论文相关指标在美国科技评价中的作用研究

刘克佳
（中国科学技术部，北京 100862）

摘 要：学术论文相关指标是科技评价的重要参考标准之一。本文通过分析美国政府部门和研究机构开展科技评价的具体案例，发现论文相关指标在美国科技评价体系中发挥着无可替代的作用，但同时也引发了诸如过于看重论文数量等问题，并形成了不良学术风气。美国学术界近年对此进行了深刻反思，并努力矫正，但仍没有从根本上解决这一问题。

关键词：美国；科技评价；论文相关指标；研究机构
中图分类号：G321 **文献标识码**：A DOI：10.3772/j.issn.1009-8623.2021.05.008

对科研成果和科研人员进行客观公正的评价一直是学术界的热点议题。学术著作、项目产出及其经济影响、研究生培养、奖励、头衔、学术活动等各项评价指标都是体现研究人员和科研机构实力的重要内容。随着学术界对客观、公正、透明的要求日益强烈，可量化的论文相关指标，如论文数量、被引次数、期刊影响因子等成为学术界关注的核心评价指标，发挥着无可替代的重要作用。

1 论文相关指标在科技评价中的现状

按照评价对象的不同，如国家科技发展总体水平、政府科技计划的绩效、特定研究机构或研究人员等，所采用的评价方法、指标设置以及各指标的权重都有很大差别。

1.1 针对国家和机构的评价

美国是最早进行科研绩效评价的国家之一。20 世纪初，美国国会成立了咨询服务部，对科技进行研究、分析和评价，成为科技评价的雏形。二战后，随着政府不断加大科研投入，为了加强科研质量和效率管理，政府对科研投入、研究过程与结果等进行绩效评价的规定也愈发完善。美国国会 1993 年通过了《政府绩效与结果法案》，作为开展政府部门科研绩效评价的法律依据，并制定了绩效评价程序。

1.1.1 开展科技评价工作的机构

一是国会、联邦政府的科技评价机构。美国政府中没有负责全面管理科技评价的独立部门，但国会预算办公室、国会研究服务部、审计署等机构具有科技评价的职能，主要为国会和政府机构提供服务。此外，国家航空航天局、国立卫生院、国家科学基金会、国家标准与技术研究院等联邦科研资助机构都有自己的评估体系。

二是开展科技评价的社会机构。世界技术评估中心（World Technology Evaluation Center）是一家有代表性的非营利评估组织，发布了数十份宏观科研评价报告，涵盖信息、电子、生物、制药、核能、深海等领域。

三是学术机构。各类学术团体通常具有完善的组织网络、雄厚的研究资源、成熟的运作机制和显著的社会影响力。美国科学、工程和医学院是典型代表，一般只接受国会或联邦政府的任务委托，对重大研究项目进行评价。此外，美国知名智库，如战略与国际问题研究中心、布鲁金斯学会、信息技术与创新基金会、兰德公司等也会设计各种指标体

作者简介：刘克佳（1987—），男，硕士，主要研究方向为科技政策、国际科技合作。
收稿日期：2021-02-17

系，从各种角度对不同国家、不同领域的科技创新进行评价。

1.1.2 常见评价方法

美国科学院在1999年发布的《评价联邦研究计划：科学研究与政府绩效和结果法案》报告对最常用的科技评价方法进行了总结，主要有以下几种[1]。

文献计量分析：包括对出版物、论文引用情况及专利等的计数分析。该方法的前提是认同文献发表和引用的价值，最大优势是其定量的特性。一项研究表明文献计量分析与同行评议结果的吻合度大概在60%左右，说明文献计量分析具有较高的准确性。但其最大的问题是无法准确区分不同文章和引用的重要程度，例如谷歌创始人拉里·佩奇关于搜索引擎算法的论文发表在影响因子只有1.2的《计算机网络》期刊上，与论文的实际影响力显著不对称。

经济回报率分析：从经济学角度出发，通过多种手段建模，对科学研究的经济回报进行分析。优势是可以量化科技成果的价值，便于计算与比较，但缺陷在于只适用于某些目标比较明确的项目，而对基础科学以及需要很长时间才可能表现出社会、经济和科学影响的研究，很难进行准确评价。

同行评议：同行评议是科学界进行自我评价的一种方法，是很多联邦机构对科学研究进行事前、事中和事后评价的主要方法，缺点在于会受到人的主观判断影响。如果进行严格的专家选择与科学的流程设计，再辅以文献计量等定量指标，可能使主观影响降到最小。但同行评议往往难以使某些高风险、突破性创新脱颖而出，且只适合某一较窄领域，不适用于宏观整体评价。

案例分析：对重大科学事件和科学成果产生过程进行回顾式分析，优点是能发现某些潜在的、具有启示性的一般规律，帮助确定某些重要的非知识性产出，如机构或人员间的合作关系等。缺点是成本较高，难以大量统计，且容易受到分析者主观因素和分析能力影响。

定量指标分析：针对某一特定的评价目的，设计某些定量指标，通过统计学方法进行采样分析。其可靠性取决于指标反映评价目的的相关程度，以及可获取的采样数据质量。

1.1.3 典型案例

国家科学基金会下设的国家科学与工程统计中心（NCSES）是美国政府中唯一法定的科技创新相关指标统计机构，其两年发布一次的《科学和工程指标》涵盖了美国在科学、技术、工程、教育及创新等相关领域的权威统计数据，并对各领域发展态势进行总结。该报告由国家科学与工程统计中心在众多统计调查的基础上通过定量分析得出，只提供客观的基础数据，不强调数据所代表的意义，也不提供相关政策建议，为各类科技评价提供了坚实的基础。2020年度报告主要包括中小学科学教育、科学和工程高等教育、科学和工程劳动力、论文与专利、研发经费、成果转化、高技术产业产出与贸易、公众对科学的态度等内容，采用指标涵盖了在校生数量、科学和工程劳动力总量与薪酬水平、论文数量、经费投入、贸易量等，其中论文相关指标仅作为衡量科学产出的众多指标之一。

美国科技类智库信息技术与创新基金会从2009年起对世界主要经济体的创新竞争力进行评价和比较研究，主要采用人力资本、创新能力、创业、信息技术基础设施、经济政策、经济绩效6个一级指标，毕业率、单位就业人口的科研人员数量、企业研发投入、政府研发投入、出版物的数量和质量、风险资本、新企业数量等16个二级指标。

1.2 针对研究人员的评价

1.2.1 论文在科技评价中的重要作用

自18世纪末研究型大学诞生以来，通过出版物获得声誉和认可一直是学术界职业发展的重要途径。发表研究成果的能力也有助于在大学或研究机构获得终身职位。除论文外，获得项目资助、研究生培养、受邀的学术会议演讲和访问、奖励和头衔、成为国际期刊编辑委员会成员、以推荐信和评语为代表的同行评价等也可以体现研究能力，但这些指标具有较大的主观性。

根据Schimanski等[2]针对美国高校研究人员晋升和评价过程的研究，多数美国高校的指导方针和政策中通常没有明确规定各类指标的权重，但大多数高校研究人员都认为一个优秀的研究和论文发表记录是必要的，而且近几十年其重要性愈发突出：20世纪80年代，大多数高校希望员工在研究、教学或学校服务三者之一中取得优异成绩；

90 年代开始出现转变，卓越的教学和学校服务不再足以获得终身教职，到 21 世纪初，研究和论文发表已明确成为最重要的标准。

在各类学术出版物中，经过评审的期刊论文是证明研究质量的关键，权重明显高于书籍、报告等[3]。根据 Foos 等[4]的研究，美国高校的地球科学系在开展科技评价时一般会依据影响力和重要性将各类学术出版物进行打分，1 分为“非常重要”，5 分为“不考虑”。国际和国家级期刊论文为 1.22 分，书籍章节和地区性期刊为 2 分，会议论文为 2.3 分，政府出版物、教科书、实验室手册、实地指南和技术报告等为 2.5~5 分。

1.2.2 论文相关指标的应用实践

随着学术界对客观、公正、透明的要求日益强烈，一系列论文相关指标，如论文数量、被引次数、期刊影响因子、作者顺序、H 指数等，得以产生和发展。论文指标涵盖了基于研究和学术活动的广泛内容，其中一些数据可以展示学术能力和论文影响力，成为评估个人或机构科学能力的关键工具之一[5]。但是没有一个单一的指标能够衡量作者的业绩、质量或影响力，也不能预测作者的研究潜力。这些指标在实践中通常需要综合使用。

论文数量。由于美国高校在求职和晋升方面的要求日益严格，研究人员普遍感到发表特定数量论文的压力[6]，但有的高校针对论文数量要求做出正式规定，有的则是以潜规则或口头通知的方式。例如在美国的地球科学领域，27% 的院系制定了获得终身教职所需论文的数量要求，范围在 1~12 篇之间，平均 3.7 篇。在加州大学伯克利分校化学工程系[7]，虽没有明文规定，但研究人员认为一年需要 3~4 篇论文才能成功获得终身教职，而晋升正教授需要在主要期刊上发表大约 20 篇论文，并在研究领域获得广泛的国际认可。目前，我们很难准确量化不同学科对论文数量的要求，因为高校的指导政策并不具体，并且有一定的灵活性，比如当研究人员拥有高质量论文时，数量要求就不再重要。

期刊影响因子。尽管影响因子一直受到争论和批评，尤其是不适合判断某一具体论文或研究人员的水平，但学术界仍将其视为质量的代表。美国大学的地球科学系在评价期刊时，一般认为国际和国家级期刊优于地区性期刊。在美国的信息系统领域[8]，研究人员预计至少在“顶尖期刊”上发表 4 篇论文才能获得终身职位。美国大学高度重视优质论文的发表，但大多没有给出关于影响因子的明确规定，而是强调论文应是“开创性的”“可持续学术的象征”和“受到广泛称赞”。

各指标所占权重在规则层面的模糊性。美国研究机构通常不会规定出明确的权重来说明科技评价中哪一个指标是最重要的，政策的措辞往往非常宽泛。有研究发现[9]，美国研究型大学在教师的任期和晋升决策方面具有相当的灵活性，评价人员多使用自己的加权判断做出决定，或通过查看所有的申请材料来预测申请人是否具备持续和优秀的学术潜质。一般而言，研究成果质量是最重要的，可以推翻关于论文数量或被引次数的规定。这一规则的模糊也导致不同机构之间，甚至机构内各部门之间的差异。相比顶尖院校，二线院校对出版的要求不那么严格，可接受的期刊更多，所需论文数量较少，而且更加重视教学。此外，在不同的职业生涯阶段，晋升和评价的标准也有变化，申请终身教职的助理教授压力更大，倾向于在影响力大的期刊发表文章，而已获得终身教职的副教授可选择的期刊范围更广。

2010 年，Nature 杂志对 150 名研究人员和近 30 家研究机构的教务长、部门负责人和其他管理人员做了一项调查[10]，了解他们如何看待和使用各类定量指标。结果表明，研究人员和管理者对指标价值的看法存在脱节。研究人员表示，在录用、晋升、评估中，机构最重视项目拨款、论文数量、高影响因子论文以及被引次数，而同行评价在所有 34 项指标中仅排在中间位置，有 20%~30% 的受访者表示机构根本不重视同行评价。然而，大多数管理者坚持认为，定量指标确实会产生影响，但对录用、晋升、评估的影响并不像研究人员认为的那么重要。一些管理者甚至表示在决策时完全忽略论文相关指标，而主要依赖同行评价。但来自不同机构的管理者也显示出了相当大的差异。例如，亚利桑那州立大学等许多机构密切关注候选人争取项目资助的能力，而麻省理工学院则表示，对研究人员的晋升几乎从来没有考虑过项目资助，同时很少关注论文数量和被引次数。调查还发现，尽管 63% 的受访者对定量指标的使用方式表示不满，但他们仍

然欣赏定量指标的透明度和客观性。量化指标可以让年轻研究人员确信，学术网络中实际成就比个人关系更重要。所以管理者面临的挑战不是减少对指标的依赖，而是更清晰、一致和透明地应用这些指标。

1.2.3 新兴指标的出现

传统的基于计数的方法已无法展示更有意义和透明的成果，一种超越简单计数的指标变得越来越重要。H 指数和文件级分析（Document-level Unit of Analysis）是近年新兴的两个指标。

2005 年，加州大学圣地亚哥分校物理学家乔治·赫希（Jorge Hirsch）提出 H 指数，学术界越来越多地将其作为衡量研究人员、学术机构甚至资助机构业绩的基准。H 指数摆脱了传统上对期刊进行评价的思路，而是更为关注研究人员本身。H 指数借助一个公式，通过论文和引用情况估算出某一研究人员长期累积的研究成果的重要性和影响力。目前，包括 Google Scholar 在内的许多数据库已提供研究人员的 H 指数。

随着出版平台和社交媒体的发展，文件级分析作为新的衡量标准，已经允许社会公众参与衡量出版物的使用情况。文件级分析的衡量依据包括：在线下载次数、在线浏览次数、在线阅读书签标记次数、在 Twitter 或 Facebook 等社交网站上提及次数、博客或大众媒体中的讨论、分享平台如 Slideshare、Figshare 或 YouTube 中的收藏或推荐、在线评论平台如 PubMed Commons 中的评论或注释等。公共科学图书馆（PLoS）于 2009 年首次提供文件级分析数据，是目前最成熟的发布平台。其他发布者还有 ScienceDirect、PubMed Central、BioMed Central 等。然而，这些基于社会关注度或公众参与度的指标也可能代表了强烈分歧、研究错误或科研不当[11]。

1.2.4 典型案例

美国弗吉尼亚理工大学规定，终身教职候选人的评估主要根据三部分内容：教学、研究和推广，应基于某一领域的杰出成就和对个人职业生涯中未来成就的预测[12]。被提升为终身副教授的个人，必须在教学、创造性学术以及学校服务方面明确表现出突出的专业成就。评估过程所需的材料包括：学院院长、系主任等对候选人专业能力的评估，至少 4 封同行机构外部评审员对候选人学术能力的评审函，个人介绍性陈述，教授的课程和研究生指导工作，大学服务，国际活动等。

弗吉尼亚理工大学强调，在考察候选人研究成绩的质量和数量时，应首先考虑质量。质量主要是指对一个学科的推动、学科间关系的建立、思想和方法的创造性。相比候选人对熟悉的课题进行细微变化或重复的工作，在概念框架、结论和方法上的原始创新应更受重视。质量是评价候选人成绩的首要标准，尽管不是唯一的标准。此外，经同行评审的论文、会议论文和主题演讲、同行评价、担任期刊编辑、知识产权、获得项目资助情况、奖励和表彰等也可以展示候选人的专业素质。

2 论文相关指标引发的问题和反思

随着论文相关指标与个人求职、晋升、获得资助和荣誉等挂钩，论文发表已成为一种压力，在美国学术界引发了诸多问题，主要表现为对论文相关指标的过分强调、论文欺诈、期刊行业畸形繁荣，以及形成“不发表就毁灭”（Publish or Perish）的学术文化等。美国学术界目前已有一些反思的声音，提出一些倡议，但尚未见到明显效果。

（1）衡量标准模糊。

使用文献计量工具来评估实验室、研究项目或个人一直存在争议。虽然一个科学组织的科学生产力可以被看成投资的产物，但是在行政、经济或文献计量方面，没有一个线性模型能够比较组织架构、行政和财务状况非常不同的机构[13]。对于研究人员的工作表现，目前也没有一种算法或衡量科学生产力的基准可以用在评估中，这就造成不同机构采取的标准千差万别，无法形成统一明确的规则。

（2）对于数字的片面强调导致数字本身成为目的。

随着科学界对自身和组织的理解发生了深刻变化，招聘、项目评价和机构评估等逐渐从对同行评审的依赖转为对指标的依赖。这种对科学活动量化的强调提供了对复杂现实进行简化和客观化的方法，但数字本身变成了目的，丧失了对科学政策和资源分配的有效指导。此外，对数字的过分强调导致了一些扭曲结果，包括频繁的抄袭、自我引用、多余的论文、不当的作者归属和搭便车的研究人员。

科学评价不能沦为对论文的分析，一篇文章的有用性不能局限于它在科学界的可见度，科学成果的卓越性也不能降低到国际上的相似性[14]。

（3）指标设置和使用上的问题。

将期刊影响因子等同于某一具体文章的影响力是指标误用的常见表现之一。期刊上发表的论文中 10%~20% 的论文对期刊影响因子的 80%~90% 负责[15]。有研究显示 4 500 种顶级科学期刊发表的文章中，只有 45% 在发表后的 5 年内被引用[16]。期刊影响因子能否衡量论文优劣是应当存疑的。目前，学术界采用更精确的文档级指标以取代粗放的期刊级指标的工作进展缓慢。此外，论文影响力和质量的其他体现方法，因为没有数字化的处理和表现形式，也被忽略了。

（4）学术界自我拯救的努力。

学术界的反思和自我拯救的努力，除了上文提及的建立新的绩效衡量指标外，还以研究人员、机构和协会提出倡议的形式出现。然而，这些倡议对改变当前实践的全球影响仍然未知。

2012 年 12 月在美国细胞生物学学会年会上起草的《旧金山研究评估宣言》（San Francisco Declaration on Research Assessment）对期刊影响因子和研究人员绩效等指标之间的相关性提出了质疑[17]。截至2020年9月，已有2 033个组织和近16 400名个人签署了该宣言。《宣言》提出三项基本原则，并分别针对资助机构、科研机构、出版机构、计量指标提供方和研究人员提出具体建议。三项基本原则包括：一是在资助、聘用和晋升科研人员时，需要摈弃基于期刊的计量指标，例如期刊影响因子；二是需要评价研究工作本身的价值，而非出版物所在期刊的价值；三是要充分利用在线出版的机会，例如放宽对于论文字数、数据和参考文献的不必要限制，并探索评价重要性与影响力的新指标。

2015 年，Hicks 等[18]在 Nature 杂志上发表《莱顿宣言》（Leiden Manifesto），提出了适用于研究评估的十项新原则，凝练了基于指标的科研评估的规范。《莱顿宣言》提出定量和定性的证据不可或缺，并且这二者都是客观的。十大原则包括：量化评估应当支持定性的专家评审；根据机构、团队或研究人员的研究任务衡量绩效；保护当地相关研究的优秀成果；数据采集和分析过程应公开、透明、简单；允许被评估者检验相关数据和分析；在发表和引用中考虑到学科差异；对于学者个人的评估应基于对其全部研究成果的定性评判；避免不当的具体性和虚假的精确性；认识到科技评估和指标的系统性影响；定期审查指标并更新。

随着学术界对客观、公正、透明的要求日益强烈，科技评价经历了从定性评价到定量评价的演变。论文相关指标作为重要的学术能力标志，在美国科技评价体系中发挥着无可替代的作用，但同时也引发了诸如过于看重数据指标、学术风气不良等问题。美国学术界对此进行了深刻反思和努力矫正，虽然目前仍没有根本解决这一问题，但其提出的倡议激励着后来的研究人员继续不断地探索和前行。■

参考文献：

[1] 陈宁 . 美国的科技评价与科研事后评价概况 [J]. 全球科技经济瞭望，2007，22（12）：25-31.

[2] Schimanski L A, Alperin J P. The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future[J]. F1000Research, 2018(7): 1 605.

[3] Acord S K, Harley D. Credit, time, and personality: the human challenges to sharing scholarly work using Web 2.0[J]. New Media & Society, 2013, 15(3): 379-397.

[4] Foos A, Holmes M A, O'Connell S. What does It take to get tenure?[J]. Geotimes. 2004(5): 38-39.

[5] Hicks D, Wouters P, Waltman L, et al. The Leiden Manifesto for research metrics[J]. Nature, 2015(520): 429-431.

[6] Walker R L, Sykes L, Hemmelgarn B R, et al. Authors' opinions on publication in relation to annual performance assessment[J]. BMC Med Educ, 2010(10): 21.

[7] King C J, Harley D, Earl-Novell S, et al. Scholarly Communication: Academic Values and Sustainable Models[R]. Berkeley: Center for Studies in Higher Education, 2006.

[8] Dennis A R, Valacich J S, Fuller M A, et al. Research standards for promotion and tenure in information systems[J]. MIS Quarterly, 2006, 30(1): 1-12.

[9] Harley D, Acord S K, Earl-Novell S. et al. Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven

Disciplines[R]. Berkeley: Center for Studies in Higher Education, 2010.

[10] Abbott A, Cyranoski D, Jones N, et al. Metrics: do metrics matter?[EB/OL]. [2020-11-07]. https://www.nature.com/articles/465860a#Sec1.

[11] Miller A B, Wall C, Baines C J, et al. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial[EB/OL]. [2020-11-07]. https://www.altmetric.com/details/2114071?src=bookmarklet.

[12] Virginia Tech. Promotion & Tenure[EB/OL]. [2020-11-07]. https://faculty.vt.edu/promotion-tenure.html.

[13] Godin B. The Making of Science, Technology and Innovation policy: Conceptual Frameworks as Narratives, 1945-2005[R]. Quebec: Centre Urbanisation Culture Société, 2009.

[14] Mattedi M A, Spiess M R. The evaluation of scientific productivity[J]. Hist. Cienc. Saude-Manguinhos, 2017, 24(3): 623-643.

[15] Garfield E. The history and meaning of the journal impact factor[J]. JAMA, 2006, 295(1): 90-93.

[16] Bauerlein M, Gad-el-Hak M, Grody W, et al. We must stop the avalanche of low-quality research[EB/OL]. [2020-11-07]. https://www.chronicle.com/article/we-must-stop-the-avalanche-of-low-quality-research/.

[17] Alberts B. Impact factor distortions[EB/OL]. [2020-11-07]. https://science.sciencemag.org/content/340/6134/787 .

[18] Hicks D, Wouters P, Waltman L, et al. Bibliometrics: the Leiden Manifesto for research metrics[EB/OL]. [2020-11-07]. https://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351.

Research on the Role of Paper Related Indicators in US Science and Technology Assessment

LIU Ke-jia

(Ministry of Science and Technology of the People's Republic of China, Beijing 100862)

Abstract: The research paper related indicators are important standards for science and technology assessment. By analyzing the specific cases of science and technology assessment carried out by the US government and research institutions, this paper finds that paper related indicators play an irreplaceable role in the science and technology evaluation system in the US, but at the same time, it also causes problems such as over emphasis on numbers and bad academic atmosphere. In recent years, US academic community has made profound reflection and correction efforts, but problems have not been fundamentally solved yet.

Keywords: the U.S.; science and technology assessment; paper related indicators; research institutions