

●邱均平 马瑞敏 程 妮

利用SCI进行科研工作者 成果评价的新探索*

摘 要 SCI是一个国际化的引文索引系统,收录的期刊有严格的审核标准,发文质量较高。目前国外利用SCI对科研工作者成果评价的最新方法是建立在h指数和g指数的考核基础上。虽然二者在科研工作者成果评价中都有很重要作用,但也存在一些缺陷。论文质量指数(Paper Quality Index)的引入可以解决不同学科之间比较和时效性方面的问题。进行科研成果评价时,应将h指数、g指数和论文质量指数结合使用,将短期评价和长期评价结合使用。表6。参考文献7。

关键词 SCI h指数 g指数 论文质量指数 科研成果评价

分类号 G350

ABSTRACT The authors introduce basic facts about SCI and its strict review standards, and the newly-developed methods for the evaluation of scientific researches based on h-index and g-index by using SCI. Both of the methods have their advantages and disadvantages. They think that the introduction of Paper Quality Index can solve the problems in the comparisons of different disciplines and in the timeliness. They propose to integrate h-index, g-index and paper quality index and to integrate short-term evaluation and long-term evaluation for the evaluation of scientific researches. 6 tabs. 7 refs.

KEY WORDS SCI. h-index. g-index. Paper quality index. Evaluation of scientific researches.

CLASS NUMBER G350

1 利用SCI进行科研工作者成果评价的现实必要性

1963年,加菲尔德博士出版了《科学引文索引》(Science Citation Index, SCI),SCI这个名词从此便风靡全球。几十年后,SCI成为中国学术界最时髦的词语之一,科研工作者们的职称评定、奖金、评奖、经费申请乃至院士评选,无不与SCI挂钩。在国内从1990年代中期便开始了对SCI评价功能的探讨,一直持续至今。有的认为“SCI正成为现阶段科学管理和科学评价不可替代的强有力科技评估工具”;有的认为“SCI成了Stupid Chinese Idea(中国人愚蠢的想法)的缩写”;有的认为“好坏参半,应谨慎使用”……不管何种看法,我们认为,在当前中国科研环境下,利用SCI进行科研工作者成果评价有其现实必要性。

(1)科研工作者应有国际化的科研眼光,立足于一个良好的国际学术交流平台并得到国际同行的认

可。SCI是一个国际化的引文索引系统,收录的期刊有着严格的审核标准,发表在这些期刊的论文有着较高的质量,容易引起较大关注。

(2)同行评议在中国的可操作性如何?一些学者认为中国的科研工作者成果评价应该废除SCI指标,而利用同行评议(peer review)。同行评议是定性评价的一种方法,一般是科研工作者向评估委员会提交规定数量内的代表作,同行专家对这些代表作进行审核和打分以及排序,最终确定优劣。这是国外学术界非常通行的做法,于是一些国内的学者便认为我们也应该这样。但是中国的学术界受传统文化、政治环境等的影响,关系网络非常繁杂,要想用定性的方法来评估,争议会比利用SCI进行合理定量评价更广泛。这些年来,我们可以发现中国学术界的各种评奖备受争议,并且时有学者揭出评奖黑幕的事件发生。各种因素的干扰使中国进行同行评议很难彻底实行下去,可以说,同行评议的大环境一直还没在中国形成。

(3)中国自己的引文索引系统是否能满足国内

* 本文系国家社科基金重点项目“我国人文社科研究评价体系的构建与实证分析”(05AZX004)的研究成果之一。

学术界的评价?要解决这个问题,首先要明确国内学术界现在进行科研定位是什么,是仅满足于国内学者交流,还是要融入国际学术交流?如果仅要实现前者目标,国内比较著名的《中国科学引文索引》(CSCD)、中国科技论文与引文数据库(CSTPC)和中国社会科学引文索引(CSSCI)便基本可以满足。但是要实现后者目标,国内没有一个索引数据库可以实现,因为它们收录的绝大多数都是中文期刊,影响力有限。定量评价工作自身的不足使我们对SCI更加依赖。

(4)中国现在到了终止使用SCI进行科研工作成果评价的恰当时刻了吗?事实上,现在终止SCI来进行科研工作者评估时机尚不成熟。除了上述同行评议在中国难以实行及中国引文索引系统自身过于本土化问题等原因,我们认为国内对于SCI中的一些文献计量术语有着较大的误解,由误解而排斥或盲目评价。比如有些学者往往在成果描述时写成“当年发表了一篇影响因子为8的论文”,这是完全错误的表达方式。错误一是把期刊影响因子和论文影响力混为一谈;错误二是没有弄明白期刊影响因子的含义,某期刊当年的影响因子是前两年在该期刊发文在当年的平均被引情况,和当年的论文没有关系。另外还有一个取消SCI评价的理由便是收录中国期刊过少,这的确是个事实,但是我们更应该认真反思自己的期刊到底是否具有国际竞争力并如何改进,这才是解决问题的根本之法。

2 国外利用SCI对科研工作者成果评价的最新方法

2.1 h指数

2005年11月,美国加州大学圣地亚哥分校的物理学家赫希(J. H. Hirsch)教授在《美国科学院院刊》上发表《一种量化个人科研成果的指数》一文,正式提出h指数这一概念。h指数(h-index)的定义是:一个人的h指数就是指他发表的所有论文中,有h篇论文分别被引用了至少h次。h确定的方法是:按照论文被引次数降序排列,当 $TC_r < r$ 时,则 $h = r - 1$,其中 r 为按论文被引次数降序排列的论文序号, TC_r 表示排名第 r 的论文的被引次数。

这种方法非常简单,可操作性强。用赫希的话说,只需要花30秒钟。下面我们给出科学计量学界最高奖普赖斯奖获得者的h指数。因为武汉大学数据库订购时间从1999年开始,所以不能够反映这些学者真实的h指数。这里引用比利时科学计量学家Leo Egghe论文数据(见表1)。

表1 普赖斯奖获得者的h指数(按h指数降序排列)

Name	h-index
Garfield	27
Narin	27
Braun	25
Van Raan	19
Glönnzel	18
Moed	18
Schubert	18
Small	18
Martin	16
Egghe	13
Ingwersen	13
Leydesdorff	13
Rousseau	13
White	12

2.2 g指数

2006年初,科学计量学家Leo Egghe在ISSI Newsletter上发表了《h指数的改进—g指数》一文,正式提出g指数这一概念。我们从他的题目也可以看出,g指数是在h指数的基础上演化而来,和h指数有着紧密联系。g的定义为:按论文被引次数降序排列,g为排名前g位的论文至少被引用 g^2 次的最大排列号。很明显, $g \geq h$ 。g的确定方法是:按照论文被引次数从高到低排列,当 $\sum TC_r < r^2$ 时,则 $g = r - 1$;其中 r 为按论文被引次数降序排列的论文序号, TC_r 表示排名第 r 位论文的被引次数, $\sum TC_r$ 表示排名前 r 位所有论文被引次数累计和。

下面我们给出科学计量学界最高奖普赖斯奖获得者的g指数。这里仍引用比利时科学计量学家Leo Egghe论文数据(见表2)。

表2 普赖斯奖获得者的g指数(按g指数降序排列)

Name	g-index
Garfield	59
Narin	40
Small	39
Braun	38
Schubert	30
Van Raan	27
Glönnzel	27
Moed	27
Martin	27
Ingwersen	26
White	25
Egghe	19
Leydesdorff	19
Rousseau	15

2.3 g 指数与 h 指数的关系

g 指数是 h 指数的改进,是在 h 指数的基础上发展而来。它们方法简单,操作起来也很方便。h 指数注重科研工作者被引论文的数量,对于从质量角度评价一个科研工作者的贡献有积极意义。但是我们也可以发现 h 指数的确定具有静态性,一篇文章一旦被确定为高被引论文,不管它以后几年的被引次数如何变化,对于某个科研工作者 h 指数的确定都没有影响。所以它不能反映这些高被引论文的变化情况。g 指数的提出主要是解决 h 指数这一缺憾,在保持 h 指数优点的基础上,更多考虑高被引论文的变化情况。根据 2.1 和 2.2 的相关介绍,它们算法联系具体如表 3 所示。这里仍引用 Leo Egghe 论文数据。从表 3,结合 h 指数和 g 指数的定义,可以得到 L. Egghe 的 h 指数为 13, g 指数为 19, $g > h$ 。

表 3 L. Egghe 的 h 指数和 g 指数

TC	r	$\sum TC$	r^2
47	1	47	1
42	2	89	4
37	3	126	9
36	4	162	16
21	5	183	25
18	6	201	36
17	7	218	49
16	8	234	64
16	9	250	81
16	10	266	100
15	11	281	121
13	12	294	144
13	13	307	169
13	14	320	196
13	15	333	225
12	16	345	256
12	17	357	289
12	18	369	324
12	19	381	361
11	20	392	400

2.4 g 指数与 h 指数的局限性

虽然 g 指数和 h 指数在科研工作者成果评价中都有重要作用,但是两者都有一些缺陷,主要有两点:

(1) 局限在同一学科内比较,无法比较不同学科科研工作者的成果。不同学科的被引情况不尽相同。有些学科引用次数很高,必然导致 h 指数和 g 指数相对其他学科高,但并不能表明这些科研工作者是优秀的,当然也不能说他们比其他 h 指数和 g 指数比较低的学科的学者优秀。

(2) 对青年学者的评价不利。h 指数和 g 指数都偏向高被引论文,这些论文往往需要时间的积累,这是对青年学者的客观不利因素。也就是说 h 指数和 g 指数的即时性不是很好。即使一些年老的学者近几年不写论文了,他们的 h 指数和 g 指数仍然比较高。h 指数和 g 指数不适宜评价近几年科研工作者的成果情况,不利于发掘新人的知识热点。

3 我们的方法

3.1 基本思路

如何弥补 h 指数和 g 指数的缺憾呢? 我们提出下面的思路。首先我们认为决定一篇论文质量的因素有两点:一是它所在的期刊;二是其自身表现。期刊是论文的载体,一个好的载体可以提升一篇论文的档次,比如在 *Science* 发一篇文章的确不易。最关键的是杂志的评委对该论文的认可,属于一种潜在的定性评价,所以发在质量较高期刊上的论文很有潜力,当然我们可以用影响因子来量化该杂志的档次。自身表现则是该论文的被引情况,被引多表明受关注度高,有较大影响,是定量评价。当然,如果一篇论文被引次数为零,即使载体再好也无大的实际意义。

要做到不同学科间的比较,必须消除被引情况的学科差异,包括不同学科期刊影响因子差异和不同学科论文被引次数差异。同时期刊影响因子本身就有时效性的特点,主要是考察论文发表后两年之内的质量,属于短期质量评价。这样我们便解决了不同学科的比较和时效性两个问题。

3.2 具体方法

在基本思路指导下,我们首先提出某论文在某年的质量指数(Paper Quality Index, PQI)这一概念。其数学表达式为:

$$PQI_{ij} = \frac{IF_{ij}}{IF} \times \frac{TC_{ij}}{TC}$$

其中, IF_{ij} 表示刊登 i 论文期刊在 j 年的影响因子; IF 表示 i 论文所在期刊所归属学科的所有期刊 j 年的平均影响因子; TC_{ij} 表示 i 论文在 j 年的被引次数, TC 表示 i 论文所在期刊所归属学科的所有论文在 j 年的平均被引次数,即 JCR 中该学科的 Aggregate Impact Factor 值。从公式也可看出,如果 $TC_{ij} = 0$, 则 $PQI_{ij} = 0$ 。 $\frac{IF_{ij}}{IF}$ 是消除不同学科期刊影响因子的差异,同时可

评价论文的载体质量; $\frac{TC_{ij}}{TC}$ 可消除不同学科论文被引的差异,同时可评价论文的自身质量。两者用乘号连接是因为两者都为正数,相互牵制,用乘号最合适。

对于科研工作者的评价,则是这些质量指数的累加,既可以按某年累加来看他发表论文在某年总的质量指数,也可以按照某几年累加看他发表论文在这几年的质量指数。

3.3 具体操作需注意的问题

SCI 和 JCR 有自己的一些特点,所以在我们具体操作时会遇到一些问题。相应具体处理规则如下:

(1) 某文章所在期刊属于多个学科,就需要分析文章,将它归属到最符合的期刊。即由文章决定期刊的唯一归属。

(2) 对于综合学科 (MULTIDISCIPLINARY SCIENCES),最好的办法是顺其自然,直接归属综合学科,以保证论文的唯一归属。目前为止,还很难用其他方法解决这一问题。

(3) 因为消除了学科差异,所以一个科研工作者归属不同学科的论文都应该计算在内。

(4) 对于 $\overline{TC}$, JCR 自 2003 年开始才提供该值,评价论文的时段应该选择在 2001 年后。

(5) 对于合作者的情况,只将该成果计入第一作者。

(6) 文献的形式在 SCI 中有多种,这里只选择论文 (Article) 这一形式。

3.4 具体实施步骤

这里以 L. Egghe 的 “A proposal to define a core of a scientific subject :A definition using concentration and fuzzy sets” 这篇文章为研究对象,详细论述该方法的实现步骤。

(1) 在 SCI-E 中可很容易找到该文章发表在 2002 年的 *Scientometrics* 杂志上

(2) 在 SCI-E 中根据字段 Times Cited ,发现它至今共被引了 6 次。然后根据链接进入被引入详细情况,统计出在 2003 和 2004 年分别被引了 2 次。如果该文章被引次数很多,可以借助 SCI-E 的 “ANALYZE” 分析统计功能,这样可以提高效率。

(3) 在 JCR 中查找 *Scientometrics* 在 2003 和 2004 年的期刊影响因子分别为 1.251 和 1.120。

(4) 在 JCR 中发现 *Scientometrics* 属于 computer science, interdisciplinary applications 这一学科,点击 “View Journal Summary List” 进入,将该学科所有期刊的相关指标下载到 Excel,然后求出该学科所有期刊的平均影响因子。2003 年和 2004 年的值分别为 0.985 和 1.015。

(5) 在进行步骤 (4) 的同时,点击 “View Category Data” 进入,发现 “Aggregate Impact Factor” 这一指标,即该论文所在期刊所归属学科的所有论文在某年的平均被引次数 $\overline{TC}$ 。2003 年和 2004 年的值分别为 1.326 和 1.379。

(6) 根据公式 $PQI_{ij} = \frac{IF_{ij}}{\overline{IF}} \times \frac{TC_{ij}}{\overline{TC}}$, 得到 $PQI(03) = \frac{1.251}{0.985} \times \frac{2.000}{1.326} = 1.916$, 同理可得 $PQI(04) = \frac{1.120}{1.015} \times \frac{2.000}{1.379} = 1.600$, 则这篇文章总的 $PQI = PQI(03) + PQI(04) = 3.516$

按照上面的步骤可以找出 L. Egghe 发表的其他论文的质量指数。需注意的是,一个人或该学科活跃的作者群发文大部分都属于同一个学科,因此 $\overline{IF}$ 和 $\overline{TC}$ 无需重复查找。表 4 给出普赖斯奖获得者 L. Egghe 在 2001 ~ 2004 期间以第一作者身份发表论文 (只包括 Article) 在 2003 ~ 2005 各年的质量指数,其中 PY 为认论文发表年代, $PQI1$ 为该论文在发表两年后质量指数之和, $PQI2$ 为该作者前两年发的文章在该年的质量指数之和。

从表 4 可以看出, L. Egghe 所发论文总的质量指数为 26.298, $PQI2(03) = \sum PQI(03) = 4.470$ 。表明他 2001 或 2002 年发的文章在 2003 年的总的质量指数为 4.470, 其他依次类推。 $PQI1$ 的含义同本节步骤 (6)。

表 4 L. Egghe (第一作者) 所发论文各年的质量指数

论文	所在期刊	归属学科	PY	PQI(03)	PQI(04)	PQI(05)	PQI1
1	JASIST	COMPUTER SCIENCE, INFORMATION SYSTEMS	03	0.000	3.686	2.425	6.111
2	SCIENTOMETRICS	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	04	0.000	0.000	3.850	3.850
3	SCIENTOMETRICS	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	02	1.916	1.600	0.000	3.516

续表							
论文	所在期刊	归属学科	PY	PQI(03)	PQI(04)	PQI(05)	PQI1
4	SCIENTOMETRICS	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	04	0.000	0.000	2.888	2.888
5	JASIST	COMPUTER SCIENCE, INFORMATION SYSTEMS	02	0.000	1.843	0.000	1.843
6	INFORMATION PROCESSING& MANAGEMENT	COMPUTER SCIENCE, INFORMATION SYSTEMS	03	0.000	0.000	1.826	1.826
7	JASIST	COMPUTER SCIENCE, INFORMATION SYSTEMS	02	1.238	0.000	0.000	1.238
8	JASIST	COMPUTER SCIENCE, INFORMATION SYSTEMS	03	0.000	0.000	1.213	1.213
9	INFORMATION PROCESSING& MANAGEMENT	COMPUTER SCIENCE, INFORMATION SYSTEMS	02	0.991	0.000	0.000	0.991
10	SCIENTOMETRICS	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	04	0.000	0.000	0.963	0.963
11	SCIENTOMETRICS	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	02	0.000	0.800	0.000	0.800
12	MATHEMATICAL AND COMPUTER MODELLING	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	02	0.000	0.342	0.000	0.342
13	MATHEMATICAL AND COMPUTER MODELLING	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	03	0.000	0.342	0.000	0.342
14	MATHEMATICAL AND COMPUTER MODELLING	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	02	0.325	0.000	0.000	0.325
15	PROGRESS IN NATURAL SCIENCE	MULTIDISCIPLINARY SCIENCES	03	0.000	0.024	0.025	0.050
PQI2				4.470	8.638	13.190	26.298

下面给出普赖斯奖获得者和 1990 年代以来菲尔兹奖获得者的质量指数。具体如表 5, 表 6 所示。PQI 为某学者 2003, 2004, 2005 这三年总的质量指数。

受到样本的局限, 我们只能人为地设定一个 PQI 阈值。如果把阈值设置为 20, 那么普赖斯获得者 White、Glänzel 和 Egghe, 菲尔兹奖获得者 Bourgain、McMulle、Kontsevich、Lafforgue、Lions、Witten 和 Gowers 在 2001 ~ 2004 年都是有较高质量发文的作者。这和我们不同科的科学家可以进行类比的目的

相符。至于阈值到底应该选择多少, 则需要更大数据样本的试验。

结合表 1 和表 2 来看, h 指数和 g 指数较大的 Garfield 博士的 PQI 处于中等水平, 说明近期研究不是很活跃; Van Raan 基本保持名次不变, 说明他的学术研究一直都很活跃, 且质量较高。White 的名次有较大提高, 跃居第一, 说明他近期研究活跃且引起广泛关注, 质量很高, 并且随着时间的积累, 他的 h 指数和 g 指数也会随之增大。

表5 普赖斯奖获得者质量指数情况

学者	PQI(03)	PQI(04)	PQI(05)	PQI
White	5.911	22.115	13.175	41.201
Glänzel	12.451	7.202	21.176	40.829
Egghe	4.470	8.638	13.190	26.298
Leydesdorff	6.191	3.443	6.988	16.622
Van Raan	5.747	0.000	4.813	10.560
Braun	0.000	6.565	1.925	8.490
Moed	3.831	2.401	1.925	8.157
Garfield	0.000	3.686	2.798	6.484
Small	0.000	1.843	2.888	4.731
Schubert	2.873	1.600	0.000	4.473
Ingwersen	0.000	0.800	0.000	0.800
Rousseau	0.000	0.000	0.000	0.000
Narin	0.000	0.000	0.000	0.000
Martin	0.000	0.000	0.000	0.000

表6 菲尔兹奖获得者(1990-2002)质量指数情况

学者	PQI(03)	PQI(04)	PQI(05)	PQI
Bourgain	19.937	111.348	68.393	199.678
McMullen	17.402	52.940	75.163	145.505
Kontsevich	0.459	44.193	52.076	96.728
Lafforgue	26.884	60.832	0.000	87.716
Lions	12.937	13.376	26.294	52.607
Witten	0.784	0.000	47.136	47.920
Gowers	9.760	36.421	0.000	46.181
Voevodsky	3.516	7.154	4.221	14.891
Yoccoz	0.000	0.000	0.000	0.000
Zelmanov	0.000	0.000	0.000	0.000
Borcherds	0.000	0.000	0.000	0.000
Drinfeld	0.000	0.000	0.000	0.000
Vaughan	0.000	0.000	0.000	0.000
Mori	0.000	0.000	0.000	0.000

4 结论

我们提出的方法因为考虑了期刊载体质量因素,所以受到期刊影响因子的影响,只能够考察一篇论文发表后两年之内的质量指数,如果超出两年或者是当年的话,比如有的论文可能在发表3年或更多年后才被引用,便无法计算相应的质量指数。这里还要强调,我们提出的方法只是用来评价学者论文发表后两年之内的质量,是一种短期评价方法。但是我们消除了学科差异,并考虑了期刊载体的质量问题,因此有着独特的实践意义。

我们认为短期评价和长期评价应该结合,既考虑一个科研工作者的现实成绩也考虑其历史贡献,应该将我们的方法和h指数以及g指数结合使用。我们提出的方法和h指数以及g指数一样是建立在引文规范基础上的,该方法对于被引次数更加敏感,这就更要求引用行为的规范,但在我国这是一个大问题,自引不规范,过高自引无疑会提高论文的质量。但是在SCI系统中,只有Times Cited这样一个字段,没有区分自引和他引。使用该方法对于中国科学家的评价要谨慎一些,可以排除自引,防止学术投机现象的发生。基于以上思考,建议ISI公司考虑建立排除自引的论文Times Cited和期刊影响因子等指标,这不论对于h指数和g指数,还是我们建议的方法,在不同国度采用不同标准进行时,都有很强的现实意义。

参考文献

- 1 Egghe L. Theory and practise of the g-index. *Scientometrics*, 2006,69(1).
- 2 Egghe L, ROUSSEAU. An informetric model for the Hirsch-index. *Scientometrics*, 2006,69(1).
- 3 Egghe L. AN IMPROVEMENT OF THE H-INDEX. THE G-INDEX. *ISSI Newsletter*, 2006,2(1).
- 4 Hirsch J. E. . An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*. 2005,102(46).
- 5 包丽敏. 学术评价告别SCI?. *北京青年报*, 2006-01-04
- 6 孙亦樑,徐克敏. 合理利用SCI对个人做学术评价. *图书馆论坛*, 2002,22(5)
- 7 邱均平. *文献计量学*. 北京:科学技术文献出版社,1998

邱均平 武汉大学信息管理学院教授、博士生导师。通讯地址:武汉大学信息管理学院。邮编430072。

马瑞敏 程 妮 武汉大学信息管理学院情报学博士研究生。通讯地址同上。(来稿时间:2006-11-09)