

以 *SCI* 文章作为评估指标 局限性的探讨

张鸿翔*

(中国科学院资源环境科学与技术局 北京 100864)

摘要 作为科研绩效评估和科研管理的工具,科学引文索引(*SCI*)对于科学的发展和科研绩效评估体系的建立具有重要的意义,是现阶段不可替代的科技评估工具。在充分肯定 *SCI* 文章科学价值的前提下,分析了以 *SCI* 文章作为评价指标存在的问题和局限性,对于建立合理的评估体系具有积极的作用。

关键词 *SCI* 文章,评价体系

一个国家的科技走向世界并影响世界的一个重要标志是通过科技文献向世界介绍该国的研究成果、科学发现与学术思想,尤其表现在基础科学研究领域。科学研究成果的影响力与科技文献发表的出版物的传播范围与引用率有相当的关系。因此文献索引的重要性已普遍为科技界所接受。

1 *SCI* 的重要性

SCI(Science Citation Index 科学引文索引)与 *EI*(工程索引)和 *ISTP*(科技会议录索引)是世界著名的 3 大科技文献检索系统。*SCI* 是由美国科学信息研究所(ISI,美国汤姆逊集团下属的商业化运作的研究机构)的 E·Carfield 于 1961 年创立并逐步向全世界出版发行的科学引文数据库^[1],在学术界和商业界被公认为是最重要、最具影响力的科技文献检索系统。它的作用主要体现在科研绩效评价、科学前沿趋势预测和科研管理等方面。其最基本的功能是评价。其重要性毋

庸置疑,本文不再论述。

近些年关于 *SCI* 的作用产生了激烈的争论^[2-6],在肯定的基础上出现更多的质疑,但质疑的初衷是为了如何更科学地发挥 *SCI* 的功能和评价作用。

本文的论述对象主要为 *SCI* 文章,因为 *SCI* 文章的是非功过,最能体现焦点问题。目前 *SCI* 文章已成为科研工作者个人与科研单位的科学价值商标,但与论文发表相关的各种学术不端行为也较普遍地存在,致使以 *SCI* 衡量学术绩效的评价制度也饱受诟病。

2 *SCI* 文章作为评估指标的局限性和问题探讨

2.1 *SCI* 编制的理论局限性与期刊来源的不平衡性造成评价结果的不确定性

SCI 编制的理论是引文索引法,理论上的局限性是 *SCI* 有失公允的基本来源。

(1)基本假设不够严密的缺陷:造成引用文献动机的复杂性和多样性,例如论文作者的自引、反引、友情互引等等,使得完全建立在引文分析基础上的 *SCI* 评价体系存在一定的不确定性。

(2)索引过程中对单位全、简称的区别、

* 中国科学院资源科学与技术局固体地球科学处副处长,副研究员,E-mail:hxzhang@cashq.ac.cn
修改稿收到日期:2008 年 7 月 18 日



中国科学院

姓名鉴别区分能力的问题(尤其中国名字重叠太多、字母过少)造成大量的索引错误。

(3) 期刊影响因子计算方法的缺陷:期刊影响因子是该刊前 2 年所发表的论文在该年的平均被引次数,由于统计时间的局限性,因而一些特定学科的论文、一些超前意识的文章,它们的引证高峰期并没有反映到影响因子中,从而降低了期刊的影响因子。

SCI 收录的来源期刊在学科、地域、语言上的不平衡性是 SCI 评估不确定性的另一个原因。

(1) 语言差异:SCI 收录的源刊绝大多数是英文文献,因此传统的英语国家占有了先天优势。非英语国家受语言能力不足的限制,造成文章词不达意,晦涩难懂,大大影响了文章的引用。

(2) 地域差异:SCI 收录的来源期刊中发达国家的期刊占 90% 以上,尽管确实存在国与国学术水平的差距,但游戏规则的制定者自然拥有得天独厚的竞争优势。

(3) 学科差异:SCI 适合评价基础研究和部分应用基础研究工作,但对于应用研究,它关注的对象是科学价值链偏下游的产出。SCI 在运用于科研绩效评价时,不考虑不同学科特点的一刀切式的评价是一种简单化的做法。

2.2 SCI 论文数量和质量,哪一个更重要

数量易于统计,质量鉴定较为复杂。但对于一个评价对象的鉴定,二者缺一不可。随着中国科技界对 SCI 论文的重视与日俱增,2006 年 SCI 所收录的中国论文为 7.1 万篇,占世界的 5.9%,排在世界第 5 位^[7],单以 SCI 论文的数量作为评价指标,中国已经成为国际科学大国。但这能说明我国基础科学的真正地位吗?

1997—2006 年,我国科技人员作为第一作者发表的国际论文中,有 40% 的论文在 10 年间没有被引用过 1 次,过去 10 年间我

国论文被引用次数在世界上排在第 13 位^[7]。

说明我国 SCI 论文的质量提升明显滞后于论文数量的增加。论文质量的提高有赖于学术规范和监督机制的加强和完善,另一方面则应注重评价体系的导向作用。作为科技起步、走出国门的初始阶段,以数量作为特定时期的评价标准并不为过,但现在我国已是 SCI 论文的产出大国,再不要论文质量只能导致“垃圾论文”比例的日益增加,因此要摒弃单纯以论文数量来评定科研绩效的片面做法,要提倡数量与质量相结合的评价方法。而如何科学地选定一套评价标准是科研管理和评价机构需要认真思考的问题。

2.3 SCI 期刊影响因子与文章质量是否正相关

从理论上讲,影响因子高的期刊,文章的质量要高,论文被引用的频次以及是否被该领域权威专家所引用反映了研究的价值和水平。只有那些有价值的研究工作或新的研究方法才有可能被大量地、长期地引用,特别是被 SCI 论文所引用。因此,在很多评价活动中,期刊影响因子的简单数字对比被用来决定论文质量的高低,甚至作为跨学科跨领域论文成果类比的尺子。但事实上不同学科 SCI 源刊的影响因子分布极不均衡,不同学科的期刊数目比例、平均参考文献数、引证半衰期、文章产出周期、学科发展特点、合作规模、科学研究的独立性程度、文章发表难易程度、研究群体大小等不同而决定了期刊影响因子的差异,也决定了论文的影响力和范围的差异,因此将期刊影响因子作为不同学科、边缘学科和多学科期刊间评价的标准,从而由期刊影响因子演绎出的论文影响因子就不可避免地导致文章质量的误判。

例如,资源环境领域的很多研究工作都建立在长期观测与数据积累之上,几年甚至几十年的研究数据才能发表一篇高水平文章,这与某些学科发表一些实验方法和实验

现象显然不同,但方法学研究往往得到不同学科的关注从而导致该学科文章影响因子要远高于资源环境科学领域文章的影响因子,但究其价值,孰高孰低,恐怕难以一概而论。所以,SCI 源刊的影响因子只能收敛于某一学科内部可比的环境下进行比较。

2.4 同样期刊上的文章质量相同吗

答案显而易见,不尽相同。以英国的 *Nature* 和美国的 *Science* 为例,它们的发行量和影响力高于其他任何专业性期刊。在 *Nature* 和 *Science* 上科技论文基本以 3 种形式出现:(1) 学术论文:*Nature*:Article;*Science*:Research article;(2) 研究报告:*Nature*:Letter;*Science*:Report;(3) 通讯:*Nature*:Correspondence;*Science*:Letter,从事科研写作的人都知道,这几类文章的质量还是有显著差别的。更多期刊上的论文分为基础研究、应用研究、快报、评论等,这些文章的质量也不尽相同。在我国目前的评价体系中,只看期刊的类别,而忽略文章的类别,甚至将“回复”(reply)的文章也会被评价体系认为是等同于 article 的文章。

2.5 SCI 期刊上以论文集发表的国际会议论文摘要应如何定位

随着我国科学界国际交流的日益增强,我国科研人员越来越频繁地参加国际会议,2006 年我国在主要国际会议上发表论文 3.6 万篇,占世界总数 9.0%,数量比 2005 年增加了 15.8%。诚然,在发展迅速的学科领域,许多创新的思想经常会首先在国际学术会议上进行交流,也在一定程度上反映了科学前沿和最新研究动向。但国际会议论文摘要通常会以国外某著名期刊为载体作为增刊发表,在一定程度上促使国内一些科研工作者踊跃以摘要的形式向国际大会投稿,“搭个顺风车”,不担心拒稿和评审修改的艰辛,还能多出一篇 SCI 文章。这样文章的真实价值需要客观进行评价,应加以界定。

2.6 关于 SCI 文章产生的一些噪音

SCI 文章的考核体系对广大科研工作者勤奋工作起到了积极的作用,但同时衍生出一些有悖于学术道德的行为。有些科学家 20 多年发表了三四百篇文章,从事自然科学的人都知道一篇科学文章产出的周期。文章如此众多,道理也浅显易懂,下面简单列举几种 SCI 所衍生的噪音:

(1)A+B 模式(文章互挂署名)。文章的合作很正常,学术思想的互补对于一篇好文章的完成非常重要。但出于多出文章的动机,把不相干的人挂在文章上面就是一种不正常的现象了。某些科学家要求所有学生的文章不论与他是否相关,都要挂上导师的名字,这样统计的文章数量只能让人感觉到离“大家”还欠点火候。A+B 的方式虽然为 A 或 B 快速产出大量文章,但是否能代表自己的学术贡献相信自有公道。

(2)买论文(雇用枪手)。现在某些科学家,甚至某些领导同志,花钱去买论文。纯粹用钱去买,这样做的层次太低,太不含蓄,而通过科技界也存在的“包工头”的方式,将手中掌控的科研经费拨出去换文章,挂上第一作者。更有甚者,将宝贵的数据和标本交给外国人,充分发挥外国人英语文字功底好、期刊人脉顺的优点去发表高质量的文章,甚至在让科研工作者顶礼膜拜的 *Nature* 上都出现过这类现象。

(3)通讯作者和第一作者的游戏(1=2 的游戏)。通讯作者是在科研文章中提出科学思想并且承担文章责任的作者,可能不是第一作者,这是国际科技界通行的做法。通讯作者对论文成果产出贡献的重要性近年来也已为我国科技评价体系所接受,这本来是充分尊重科学规律的做法。但近些年出现一种怪现象,由于通讯作者在评价体系中的权重等同于第一作者,因此,一些文章纷纷出现不相干的通讯作者。1=2,“倍增效应”



中国科学院

足实让一些人尝到好处。

(4)文章引用动机。引用文献的初始动机是好的,一篇科研文章应建立在严格的文献调研基础上,但仔细分析,还存在一系列的“假引”。例如,论文作者的“自引”(不排除自身科研工作延续性所进行的历史引用,这种自引无可厚非),但无所关联的自引有悖于科学的严肃性;“友情互引”,这雷同于文章的A+B合作模式,纯粹的投机主义;“转引”与“虚引”,对没有亲自查阅过的文献从别人论文的引文中加以转引而作为基础调研资料;为增强文章的“权威性”,而盲目地搬用“权威名著”的“崇引”;为了使文章顺利发表,将不同观点的历史文献有意“漏引”。这些引用的复杂性直接威胁到SCI评价体系的公正性。

以上举例不能全部枚举SCI出现的种种浮躁现象,但确为科技界普遍存在的现象。急功近利是不能造就真正的科学,也不能诞生真正的大家,SCI文章出现的种种浮躁现象只是学术界浮躁特征的一个表象。

深层次思考这些噪音出现的根源,和管理部门及评价机构简单化“一刀切”的评价体系是息息相关的,如何避免这些噪音的出现,使SCI充分发挥其科学价值,其责任不能简单地通过科学家的自律完成,管理部门政策导向的科学性才是根本所在。

3 同行评议和SCI定量指标的联合是科学的评价体系

将SCI作为科研管理和评价的工具,比较适用于宏观层次(国家、地区)和中观层次(研究机构、学科领域)的评价。对于微观层次(科技期刊、研究者个人)的评价,应十分慎重。唯SCI而论,纯粹将其作为评价标杆往往会导致一些错误,这种错误如不加以克服,将产生一定误导,将会有损于一些重要的冷门学科、交叉学科的正常发展,影响到我国核心期刊的健康成长,也会导致科研工

作者急功近利的思想,造成“高产低能儿”和“垃圾文章”的出现。

开展科技评估的另一种重要方法是始于19世纪的同行评议方法^[9],该方法一直是科技界进行科技评估的基本程序,在科研绩效评估、科技奖项评定、科技项目评审与科研经费分配等方面发挥着重要作用。由于科学评价体系中存在的“部分信用缺失”,也导致同行评议制度存在着一定的缺陷,评议结果可能会受到人为因素的干扰,一方面有来自专家违反科学信用的行为,另一方面也有可能来自管理部门的干涉,最终导致同行评议的公正性和客观性受到质疑。

因此,在评价体系中为了尽可能地减少人为因素导致的偏差,应坚持同类相比、同行评议、辅以科学引文计量的原则,建立同行评议和SCI定量指标的联合评价体系。而实际上SCI论文的发表和被引用的过程实际上也是一个同行评价的过程,并不能简单地将SCI标准与同行评价割裂开来。

主要参考文献

- 1 唐品.《科学引文索引》及其在科研评价中的作用.图书与情报,2005,(1):68-70.
- 2 杨红珊,陈图文.SCI一种全新的文献评价标准,现代情报,2002,9:67-69.
- 3 吴顺荣,任玉珍.SCI在评定科研成果中的局限性,现代情报,2003,9:13-14.
- 4 刘鸿翔.SCI在科研成果评价体系中的作用与局限性浅探,湖北教育学院学报,2006,23(12):130-133.
- 5 刘则渊,梁永霞.SCI的科研管理功能.大连理工大学学报(社会科学版),2007,28(1):63-67.
- 6 卢秉福,霍丽华.刍议SCI在科研评价体系中的作用.科研管理,2007,28(ZK):180-183.
- 7 中国科技信息研究所.中国科技论文统计结果,2007.
- 8 梁立明,武夷山等著.科学计量学:理论探索与案例研究.北京:科学出版社,2006.