

期刊评价：“核心期刊”是个什么东东？

--关于《中文核心期刊要目总览》核心期刊遴选方法的质疑（七）¹

《现代图书情报技术》编辑部

2015年11月2日

《中文核心期刊要目总览 2014》（简称《总览》）发布后，我们针对它在计算期刊学科影响力时的不科学不合理做法，已经做了若干评论²。看起来，我们是在“争取自己作为核心期刊的地位”。但通过这些分析中，我们开始意识到，问题的根源可能就在于“核心期刊”概念本身，在于我国学术界、期刊界、乃至科技管理领域的“核心期刊”情结。因此，有必要对“核心期刊”正本清源，避免围绕“核心期刊”及“核心期刊目录”的误解甚至迷信。

1、中国特色的“核心期刊”现象

在我国，“核心期刊”成为一个重要的现象、指标和名誉。许多机构都明确和公开地选用某个核心期刊目录（或者自己选定一组期刊组成“核心期刊”），规定在应聘、晋升、考核、申请项目、奖励时只接受发表在“核心期刊”中的论文。这似乎在说，一篇论文发表在核心期刊上就是好论文，一种核心期刊上的所有论文都达到了一定的学术水平；反之，一篇论文无论其真实学术水平如何、实际效果有多大，只要没有发表在核心期刊上就不能算数。笔者多年在管理岗位和参加各种评聘中的经验也证明，没有发表在核心期刊的论文往往没有资格进入评审，而对于那些“核心期刊”论文，实际评聘过程中也没有多少人仔细阅读和评价。

有需求，就有许多著名机构长期研究、大规模投入、或多年积累，编选“核心期刊目录”，就有许多期刊把进入核心期刊目录作为努力的目标、在进入某个目录后就在期刊封面显著注明自己的核心期刊地位，也就有许多作者费心费力、采取各种方式要把论文“发表”在“核心期刊”上。

但是，2006年国际数学家大会将当年的菲尔兹奖授予了证明庞加莱猜想的俄罗斯数学家格里戈里 佩雷尔曼³，美国克雷数学研究所也在2010年将100万美元的千禧年数学大奖给予了佩雷尔曼，而佩雷尔曼证明庞加莱猜想的论文从来没有发表在任何期刊上、更不用说什么核心期刊了，而是发布在开放预印本库arXiv.org上。如果按照我国许多机构的规定，佩雷尔曼的论文根本就不能算“学

¹ 执笔：张晓林

² 期刊评价：鼓励还是阻碍跨学科研究？<http://ir.las.ac.cn/handle/12502/7897>；期刊评价：不能狭隘地看待研究基金的作用。<http://ir.las.ac.cn/handle/12502/7893>；期刊评价：与时俱进和实事求是地看待文摘索引刊物的作用。<http://ir.las.ac.cn/handle/12502/7894>；期刊评价：压低跨学科影响力会伤害一大批期刊。期刊评价：压低跨学科影响力会伤害一大批期刊。<http://ir.las.ac.cn/handle/12502/7907>；期刊评价：从SCI和SSCI数据看跨学科影响力的重要性。<http://ir.las.ac.cn/handle/12502/7908>

³ 俄破解庞加莱猜想数学家获世界数学最高奖。<http://tech.sina.com.cn/d/2006-08-22/11561097423.shtml>

术论文”呢！难道在我国学术生活和职业发展中如此举足轻重的“核心期刊”概念，在真正的科学大奖面前就如此不堪一击了吗？

我们仔细想想，几乎所有的科技发达国家里，很少有严肃的科研教育机构会事先规定一个核心期刊目录，要求人们在应聘、晋升、考核、奖励时呈交的论文必须来自这个核心期刊目录中的期刊。人们会根据论文本身而不是它们的“出身”来判断论文的学术水平。英国英格兰高等教育资助理事会（HEFCE）等组织的英国大学学术卓越性评价（Research Excellence Framework）⁴并不要求提交评价的论文的“出身”，澳大利亚学术评价（Excellence in Research for Australia）在2012年及以后的评审中取消了原来依靠一个期刊排序表的做法而依靠评审专家的专业判断⁵。

那么，我们是对评审专家的学术水平或者学术公正或者学术诚信不自信、因此才会依赖一个核心期刊名单？或者是实在没有时间或不愿意花时间去评价论文本身、才让一个“核心期刊”目录作为“拦路虎”来减轻负担？其实，问题可能出在对“核心期刊”及其遴选机制的误解和误导上。

2、“核心期刊”概念可能是笔糊涂账？

我们也许应该看看“核心期刊”的历史内涵。

从某种意义上讲，“核心期刊”是纸本时代的一种检索手段。在纸本时代，人们面对发表在众多期刊上的众多论文，自然希望先选择少数“可能刊登了较多较高水平论文的期刊”作为一个遴选代理，帮助自己缩小需要阅读的范围来“找到值得读的论文”。这也许就产生了评价期刊质量、遴选“核心期刊”的需求。另一方面，“核心期刊”也是图书馆在纸本时代的一种选刊手段，图书馆因为采购经费限制，只能选订一部分期刊，因此也需要遴选出那些平均论文水平比较高的期刊作为“值得订阅的期刊”。

为遴选这样的核心期刊，有的机构依靠自己的学者进行遴选，也有的机构依靠某个客观机制来选择，例如汤森路透公司的期刊影响因子。期刊影响因子通过期刊在一个时间段内发表论文的被引次数来反映“期刊学术质量”，这种指标本身有一定的合理性。一种期刊通过自己的学术标准、同行评议专家水平、学术诚信控制能力、编辑能力等，努力保障自己所发表论文的学术水平，而论文的学术水平可以在一定程度上通过论文的被引用来体现，因此一般来说，期刊的学术质量水平与所发表论文的平均学术水平（因此平均被引频次）正相关。这样，利用以前发表论文的平均被引频次就可以在一定程度上反映该刊的学术质量水平。所以，人们用如此计算得来的期刊影响因子在一定程度上反映该刊的学术水平。

⁴ REF2914 Assessment framework and guidance on submissions.
<http://www.ref.ac.uk/pubs/2011-02/>

⁵ Why is the Ranked Journal List no longer used. <http://www.arc.gov.au/era-faqs>

但是，即使我们承认引用统计能在一定程度上反映学术期刊的学术影响力，但影响因子本质上是且只是期刊质量的评价指标，期刊的高影响因子从来就不能简单等同于发表在这个期刊上的某篇具体论文的高质量，高影响因子期刊上也有许多论文无人引用或者所发表论文也会因为学术不端或失误等被撤销。例如，Nature 不得不撤销弄虚作假的小保方晴子的论文⁶；有研究指出，多种著名期刊都有多篇论文（其中不乏高引用率论文）因为科学不端行为而被撤销⁷。因此，不能因为一篇论文发表在某个高影响因子的期刊上就认为这篇论文一定是高水平的。那种通过发文期刊的影响因子来代表具体论文水平的做法本身是不科学的，那些宣传“论文影响因子”、计算某个研究人员的“发文平均影响因子”、或者统计某个研究团队或研究机构的“论文平均影响因子”或“论文累计影响因子”的做法更是已在荒唐的边缘。

值得关注的是，有研究发现，来自高影响因子期刊的高被引论文比例在不断下降，来自非高影响因子期刊的高被引论文比例在不断上升⁸。其实，这反映了学术期刊网络化检索利用的现实。在 Google、Bing、百度、CNKI 以及大规模集成检索系统已经成为人们检索文献的主要工具时，以期刊为主的文献检索已经让位于以论文为主的文献检索，人们能够很方便地通过主题词、利用搜索引擎或集成检索系统方便地检索到来自几乎任何期刊的与自己所需要主题相关的论文，期刊本身作为一种遴选机制的作用迅速下降，而图书馆采用的 Big Deal 采购机制（购买大规模期刊数据库而非单独期刊）也明显淡化了“核心期刊”的作用。

3、“核心期刊目录”及其遴选可能出了什么错？

但是，在期刊遴选作用实际上不断弱化的同时，我国的“核心期刊”评选依然方兴未艾，而且我国的“核心期刊”名单何其多也！不仅多个研究机构编制了多个核心期刊目录，国务院学位委员会等也提出了自己的核心期刊目录，许多教育科研单位又有自己的目录。这些目录一方面都是在一定程度上依据论文引用数据作为基本客观数据，另一方面又“各有创新”引入许多间接指标，再加上出于种种原因各自设计出不同的分类体系，因此造成一种期刊在不同评价体系中出现在不同类别、出现在不同排名位置，甚至在有的体系中排名很高的期刊在另一个体系中名落孙山。而且，最近些年又有许多新的“评价指标”被不断地“研究”出来，使得几乎每一个稍稍正经点的期刊似乎都能找到一个有利指标来彰显自己的“高水平”。这样，“核心期刊”遴选成为研究热点，相关“市场”也持续火爆。

⁶ STAP retracted 02 July 2014. <http://www.nature.com/news/stap-retracted-1.15488>

⁷ Fang, F.C., et al. Misconduct accounts for the majority of retracted scientific publications, PNAS, January 15, 2013, vol. 110(3)

⁸ Uprising: Less prestigious journals publishing greater share of high-impact papers. <http://news.sciencemag.org/scientific-community/2014/10/uprising-less-prestigious-journals-publishing-greater-share-high-impact>

而且，这种局面往往引导学术期刊为获得某个体系的青睐而采取种种“有利于提高期刊影响力”的非学术做法。例如，《总览》惩罚那些发表跨学科论文的期刊，就会助长期刊只关注和发表“本学科”的内容，因为在有限发文量内发表跨学科论文将导致自己的“本学科论文的引用量”的减少，这实际上导致“画地为牢”和必然的“坐井观天”。这还容易驱使部分期刊建立“引用联盟”、相互“友情”引用来提高自己被引量，毕竟“引用联盟”只有在“本学科”中才能有效建立。另外，这往往造成对特色化专门化期刊发展的阻碍。例如，我们注意到，我国图书馆学情报学领域的多数期刊都是综合类期刊，但在国际图书馆学情报学领域、在美国这样的图书馆学情报学发达国家，专门化期刊占了大多数存在。为什么专门化期刊在我国难以发展呢？除了管理体制的原因（例如许多属地办刊）外，要获得较多“学科内引用”，就要发表这个学科领域内不同人员都读得懂的论文，或者在本学科所有主题领域都发表论文来覆盖各种引用可能，因此涉及宏观主题、宏大叙事和“热点主题”的论文以及综述等可能就受欢迎，因此许多期刊发表了上至宏观哲学般内容、下至非常高深细粒的技术内容，不管自己的评审专家或编辑是否能看懂。

虽然，有的评价体系不得不承认自己的遴选方法对一部分期刊不公平，却宣称自己本身就没打算进行全面评价，因此是否客观和公平似乎就没关系。但是，当这些体系采用的核心评价方法与科学研究主要趋势不一致、与学术期刊本身推进学科发展的宗旨不相符、与真正的符合发展趋势的学术质量无关时，它就已经失去了整体上的可信度。何况实际上，没有一个体系会说自己只是“部分期刊的核心期刊目录”，人们也会在实践中把它作为一个普适的体系来应用。

还有，许多“核心期刊目录”的评选过程并不透明，数据来源及其计算不清楚，往往还人为加入了许多主观“调整”，而“调整”的依据以及“调整”操作者信息也不公开。我们没有证据说明其中存在“任性”，但我们也没有数据可以对其进行重复验证。我们呼吁所有的“核心期刊目录”体系公开自己的所有数据及其计算方法，公布自己主观评时的定性指标及其判断依据，公布自己“调整”遴选结果的依据。如果这些评选及其调整是科学、严谨、规范的，其结果就应该在同样的数据、标准和流程条件下可重复验证。将数据、过程和责任人员予以公开，这已经成了科研领域和出版领域的基本要求⁹，也是公信力的基础。

其实，国际科技界《研究评价的旧金山宣言》¹⁰、《关于研究评价原则的莱顿宣言》¹¹和中国科学院学部主席团《追求卓越科学》宣言¹²中都要求科学地应

⁹ Journals unite for reproducibility <http://www.sciencemag.org/content/346/6210/679>

¹⁰ The San Francisco Declaration on Research Assessment. <http://www.ascb.org/dora/>

¹¹ Leiden Manifesto for Research Metrics. <http://www.leidenmanifesto.org/>

¹² 中国科学院学部主席团发布《追求卓越科学》宣言
http://www.cas.cn/xw/zyxw/yw/201405/t20140525_4126367.shtml

用评价指标和评价体系。如果仍然坚持对“核心期刊”的迷信和对“核心期刊目录”的迷信，本身就说明在学术评价和期刊评价上的肤浅和不负责任。

4、都是“指标驱动”惹的祸？

也许应该指出，“核心期刊”神圣化以及“影响因子”神圣化本身还不是问题的全部。问题还在于我们对科学的评价过度地依赖某种显性的、最好是量化的指标。这种对指标的依赖似乎已经在很大程度上绑架了我们的意识，造成“指标驱动的评价”、甚至“指标驱动的研究”。

笔者曾在 Nature 举办的一个讨论会上反复听到我国研究人员问“我怎样才能能在 Nature 上发表论文”，尽管 Nature 编辑回答说“你需要做一个高水平的研究”，但提问者明显感到“不解决问题”。我们看到，问题和答案的出发点有明显的差异，很可能导致人们行为的差异。对于提问者来说，发文→在核心期刊发文→在影响因子尽可能高的期刊发文，已经成了科研是否成功、工作能否“交账”和“职业生涯”能否发展的主要（甚至全部）依据，成了各类机构的宣传重点、成就象征和评价条件。这就在潜意识上、甚至在实际评价和管理中造成只要发了文章就“功成名就”，至于是否真正解决了问题、是否解决了对科学或对发展真正重要的问题，那已经无关紧要了。那么，为了快发文、多发文、发那些高影响因子期刊“喜欢发”的论文，可能就直接导致追风式科研、短平快科研、碎片化科研等，也容易诱导科学不端行为。

对指标的痴迷也对办刊理念和办刊方法带来负面影响。例如，前面已经提到的对跨学科内容的回避、对特色化办刊的回避、对“热点”主题的盲目追逐、对冷门问题的冷漠，等等。而且，对指标的痴迷还会驱使期刊“想方设法”把引用量“做上去”，这在一些圈子里已经成了交流的热点、密室交换的利器、甚至专业化“杀手锏”。有些期刊在这方面颇有“创新”意识，例如少发文章、发表综述文章、奖励作者引用自己刊物、建立“引用联盟”等等。

其实，许多看似神圣的指标本身应该走下神坛。前面我们对“核心期刊”去神圣化，现在来分析“引用”这个指标本身。“引用”作为影响力指标，是假定“引用”代表了使用、使用代表了影响、影响促进了科学发展。在大量文章的大量引用情况下，从平均来说“引用”确实能够从一个方面体现“关注”和“影响”，但是，引用本身有很多原因，热门引用论文不一定代表了杰出科学思想，引用量并非一定与论文的学术水平严格正相关。Nature 在 2014 年发文指出¹³，许多提出了后来获得诺贝尔奖的发现或创造的论文并不在高被引论文名单上，“史上引用量最高”的论文往往是与方法、数据等有关。尽管我们承认这些论文的效用，但它们往往不能代表所在学科的学术创造方向和一流学术水平。

¹³ The top 100 papers. Nature 29 October 2014. <http://www.nature.com/news/the-top-100-papers-1.16224>

而且必须看到，引用指标严重偏向基础研究论文，因为人们必须写一篇论文来引用他读过的另一篇论文才产生“引用”，因此“引用”作为指标有利于把学术论文作为主要产出的基础研究领域。但是，技术应用、科学普及、政策研究等方面的学术和研究，它们的读者受到“影响”后往往不是再写一篇研究论文，而是开发技术和工具、改变或完善政策、组织生产或管理、开展教育等。因此对这些领域，即使很大的“影响”也往往不是通过其他新的论文来体现，也可能带不来很多的引用量，因此在那些以引用量为基础的评价体系中就缺乏“价值”。另外，对于区域性主题、细节化问题、甚至高深的突破性的创新，往往比那些“全球性”问题或“宏观性”问题的“受众”少，无论研究水平和创新程度有多高，自然就不会有多少引用。试想，如果霍金没有写《时间简史》，有多少人知道他呢？

我们不反对采用指标、包括采用量化指标。许多指标，如果正确地计算和正确地应用，有其客观的有限的作用。但是，把任何指标用到它本身力所不逮的程度就变成了荒唐，真理超过一步就是谬误。

最后需要说明，我们批判某些指标体系的错误，并不指望甚至不希望又出现一个超级精细和“全面”的指标体系。任何指标都无非是从某一个角度观察复杂世界的投影，而异常复杂纠缠的指标体系往往可能存在更多的问题。我们的目的是排除对指标的迷信，不被指标（更不用说其中错误百出者）所裹胁。我们会坚持自己的办刊理念和职业操守，走科学的路，尽管荆棘满地。我们自身还要在发现创新思想、挖掘重大问题、引导前沿研究、夯实经验知识、建立最佳规范等方面努力。同时，作为学术期刊，更要在学术诚信方面筑牢底线，例如我刊早已实行“作者贡献声明”制度，并将在2016年实行支撑数据公开制度和潜在利益冲突声明制度，将公布学术诚信申诉查核和学术不端公开处理机制等。没有这些机制，仅仅依靠编辑部的善意和同行专家的能力，不足以保证期刊的学术质量，这已经是国际学术期刊界的共识和规范。我们希望，所有期刊都这样做，并作为学术期刊的基本合规门槛¹⁴。

¹⁴ 提高学术期刊的合规标杆. 《中国科技期刊研究》 2015年01期。
<http://ir.las.ac.cn/handle/12502/7983>