출해둔 후, h-지수 계산과 마찬가지로 실질 순위가 인용빈도보다 작거나 같은 가장 큰 순위를 파악한다. <표 8>에서 실질 순위가 3인 일곱 번째 논문까지가 인용빈도가 실질 순위보다 크거나 같으므로 h_m -지수는 7이 된다.

<표 8>의 연구자의 h-지수는 5이다. 5위 이내의 공저자 수 평균은 $12/5 = 2.4$ 이므로 h_I -지수는 $5/2.4 = 2.08$ 이다. 이런 과정에서 인용빈도가 3회로 h_I -지수인 2.08보다 높은 6위와 7위 논문은 지수 산출에 반영되지 못한다. 이처럼 h_m -지수는 h_I -지수보다 더 많은 논문을 고려하여 지수를 산출하게 된다는 장점이 있다. 그러나 h_m -지수 산출 방식은 공저자가 많은 논문일수록 실질 순위의 증가가 더디게 되어 오히려

<표 8> h_m -지수 산출 사례

논문 번호	인용빈도 순위 r	인용빈도 C	저자 수 Na	저자 당 인용빈도 C/Na	실질 순위 r_{eff}
D1	1	16	2	8	0.5
D2	2	15	2	7.5	1
D3	3	14	3	4.67	1.33
D4	4	12	3	4	1.67
D5	5	10	2	5	2.17
D6	6	3	2	1.5	2.67
D7	7	3	3	1	3
D8	8	2	1	2	4

〈표 9〉 상위 논문의 공저자가 매우 많은 경우의 h_m -지수 산출 사례

논문 번호	인용빈도 순위 r	인용빈도 C	저자 수 Na	저자 당 인용빈도 C/Na	실질 순위 r_{eff}
D1	1	16	10	1.6	0.10
D2	2	15	10	1.5	0.20
D3	3	14	10	1.4	0.30
D4	4	12	10	1.2	0.40
D5	5	10	10	1	0.50
D6	6	3	10	0.3	0.60
D7	7	3	3	1	0.93
D8	8	2	1	2	1.93

〈표 10〉 상위 논문이 모두 단독 저술일 경우의 h_m -지수 산출 사례

논문 번호	인용빈도 순위 r	인용빈도 C	저자 수 Na	저자 당 인용빈도 C/Na	실질 순위 r_{eff}
D1	1	16	1	16	1.00
D2	2	15	1	15	2.00
D3	3	14	1	14	3.00
D4	4	12	1	12	4.00
D5	5	10	1	10	5.00
D6	6	3	2	1.5	5.50
D7	7	3	3	1	5.83
D8	8	2	1	2	6.83

지수가 높게 판정될 수 있다는 치명적인 약점이 있다. 예를 들어 〈표 8〉에서 인용빈도 순위 1위부터 6위까지의 논문들의 공저자를 모두 10 명씩으로 〈표 9〉처럼 변경해보면, 실질 순위는 $1/10=0.1$, $1/10+1/10=0.2$, ...와 같이 느리게 증가하여 h_m -지수가 오히려 7에서 8로 높아지는 결과를 가져온다. 즉, 공저자가 많을수록 오히려 지수가 높게 산출되는 것이다.

반대로 〈표 8〉의 h_m -지수 산출 사례에서 인용빈도 순위 1위부터 5위까지의 논문들을 모두 〈표 10〉과 같이 단독 논문으로 변경해보면, 실질 순위는 1, 2, ...와 같이 빠르게 증가하여 h_m -지수가 오히려 7에서 5로 낮아지는 결과를 가져

온다. 즉, 단독저술을 많이 할수록 오히려 지수가 낮게 산출되는 것이다.

3.4 h_c -지수와 g_c -지수

h_c -지수는 이 연구에서 이름 붙인 것으로서 공저 보정 h-지수(co-authorship corrected h-index)의 약자이다. 유사한 내용의 지수를 Schreiber (2008a)는 h_f -지수, Harzing(2010)은 hI_{norm} 이라고 각각 지칭하였는데, 지수의 내용을 잘 반영할 수 있도록 이 연구에서는 공저보정 h-지수로 일관되게 부르기로 한다. h_c -지수, 즉 공저 보정 h-지수는 다음과 같은 절차로 산출한다.

- 1단계: 각 논문의 인용빈도를 저자 기여도에 비례하도록 나눈 보정 인용빈도를 산출한다.
- 2단계: 보정 인용빈도 내림차순으로 논문의 보정된 순위를 동률 순위 없이 산출한다.
- 3단계: 보정된 순위가 보정 인용빈도보다 작거나 같은 가장 큰 보정된 순위를 공저 보정 h_c -지수(h_c -지수)로 삼는다.

이때 1단계에서 저자 기여도는 가장 단순한 동등기여방식(1/저자수)에서부터 제1저자나 교신저자를 우대하는 다양한 차등기여방식(이종욱, 양기덕, 2011; Lee & Chung, 2014)을 적용하여 산출할 수 있다. 예를 들어 앞의 <표

3>에 제시되었던 사례에 대해서 동등기여방식으로 h_c -지수를 산출해보면 <표 11>과 같다. 이 표는 저자 수로 인용빈도를 나눈 보정 인용빈도 순위로 정렬한 표이므로 논문번호와 순위가 일치하지 않는다. <표 11>에서 보정 인용빈도 5위까지의 논문은 보정 인용빈도가 보정 인용빈도 순위보다 크거나 같은 반면, 6위 이하의 논문은 보정 인용빈도가 보정 인용빈도 순위보다 작으므로 연구자 A의 h_c -지수는 5가 된다. 이런 결과는 인용빈도가 높고 공저자가 많은 D3번 논문을 포함하거나 제외하거나 동일하다. 반면에 h_i -지수는 3이었다가 D3번 논문을 제외하면 5.14로 비정상적으로 급증하는 문제가 발생한다.

<표 11> 연구자 A의 h_c -지수 산출

보정 인용빈도 순위	논문번호	인용빈도 C	저자 수 Na	보정 인용빈도 $C_c = C / Na$
1	D1	20	1	20
2	D2	17	1	17
3	D5	9	1	9
4	D6	8	1	8
5	D7	6	1	6
6	D4	10	2	5
7	D9	5	1	5
8	D10	5	1	5
9	D8	6	2	3
10	D3	12	6	2

<표 12> 예외적으로 공저자가 많은 논문을 포함할 경우와 제외할 경우의 h -지수, h_i -지수, h_c -지수 산출 결과

	3번 논문 포함할 경우	3번 논문 제외할 경우
h -지수	6	6
h_i -지수	3.00	5.14
h_c -지수	5	5

동등기여방식의 h_f -지수를 Schreiber(2008a)는 분수 계상(fractional counting)을 적용했다는 의미로 h_f -지수라고 불렀고, 인용 분석 소프트웨어 Publish or Perish(Harzing, 2010)에서는 $h_{f,norm}$ 이라고 표기하고 있다. 이 방식에 대해서 Schreiber(2008a)와 Harzing(2010)는 상반된 견해를 제시하고 있다. Schreiber(2008a)는 논문의 순위를 인용빈도가 아닌 공저자 수를 고려한 값으로 새롭게 산출하게 되면 인용빈도가 높은 논문이 공저자 수가 많다는 이유로 지수 산출에서 제외되는 이상한 현상이 발생한다고 부정적인 의견을 제시하였다. 반면에 Harzing(2010)은 이 방법이 Batista 등(2006)의 h_f -지수나 Schreiber(2008a)의 h_m -지수보다 정교하고 저자 1인당 영향력을 더 근사하게 산출해준다고 옹호하였다.

본 논문에서도 h_f -지수는 공저자가 많은 고 인용 논문에 지나치게 좌우되는 문제가 있고, h_m -지수는 단독저술을 많이 할수록 오히려 지수가 낮게 산출되는 문제가 발생하는 것으로 파악되었다. 반면에 h_c -지수는 이같은 문제가 발생하지 않고 저자 1인당 영향력을 반영하므로 공저를 고려한 연구성과 평가에 더 적합한 지수라고 판단된다.

공저자 수 이외에도 Wan, Hua, 그리고 Rousseau(2007)는 저자 순서까지 고려하자는 주장을 하였으나 현실적으로 저자 순서가 반드시 공헌도에 비례하는 것은 아니므로 다소 무리가 있다. Hu, Rousseau, 그리고 Chen(2010)은 주 저자, 즉 제1저자나 교신저자로 발표한 논문만 고려하는 h-maj-index를 산출하여 h-지수를 보완하자고 주장하였다. 이 방법은 주 저자 이외의 공헌은 아예 제외하는 것이 단점이며, 인용 데

이터베이스에서 교신저자를 별도로 표기하지 않는 경우에는 적용할 수 없다는 한계가 있다. 국내 인용색인 데이터베이스인 KCI에서 h-지수 방식으로 연구자를 평가하려면 주저자만이 아닌 모든 저자의 기여를 반영하는 동등기여방식의 h_f -지수를 적용하거나, KCI에서 교신저자가 별도로 표시되지 않으므로 제1저자만 우대하는 차등기여방식의 h_f -지수를 적용하는 것이 바람직할 것이다.

g_c -지수는 공저 보정 g-지수(co-authorship corrected g-index)의 약자로서 h_c -지수처럼 동등기여방식이나 차등기여방식으로 보정인용빈도를 산출한 후에 Schreiber(2010)의 g-지수 산출 방식을 적용한 것이다. 공저 보정 g_c -지수는 다음과 같은 절차로 산출한다.

- 1단계: 각 논문의 인용빈도를 저자 기여도에 비례하도록 나눈 보정 인용빈도를 산출한다.
- 2단계: 보정 인용빈도 내림차순으로 논문의 보정된 순위를 동률 순위 없이 산출한다.
- 3단계: 보정된 순위까지 각 논문의 보정 인용빈도 평균이 g 회 이상인 가장 큰 보정된 순위를 공저 보정 g-지수(g_c -지수)로 삼는다.

예를 들어 앞의 <표 11>에서 동등기여방식으로 h_c -지수를 산출했듯이 동등기여방식으로 g_c -지수를 산출해보면 <표 13>과 같다. <표 13>에서 보정 인용빈도 8위까지의 논문은 보정 인용빈도의 누적 합계가 75이고 평균이 $75/8=9.375$ 로서 보정 인용빈도 순위 8보다 크거나 같은 반면, 9위까지의 보정 인용빈도 누적 합계가 78이

〈표 13〉 연구자 A의 g_c -지수 산출

보정 인용빈도 순위	논문번호	인용빈도 C	저자 수 Na	보정 인용빈도 $C_c = C / Na$	보정 인용빈도 누적 평균
1	D1	20	1	20	20.0
2	D2	17	1	17	18.5
3	D5	9	1	9	15.3
4	D6	8	1	8	13.5
5	D7	6	1	6	12.0
6	D4	10	2	5	10.8
7	D9	5	1	5	10.0
8	D10	5	1	5	9.4
9	D8	6	2	3	8.7
10	D3	12	6	2	8.0

고 평균이 $78/9 \approx 8.667$ 로서 보정 인용빈도 순위 9보다 작으므로 연구자 A의 g_c -지수는 8이 된다.

4. KCI 연구자 h-지수 및 변형 지수 측정

4.1 KCI 데이터 및 공저 보정 방법

이 연구에서는 KCI 연구자 인용 데이터를 대상으로 단순한 h-지수와 함께 h_c -지수(공저 보정 h-지수), g_c -지수(공저 보정 g-지수)를 측정해보았다. 분석에 사용된 인용 데이터는 KCI에 구축된 2004년부터 2013년까지 발행된 논문의 인용 데이터이다.

공저자 수를 보정하는 방식 중에서 Schreiber (2008a)의 h_m -지수는 공저자 수가 많을 경우에 오히려 유리하게 측정되는 문제가 발생하므로 제외하였다. 또한 Batista 등(2006)의 h_r -지수도 h-core 내 논문의 평균 저자 수만 사용하

여 보정하기 때문에 h보다 낮은 순위에 있지만 1인당 인용빈도가 높은 논문이 고려되지 않으며, 공저자 수가 예외적으로 많은 논문이 h-core에 포함되어 있으면 공저자 수 평균을 지나치게 높여서 이 논문을 제외하였을 때 오히려 h_i -지수가 커지는 현상이 발생할 수도 있는 치명적인 문제가 있으므로 제외하였다.

이 연구에서 적용하는 h_c -지수(공저보정 h-지수)의 산출 절차는 다음과 같다.

- 1단계: 동등기여방식에 따라 각 논문의 인용빈도를 공저자 수로 나눈 보정 인용빈도를 산출한다.
- 2단계: 보정 인용빈도 내림차순으로 논문의 보정된 순위를 동률 순위 없이 산출한다.
- 3단계: 보정된 순위가 보정 인용빈도보다 작거나 같은 가장 큰 보정된 순위를 해당 연구자의 h_c -지수로 삼는다.

동일한 방식으로 g-지수를 변형한 g_c -지수(공저보정 g-지수)의 산출 절차는 다음과 같다.

- 1단계: 동등기여방식에 따라 각 논문의 인용빈도를 공저자 수로 나눈 보정 인용빈도를 산출한다.
- 2단계: 보정 인용빈도 내림차순으로 논문의 보정된 순위를 동률 순위 없이 산출한다.
- 3단계: 1위부터 보정된 순위 g 까지 각 논문의 보정 인용빈도 평균이 g 회 이상인 가장 큰 g 값을 해당 연구자의 g_c -지수로 삼는다.

또한 비교를 위해서 각 연구자의 논문 수(P), 총 인용빈도(TC), 논문별 평균 인용빈도(CPP)를 산출하고 각각에 대해서 논문별 공저자 수를 나누어 보정한 공저 보정 논문 수(보정 P), 공저 보정 인용빈도(보정 TC), 공저 보정 논문별 평균 인용빈도(보정 CPP)를 함께 산출하여 분석하였다.

지수 간 비교를 위해서 논문 수가 5편 미만인 연구자는 분석 결과를 왜곡시킬 수 있으므로 제외하였다. 또한 소속 연구자가 10명 미만인 학문분야(한국연구재단 학문분야분류표 상의 중분류)도 한 두 연구자에 의해 분야 수치가 지나치게 좌우될 수 있으므로 분석에서 제외하였다.

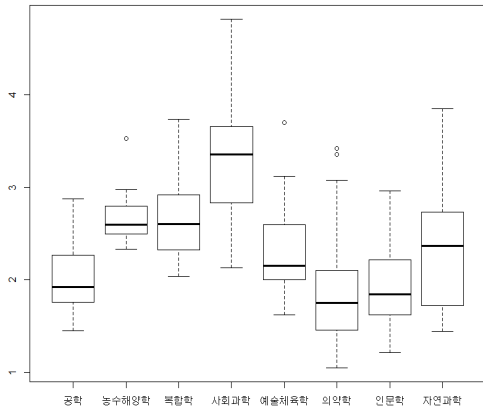
4.2 학문분야별 지수 평균값 비교 분석

국제 학술지에 논문을 투고하는 비율이 높은 이공 계열 국내 연구자와 달리 인문사회 계열의 국내 연구자들은 주로 국내 학술지에 논문을 발표하고 있다. 인문사회 계열 중에서도 특히 인문학 분야 연구자들은 단독 저술의 비중이 높다. 이와 같은 경향이 h-지수의 측정에 미

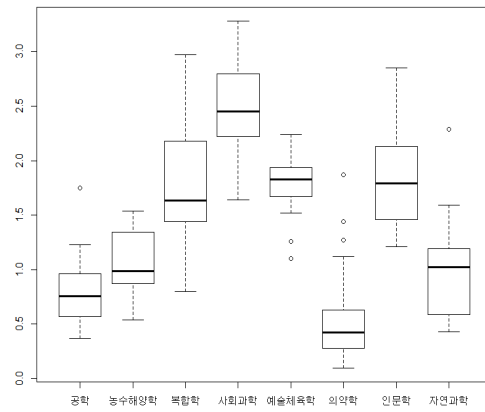
치는 영향을 살펴봄과 동시에, 공저 보정에 따라 학문분야별 지수 평균이 어떻게 달라지는지를 살펴보았다.

공저를 보정하기 전과 보정한 후의 지수 측정 결과를 비교하기 위해서 소속 연구자가 10명 이상인 155개 학문분야마다 소속 연구자들의 h-지수와 h_c -지수 평균을 산출하였다. h-지수의 155개 학문분야별 평균값을 학문분야분류표의 8개 대분류별로 나누어 분포를 살펴본 결과는 <그림 1>과 같다. 이를 보면 같은 인문사회 계열에 속한 사회과학과 인문학의 h-지수 수준이 상반되게 나타난다. 사회과학에 속한 학문분야의 h-지수 평균이 전체 8개 대분류 중에서도 가장 높은 수준 반면, 인문학에 속한 학문분야의 h-지수 평균은 의약학에 속한 학문분야들과 함께 가장 낮은 수준을 보이고 있다. 이공 계열에 속한 학문분야의 연구자들은 KCI 등재지 이외에 해외 학술지에도 많은 논문을 투고하고 국내 학술지 논문을 인용하는 경우는 상대적으로 적음을 감안하면 인문학 분야 연구자들의 h-지수가 상당히 저평가되고 있다고 할 수 있다.

<그림 2>는 공저를 보정한 h_c -지수의 155개 학문분야별 평균값을 KCI 분류표 상의 8개 대분류별로 나누어 분포를 제시한 결과이다. 공저를 보정하기 전과 달리 인문학에 속한 학문분야의 지수값이 이공계열에 해당하는 공학, 농수해양학, 의약학, 자연과학 분야에 비해서 뚜렷하게 높은 수준으로 나타난다. 공저를 보정하여 지수를 측정한 결과 국내 저술 및 단독 저술의 비중이 높은 인문학 분야의 연구자들에 대한 저평가 문제가 해결된 것이다. 전체적으로 공저를 보정한 이후에는 인문사회 계열에



〈그림 1〉 155개 학문분야별 소속 연구자들의 h-지수 평균을 8개 대분류로 나누어 파악한 분포



〈그림 2〉 155개 학문분야별 소속 연구자들의 hc-지수 평균을 8개 대분류로 나누어 파악한 분포

해당하는 인문학, 사회과학, 예술체육, 복합학 분야의 평균 지수가 이공 계열에 해당하는 분야보다 뚜렷하게 높은 것으로 나타났으며, 복합학, 예술체육, 인문학에 속한 분야들의 공저 보정 h-지수 값은 서로 비슷한 수준이었다. 결국 공저를 보정하지 않은 h-지수로는 국내 저술 비중이 높은 인문사회 계열 중에서도 단독 저술의 비중이 높은 분야가 저평가되는 문제가 있으나, 공저를 보정한 hc-지수로는 이런 문제가 해소됨이 확인되었다.

h-지수와 g-지수는 연구의 양적인 측면과 질적인 측면을 반영하는 복합 지표이다. 일반적으로 연구의 양은 논문 수로 측정하며 연구의 질은 논문 당 평균 인용빈도로 측정한다. h-지수와 g-지수가 연구의 양과 질을 복합적으로 반영하므로 논문 수 및 논문 당 평균 인용빈도와는 상관관계가 모두 통계적으로 유의하게 나타나야 한다. 이를 확인하기 위해서 155개 학문분야 별로 해당 연구자들의 논문 수, 논문 당 인용빈도, h-지수, g-지수 평균을 산출한 다

음, 4개 변수의 155개 학문분야 별 평균 값 사이의 상관계수를 산출해보았다. 공저를 보정하지 않은 4개 변수 간 상관계수는 〈표 14〉에, 공저를 보정한 4개 변수 간 상관계수는 〈표 15〉에 제시하였다.

공저를 보정하지 않았을 때에는 〈표 14〉와 같이 h-지수와 g-지수가 연구의 질을 반영하는 논문 당 인용빈도와는 매우 높은 상관관계를 보인다. 그러나 연구의 양을 반영하는 논문 수와의 상관관계는 h-지수의 경우 약한 상관관계를 보이며, g-지수의 경우 아예 통계적으로 무관한 것으로 나타났다. 즉, 연구의 양이 많은 학문분야라고 해서 해당 분야 연구자의 g-지수도 높지는 않음을 의미한다. 결국 공저를 보정하지 않은 경우에는 g-지수가 연구 성과의 양적인 측면을 제대로 반영하지 못하고 있으며, h-지수가 연구 성과의 양적인 측면을 반영하는 정도도 강하지 않음이 확인되었다.

반면에 공저를 보정한 후에 상관관계를 측정 한 〈표 15〉에서는 hc-지수와 gc-지수가 연구

〈표 14〉 공저를 보정하지 않은 지수 간 상관계수

		h-지수	g-지수
논문 당 인용빈도	Pearson 상관계수	.939**	.966**
	유의확률 (양쪽)	.000	.000
	N	155	155
논문 수	Pearson 상관계수	.165*	.131
	유의확률 (양쪽)	.040	.103
	N	155	155

** . 상관계수는 0.01 수준(양쪽)에서 유의.

* . 상관계수는 0.05 수준(양쪽)에서 유의.

〈표 15〉 공저를 보정한 지수 간 상관계수

		hc-지수	gc-지수
보정된 논문 당 인용빈도	Pearson 상관계수	.964**	.982**
	유의확률 (양쪽)	.000	.000
	N	155	155
보정된 논문 수	Pearson 상관계수	.811**	.766**
	유의확률 (양쪽)	.000	.000
	N	155	155

** . 상관계수는 0.01 수준(양쪽)에서 유의.

의 양을 반영하는 논문 수 및 연구의 질을 반영하는 논문 당 인용빈도와 매우 강한 상관관계를 보이는 것으로 나타났다. 이로써 공저를 보정하여 산출한 hc-지수와 gc-지수는 연구 성과의 양적인 측면과 질적인 측면을 모두 강하게 반영하고 있음을 알 수 있다.

4.3 지수별 상위 연구자 다양성 분석

연구 성과 평가를 위한 지수가 공저를 고려하여 산출되지 않는다면, 공동연구를 함께 자주 수행한 연구팀이 상위 순위에 동반하여 자리잡는 상황이 발생할 수 있다. 또한 공동연구가 활발한 분야의 연구자들이 단독 연구를 주로 하는 분야의 연구자들을 제치고 상위 순위

를 과점할 가능성도 높다. 이를 확인하기 위해 KCI 데이터에서 소속과 연구분야가 등록된 109,551명의 데이터를 분석하여 각 지수별 상위 50위 이내에 자리한 연구자의 소속 분야와 소속 기관을 살펴보았다. 지수별 순위가 동일할 경우 논문 당 평균 인용빈도(CPP)로 순위를 구분하였다.

〈표 16〉에서 보듯이 h-지수 기준으로는 50위 이내에 포함된 연구자의 소속 분야가 10개에 불과하였으나, 공저를 보정한 hc-지수 기준으로는 15개 분야의 연구자가 50위 이내에 포함되었다. g-지수 기준으로는 14개, 공저를 보정한 gc-지수 기준으로는 16개 분야의 연구자가 50위 이내에 포함되었다. h-지수 기준 상위 50위 이내에 가장 많이 포함된 분야는 관광학 분야로서

〈표 16〉 전체 상위 50위 이내 연구자의 소속 분야와 기관 종수

	50위 이내 연구자의 소속 분야		50위 이내 연구자의 소속 기관	
	분야 종수	분야 Entropy	기관 종수	기관 Entropy
h-지수	10	2.94	29	4.57
g-지수	14	3.44	26	4.33
hc-지수	15	3.47	37	4.86
gc-지수	16	3.38	34	4.70

〈표 17〉 관광학 분야 지수별 10위 이내 연구자 소속기관 분석

순위	h-지수	g-지수	hc-지수	gc-지수
1	경희대학교	세종대학교	경주대학교	동명대학교
2	세종대학교	경희대학교	경희대학교	제주지역사업평가원
3	배재대학교	경희대학교	동명대학교	한양대학교
4	세종대학교	경희대학교	진주대학교	경희대학교
5	경희대학교	제주지역사업평가원	제주지역사업평가원	경희대학교
6	경희대학교	경희대학교	경희대학교	경기대학교
7	경희대학교	경희대학교	경기대학교	경주대학교
8	우송대학교	세종대학교	경남대학교	경남대학교
9	세종대학교	배재대학교	대구대학교	대구대학교
10	경희대학교	동명대학교	대구대학교	동국대학교
2인 이상 포함 기관	경희대(5명), 세종대(3명)	경희대(5명), 세종대(2명)	경희대(2명), 대구대(2명)	경희대(2명)

13명이 포함되었으며, 그 다음으로 체육이 9명, 교육학이 8명 순이었다. 공저를 보정한 hc-지수 기준 상위 50위 이내에 가장 많이 포함된 분야는 교육학 분야로서 12명이 포함되었으며, 그 다음으로 한국어외문학 분야, 체육 분야, 관광학 분야가 각각 5명씩 포함되었다. 이처럼 공저를 보정한 결과 상위에 포함된 연구자의 소속 분야가 다양해지고 특정 분야가 과점하는 정도가 완화되었음을 확인할 수 있었다.

소속 기관 측면을 살펴보면, h-지수 기준으로는 50위 이내에 포함된 연구자의 소속 기관이 29개에 불과하였으나, 공저를 보정한 hc-지수 기준으로는 37개 기관의 연구자가 50위 이내에 포함되었다. g-지수 기준으로는 26개, gc-지수

기준으로는 34개 기관의 연구자가 50위 이내에 포함되었다. 이처럼 공저를 보정한 결과 상위에 포함된 연구자의 소속 기관이 다양해지는 결과를 확인할 수 있었다.

h-지수 기준 상위 50위 이내에 가장 많은 연구자가 포함된 관광학 분야로 제한하여 상위 연구자의 소속 기관을 살펴본 결과는 〈표 17〉과 같다. 지수별 관광학 분야 상위 10위 연구자의 소속 기관을 보면, 동일 기관에 속한 저자가 10위 이내에 2명 이상이 함께 포함되는 경우는 공저를 고려한 gc-지수를 기준으로 한 경우가 가장 적은 것으로 나타났다. 반면에 h-지수 기준일 때에는 10위 이내에 경희대 소속 연구자가 5명, 세종대 소속 연구자가 3명으로 상위권

을 과점하여서 다양성이 매우 적게 나타났다. h-지수와 g-지수 모두가 공저를 보정한 경우에 상위권 연구자의 소속 기관이 다양하게 됨을 알 수 있다.

5. 결론

한국연구재단 KCI에서 서비스되고 있는 연구자 평가 지수인 h-지수는 동등한 수준의 연구 성과가 추가된 두 연구자의 지수값이 일관되게 증가하지 못하고 연구자 간 우열도 뒤바뀔 수 있다는 단점이 있다. 또한 공저를 많이 하는 연구자에게 지나치게 유리하다는 단점이 있어서 국내 연구자의 연구 성과 평가 척도로 사용하기에는 심각한 문제점을 가지고 있다. 이를 극복하기 위해 h-지수와 g-지수, 그리고 공저 보정 방안을 검토하고 실제 KCI 데이터를 대상으로 분석해본 결과는 다음과 같다.

첫째, h-지수와 g-지수 중에서 국내 연구자 성과 평가를 위해서는 g-지수를 사용하는 것이 더 바람직하다고 판단된다. g-지수가 상위권 인용논문의 인용빈도를 반영하므로 h-지수보다 변별력이 높다는 점은 잘 알려져 있었으나, 이 연구에서 검토한 바에 따르면 h-지수의 단점으로 최근 Waltman과 Van Eck(2012)이 지적한 일관성 결여 문제도 g-지수를 사용하면 대부분 극복할 수 있기 때문이다. 최상위 논문의 인용빈도가 지나치게 크게 반영된다는 g-지수의 잘 알려진 단점도 연구자 집단 규모가 작은 국내 상황에서는 전체 1위 논문의 인용빈도가 400회에도 미치지 못하므로 큰 문제가 되지 않는다.

둘째, 공저를 보정하는 지수 산출 방식 중에서는 공저자 수로 나눈 인용빈도를 사용하는 h_C -지수와 g_C -지수가 바람직하다. Batista 등(2006)의 h_1 -지수는 공저자가 많은 고인용 논문에 지나치게 좌우되는 문제가 있고, Schreiber(2008a)의 h_m -지수는 단독저술을 많이 할수록 오히려 지수가 낮게 산출되는 문제가 발생하는 반면에, h_C -지수와 g_C -지수는 이같은 문제가 발생하지 않으며 비교적 간단하게 산출할 수 있다.

셋째, 연구 성과의 양적인 측면과 질적인 측면을 한꺼번에 반영하는 복합 지수라는 h-지수와 g-지수의 고유한 특성을 유지하기 위해서는 반드시 공저를 보정하여 지수를 측정해야 한다. 공저를 보정하지 않은 h-지수와 g-지수는 연구 성과의 질적인 측면은 강하게 반영하지만 양적인 측면은 미약하게만 반영하거나 제대로 반영하지 못하는 것으로 나타났다.

넷째, 공저를 반영한 h_C -지수를 기준으로 평가하면 단독 저술 비중이 높은 인문학 분야 연구자의 저평가 문제를 피할 수가 있다. 이는 공저를 반영한 g_C -지수도 마찬가지이다.

다섯째, 특정 분야나 특정 기관에 속한 연구자가 상위 순위를 과점하는 현상을 방지하고 다양성을 유지하기 위해서는 h_C -지수나 g_C -지수를 기준으로 평가하는 것이 바람직하다.

이상과 같이 학계에 널리 알려진 h-지수는 국내 연구자의 연구 성과를 평가하기에는 일관성 결여와 공정성 부족이라는 한계를 지니는 것으로 드러났으므로 두 가지 문제를 동시에 극복할 수 있는 g_C -지수의 적용을 정책적으로 시급하게 고려해야 할 것이다.

참 고 문 헌

- 고영만, 조수련, 박지영 (2013). 학술지의 피인용횟수 순위를 적용한 tapered h-지수의 변형지표 “Kor-hT”에 관한 연구. *정보관리학회지*, 30(4), 111-131.
<http://dx.doi.org/10.3743/KOSIM.2013.30.4.111>
- 김판준, 이재운 (2010). 학술지 영향력 측정을 위한 h-지수의 응용에 관한 연구. *정보관리학회지*, 27(1), 269-287. <http://dx.doi.org/10.3743/KOSIM.2010.27.1.269>
- 박지연, 김정은, 민윤경 (2010). 기관단위 연구성과 평가방법에 관한 연구: h-지수 및 변형지수를 중심으로. *정보관리학회지*, 27(1), 249-267. <http://dx.doi.org/10.3743/KOSIM.2010.27.1.249>
- 유소영, 이재운 (2008). 학제적 분야의 정보서비스를 위한 학술지 인용 분석에 관한 연구: Y대학교 생명공학과를 중심으로. *정보관리학회지*, 25(4), 283-308.
<http://dx.doi.org/10.3743/KOSIM.2008.25.4.283>
- 이재운 (2006). 연구성과 평가를 위한 h-지수의 개량에 관한 연구. *정보관리학회지*, 23(3), 167-186.
<http://dx.doi.org/10.3743/KOSIM.2006.23.3.167>
- 이재운, 최상희 (2015). 논문 인용 영향력 측정 지수의 편향성에 대한 연구: KCI 수록 논문을 대상으로. *정보관리학회지*, 32(4), 205-221. <http://dx.doi.org/10.3743/KOSIM.2015.32.4.205>
- 이종욱, 양기덕 (2011). 교수연구업적 평가법의 계량적 분석: 국내 문헌정보학과 교수연구업적을 중심으로. *정보관리학회지*, 28(4), 119-140. <http://dx.doi.org/10.3743/KOSIM.2011.28.4.119>
- Banks, M. G. (2006). An extension of the Hirsch Index: Indexing scientific topics and compounds. *Scientometrics*, 69(1), 161-168. <http://dx.doi.org/10.1007/s11192-006-0146-5>.
- Bar-Ilan, J. (2008). The h-index of h-index and of other informetric topics. *Scientometrics*, 75(3), 591-605. <http://dx.doi.org/10.1007/s11192-007-1880-z>.
- Batista, P. D., Campiteli, M. G., Kinouchi, O., & Martinez, A. S. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68(1), 179-189.
<http://dx.doi.org/10.1007/s11192-006-0090-4>
- Braun, T., Glänzel, W., & Schubert, A. (2005). A Hirsch-type index for journals. *The Scientist*, 19(22), 8.
- Chai, J. C., Hua, P. H., Rousseau, R., & Wan, J. K. (2008). The adapted pure h-index. In: H. Kretschmer and F. Havemann (eds), *Proceedings of WIS 2008, Berlin. Fourth International Conference on Webometrics, Informetrics and Scientometrics and Ninth COLLNET Meeting*. Retrieved from <http://www.collnet.de/Berlin-2008/ChaiWIS2008aph.pdf>
- Cronin, B., & Meho, L. (2006). Using the h-index to rank influential information scientists.

- Journal of the American Society for Information Science and Technology, 57(9), 1275-1278.
<http://dx.doi.org/10.1002/asi.20354>
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69, 131-152.
<http://dx.doi.org/10.1007/s11192-006-0144-7>
- Egghe, L. (2007). Mathematical theory of the h-index and g-index in case of fractional counting of authorship. *Journal of the American Society for Information Science and Technology*, 59(12), 1608-1616. <http://dx.doi.org/10.1002/asi.20845>
- Hagen, N. T. (2008). Harmonic allocation of authorship credit: Source-level correction of bibliometric bias assures accurate publication and citation analysis. *PLoS ONE*, 3(12), e4021.
<http://dx.doi.org/10.1371/journal.pone.0004021>.
- Harzing, A.-W. (2010). *The publish or perish book: Your guide to effective and responsible citation analysis*. Melbourne, Australia: Tarma Software Research Pty Ltd.
- 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(46), 16569-16572. <http://dx.doi.org/10.1073/pnas.0507655102>
- Hu, X. J., Rousseau, R., & Chen, J. (2010). In those fields where multiple authorship is the rule, the h-index should be supplemented by role-based h-indices. *Journal of Information Science*, 36(1), 73-85. <http://dx.doi.org/10.1177/0165551509348133>
- Kelly, C. D., & Jennions, M. D. (2006). The h index and career assessment by numbers. *Trends in Ecology & Evolution*, 21(4), 167-170. <http://dx.doi.org/10.1016/j.tree.2006.01.005>
- Lee, J. Y., & Chung, E. (2014). A comparative analysis on multiple authorship counting for author co-citation analysis. *Journal of the Korean Society for Information Management*, 31(2), 57-77. <http://dx.doi.org/10.3743/KOSIM.2014.31.2.057>
- Lee, J. Y., & Choi, S. (2011). Intellectual structure and infrastructure of informetrics. *Journal of the Korean Society for Information Management*, 28(2), 11-36.
<http://dx.doi.org/10.3743/KOSIM.2011.28.2.011>
- Rousseau, R. (2006). A case study: Evolution of JASIS' Hirsch index. *Science Focus*, 1(1), 16-17.
- Schreiber, M. (2008a). To share the fame in a fair way, hm modifies h for multi-authored manuscripts. *New Journal of Physics*, 10, p.040201.
<http://dx.doi.org/10.1088/1367-2630/10/4/040201>
- Schreiber, M. (2008b). A modification of the h-index: The hm-index accounts for multi-authored manuscripts. *Journal of Informetrics*, 2(3), 211-216.
<http://dx.doi.org/10.1016/j.joi.2008.05.001>

- Schreiber, M. (2010). Revisiting the g-index: The average number of citations in the g-core. *Journal of the American Society for Information Science and Technology*, 61(1), 169-174. <http://dx.doi.org/10.1002/asi.21218>
- Schubert, A. (2009). Using the h-index for assessing single publications. *Scientometrics*, 78(3), 559-565. <http://dx.doi.org/10.1007/s11192-008-2208-3>
- Van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67(3), 491-502. <http://dx.doi.org/10.1556/Scient.67.2006.3.10>
- Vanclay, J. (2007). On the robustness of the h-index. *Journal of the American Society for Information Science and Technology*, 58(10), 1547-1550. <http://dx.doi.org/10.1002/asi.20616>
- Waltman, L., & van Eck, N. J. (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology*, 63(2), 406-415. <http://dx.doi.org/10.1002/asi.21678>
- Wan, J.-K., Hua, P.-h., & Rousseau, R. (2007). The pure h-index: Calculating an author's h-index by taking co-authors into account. *Collnet Journal of Scientometrics and Information Management*, 1(2), 1-5. <http://dx.doi.org/10.1080/09737766.2007.10700824>
- Zhang, C. T. (2009). A proposal for calculating weighted citations based on author rank. *EMBO Reports*, 10(5), 416-417.

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Kim, Pan-Jun, & Lee, Jae Yun (2010). A study on journal impact measurement with Hirsch-type indices. *Journal of the Korean Society for Information Management*, 27(1), 269-287. <http://dx.doi.org/10.3743/KOSIM.2010.27.1.269>
- Ko, Young Man, Cho, Soo-Ryun, & Park, Ji Young (2013). A study on the "Kor-T", a modified tapered h-index, by applying the ranking according to the number of citations of journals in evaluating Korean journals. *Journal of the Korean Society for Information Management*, 30(4), 111-131. <http://dx.doi.org/10.3743/KOSIM.2013.30.4.111>
- Lee, Jae Yun (2006). Some improvements on h-index: Measuring research outputs by citations. *Journal of the Korean Society for Information Management*, 23(3), 167-186. <http://dx.doi.org/10.3743/KOSIM.2006.23.3.167>

- Lee, Jae Yun, & Choi, Sanghee (2015). Discipline bias of document citation impact indicators: Analyzing articles in Korean Citation Index. *Journal of the Korean Society for Information Management*, 32(4), 205-221. <http://dx.doi.org/10.3743/KOSIM.2015.32.4.205>
- Lee, Jongwook, & Yang, Kiduk (2011). A bibliometric analysis of faculty research performance assessment methods. *Journal of the Korean Society for Information Management*, 28(4), 119-140. <http://dx.doi.org/10.3743/KOSIM.2011.28.4.119>
- Park, Ji-Yeon, Kim, Jung-Eun, & Min, Yoon-Kyung (2010). A study on the evaluation methods of research institution: Based on the h-index and its variants. *Journal of the Korean Society for Information Management*, 27(1), 249-267. <http://dx.doi.org/10.3743/KOSIM.2010.27.1.249>
- Yu, So Young, & Lee, Jae Yun (2008). Journal citation analysis for library services on interdisciplinary domains: A case study of Department of Biotechnology, Y University. *Journal of the Korean Society for Information Management*, 25(4), 283-308. <http://dx.doi.org/10.3743/KOSIM.2008.25.4.283>

