

文章编号: 1003-2053(2015)08-1127-07

# 国际科研评价体系现存问题及发展新趋势分析

张树良, 张志强, 王雪梅

(中国科学院兰州文献情报中心, 甘肃兰州 730000)

**摘要:** 在科学发展的新形势和新需求面前, 探索如何构建新的适应科学研究发展新模式的科研评价标准及评价机制已经成为当前科学界亟需解决的重大现实问题。基于对目前国际科研评价所存在的关键问题的剖析, 结合国际科研评价研究及科研评价体系改革的最新动态, 分析了国际科研评价发展的新趋势。

**关键词:** 科研评价; 科研评价体系; 同行评审; 科学计量评价; 现状; 发展趋势

中图分类号: G311

文献标识码: A

DOI: 10.16192/j.cnki.1003-2053.2015.08.002

1731年, 世界首份同行评审学术出版物英国皇家学会《医学短评与观察》(Medical Essays and Observations)的问世<sup>[1][2]</sup>标志着现代科学评价体系的正式诞生。同行评审是具有相同或相近学术背景或能力的研究人员对科研活动进行审核和评估的过程, 旨在保证科研活动的质量、绩效以及可信度<sup>[3]</sup>。同行评审的出现, 结束了科研活动自发和自我评估的局面, 对于维护客观公正的学术环境具有重要作用。发展至今, 同行评审机制已经成为建制最为成熟的科研评价机制, 并且是国际科研评价体系的基础和核心, 可以说现行国际科研评价体系仍是以同行评审为主导的评价体系。

随着科学研究全面步入新的发展模式即“模式2”时代<sup>[4]</sup>, 科研组织模式及知识生产方式的多元化, 使得担负引导、监督和管理科研活动、维护基本科学主旨和原则使命的科研评价体系面临着新的前所未有的挑战。在新的科研需求和不断加剧的科研资源竞争面前, 现行科研评价体系所存在的缺陷和弊端日益突显, 现行科研评价体系何去何从? 重新审视现行科研评价体系, 探索如何构建新的适应科学研究发展新模式的科研评价标准及评价机制, 以推动科学研究真正朝向服务人类福祉和社会发展的方向迈进已经成为摆在决策者和科研人员面前的重

大现实课题。

## 1 科研评价面临的新需求和新挑战: 现行科研评价体系存在的主要问题

### 1.1 缺乏统一的科研评价标准

目前, 不论在国家层面还是在科学共同体层面, 均缺乏统一的科研评价标准。不同国家之间在科研评价政策取向及科研评价体系构建方面存在显著差异。同时, 不同科学共同体之间在科研评价原则及方式上也有很大的不同。尽管, 上述差异和不同与国家体制及科研建制、社会经济发展水平以及科学共同体发展模式及运行机制的差别有关, 但这有悖于科学的“开放性”、“共通性”的本质, 也不应当成为凌驾于科学之上的构建统一的科研评价标准框架的所谓制度性因素。特别是在全球一体化进程不断加快、全球化问题和挑战日益严峻的背景下, 统一的科研评价标准对于推动科学研究的跨界合作以及加速科技发展进程显得愈发重要。

统一的科研评价标准的缺失导致国际科研比较难以获得广泛认同, 同时也有碍于客观、公正地认识不同国家科研体系及不同科学共同体的特点、发展水平及其差异, 从而影响到公众对科研价值的认知。

收稿日期: 2014-10-10; 修回日期: 2014-12-04

基金项目: 中国科学院兰州文献情报中心业务发展领域前沿扫描项目(Y300201001); 中国科学院文献情报中心情报战略调研项目(Y300281001)

作者简介: 张树良(1973-), 男, 河南荥阳人, 副研究员, 博士, 研究方向为情报研究理论与方法。E-mail: zhangsl@llas.ac.cn。

张志强(1964-), 男, 甘肃定西人, 研究员, 博士, 研究方向为科技情报与科技政策。

王雪梅(1976-), 女, 重庆永川人, 副研究员, 博士, 研究方向为资源环境科学计量评价。

## 1.2 科研评价机制无法适应现时需求

伴随科研模式的变革和科研需求的不断升级,现行科研评价机制的弊端愈发突显。

首先,尽管有着 200 多年历史的同行评审仍然是国际科研评价的主流手段(特别是在有着深厚评估文化传统的欧美发达国家),但是随着研究资源竞争的加剧,其本身所固有的主观性强、容易被人操纵、易形成学术壁垒以及开放度和透明度有限等弊端日益显现。在全球范围内,无论是评审者利用权力之便剽窃他人创新思想,还是通过评议者进行“暗箱操作”骗取研究资助等行为都屡见不鲜<sup>[5][6]</sup>。2013 年,美国哈佛大学 John Bohannon 有意伪造的学术论文竟然获得所投 OA 期刊 50% 以上的录用率<sup>[7]</sup>的事件更是引起国际学术界对现行同行评审制度的空前质疑,有学者更是一针见血地指出“这意味着同行评审体系的崩溃”<sup>[8]</sup>。

其次,就担负着“实现科研评价的客观公正”使命的科学计量评价而言,尽管,自科学计量学诞生至今,有关科研评价的计量学指标的研究层出不穷、成果纷繁,特别是自 20 世纪 90 年代初以来的近 20 年间,有关科研评价的研究产出总量增加了 10 倍<sup>[9]</sup>,但是这并未带来科研评价体系的根本改观:一方面,绝大多数研究仅仅局限于理论或学术层面的探讨,真正能被实际用于科研评价的屈指可数,于是科学界陷入了“面对评价指标的爆炸却莫衷一是”的尴尬境地;另一方面,由于数据标准缺失、数据垄断以及数据失准和不完整等因素,也使得基于科学计量的科研评价的置信度及其效力大打折扣。不仅如此,科学计量评价体系在数据选择、指标设计等方面存在的主观性,以及用于标准计量规程的研究成果类型的有限性也影响到评价结果的公正性。

同行评审机制是否该“寿终正寝”?科学计量评价出路何在?究竟什么样的评价机制才能真正推动科学研究的进步?如何构建符合科学研究发展实际、更为科学合理的科研评价机制已成为科学界必须深思的重要问题。

## 1.3 科研评价管理与运行模式亟需改进

在管理与实施层面,现行科研评价体系因其“功利性”和程序的复杂性而正在成为阻碍创新的制度性因素。一方面,由于现行科研评价体系通常都与研究资助和项目审批相挂钩,基于成本及风险防控的考虑,无论对于科研人员个人的考评,还是对于研究机构的评估而言,评估更看重已有研究成果

所反映的所谓“研究积累”,因而从这个角度讲,现行评价体系实际上是“产出导向的评估”<sup>[5]</sup>,这使得刚刚加入科研队伍的年轻学者和处于职业生涯早期的青年科学家很难获得重要竞争性科研项目的支持或研究机构及团队的认可,从而使其创新思想的展示及创新热情受到抑制。

另一方面,为应对程序日益繁杂的科研评价规程,科研人员不得不为此付出更多的时间成本,这势必影响到其科研的专注度和科研进程。目前在英国,科研人员因此而花费的时间有的甚至占到其全部科研工作时间的 1/3,为此,英国国家科研评价制度越来越受到科研人员的集体质疑。而在美国,根据对美国顶尖研究机构的调查,研究人员超过 40% 的时间用于科研管理及评估相关的行政事务<sup>[10]</sup>。牛津大学实验心理学家 Dorothy Bishop 认为,如今,越来越多的科研人员把精力都用在如何争取科研资助以及算计自己未来的科研评价结果上,科研人员需要做的科研之外的事情太多,这导致了目前大量科学研究有失水准<sup>[11]</sup>。由此看来,科研人员“整天忙于‘跑’项目、‘跑’经费”的现象并非中国科研现状个别。

## 1.4 亟待建立更为全面、公正的科研评价指标体系

目前,在科研评价中“唯论文”指标或过分强调论文指标的现象十分普遍,首当其冲的表现是引文指标的滥用。引文指标最初被用于科研评价的本意只是作为政府及资助机构评判研究质量的参考性量化指标,并且“重要的论文可能被更频繁的引用”作为一般规则,仅仅是在理想条件下的一种合理假设。实际上,基于论文引证关系所派生出的诸多引文指标(如影响因子 IF、H 指数等等)其适用性有着严格的条件限制。然而,在国际上,不加区分地将引文指标作为刚性指标用于评价科学家个人贡献、机构学术地位乃至国家研究实力的现象仍在蔓延。这之中最极端的例证当属芬兰的做法。芬兰是目前唯一一个以立法形式将引文指标确定为科研绩效评价主要指标的国家<sup>[12]</sup>。

不仅如此,目前的论文及其引文指标实际是以 ISI 论文数据为主要依据的。一方面,对于明显具有公益性质和公共导向的科研评价而言,如果其评判标准建立在 ISI 这样一个非开放获取的商业数据源基础之上,其动机和评价的公正性令人质疑。另一方面,就 ISI 数据库本身而言,由于数据收录的偏颇及质量问题,显著影响到结果的可靠性。英国伦敦

城市大学 Grant Lewison 认为,面对越来越复杂的研究过程和知识的多样化,基于 SCI 的研究评价体系已经无法满足需求,以 SCI 为中心建立研究评价体系的年代即将终结,应当建立在 SCI 之上的新的科研评价体系<sup>[13]</sup>。

基于更普遍的意义,以期刊论文指标为核心的科研评价体系不符合科研实际。首先,从科研过程来看,研究所依据和参考的知识来源并不只有论文一种形式,应当关注非论文知识形式的共享情况<sup>[14]</sup>。其次,从科研产出来看,论文实质是对学术思想或科研活动的特定表达和记录,只不过是众多科研成果类型中的一种,因此忽视非论文科研成果的科研评价体系显然有失科学性。

## 2 国际科研评价体系发展新趋势

针对现行科研评价体系存在的问题,为应对科研评价所面临的新需求和新挑战,世界范围内,对科研评价的理性思考不断涌现,在不同层面所开展的有关科研评价理论、方法及实施举措的探索和实践不断深化,这在促进国际科研评价体系不断完善的同时,也反映出其未来发展的新趋势。

### 2.1 科研的社会经济效应将成为科研评价的核心标准

如前所述,目前以产出为导向的科研评价体系,无法如实反映科研活动所产生的实际作用和成效,因而无法揭示科学研究的价值所在,这显然不符合科学研究的初衷以及科研评价的主旨。因此,有越来越多的研究提出“社会需求导向”、“社会价值导向”或“公共价值导向”<sup>[15-17]</sup>的科研评价概念,各国政府也纷纷推出相应的国家科研评价体系改革举措。其基本思想是:科学研究的公益性质决定了科研评价应当以考量科学研究的公共价值为核心。对于拥有深厚“评估文化”传统的西方发达国家而言,关注科研的社会经济效益的现实依据是:科学研究必须向作为纳税人的公众证明其对科研的投入是“物有所值”的,而对于公众而言,科研活动所产生的社会经济效应才是其所关注的科研价值的核心所在。

在国家科研体系层面,英国成为率先开展科研效应评估探索 and 尝试的国家之一。2014年,英国全面推行新的国家科研评价体系“研究卓越性评估体系”(Research Excellence Framework, REF)<sup>[18][19]</sup>,以

取代实施近30年的原科研评价体系 RAE。新的评价体系 REF 将首次引入“研究效应”指标,并将该指标的权重设定为20%的较高水平(计划未来将进一步提高至25%)。作为英国科研体系改革的重要内容,英国国家科研评价体系关注研究效应的目的在于响应社会日益高涨的对科研体系实际作用与科研成效进行考评的呼声,改革现行的以学术产出为重的科研评价标准,引导和推动科学研究朝向真正符合其致力于社会进步的使命的方向发展。REF 对研究效应的界定是:科学研究对学术之外的经济、社会、文化、公共政策或服务、医疗、环境或生活质量等的变革或改善所产生的影响。

由于英国原有科研评价体系完全奉行的是以学术为中心的关注科研质量和学术产出的标准,因而,将非学术成果即学术之外的研究所产生的社会经济等成效正式纳入国家科研评价体系无疑表明其改革的力度空前和决心之大,同时更为重要的是,作为“同行评审”的发源地,英国将再次引领科研评价制度的改革。

### 2.2 同行评审与科学计量评价相结合科研评价机制将成为主流

如上所述,无论是同行评审还是基于科学计量的单一评价机制均已无法满足科学发展及科研管理的现实需求,因此,如何克服现行评价机制的缺陷,探索新的科研评价手段或解决方案就成为当前及未来科研评价领域的焦点议题。其中,受关注程度最高的是通过将同行评审与科学计量评价相结合来改进现行评价机制的做法,从现实性和可操作性来讲,这也是现行科研评价机制改革的最便捷路径。

自2010年起,澳大利亚正式启用新的国家科研评估体系“研究卓越性评估”即 ERA 评估体系<sup>[20][21]</sup>,旨在建立澳大利亚国家科研体系研究卓越性评估框架,强化澳大利亚研究实力。澳大利亚 ERA 评估体系最令人关注的亮点即在于:探索如何将同行评审和文献计量指标有机结合。

首先,从整个评估体系架构来看,确定了同行评审与科学计量评价具体应用的“领域适用”原则即基于“领域差异决定了评价机制的不同”的思想,对两种评价机制的应用进行了明确的领域范围及类型划分。在评估体系所涉及的全部22个领域157个学科中,采用同行评审机制的学科为66个(涉及13个领域),主要为社会科学领域及纯理论学科;采用科学计量评价的学科为91个(涉及11个领域),主

要来自自然科学领域。

其次,基于“不同成果所对应的评价机制也应不同”的思想,特别对评估所认可的科研成果类型予以区分即将研究成果分为 2 大类:(1) 传统研究成果,包括书籍著作、书籍著作章节、会议出版物和期刊论文;(2) 非传统研究成果,包括专门策划的活动(如展览)、现场表演、原创作品、记录或呈交的作品以及非传统研究成果集。对传统研究成果的评估以科学计量分析结果为主要依据,而对非传统研究成果的评估则主要通过同行评审程序完成。

同时,为确保评估的公正性,规定了被纳入评估程序的研究成果接受同行评审的比例阈值:在评估过程中,接受评估的研究单元必须至少提交其研究成果的 30% 用于同行评审。

### 2.3 科研评价将实现系统化与自主化

为有效减轻科研评价给科研人员造成的负担,减少科研管理事务对科学研究的干扰,如何实现科研评价的系统化以提高科研评价的效率日益受到科学界的广泛关注。

为配合新的国家科研评价体系 REF 的顺利实施,英国高等教育资助理事会(HEFCE)委托美国兰德公司(RAND)开发了研究效应分析系统 Impact-Finder<sup>[22]</sup>,其核心是建立研究单元的研究效应矩阵(IA)。矩阵的纵向指标为不同的研究项目,横向指标为研究项目所对应的不同研究产出及其效应。有关信息通过专门的网络问卷调查子系统收集。分析结果以可视化的方式展示,据此,评估对象或用户可以即时且一目了然地获得其研究效应的具体分布及其强度等信息。同时,系统还提供针对特定研究的深度定量分析功能。该系统设置了两种应用情景:一种面向单一研究单元或评估对象,可用于其内部分析或自评;另一种则面向多研究单元或评估对象,可实现对多目标对象的整体分析和比较。该系统不仅将有效减轻新的科研评价程序给研究人员和研究机构所带来的负担,同时也将有助于降低在整个科研体系评估过程中对不同研究单元研究效应的分析难度及人为干预程度。

在机构层面,比较有代表性的实践案例为瑞典斯德哥尔摩环境研究所评估系统 PMEC<sup>[23]</sup>。2013 年 7 月,斯德哥尔摩环境研究所宣布建成研究评估系统 PMEC。该评估系统旨在实现对机构研究成果质量及影响的系统化评估。PMEC 研究评估系统的先进性及特色在于评估过程的自主化和自适应性。

目前该系统已被嵌入斯德哥尔摩环境研究所项目及机构管理体系。

PMEC 是一个集研究规划、监测、评估和交流于一体的综合研究评估系统。研究规划子系统基于研究项目的基本信息界定研究的直接影响范围,交流子系统据此确定研究具体发挥作用和影响的最佳路径。研究监测子系统则根据项目组或研究单元的科研活动日志跟踪其研究过程及研究作用和影响范围的变化,揭示变化原因及其趋势,并对项目组或研究单元的未来响应举措的选择提供依据。

在整个系统化评估过程中,要求项目组或研究单元向系统提交所有与其研究有关的研究产出信息,包括学术出版物(著作、著作章节、期刊论文、研究报告等)、学术活动(专题研讨会、学术会议等)以及媒体报道等。

如果研究项目得以成功实施,PMEC 评估系统则要求项目组或研究单元提供所有有关研究成效及研究进展的指标,用于研究成果评估和研究监测。所有指标都必须是可用于定量或定性评估的,如出版物数量、所使用的模型的类型等。同时项目组或研究单元还必须对每一指标的基准进行描述,以便监测其变化。

PMEC 评估系统采用“自适应项目管理”方法,使得项目组或研究单元可以随时根据系统监测评估结果,对其研究设计或部署做出调整,以使研究进展和成效符合预期。

PMEC 评估系统同时还可以定制的方式,实现对环境管理体系、气候风险减轻、治理方式转型和发展的重新审视等 4 个主题领域研究的专门评估。

PMEC 评估系统不仅可用于单个研究项目的监测和评估,而且同时可以满足更高要求的研究报告、监测及评估需求。此外,在整个机构层面,PMEC 评估系统还能够定期向斯德哥尔摩环境研究所研究的资助方(如瑞典国际发展合作署)提供更为详尽有效的高层管理报告,向其汇报有关研究所项目进展及科研管理等方面的信息,从而实现了从研究实施方(项目组或研究单元)到科研管理方(科研机构)直至研究资助方之间的实时有效互动,使评估结果更加全面、客观。目前该系统已被嵌入斯德哥尔摩环境研究所项目及机构管理体系。

### 2.4 网络指标日益成为科学评价的重要参考指标

随着网络应用在社会生活各方面的渗透和普及,科研模式和学术交流模式正在发生深刻变化,网

络平台正在成为科学传播和研究影响扩散的重要渠道,传统文献数据库已经无法满足 web2.0、web3.0 及至 webx.0 环境和大数据背景下的科学评价,网络指标作为科学评价指标体系的补充成为必然。

目前,在国家层面,欧盟成为积极应对上述新挑战,探索将网络指标纳入科研评价体系的代表。作为第 7 框架计划(FP7)科研体系改革的重要组成部分,欧盟于 2010 年启动新的科研人员评估标准联合研究计划 ACUMEN<sup>[24]</sup>,旨在完善科研评估标准,克服当前科研体系所存在的无法反映科学研究服务社会经济发展的基本职能、定量指标不适于科研人员个体评估以及定量化指标严重不足等缺陷。

与传统科研人员评价标准相比,ACUMEN 科研人员评价体系最大的特点即在于正式引入网络指标及其相应的评价工具。

#### (1) 引入学术搜索引擎数据

引入 Google Scholar 和 Google Books 引文指标以及基于 Google Scholar 的 H 指数。事实上,目前 Google Scholar 学术文献传播应用和引文影响已经超过了以 WoS、Scopus 等为代表的传统文献数据库。同时,研究表明,在社会科学和人文科学领域,Google Scholar 引文同 WoS 引文之间存在强关联关系。而 Google Books 数据则弥补了传统数据库书籍被引记录缺失的不足,并且更符合社会科学和人文科学领域以专著产出为重的客观实际。不仅如此,基于网络学术搜索引擎数据的 H 指数不仅能够更为全面地反映科研人员的学术影响,而且能够改善传统引文分析对社会科学不利的情况<sup>[25]</sup>。

#### (2) 网络计量指标的应用

ACUMEN 评价体系所采用的网络计量指标(Webometric indicators)主要包括网络引文、网络链接和网络报道指标。网络引文指标关注由科研人员完成或提供的网络学术资源与网络课件的被利用情况;而网络链接和网络报道指标则通过分析科研

人员科研成果的网络被引用情况揭示科研人员的网络学术影响。

#### (3) 基于社交媒体数据的补充计量指标的纳入

随着基于 web2.0 技术的网络应用的推广,社交媒体正在改变包括科研人员在内的社会主体的行为模式以及包括科研成果在内的信息传播方式,因而基于社交媒体数据的补充计量指标(Altmetric indicators)的兴起已不可阻挡。

ACUMEN 评价体系主要采用以下补充计量指标:

① 在线学术成果的被使用情况:通过数字图书馆、数字期刊、在线数据库及机构网站,或者通过学术社交网站,科研人员的学术成果的被浏览量、下载量和读者数量。

② 学术博客被引用情况:科研人员以博客形式发表的论文或学术信息的被引用量。

③ 科研成果所获得的网络评论情况:科研人员发表的学术成果所获得的积极的网络评论数量。

④ 学术成果的微博客引用情况:科研人员的学术成果在以 Tweet 为代表的微博客中的被引用量。

⑤ 科研成果被社交网站的引用情况:科研人员的学术成果被以 Facebook 为代表的社交网站的发帖引用量。

⑥ 学术成果的网络视频受关注程度:科研人员的学术讲座等视频信息在 YouTube 和 TedTalks 等视频网站的浏览量及得到的评论数。

在机构评价方面,美国全球发展中心(Center for Global Development, CGD)在智库绩效评估<sup>[26]</sup>中综合考虑分析成本、评估结果的客观性和代表性以及评估过程的透明性等因素,提出了以网络和社交媒体指标为核心的新的机构影响力评估方法,旨在客观反映智库的公共特征即社会及决策影响的功能实质,揭示智库因其职能的有效发挥所赢得的“公共影响力”。具体指标构成如表 1 所列。

表 1 CGD 机构公共影响力评价指标构成

序号	评估指标	测算标准	测算方法	数据来源
1	机构受社交网站用户的关注程度	机构的 Facebook 和 Twitter 用户关注量	网页快照: Facebook like + 和 Twitter followers 数量	网站数据
2	机构网站流量	机构的相关全球网络流量排名情况	特定时间平均值	网站数据
3	机构网站被链接量	机构网站被其他网站链入数量	特定时间累计值	网站数据
4	机构被媒体报道量	机构被国际新闻报道的次数(全部语种)	特定时间累计值	网站数据
5	机构被媒体报道量	机构成果在 Google Scholar 中的被引用量	特定时间成果的被引用量	网站数据

科研成果通过网络的传播和扩散速度是传统媒体所无法企及的,同时,网络指标的即时性可以有效克服传统文献指标的时滞性,特别在研究效应的评估方面,网络指标不仅可以揭示传统文献计量指标所无法反映的研究效应,而且利用网络指标还能够实现对研究效应的实时监测与跟踪。

### 3 总结与展望

现行的以同行评审机制为主的科研评价体系已经无法适应科学发展新模式的需求,其在评价标准、评价机制、评价的管理与运行模式以及评价指标体系等方面所暴露出的问题正在成为科学研究实现其服务人类福祉和社会进步根本使命的障碍。为应对科研评价的新挑战,以英国、澳大利亚、欧盟等为代表的国家和地区以及相关科研机构正在积极探索完善和改进科研评价体系的政策举措,与此同时,相关研究也持续受到关注。受此促动,国际科研评价开始向以社会经济价值为导向、不同评估手段相结合、评估系统化、自主化以及评价指标多元化的方向发展。

可以预见,在国家、国际组织和科学共同体的协同努力并在公众的监督下,国际科研评价体系将不断得以完善,科研评价将回归其科学之本真,实现其引导和监督科研活动、维护基本科学主旨和原则的使命,推动科学研究的良性发展。展望未来的科研评价体系,多种评价手段互为补充、定性与定量方法相结合的评价模式将成为主流,科研评价机制及评价指标体系构建将更加体现出系统化、动态性、跟踪型的特点。

#### 参考文献:

- [1] Spier R. The history of the peer - review process [J]. *Trends in Biotechnology*, 2002, 20(8): 357 - 358.
- [2] Gupta M N. Peer review: Past, present and future [J]. *Current Science*, 2013, 105(2): 159 - 161.
- [3] Zuckerman H, Merton R K. Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system [J]. *Minerva*, 1971, 9(1): 66 - 100.
- [4] Faggiolani C. Research evaluation and Mode 2 science [J]. *Lett Mat Int*, 2014, 1: 191 - 197.
- [5] Fedderke J W. The objectivity of national research foundation peer review in South Africa assessed against bibliometric indexes [J]. *Scientometrics*, 2013, 97: 177 - 206.
- [6] Gould T H P. The church and peer review: Was “peer” review fairer, more honest than now? [J]. *Journal of Scholarly Publishing*, 2012, 44(1): 36 - 60.
- [7] John B. Who’s afraid of peer review? [J]. *Science*, 2013, 342(6154): 60 - 65.
- [8] 林小春. 免费电子学术期刊被“钓鱼”引热议 [EB/OL]. [http://news.xinhuanet.com/2013-10/08/c\\_117620745.htm](http://news.xinhuanet.com/2013-10/08/c_117620745.htm). 2013-10-08.
- [9] Van Noorden R. A profusion of measures [J]. *Nature*, 2010, 465: 864 - 866.
- [10] Penfield T, Baker M J, Scoble R, et al. Assessment, evaluations, and definitions of research impact: A review [J]. *Research Evaluation*, 2014, 23: 21 - 32.
- [11] Owens B. Research assessments: Judgement day [J]. *Nature*, 2013, 502: 288 - 290.
- [12] Adam D. Citation analysis: The counting house [J]. *Nature*, 2002, 415: 726 - 729.
- [13] Lewison G. Beyond SCI citations: New ways to evaluate research [J]. *Current Science*, 2005, 89(9): 1524 - 1530.
- [14] Hahnel M. Referencing: The reuse factor [J]. *Nature*, 2013, 502: 298.
- [15] Edler J, Georghiou L, Blind K, et al. Evaluating the demand side: New challenges for evaluation [J]. *Research Evaluation*, 2012, 21: 33 - 47.
- [16] de Jong S, Barker K, Cox D, et al. Understanding societal impact through productive interactions: ICT research as a case [J]. *Research Evaluation*, 2014, 23: 89 - 102.
- [17] Bozeman B, Sarewitz D. Public value mapping and science policy evaluation [J]. *Minerva*, 2011, 49: 1 - 23.
- [18] REF 2014. Assessment Framework and Guidance on Submissions [EB/OL]. [http://www.ref.ac.uk/media/ref/content/pub/assessmentframeworkandguidanceonsubmissions/02\\_11.pdf](http://www.ref.ac.uk/media/ref/content/pub/assessmentframeworkandguidanceonsubmissions/02_11.pdf). 2012-07-06.
- [19] REF 2014. Decisions on Assessing Research Impact. [EB/OL]. [http://www.ref.ac.uk/media/ref/content/pub/decisionsonassessingresearchimpact/01\\_11.pdf](http://www.ref.ac.uk/media/ref/content/pub/decisionsonassessingresearchimpact/01_11.pdf). 2012-11-13.
- [20] ARC - ERA. ERA 2012 Evaluation Handbook [R]. Canberra: Commonwealth of Australia, 2012.
- [21] ARC - ERA. ERA 2012 Peer Reviewer Handbook [R]. Canberra: Commonwealth of Australia, 2012.
- [22] Rand. Impact and the research excellence framework:

- new challenges for universities [EB/OL]. www.rand.org/content/dam/rand/pubs/. . . /RAND\_CP661.pdf? 2012-01-18.
- [23] SEI. Measuring the Outcomes of SEI's work: The Planning, Monitoring, Evaluation and Communication (PMEC) System [EB/OL]. <http://www.sei-international.org/mediamanager/documents/Publications/SEI-FS-2013-PMEC.pdf>. 2013-07-26.
- [24] ACUMEN Consortium. Guidelines for Good Evaluation Practice with the ACUMEN Portfolio [EB/OL]. [http://www.acumen.eu/wp-content/uploads/Guidelines-Portfolio-TOC-document.v13dx.pdf](http://research-acumen.eu/wp-content/uploads/Guidelines-Portfolio-TOC-document.v13dx.pdf). 2014-05-02.
- [25] Anne-Wil H. A preliminary test of Google Scholar as a source for citation data: A longitudinal study of Nobel prize winners [J]. *Scientometrics*, 2013, 94: 1057-1075.
- [26] CGD. Measuring Think Tank Performance: An Index of Public Profile [EB/OL]. [http://www.cgdev.org/sites/default/files/think-tank-index\\_0\\_0.pdf](http://www.cgdev.org/sites/default/files/think-tank-index_0_0.pdf). 2013-08-08.

## The analysis on the existing problems and new development trends of the international research evaluation system

ZHANG Shu-liang, ZHANG Zhi-qiang, WANG Xue-mei

(Lanzhou literature and information Center of Chinese Academy of Sciences, Lanzhou 730000, China)

**Abstract:** Faced with the new circumstances and requirements of science development, exploring how to develop the standards and mechanisms of research evaluation adapting the new development mode of scientific research has become the significant practical problem. Based on profiling the crucial issues of international research evaluation and combined the recent advances in international studies of research evaluation and the innovation of research evaluation systems, the trends of international research evaluation are analyzed.

**Key words:** research evaluation; research evaluation system; peer review; scientometrics evaluation; current situation; development trend

(上接第 1126 页)

## The limitation of the peer review and ethical assessment

WANG Guo-yu, ZHU Xiao-lin

(Department of Philosophy, Research Center of S&T Ethic and Management, Dalian University of Technology, Dalian 116024, China)

**Abstract:** As one of the scientific evaluation method, peer review reflects science autonomy and self-discipline since the Enlightenment. Yet, with the changes of means and ends in the science research, scientific knowledge is now defined as uncertainty rather than certainty, along with convergence and interaction of science, technology, economy, society and politics. The scope of the scientific activities has been far beyond the range of scientific community. The influence of scientific activities, including its negative elements, has spread to the nature and various social fields. In this context, it is important to evaluate the social consequences to scientific applications. Peer review, as one of the important scientific evaluation methods, still has its limitations. Based on the review of the origin and limits of the peer review method, this paper proposes the constructing a multi-evaluation mechanism, in particular the ethical assessment should be included on the scientific assessment system, and the public acceptability of the scientific and technological researches should be considered.

**Key words:** peer review; science autonomy; social construct; multi-evaluation; ethical assessment