中国教育科研参考

2017 年第 4 期
总第（398）期

中国高等教育学会编 **2017 年 2 月 28 日**

目　录

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（02）
国外哲学社会科学学术评价制度及其启示.....袁　清　姚　威（07）
英国大学科研评价改革：从 RAE 到 REF .....刘　莉（10）
澳大利亚 ERA 分类科研评价体系及其启示.....刘兰剑　杜向民（15）
日本大学教师学术评价的理念、内容及方法.....韩小娇　高　军（23）
韩国高校学术评价制度体系的借鉴与反思.....王晓杰　金佳律　付瑞萱（28）

编者的话：学术评价是衡量学者、科研成果、学术机构的学术水平的主要尺度。近年来，我国各级政府和高校、科研院所都制定和出台了具有针对性和指向性的学术评价政策和规则，对于引导学术评价的健康有序发展具有积极意义，但在实践中也出现一些备受争议的问题。因此，如何创新理念、突破既有思维模式，优化并重构符合学术发展内在逻辑的学术评价制度是我国高校面临的重大课题。知识经济时代世界主要发达国家的学术评价制度的演进模式和发展经验可为我国高校构建符合中国国情和国际发展趋势的学术评价制度提供一定的借鉴。为此，本刊以“高校学术评价”为选题，集中选编若干文章，供读者参阅。

主编：王小梅　　本期执行主编：范笑仙　　责任编辑：王者鹤
地址：北京市海淀区文慧园北路 10 号中教仪楼中国高等教育学会《中国高教研究》编辑部
邮编：100082　　电话：（010）59893297
电子信箱：gaoyanbianjibu@163.com
网址：www.hie.edu.cn（中国高等教育学会——**学术观点**栏目）

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

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。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，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。

一、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

（一）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。

（二）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、容易被人为操纵、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。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”。

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，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。

（三）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”，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。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。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例。

（四）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。

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时代即将终结，应当建立在SCI之上的新的科研评价体系。

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。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合理性。

二、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。

（一）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”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和尝试的国家之一。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)，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正真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。

（二）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，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。

（三）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，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估；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。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。

（四）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、web3.00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，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。

2. 网络计量指标的应用。ACUMEN评价体系所采用的网络计量指(Webometric indicators)标主要包括网络引文、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。

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

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。

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。

三、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。

（张树良，中国科学院兰州文献情报中心副研究员，甘肃兰州　730000；张志强，中国科学院兰州文献情报中心研究员，甘肃兰州　730000；王雪梅，中国科学院兰州文献情报中心副研究员，甘肃兰州　730000）

（原文刊载于《科学学研究》2015年第8期）

国外哲学社会科学学术评价制度及其启示

袁　清　姚　威

一、引言

近年来，随着国家对哲学社会科学研究投入的大幅增加，哲学社会科学学术评价已成为学术界普遍关注的问题，也是哲学社会科学研究领域异常关注的一个热点话题。哲学社会科学的学术评价将事关学术发展的前途和方向。

由于哲学社会科学的研究成果存在着一定程度的主观性、经验性和价值判断等特点，哲学社会科学评价标准客观上存在着学术标准与非学术标准的冲突;基础理论研究标准与应用研究标准的冲突;本土化标准与国际性标准的冲突;创新性标准与规范性标准的冲突;不同评价对象标准的冲突等悖论，进一步增加了成果评价的复杂性。

关于哲学社会科学学术评价制度，在西方高等教育发展的漫长进程中，已经形成一整套相对成熟的制度规范和操作办法。他们普遍运用同行评议制度作为哲学社会科学学术质量评价的主要手段。由于同行评议评价具有主观性、耗时长、成本高、局限于本领域评价而不利于跨学科成果评价等问题，为弥补这一缺陷，匈牙利学者Beck开发了客观测评科学成果质量的测量标尺方法，该方法有42个测度等级，将科学认识论结构分为6个阶段：设立公理、建立定理、联结理论模型和现实、获得经验事实、探索性实践和程序性实践。此外与同行评议相比，文献计量方法具有客观具体、可操作性强等优点，也受到了广泛的应用，除了常见的引用次数和影响因子法，最新的文献计量方法还包括美国学者赫希(J．E．Hirsch)提出的h指数(h－index) 法和埃赫(Leo Egghe)提出的g指数(g－index)法等。

就我国哲学社会科学学术评价制度而言，目前评价制度的缺陷主要体现在重视量化考核的同时忽略了成果质量，注重引文数量的同时忽视了经济效应，重人情而欠科学公正，以导致抄袭剽窃风气盛行。虽然在理论和实践上已经进行了很多有益的探索，但总体看仍然存在着评价理论薄弱、评价主体不清、导向不明（重数量轻质量和创新）、评价标准较为模糊、评价方法不够多样化、评价过程过分行政化和人情化、评价结果公信度不高等问题，尚未形成具有国际影响和较高公信度、突出创新和质量导向的哲学社会科学学术评价制度。

2011年11月教育部发布了《教育部关于进一步改进高等学校哲学社会科学研究评价的意见》（以下简称《意见》），该《意见》指出，要开展科学有效的科研评价，是推动科研管理创新，优化研究资源配置，构建现代科研管理制度的重要内容。应该“充分认识改进哲学社会科学研究评价的重要意义”，“确立质量第一的评价导向”，“实施科学合理的分类评价制度”。因此，完善以创新和质量为导向的科研评价机制，建立健全成果分类评价标准，探索建立政府、社会组织、公众等成果受益者参与的多元多方评价机制，对于稳固提升哲学社会科学研究成果的国内外影响力，实现哲学社会科学的长期繁荣具有重大理论和现实意义。

本研究通过对相关文献的梳理，归纳学术界已经研究的主要观点，并选取有代表性的大学进行较为深入的案例分析，以期提供关于国外学术评价的系统观点，更加客观、全面地把握国外学术评价制度背后所体现的核心理念及制度基础，从而为我国哲学社会科学学术评价制度设计提供有针对性的参考。

二、发达国家哲学社会科学学术评价制度比较

由于各国本国学术界的政治体制、学术发展状况和特点、历史背景、学术制度等差异，各国在对哲学社会科学成果进行评价时的评价目标、评价主体、量化方法等方面都存在着差异。何燕玲，蓝满榆(2011)将西方主要国家的学术评价模式归纳如下(见表1)。

表1　若干国家哲学社会科学学术评价的基本模式比较

	美国	英国	法国	日本	中国
评价目的	聘任、项目资助、成果奖励	大学拨款	人事聘用、学术价值和市场价值	人员聘用、政策效果	人事聘用、奖励及项目鉴定等
评价主体	联邦政府、哲学社会科学研究者、民众	英国高等教育和科研项目办公室	政府、民众	专家	专家、行政部门
评价标准	学术性、效益性、真理性	按欧洲研究区(ERA)相关政策（注重产量、影响、质量、实效）	欧洲研究区（ERA）相关政策（注重学术价值与社会影响）	创新性	学术性、效益性、影响力
量化方法	侧重于 SCI、SSCI，A&HCI 计量	科学计量、文献计量	缺乏标准化、集中化，侧重于市场（民众/媒体）评价	论文发表数量统计、引用量统计及数据库	SSCI，CSSCI，A&HCI 等指标
评价模式	专业模式	管理模式	后社团主义模式	文献计量模式	同行评议与文献计量相结合
评价制度	同行评议为主，把SSCI、A&HCI 等作为文献计量首选	以同行评议为主，所有研究都给予相同的权重	同行评议与公众评价相结合的多元性和弹性评价	专家评议、文献计量	同行评议与文献计量相结合

通过上表，我们不难看出，从评价的目的来看，学术评价一般与人员聘任及激励行为直接关联，具有明显的筛选功能和资源配置的功能;从评价的主体来看，虽然一般的做法是由同行、专家进行评价，但政府部门的参与已经成为非常普通的现象，英国的高等教育和科研项目办公室、法国的教育部，是其学术评价体系中的重要参与者;在评价标准方面，包括了学术成果的创新性、影响力、社会效益等维度;在量化方法的运用方面，既包括了较为简单、原始的成果数量统计，也包括了SCI、SSCI、A＆HCI的收录情况及被引用情况的分析。在总体的评价模式上，上述各国也呈现出各自的特征，如美国的学术评价模式是较为典型的专业模式，而英国则是管理模式，法国采取的是后社团主义模式，日本具有文献计量模式的特征。

（一）美国大学哲学社会科学学术评价制度的基本要素

美国现代哲学社会科学研究经历了一百多年的发展，不但在学术上取得了令世人瞩目的成就，而且形成了一套建立在学术自由、学术自治、学术独立等核心理念基础之上的相对完整、行之有效的学术管理制度，而学术成果的评价制度是这一制度中的重要组成部分。具体而言，美国大学哲学社会科学学术评价制度包括了以下若干基本要素:

1. 同行评议制度。同行评议制，指的是某一或若干领域的专家采用一种评价标准，共同对涉及上述领域的一项事物进行评价的活动。哲学社会科学系统的评价对象则是指研究者进行哲学社会科学研究活动所获得的知识产品。对哲学社会科学研究成果进行同行评议采用的主要标准是该科研成果的学术水平。同行评议的专家根据此标准可以对每个待评议的研究成果的优劣作出评判。通常的做法是对成果进行打分。得分的高低即为对此科研成果水平的一个评价。

在美国，对哲学社会科学研究成果进行同行评议有多种不同的操作方法，其中较为流行的做法主要有两种:一是同行专家组或委员会评议，即小组评议;二是通信同行评议。在进行小组评议时，通常的做法是聘请本领域内的数位专家组成讨论小组，以会议的方式进行评议。会议的讨论结果即是对该科研成果水平的一个评定。在进行通信评议时，一般会将待评议的成果寄给同一领域内的多个专家，请他们提供书面的评价意见，然后通过邮件反馈回来。

2. 匿名审稿制度。美国的学术期刊普通采用匿名审稿制对学术论文进行审核。在匿名审稿过程中，普遍实行的是稿件作者和审稿人双向匿名的做法。在学术期刊实施的匿名评审制度中，由于评议人往往是该研究领域的专家或资深研究人士，因此评议人在客观上扮演了期刊学术质量的“守门人”的角色，这一做法能够较好地保证拟发表的学术论文的质量。

3. 学术规范制度。美国高校均制定了较为完善的学术规范制度体系，明确了进行学术活动的基本学术准则。与此同时，为处理各种学术违规现象，美国政府和大学成立了专门机构负责处理，并制订了相应的惩戒措施，努力将学术不端和学术违规行为降到最低程度。

（二）美国高校对哲学社会科学的学术评价

美国高校哲学社会科学教师进行学术评价时，一般考虑三个方面：教学、学术和社会服务。不过三者的受重视程度明显不同，其中学术水平是哲学社会科学教师晋升过程中最为重要的影响因素。学术水平评价的指标一般包括在高层次的刊物上发表的论文数、出版的书籍数(独著或合著)和经过同行专家评定的研究工作质量。著作被引用的次数、未公开发表的论文和自我评定等项目则占次要地位。其他指标还包括获得的研究课题资助或奖励、所获荣誉和奖励、担任专业刊物的评审人等。

1. 同行评议。为了公正、充分的对哲学社会科学教师学术水平进行评价，美国大学哲学社会科学教师学术水平评价制度一般都使用来自同行评价。尤其是进行晋升和终身聘任评价时，本系或校外同行提出的意见是判断学术水平的重要依据。不过，不同学校采用的同行评议程序有所差异，大致如下:

◆成立专门委员会。

委员会的构成各不相同，多数经选举产生，由具有终身职位的资深教授组成，也有的委员会由全系教师组成，但正在申请晋升的教师不允许在委员会任职，委员的遴选程序公开，以保证委员会能够客观公正的进行评议。

◆聘请外部评审员。

这种做法常为巨型研究型大学所使用，因为高度专业领域的专业才能，往往是同行专家的个人评判，它比一般委员会的评判更为准确。

◆发信函征求专家意见

这种形式普遍用于晋升评价中，邀请校外人员对申请晋升教师的学术成就做出客观评价。

◆本系同事评价。

这种做法的逻辑前提是同事比其他群体更具有评判某些专业方面的能力。

2. 同行专家的评议信。美国大学哲学社会科学教师学术水平评价，尤其是进行晋升和终身聘任评价时，评审委员会特别注重本系或校外同行对被评审者提出特别重要的意见，这些意见是判断学术水平的重要依据。因此，在进行评审时，评审委员会要求获得一定数量的外部推荐人及其推荐信。外部推荐人的人选是候选人和系里共同推荐的结果，委员会从名单中挑选部分推荐人(至少选择三个被评审者推荐的推荐人)，征求他们的推荐信。不同级别的晋升评审要求获得外部推荐信的数量也会有所不同，级别越高，要求的数量就越多。

三、结论与启示

合理的学术评价体系是学科成功建设的保证。美国的基础研究成果评价的基本方法采用同行评议和文献计量分析，应用与发展研究成果评价的基本方法是政策效果评估、民间测验(社会评价)、社会实验。英国高校科研评估是整个欧洲发展最为成熟的大学科研评价体系，由于受政府委托的独立中介机构——HEFCE表达政府要求，反应大学愿望，接受政府、大学和社会的监督，避免了一些人为因素的干扰。因此，我们可以借鉴美国、英国的科研评价体系，进一步改革哲学社会科学学术评价制度，主要结论和启示如下:

1. 灵活多元，进一步完善同行评议制度。进一步完善目前基于客观指标的学术评价制度，完善同行专家评议制度，对于同行评议中存在的评价主观性、局限性、片面性等问题，不应采取一刀切的办法，应逐步改变过于刚性的学术评价制度，增加评价的灵活性，把定量评价与定性评价相结合，进一步强化成果的质量，比如，对应用性成果，除了注重其成果的学术价值之外，更应考虑它的实际应用价值及其社会价值。进一步建立和完善同行评议专家数据库，吸纳国内外同行的专家参与学术评议，通过信息化的手段，完善专家库的信誉档案，在条件许可的情况下，选用小同行评议专家，从根本上杜绝外行评内行的现象。同时，同行评议制度应严格实行专家回避制度，从而减少各种非学术因素的干扰。

2. 重质轻量，加大力度推行代表性学术成果制度。代表性学术成果是能够反映学者在该学科领域中的学术水平和学术地位的成果。尤其是在哲学社会科学领域评价中，代表性学术成果将更成为哲学社会科学重要的评价指标。它最突出的特点就是克服了目前评价体系中普遍存在的“唯数量化”，这在一定程度上加强了对学术成果的质量控制，提升科研精品在评价体系中的权重。同时，代表性学术成果制度可以避免“投机主义”倾向，对于哲学社会科学工作者，尤其是对人文基础学科的学者而言，它的产出往往需要经过较长的时间周期，可以不受常规考核指标的限制，因此，对于这些基础类的研究成果，可适当延长其评价周期，激发学者学术自律与学术自觉的自我约束力，将更有利于抑制浮躁、树立良好学风，产出精品成果。

3. 公正透明，逐步试行专门评价机构制度。设立独立的哲学社会科学专门评价机构是哲学社会科学研究发展的必然趋势。我国目前哲学社会科学成果的评价由政府部门组织的居多，而从国外发达国家的经验来看，对一项学术成果的评价，是既不能由成果的完成人自己评价，也不能由成果的委托人如政府来评价，而应由独立于双方的第三方来评价。这个专门的机构独立于政府和高校，是一个非官方的中介机构，它接受政府及高校的监督，保证作为第三方的中立性，使得学术评价更为公正和透明。因此，积极探索试行由第三方中介机构参与的评价制度将是实现哲学社会科学评价科学化、规范化发展的重要途径。

4. 加强文化建设，打造良性学术生态。良好的学术生态，是确保学科实现长远、良性发展的重要基础。学术生态建设要遵循学术发展的基本规律，建立科学的评价制度更是学术生态建设的关键。应当进一步努力构建充满信任、宽容、友好、合作的哲学社会科学发展学术生态。通过定期的学术报告会、午餐会、课题研究成果交流与分享会等组织形式，充分发挥非官方学术社团在打造良性学术生态的应有作用。构建生机勃勃、富有活力的学术氛围。鼓励和支持各哲学社会科学学术单位进行学科文化或学术共同体的建设，引导哲学社会科学领域若干富有生机活力、学术成就卓越的良性学术生态群落的构建与完善。

（袁　清，浙江大学社会科学研究院副院长、副研究员，浙江杭州　310058；姚　威，浙江大学发展战略研究院副研究员，浙江杭州 310058）

（原文刊载于《科研管理》2015年第1期）

英国大学科研评价改革：从RAE到REF

刘　莉

英国科研评价(research assessment exercise，RAE)开始于1986年，是“撒切尔计划中公共开支问责制的一部分”。在过去的二十多年中，RAE分别于1986年、1989年、1992年、1996年、2001年、2008年开展了六次科研评价活动。每次科研评价都会在上一次的基础上对评价方法、程序和标准等进行一定的调整。2008年科研评价之后的改革幅度较大。

RAE是建立在同行评议基础上的。由于同行评议的主观性，越有名的机构越容易在科研评价中获得较高的等级，获得的资助也越多，由此产生马太效应，使强者更强，弱者更弱，不利于新兴交叉学科的发展，也使一些新成立的大学感到困难重重。“一些新成立的大学指出，科研评价导致科研经费过于集中于某些学校，致使他们得不到必要的科研经费的支持，科研活动受到制约”。鉴于此，“英国政府让高等教育基金会发展一个新的研究评价与资助框架，以更多地使用定量信息”。为了减少同行评议的副作用，降低评价过程中产生的成本，英国高等教育基金委员会在2008年科研评价后决定采取新的科研评价体系——研究卓越框架(research excellence framework，REF)。因此，从某种意义上说，从RAE到REF的变革是政府推动的结果。

一、卓越研究框架

REF是评价英国大学科研质量的新体系，将代替RAE，并于2014年完成。根据卓越研究框架，科研评价要定期开展，英国每所大学的院系都要依据国际标准评比，以评价研究质量。通过卓越研究框架，英国高等教育基金委员会旨在发展和维持一个充满活力和国际竞争力的研究部门，并使他们对经济繁荣、人民幸福、知识的扩张和传播做出重要贡献。REF的科研评价“是其他国家(澳大利亚、美国和欧洲其他国家)的范例”。事实上，REF已经引发了全球效应：澳大利亚、芬兰以及斯堪的纳维亚等国家纷纷效仿；法国和德国也通过大规模的改革正在走上与英国相同的轨道；著名的欧洲研究委员会也在积极追求定量评估。

2014年，英国将开始第七轮大学科研评价，并将使用REF来评价2008—2013年英国大学的科研。目前，REF的评价标准、方法等已经形成，并已公布在其官网上。在REF卓越框架中，研究被定义为形成新见解和有效分享的调查过程。研究包括与商业、工业需要直接相关的工作，与公共和自愿部门需要直接相关的工作；学术；思想、图像、行为、人工制品的产生与发明，包括导致新的或明显改善的见解的设计；试验发展中现有知识的使用产生的新的或实质性改进的材料、仪器、产品或过程，包括设计和建设。研究包括已出版、传播，或以可评价的形式公之于众的研究成果和机密报告。研究不包括常规的实验和对材料、成分过程的常规分析，比如维护国家标准，与新的分析技巧的发展不同；也不包括那些不能体现原创性研究的教学材料的发展。

（一）评价组织

REF是一个专家评价过程，以学科为基础的专家组评价高等教育机构在36个评价单元(UOAs)中提交的资料。REF设置4个主专家组。每个主专家组下设若干子专家组(详见表1)。4个主专家组将负责制定专家标准和工作方法，以确保遵守已发布的程序，整体评价标准的一致应用，并签署评价结果。专家组成员被4个英国资助机构通过公开提名的方式任命。子专家组将在4个主专家组的领导和指导下工作。每一个评价单元的子专家组将对提交的材料展开详细的评价。REF努力保证主专家组和子专家组的成员由具有从事、管理和评价高质量研究经验的个人及来自公立、私立、第三方机构的参与研究影响评价的高水平专家组成。

（二）资料提交

高等教育机构提交的材料主要包括5个部分：

(1)人员信息。人员分为A类和C类。A类人员

表1　4 个主专家组包含的子专家组

主专家组	子专家组
主专家组 A	临床医学；公共卫生、医疗服务和初级保健；联合健康专业、牙科、护理和药学；心理学、精神病学和神经科学；生物科学；农业、兽医与食品科学
主专家组 B	地球系统与环境科学；化学；物理学；数学；计算机科学和信息；航空、机械、化工、制造工程；电气与电子工程、冶金与材料；土木建设工程；通用工程
主专家组 C	建筑学、建筑环境与规划；地理、环境研究和考古学；经济学和计量经济学；商业及管理学；法律；政治和国际研究；社会工作与社会政策；社会学；人类学和发展研究；教育；体育和运动科学、休闲和旅游
主专家组 D	区域研究；现代语言和语言学；英语语言文学；历史；经典；哲学；神学和宗教研究；艺术与设计；历史、实践和理论；音乐、戏剧、舞蹈和表演艺术；通讯、文化和媒体研究、图书馆和信息管理

是指高等教育机构中的研究人员，他们只从事科研或者教学与科研并重。C类人员是指高等教育机构以外从其他组织聘用的从事研究的人员。

(2)研究产出。评价时段内(2008.1.1—2013.12.31)发表的出版物及其他可以评价研究的产出详情。每位研究人员最多提交4个成果。

(3) 影响模板和个案研究。描述评价时段内(2008.1.1—2013.12.31)被评单位使研究产生影响的方法的完整模板以及1993年1月1日至2013年12月31日这段时间的卓越研究产生影响的具体例子的个案研究。影响模板包括以下部分：背景；在评价时段内机构产生影响的方法；支持影响的战略和计划；机构的影响方法以及提交个案之间的关系。

(4)环境资料。研究型博士学位授予的资料及2008 年8 月1 日至2013 年7 月31 日这段时间内获得的相关研究经费。

(5)环境模板。描述2008年1月1日至2013年7月31日研究环境的完整模板。每份提交的材料必须包括一份单一的完整表格，该表格由以下部分组成：概况；研究战略；研究人员(包括人事战略和员工发展，研究学生)；收入，基础设施和设备；合作与对学科的贡献。

（三）评价标准和等级

从RAE到REF的转变突出表现在评价框架的变化上。从仅仅注重科研产出质量转变为以科研产出质量(65%)为主，同时重视影响(20%)和环境(15%)。这三大方面以质量为核心，从低到高分别给予5个等级：无分类、一星、二星、三星、四星。

所有子专家组使用通用的评价标准，并在共同的评价框架下运行。每个主专家组将要为子专家组研究一套通用的标准和工作方法。REF的通用评价标准包括产出、影响和环境。

产出：子专家组将从原创性、重要性、严谨性的角度，参照国际研究质量标准来评价研究产出。这个指标在总的评价体系中占65%的权重。在RAE中，产出质量建立在同行评议的基础上，而在REF中，是专家评价。

影响：子专家组将评价被评机构从事的卓越科学研究对经济、社会、文化影响的程度和重要性使研究产生影响到办法。这个指标在总的评价体系中占20%的权重。评价每个个案内描述的影响时，专家组要对作为一个整体的程度和重要性形成总体看法，而不是分别评价程度和重要性。评价影响模板时，专家组将考虑模板中描述机构的方法在多大程度上有利于实现程度和重要性的影响。

环境：子专家组将从活力和可持续性角度评价研究环境。包括对广泛的学科和研究基础的活力和可持续性的贡献。这个指标在总的评价体系中占15%的权重。专家组将考虑提交单位的活力和可持续性，以及对更广泛的研究基础的活力和可持续发展方面的贡献。

子专家组将为每个元素(产出、影响和环境)制定一个子概况，表明满足每一个等级质量标准的提交材料的比例。总的质量概况及各子概况中每个等级的定义详见表2。

（四）评价结果

子专家组在所有提交的证明材料基础上，运用专业的判断对每份提交的材料形成一个质量概况。三个子概况将被整合成一个总的质量概况。评价的基本结果是给每份提交的材料一个总的质量概况。这个总的质量概况结果将发给每个提交单位，它是REF的基本结果。以概况的形式提交结果，确保在评价结果范围被鉴别。

REF的主要结果是给每一份提交材料一个总体质量概况，将于2014年12月公布。质量概况将展示

表2　评价标准和等级定义

等级	总的质量概况	产出子概况	影响子概况	环境子概况
四星	在原创性、重要性、严谨性方面的质量都是世界领先的	在原创性、重要性、严谨性方面的质量都是世界领先的	在程度和重要性方面的影响杰出	从活力和可持续性方面，环境有利于产生世界领先质量的研究
三星	在原创性、重要性、严谨性方面的质量在国际上是卓越的，但是没有达到卓越的最高标准	在原创性、重要性、严谨性方面的质量在国际上是卓越的，但是没有达到卓越的最高标准	在程度和重要性方面的影响非常明显	从活力和可持续性方面，环境有利于产生国际卓越质量的研究
二星	在原创性、重要性、严谨性方面的质量在国际上得到认可	在原创性、重要性、严谨性方面的质量在国际上得到认可	在程度和重要性方面的影响相当大	从活力和可持续性方面，环境有利于产生国际认可的质量研究
一星	在原创性、重要性、严谨性方面的质量在国内得到认可	在原创性、重要性、严谨性方面的质量在国内得到认可	在程度和重要性方面产生了被认可的、适度的影响	从活力和可持续性方面，环境有利于产生国内认可的质量研究
没有分类	质量没有达到国内认可的工作标准，或者工作不符合为这次评价发布的研究定义	质量没有达到国内认可的工作标准，或者工作不符合为这次评价发布的研究定义	在程度和影响方面没有影响；或者影响不合格；或者影响没有得到提交单位卓越研究的支撑	从活力和可持续性方面，环境不利于产生国内认可的质量研究

被评价的活动满足四个星级质量水平的比例。除了质量概况，基金委员会将公布每个提交机构的全日制教师数，进一步的报告和反馈将在2015年初得到，评价结果还期望包括以下内容：

(1)每个主要专家组发布一份报告，在其职权范围内确认工作方法和提供观察研究状况的概况(长处、短处、活动的活力、已经产生影响的范围)。这些报告包括每个子专家组提交的部分。

(2)对每份提交材料给予简要反馈，总结质量概况的原因。这个反馈将被送到有关机构的领导那里。关于联合提交，反馈将悄悄地提供给所有相关机构的领导。

(3)每份提交材料的产出、影响和环境子概况将结合起来产生总的质量概况。总的质量概况将提交给机构领导，然后发布。

(4)子专家组和主专家组在REF评价阶段的会议记录将被发布，以提供给公众一份专家组如何开展工作的记录。

(5)REF管理人员的报告，将详细解释过程是以怎样的运行方式管理的。

(6)REF平等和多样化专家组的报告，详细解释了工作方法及在REF中实施平等措施的观察。

二、REF对RAE的继承

从RAE到REF，尽管评价框架发生了很大的变化，但是REF和RAE在很多方面还是一脉相承的，主要表现在以下五个方面。

（一）中介评价

英国政府不直接进行大学科研评价，而是委托中介机构——高等教育基金委员会对大学科研进行评价。这是从RAE到REF都没有改变的。REF是评价英国大学科研质量的新体系，是由英国四个高等教育资助机构：英格兰高等教育基金会(HEFCE)、苏格兰基金会(SFC)、威尔士高等教育基金会(HEFCW)以及北爱尔兰就业与学习部(DEL)共同开展的。

（二）评价目的相同

从RAE产生的背景、几次结果的使用来看，RAE的结果是其政府分配研究经费的重要依据。这一点，REF与其是相同的。REF指出：2014 年REF的基本目标是为每一个提交材料的机构提供评价结果；使用评价结果选择性地分配研究经费(2015—2016)；评价为投入研究中的公共资金提供问责，为公共投资产生的益处提供证据；评价结果提供基准信息，建立声誉尺度。

（三）重视产出质量

英国RAE不要求科研成果的数量，只要求科研人员最多提供4份有代表性的科研成果。在REF中，要求每位研究人员最多提交4个最有代表性的成果，并将从原创性、重要性和严谨性三个方面评价。

（四）以国际标准为标杆

RAE一直把国际标准作为重要的科研评价标准，这在2001和2008年的科研等级标准中表现得非常明显，“研究活动质量一半以上达到国际领先水平”(2001)为最高等级。REF则规定“在原创性、重要性、严谨性方面的质量都是世界领先”为最高等级。

（五）遵循公正、平等、透明的原则

RAE评价一直努力遵循公正、平等、透明的原则。现在仍能从其网站上看到1992年、1996 年、2001年、2008 年最近四次当年使用的评价标准、程序、方法以及评估结果等文件。REF的指南中列出以下原则管理REF的行为：公正，所有学科所有类型的研究和各种形式的研究成果应在公平和平等的基础上进行评价；平等，强烈鼓励高等教育机构提交他们所有优秀研究人员的工作；透明，评价过程和结果公布都较好地体现了透明性。

三、REF在科研评价上的新探索

英国大学科研评价在传承传统的基础上，进行了新的探索。主要表现在以下三个方面。

（一）引入对影响的评价

影响指标首次作为一个独特的元素被引进研究评价，符合政府所谓的“经济—影响力”议程。评价影响“将要求额外的工作，又要考虑这将会造成一定程度的负担，但同时会提供额外的收益”。“影响”将被子专家组仅用于评价高质量研究。影响必须在REF评价周期内非常明显，虽然研究可能是在10-15年前开展的。使用系的个案研究(departmantal case studies)以及系提供的影响报告书进行评价。REF“影响”的通用定义是：提交影响个案的研究和一个完整的影响模板、相关的资格准则、通用评价标准和等级定义。子专家组将依据这个框架来评价影响。

REF把影响定义为学术界以外对经济、社会、文化、公共政策或服务、健康、环境或生活质量产生的效果、变化或收益。影响包括效果、变化或收益：活动、态度、意识、行为、能力、机会、业绩、政策、实践、过程或理解；观众、受益人、社区、选区、组织或个人；在任何地理位置，无论在本地、区域、国家或国际。影响包括减少或预防伤害、风险、成本或其他的负面影响。对其他高等教育内部的影响，包括对学生和教学的影响。影响不包括对研究的影响及高等教育系统内学术知识的进步，也不包括对提交材料的高等教育机构内学生、教学以及其他活动产生的影响。

为了将影响作为REF的实质性元素，高等教育基金委员在地球系统与环境科学、物理、临床医学、英语语言文学、社会工作与社会政策等5个领域内开展试评价，以验证和改善评价影响的机制。试评价涉及29个大学，试评价的结论使用与试评价中相似的机制去评价研究有广泛的益处，评价方法已经被纳入框架。2010 年，REF小组开展了一个试评价以检验和发展REF评价影响的方法。2011年年3月，高等教育基金宣布在REF开展影响评价及赋予其权重的决定。

评价影响需要提交的申请包括一个个案研究和一个深入的个案研究，因此，需要至少提交两个案例。个案研究将使用有字数限制的通用模板提交。这样的设计使机构能通过叙述清晰地解释影响。叙述包括适合案例的指标和证据。模板包括以下信息：基础研究及其质量的证据、提交机构的研究对影响或收益做出贡献的方式、影响或收益的性质和程度、引用能被核查的独立来源。

（二）文献计量指标的使用

英国这次科研评价改革中，“最大的变化就是采用了文献计量指标，并弱化同行评议对评价结果的影响”。REF更多地建立在计量指标上，同行评议只占很小的比重。英国大学科研评价体系从基于同行评议的RAE变革为基于文献计量的REF，《高等教育纪事报》(Times Higher Education)评论说这是“进化，而不是革命”。它类似于一个同行评议为基础的RAE的演变，而不是以计量为基础的革命。英国高等教育委员会(HEFCE)建议通过使用研究的引用来明显减轻同行评价的负担，同时保留强大的对质量的判断是几乎不兼容的目标。

探索使用计量指标，如在2008—2009年的试评价中使用过以一篇文章的被引次数作为质量评价的方法。试评价的结果表明：在REF中仅用计量指标不足以代替专家同行评议。但在一定范围内，至少在一些学科，引用信息可以用来作为专家评价的补充。

（三）重视对跨学科研究或合作研究的评价

RAE时期已经注意到了对跨学科研究的评价问题。2008年RAE对程序进行了大量的完善以评价跨学科研究。REF在此基础上进行了发展：(1) REF鼓励一个评价单元内的两个或两个以上机构联合提交材料，这是描述已经开展的合作研究最合适的方式。(2) 如果研究人员共同撰写或联合研制的成果被列在一个以上提交材料里的话，将与其他成果一样在平等的基础上被评价。(3) 在环境模板内，机构应该提供给他们如何支持跨学科和合作研究的信息。专家将给出合适的分数，以表明这些安排已经

提高了研究环境的有效性与可持续性，或者提交单位提供了更广泛的研究基础。(4)凡提交的影响是以合作研究为基础的，每个提交单位都可以提交影响，因为他们的研究对影响做出了独特而重大的贡献。

四、REF存在的问题

REF虽然在RAE的基础上进行了很多新的探索和尝试，也努力避免RAE出现的一些问题，但仍然不可能是完美无缺的。尽管REF尚未实施，但是已经引发了一些讨论和质疑。

（一）大学面临的风险

伦敦大学教育研究所教授Peter Scott指出，对于评价研究表现的新制度如何运行知之甚少，这意味着大学将面临很高的风险。新制度能够改变大学思考研究的方式。REF可能重新激发关于大学的研究。结果很难预测，这增加了一个惊喜(和运动)。规则已经改变，就意味着玩游戏有危险。结果可能产生一个更开放的、有更多卓越维度的研究系统。因此，在未来10年，REF将有更多的风险。这些风险要远大于世界一流大学的命运，或者顶尖教授的自我，甚至成百上千大学院系的命运。此外，在危急关头，可能产生一种全新思考和做研究的方式。Peter Scott指出，REF将会让一些系消失。顶尖或一流大学将首先从研究角度进行定义，然后才是吸引最优秀学生的能力。最优秀的学生被吸引到顶尖的研究型大学里。对于系来说，不仅学术自尊，还有学术可持续性，甚至生存都会有危险。科研评价制度(RAE/REF)被认为对小型的院系、新兴学科、应用学科和跨学科研究不利，即使在学科内部，也有特定的学科领域和研究方法更容易得到支持。

（二）忽视教学

英国曼彻思特大学社会人类学教授Peter Wade指出，REF对影响的狭隘定义忽略了与大学的社会影响有关的教学的历史作用。REF影响的定义对日常教学事务的忽视是很明显的，然而教学是几百年来很多郡都非常关心的，也被理解为与大学广泛的社会影响相关。这种担忧现在好像被抛给了市场，像是一个被管制的、被REF狭隘影响定义驱动的介于教学和研究之间的楔子。

（三）耗时耗力

RAE实施过程中另一个非常突出的缺陷是将耗费大量的时间。RAE使用专家组评价，使整个英国的研究界浪费了大量的时间和精力。RAE每3-5年开展一次，高校在一次评价后就得开始准备下一轮的评价，耗费了大量的时间和精力。“2008年的科研评价中，学校提交申请材料的截止日期是2007年11月30日，而评价结果整整拖了一年多，直到2008年12月18日才最终揭晓。学校和教师花费了大量的时间和精力整理材料、填写申请”。REF努力克服RAE的不足，REF与2008年RAE的不同也包括一些细节性的资料要求和界定，如评价单元子专家组从67个减少到36个，主要专家组从15个减少到4个。但是科研卓越框架仍然耗时耗力。有人甚至批评说，学校在提交科研成果时，还得专门请人写“影响”。

（四）学科差异的问题仍需探索

目前的REF由一个单一的框架组成，主要在同行评议的基础上评价所有学科的研究。在REF建立过程中，曾计划建立分类评价的机制，将所有学科分为科学和其他科学两大类。在科学类学科中，将比RAE更大程度地使用计量指标。在咨询已经建立的部分发展工作后，建立一个评价所有学科的系统，将比双层系统更适合，当然在不同的专家组内部会有一些适当的变化。但在现行REF的分类评价中仍然没有实现。

（五）鼓励短期的、容易量化的科研产出

REF采用计量指标后，由于计量指标的局限，也引起了一些担忧。兰卡斯特大学教育研究系教授马尔科姆·泰特指出，REF鼓励短期的和容易产生可量化成果的研究，而不鼓励长期的尤其是成果和影响难以提前度量的研究。科研评价对科研工作的导向作用不可忽视，这种担忧不无道理。

综上所述，英国大学科研评价的卓越框架尽管没有完全解决RAE的问题，但还是做了一些新的探索和尝试，尽管也有一些不尽如人意的地方，但是却在科研评价方面有了一些新的突破，是在探索中前进的。科研评价本来就是一个比较复杂的事情，无论英国大学科研评价将何去何从，从RAE到REF毕竟是前进了一大步。

（刘　莉，上海交通大学高等教育研究院副研究员，上海　200240）

（原文刊载于《科学学与科学技术管理》2014年第2期）

澳大利亚ERA分类科研评价体系及其启示

刘兰剑　　杜向民

一、引言

澳大利亚的科研质量评价活动是整个世界科研评价运动的一部分，这场发生在澳大利亚的评价研究质量运动很快波及全球，许多国家都强调卓越科研及质量科研，实行基于绩效的资助方案。这场改革在整个世界掀起了目的在于卓越科研、质量科研的运动，此后，澳大利亚的一些做法被其他国家所借鉴。

20世纪80年代后期，澳大利亚科研成果的质量出现了明显的问题。以引用率为指标的澳大利亚出版物质量下降引起了广泛的关注。如图1和图2所示，除了农业科学以外，几乎所有的学科在80年代都有明显的衰落。发表的科研成果数量在明显增加，而以RCI衡量的质量却在明显下降。直到90年代，一些领域的绩效才有所变化和提升。

除此之外，在1988—1993年间，澳大利亚的科研影响因子在OECD的排名中从第6位降到了第11位。在这段时期，澳大利亚的研究者们发表了更多的论文，但这些论文的平均引用率却下降了。在这种情况下，探讨澳大利亚大学以简单计数为基础的科研评价方法是否适当的问题迫在眉睫。当时的科研资金分配是基于简单计数的，例如出版物的数

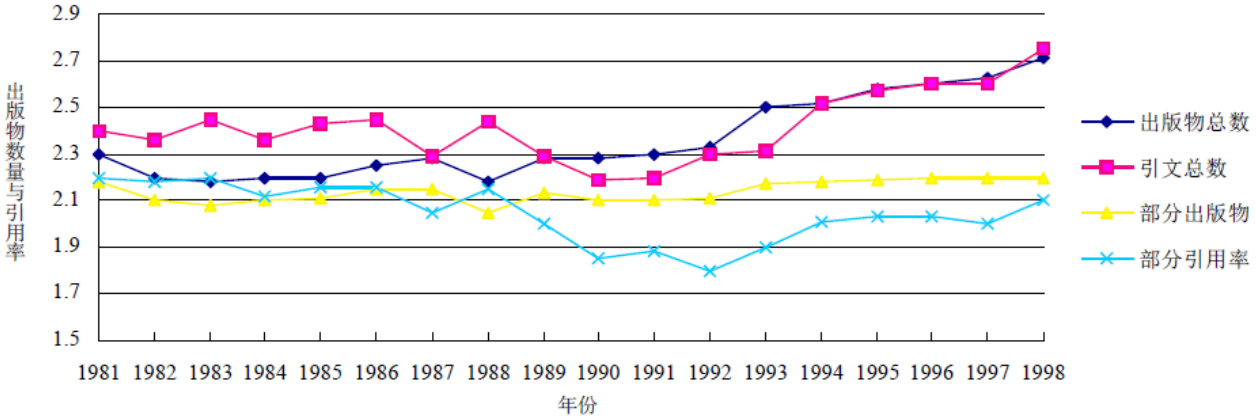

图1 澳大利亚1981—1998 年在SCI数据库中的出版物数量和引用情况

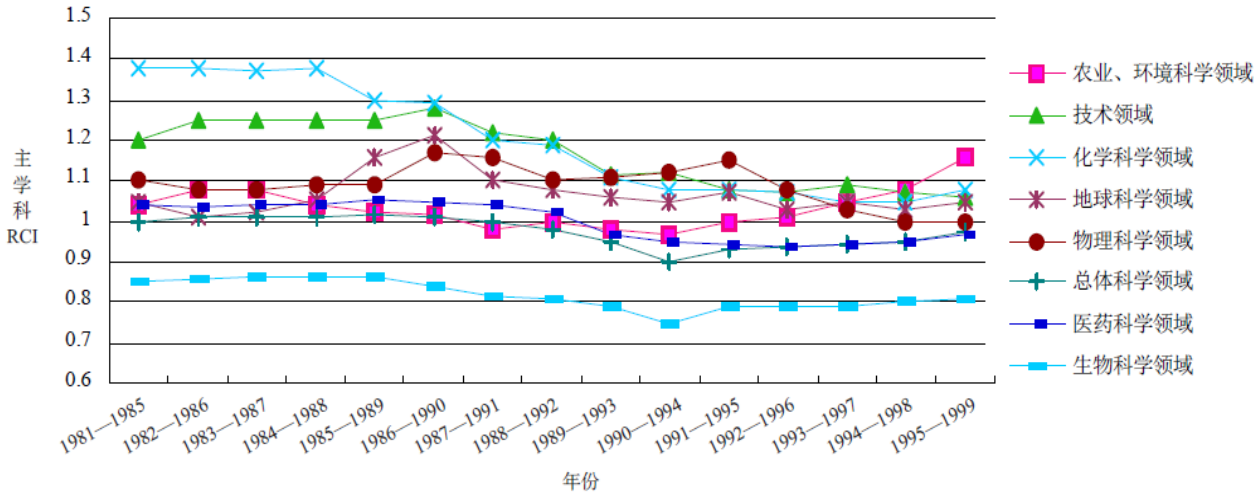

图2 澳大利亚1981—1999年主学科的RCI

量、外部科研资金的多少、高一级学位的学生数量的多少以及完成学业的学生数量等。这些简单的数量指标被用来衡量科研成果的质量，鉴别优秀的成果存在明显的缺陷，不利于鼓励广泛的社会资金投入到澳大利亚的科学研究当中。

因此，一场从科研质量框架（research quality framework，RQF）到卓越科研（excellence in research for Australia，ERA）的科研评价方法改革随后逐步展开，这些评价体系深刻地影响了科研的绩效，促进了澳大利亚科研成果质量的提升。

在此背景下建立的科研评价体系有三个特点：一是官方评价，即评价由澳大利亚官方组织实施；二是人文社科评价和其他学科采取不同的评价方法；三是建立了以质量为核心的科研评价体系。

二、澳大利亚科研评价组织及体系的演变

（一）科研评价体系的演变

在澳大利亚科研成果质量明显下降的情况下，学术界在广泛关注和讨论的基础上，当时的Hon John Howard首相在2004年宣布要建立一套公共资金资助的科研质量制度，也就是后来的科研质量框架。2005年5月，政府发表了一篇文章讨论这个问题，随后在9月份发布了政府评价模型的细节。学者Gareth Roberts 在2006年3月给教育和培训局局长写了一封建议书，接下来该局宣布成立了以局长为主席的RQF发展咨询会。发展咨询会在2006年10月向首相提交了研究报告，报告分析了RQF的一些基本问题，给政府提供了有关RQF的建议。特别强调RQF成立的目的是保证公共资金投向高质量的研究领域，并对优秀的研究予以奖励。

澳大利亚政府寻求确保公共资金投入到高质量的、真正产生效益的研究中。RQF正是源于此目的，通过RQF以提供一个全面评价科研成果及其影响的评价机制。RQF可以识别高质量、高影响力的研究，而且可以使得公共资金的使用更加透明。这反过来鼓励了更多的企业资金投入到研发上来，这些资金寻求科研的方向及其可能的应用。

在2007年第二季度，澳政府安排了一次RQF科研评价试运转，其目的是想考查该制度在一些关键和难点问题上的适应性、可行性与有效性。RQF科研评价的第一次执行时间确定为2008年。

事实上，RQF并未实施。随着2007年政府的换届，新一届政府在2008年3月宣布进一步改进科研质量。用卓越科研取代了RQF，并在2008年年底和2009年年初进行了试用。与RQF所有学科采用统一的评价指标体系不同，ERA每个学科的指标体系都不同，ERA更强调优质的科研成果，支持世界一流的科研成果。工业、创新、科技部部长在2009年7月宣布了ERA的更多细节。ERA在2010年和2012年全面展开。这两年的评估过程中，对一些细节问题进行了调整。这些调整确保了ERA2010年和2012年的可比性。

在此期间，激励创意（powering ideas）计划的颁布，进一步支持了对研究质量的测度，该项目列出了政府未来10年的创新计划，也勾勒了大学绩效的重要性。计划认为，政府如此依赖大学，如果大学的绩效滑坡，整个创新体系将遭受重创。因此，保证大学体系正常运转非常重要。不仅仅要以本国自己设计的标准，而且要以世界最高标准来进行测度，世界经济的竞争力来源于有竞争性的创新体系——创新体系的发动机是具有世界竞争力的大学。

激励创意计划也显示了政府支持增加世界一流水平研究团队数量的雄心。ERA列出了国际标准科研成果的评价机制，以推动联邦政府的资金与研究质量挂钩。

ERA是评估澳大利亚大学研究质量的一套体系，其目的是向纳税人确保其政府的科研经费投资的应用是明智的。ERA的研究报告直接引导大学支持可持续的优秀科研计划，而且ERA向工商业界等领域提供大学的科研强项等信息，以便于这些强项可以为整个国家所开发利用。ERA提供了高等教育部门科研质量的可靠信息，因此可以使管理者和投资人识别并奖励优秀的科研成果，并调整未来的科研投资机会和方向，以保证澳大利亚纳税人的投资被恰当地使用，也可以在战略规划中进一步强化科研能力，帮助澳大利亚在世界水平上提升科研优势。而且，ERA的数据是指导澳大利亚科研投资战略的理想工具，包括根据产业、区域和国家优先顺序去调整研究战略，以使公共资金在科研领域的投资效益最大化。ERA的研究报告表明，政府基于绩

效的大量投资支持了卓越的研究，这些资金鼓励大学开展直接的投资激励，以支持世界一流水平的科研项目。

第一轮ERA评估是在2010年进行的，其研究结果发表在2011年年初。这是澳大利亚第一次有组织地对学科优势和发展领域进行的一次全国性盘点。第二轮ERA是在2012年12月6日完成的，同时也出版了ERA的全国性报告。

（二）科研评价组织

澳大利亚科研局（Australian Research Council，ARC）是负责研究澳大利亚卓越科研项目的组织，其目标是在整个澳大利亚的高等教育机构中识别和全方位推进卓越科研活动。其下设科研评价委员会（Research Evaluation Committee，REC）、同行评议专家组和复审委员会。REC的职责是分配待评成果；保证在整个质量评审过程中使用统一的标准和程序；为ARC提出的关于评价的任何过程和方面的问题提供反馈和建议；向ARC汇报结果。同行评议专家组是对人文社科类等相关专业的科研成果进行评价，并将评价结果提交REC。复审委员会的职责是对评价过程、结果等任何环节可能存在的问题进行监督。组织结构如图3所示。

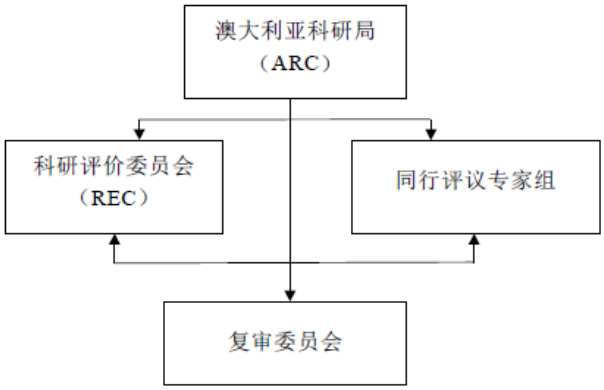

图3　澳大利亚ERA组织结构

REC成员的职责包括10个方面：（1）完全参与各自委员会内部的评价过程；（2）遵守机密规定和利益冲突的要求；（3）保守REC的审议意见和决定的秘密；（4）提前确定是否存在利益冲突或其他敏感问题，并在冲突发生之前提交给ARC；（5）在会前准备好充分的材料，避免不必要的附加行政成本以及给其他委员会成员带来的不便；（6）认真完成REC主席分配的任务；（7）必要的时候把任务分给外部同行评审；（8）评价每一份待评材料并给出初步排序；（9）对REC的所有环节全力以赴，运用自己的专长，建设性地、不徇私情地工作，参加集体决策；（10）运用技能和责任心履行自己的职责。

REC主席的职责包括13个方面：（1）确保REC在ARC建立的政策、指导方针和程序内运作；（2）遵守保密和利益冲突的规定；（3）确保和维护REC的审议意见和决定的机密性；（4） 辨识可能存在的利益冲突和其他敏感问题，并在冲突发生之前提前处理；（5）完全奉献地、建设性地、不徇私情地参与REC所有的过程，并且主持REC的集体决策工作；（6）为REC成员分配评价材料；（7）评价分给自己的材料并进行初步排序；（8）确保评审在规定时间内完成；（9）主持REC 讨论初评排序的会议，引导REC完成每一个评价单元的最终排序；（10）确保REC成员有机会在整个评价活动过程中贡献自己的力量；（11）确保REC的决策有文件备案；（12）向ARC汇报结果；（13）参与REC的终审并向ARC报告REC评价的过程。

在REC主席无法执行全部或者部分职能的时候，ARC会在REC内部委派一个执行主席，来代替主席执行全部或者部分职能，这种情况经常发生。例如，当主席发现有价值冲突时，ARC就会任命执行主席来分派待审的材料。

同行评议专家的职责包括5个方面：（1）评价分配的材料并按照规定的模板提交同行评议报告；（2）认真完成分配给自己的任务；（3）细心地运用自己的技能完成职责；（4）识别可能的利益冲突及其他敏感问题并提前报告ARC处理；（5）尊重保密性规定。

为了更好地实现ERA的评审目的，ARC会邀请REC成员对整个ERA的评审过程进行评论并提供反馈。REC会议的一个成果就是改进未来ERA 的评审。ARC也会在评审最终阶段召集REC主席会议，在这个会议上评价探讨整个评估过程。

ERA复审委员会是ERA2012年组建的，其职能是仔细检查每一个REC成员评审的每一个对象，以及整个ERA的评审过程。按照评价标准，审查每一个评价单元的评价结果。这一设计和加拿大科研评

价体系中的观察员制度非常类似。

三、人文社科与其他学科：两类不同的评价方法

澳大利亚的科研评价单位（Units of Evaluation，UoEs）采用分级编码的形式。其中，2位数编码是最高层的编码，一个2位数编码涉及广泛的学科领域，包含一批4位数编码。例如02—物理科学，包括0201—天文空间科学，0203—古典物理学以及物理科学领域的其他4位数编码。4位数编码是2位数编码下的具体学科，4位数编码下又包含一批相关的6位数编码。被评机构按照4位数编码的水平提交待评材料。学科编码矩阵中列示了学科类别，其中的人文社科类别略显宽泛，但仍然是以人文社科类为主体。具体包括数学、信息与计算机科学、环境设计、教育学、经济学、商业、管理、旅游、人类社会研究、心理认知科学、法律和法学研究、艺术创作和写作、语言文化、历史和人类学、哲学和宗教等学科。

澳大利亚的人文社科和其他学科的评价方式是不同的。其他学科的成果由REC的科研评审专家根据评审指标进行科学计量评价并排序。而人文社科门类实行的是代表作制度。这些代表作除了接受REC评审专家的科学计量评价之外，还要进行同行评审。

人文和艺术创作两个专业群需要进行同行评议，2009年第一次试评审时规定，当这2个学科群满足4位数编码各超过20份作品的最低标准时，就可以参与评价。要求被评机构选出20%的作品提交同行评议，这20%的作品就决定了所有作品的质量。如果达不到20件作品的最低要求，也可以按照20%的比例参加同行评议，评议的结果不能用来评价4位数编码学科，但可以用来综合评价2位数编码的学科。

ARC会为每个学科群指派REC科研评审专家，REC须是国际认可的科研评审专家，但不一定是同行评审专家。REC成员在评价的每一个阶段都承担着专家评审的职能，这些专家评审的依据是在学科矩阵中所列的各学科的指标体系。

而同行评议专家的评价依据是所有成果的代表作。应用同行评审的学科，被评价单位需要按照规定的比例初选出本单位的代表作。代表作占所有作品的30%，这个比例在学科矩阵中有反映。REC成员也要评价代表作，以完成这些学科的评价。

四、科研成果评价标准、等级与过程分析

（一）评价标准与指标分析

对于ERA而言，研究被定义为创造新知识和（或）以新的、创造性的方法运用现有知识以产生的新概念、新方法和新理解。这些必须综合、分析并扩展以前的研究，以达到新的和创造性的程度。这个概念包括了广义的研究与发展（R&D），由这些创造性的知识构成知识储备的基础，包括人文、文化与社会以及运用这些知识做出的新发现。各个研究机构必须以这个概念为基础审核和提交科研成果。

在评价指标开发方面，ARC成立了专门组织。2008年，ARC组建了一个指标开发小组（Indicator Development Group，IDG），小组的专家成员由科学测量专家和统计学专家组成。小组的任务是开发具体到每个学科的指标体系，包括质量的测度，应用研究的测度以及研究活动的测度。为了测试每一个学科提出的评价指标，ARC召集不同学科的专家会议，并且发布指标咨询文件以形成广泛的探讨。指标开发过程相当审慎，通过分析测试验证指标的有效性。若一个指标没有被清楚地证明是有效而且稳健的，那这个指标就不能在ERA中使用。按照ERA的指标开发原则，每项指标都需要进行详细地分析。

ERA按照国家和国际标准评估澳大利亚大学的研究质量。该评级由来自澳大利亚和海外的杰出研究人员确定，这些人员从澳大利亚及海外选拔确定。评价的单位被定义为依照澳大利亚和新西兰标准分类（ANZSRC）确定的研究领域。

ERA所使用的指标包括自然科学领域内常见的引用率等科学计量指标，在人文社会科学领域内实行的是代表作同行评审制度。ERA全面地收集资料。各个大学提交的数据包括所有符合资格的研究人员以及他们的研究成果。精确的指标体系由科研共同体密切协商确定。这种方法确保使用的指标都是适当且必要的，这就最大限度地减少了政府和大学提供资源的负担，并确保ERA的结果既稳固，又

可以被广泛接受。

ERA的指标原则指导开发不同类型的指标，指标要符合国际最好的实践标准，并且要通过ERA指标开发组的数据分析及检验。为了识别和开发每一个学科的相关指标，ARC创造了8个原则性指标来指导指标开发者。即数量标准、国际认可、与其他学科所用的指标具有可比性、可以识别出优秀的成果、与研究相关、可重复并且可检验、具有时限性、对行为有影响等8个标准。

运用这些原则性指标，收集数据形成8个学科群的4大类指标，这4大类包括：研究质量、研究数量以及活动、研究成果的应用以及公认性等。

在成果分类方面，主要可以划分成期刊论文、会议论文、专著、专著章节、非传统研究成果等。2012年，澳大利亚所有41个符合条件的高等教育机构提交ERA进行评估的研究成果超过413000个，横跨157个4位数编码的研究领域。最多的产出是期刊论文（69%），其次是会议论文（18%），非传统研究成果（NTRO）大约占2%如图4所示）。

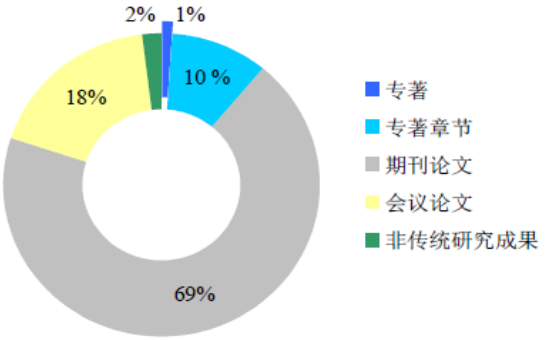

图4　研究成果类型百分比

（二）评价过程与等级分析

ERA2012由8个研究评价委员会（Research Evaluation Committees，RECs）来进行数据评价，RECs广泛代表了8个学科群，由147位来自海内外的杰出专家组成，这些专家一部分是本学科的专家，还有一些是科研评价专家。如前所述，ERA的同行评议是评议代表作，并不评议所有的作品，科研成果和专家都依据ERA的学科矩阵中所列的学科来提交和选择。需要进行同行评议的学科，被评机构在每一个学科上按照总成果30%的指定比例提交同行评议。比例和成果的类型无关。例如，专著和论文的权重都是1；对一些研究成果数量比较少的研究领域，同行评议的比例可能远远高于30%。表1是某机构提交的各学科领域的同行评审的成果数量计算表。

表1　每种类型的同行评议代表作——以某大学的人口统计学为例

成果类型	专著	期刊论文	会议论文	非传统研究成果	总的研究成果数量
成果数量（分配的成果数量）	10	26.66	31	0.7	68.36
同行评议的代表作（具体的成果数量）	3	8	10	1	22
同行评议的代表作（需要评审的百分比）	30%	30%（计算到下一个整数）	30%（计算到下一个整数）	30%（计算到下一个整数）	大于30%

注：在计算每种类型的成果代表作数量时,所有的小数必须进到下一个整数,例如,31篇会议论文的30%是9.3,上进到10

ERA的评价过程分为提交、分配、评价和报告4个阶段（见图5）。每一个阶段包含了数个子阶段或者活动。

在提交阶段，大学可以进入到ERA IT系统当中上传相关数据，这些信息随后被验证和确认以便与ERA的要求一致。这些数据接下来将被按照研究领域以4 位数以及2 位数进行编码，形成评价单元，这些编码中包括国内外相关的评价指标。提交阶段又可划分为4个子阶段（见图6）。

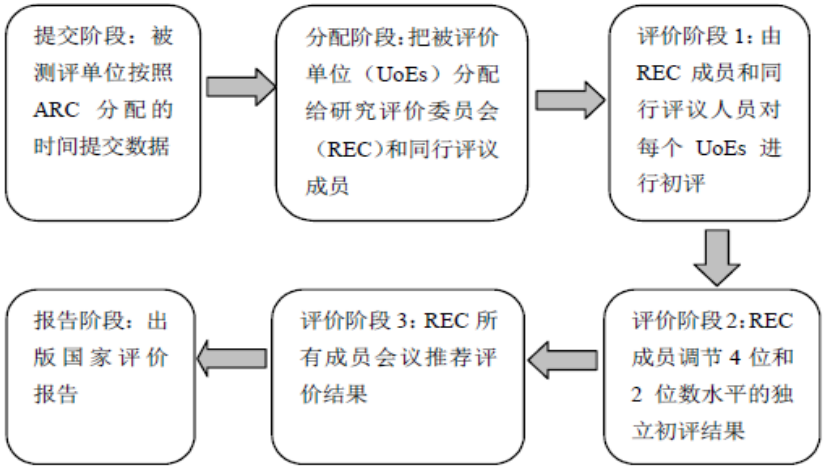

图5　ERA的评价过程图示

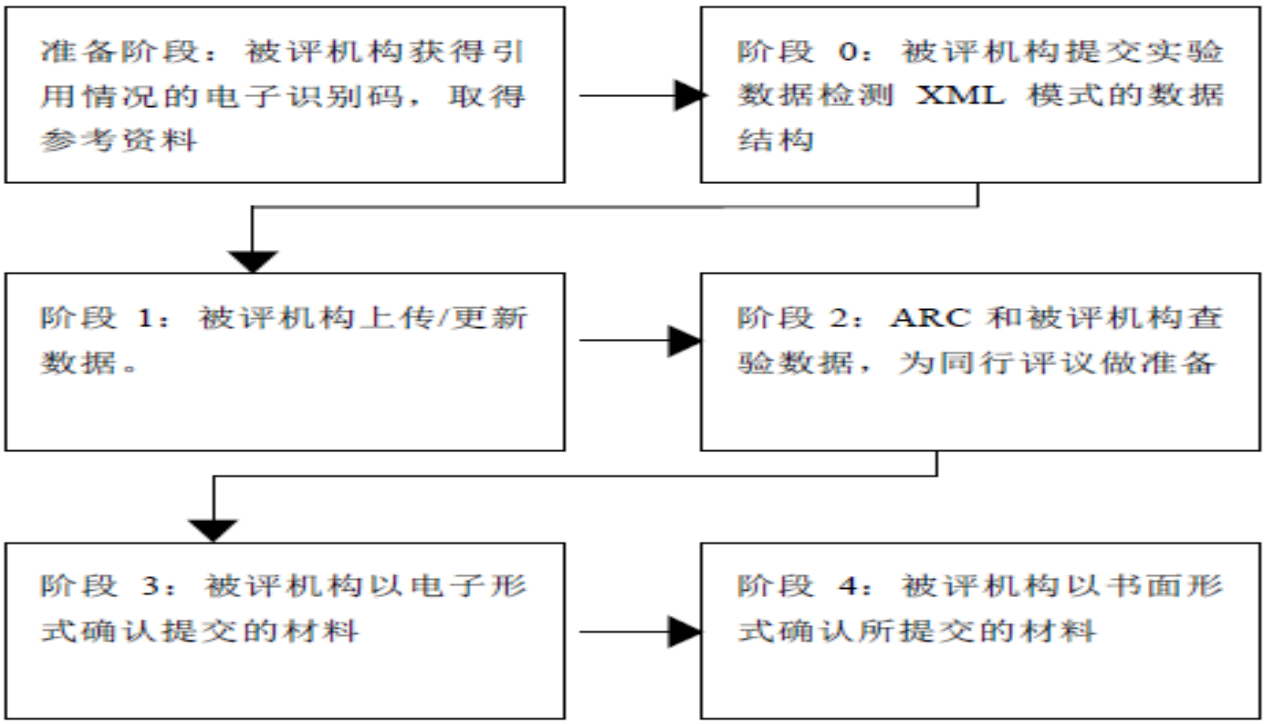

图6　资料提交各个子阶段

在分配阶段，首先是由REC主席为成员分配评价单元。如果存在利益冲突，则由代理主席实施这个分配环节。所有的4位数评价单元被分配到3位REC成员手中，接下来再按照评委的专业相关性分配到2位数评价单元。每一个评价单元有一个主评论人，该主评论人主持在REC会议上的讨论。通常待评材料分配给一个单独的委员会，也有的待评材料会从其他委员会的专家中抽调人员进行跨委员会的评审。

在同行评审中所用指标由外部的同行评审专家建构。首席评审专家把任务分配给其他专家，分配的时候要注意是否存在利益冲突、研究兴趣和工作量等因素。

评审是在网上进行，目的是把每一项评审任务的相关数据、指标和同行评审结果都可以以统一的规范输入到数据库当中。要求每个REC成员在REC会评之前把每项被评成果的初评情况输入系统。同行评议遵循类似的程序。

初次评价由REC成员独立进行。第一步由各个REC成员独立开展评价，各自评完后3位评审人交换意见，进行调整。评价过程分成几个部分，调整是整个评价过程中的重要一环。调整保证了每个被评对象在不同学科的专家之间讨论而形成结论。在最终形成网上评议结论的时候，REC综合考虑所有的初评结果，同意最终的评审意见。

在报告阶段，ARC编写国家报告，报告包括提交的总体数据以及国家层面的评价单元的总体情况，也包括每个评价单元的最终排序。每个REC承担着具体的评价过程，包括调节冲突，采用5点计分法形成的研究质量评价表（见表2）。

表2　ERA评价等级量表

等级	描述
5	成果水平远远超出世界标准
4	成果水平超过了世界标准
3	成果水平达到了世界标准
2	成果水平低于世界标准
1	成果水平远低于世界标准
n/a	由于数量过少未进行评价。研究成果的数量没有达到ERA规定的阈值

（三）回避原则及情形分析

澳大利亚保证科研评价中客观性的一个重要方面是对利益冲突的控制。利益冲突是指REC成员或者同行评议人员在履行ERA评价过程中可能存在的利益冲突，包括：第一，要评审的材料来自于自己以前工作的单位，或者有关联的单位；第二，与被评审人有比较密切的关系，包括合作伙伴、配偶、家庭成员或者亲密的朋友，还包括曾经有冲突的人；第三，被评审人来自亲密的合作伙伴，例如，评审人和被评审人有占出版总数10%以上的合作出版作品，这将

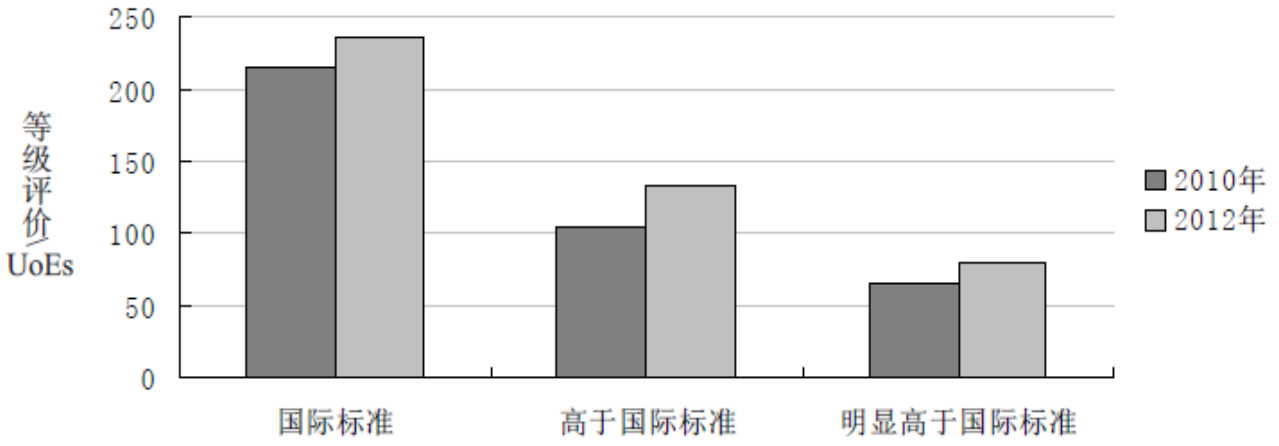

图7　2010 与2012 年澳大利亚各评价单位与世界水平的对比

构成潜在的利益冲突；第四，REC成员必须申报的其他利益冲突包括经济利益，例如，参与某企业的管理，持有某企业的股票、股票期权、专利、版税，担任企业的咨询工作等。

大多数的利益冲突必须在评审任务分配之前申报，ERA、REC成员、同行评议专家在评审过程中一旦发现利益冲突，要立即申报。在这种情况下，ARC要根据每种不同的情况处理，通常是把冲突材料分配给其他评审人员。REC成员或同行评议专家绝对不能评审他们本单位以及申报了利益冲突情况的材料。

其他敏感情况是指任何一个研究成果，如果导致REC成员或者同行评审专家产生敏感性，他们应立即向ARC报告，他们所评审的材料应该被重新分配。

五、评价体系的效果分析

从RQF到ERA的科研评价体系改革，使澳大利亚的科学研究质量有了明显的提升。从2010—2012年2年间，以2位数的评价单位（UoEs）来看，高于世界水平的科研成果增加了27%，和世界水平持平的数量增加了10%，明显高于世界水平的数量增加了22%（见图7）。可见，全国统一的科研评价体系的确推动了科研质量的提升。

ARC通过ERA评价澳大利亚大学的科研质量。澳大利亚总共有41家机构具备ERA的评价资格，ERA的目标是从这些机构提交的所有成果中识别并促进卓越的科研成果。澳大利亚科研局在2013年委托ACIL Allen 咨询公司对实施ERA体系的益处进行了评估。这项工作对ERA的货币及非货币收益进行了探索分析。

研究与发展（R&D）构成了澳大利亚经济的重要组成部分，澳大利亚在R&D方面的总支出（GERD）在2008—2009 年间达到277亿美元，商业支出占了GERD的绝大部分，并且集中在实验与发展方面。然而，高等教育部门的关注主要是纯粹的基础研究，86%的资金都投向了这类研究。2011年，高等教育部门共收到32.5亿美元的研发费用，除此之外，还收到研发及研发训练一揽子拨款，2013年的一揽子拨款总额达到16.7亿美元。

鉴于研究对澳大利亚经济产生的显著影响，对研究质量的测量显得极为重要。RQF就是如何运用最好的方法评估研究质量工作的起步。研究发现，ERA的确帮助增加了研究回报率，节约了成本，增加了大学的收入，促进了经济活动和责任心以及透明度，短期的实践表明，ERA在未来会增加更大的收益。这些收益通过ERA的影响表现在如下3个关键领域：第一是提高了研究绩效。表现在提高了研究质量、专著与科研成就，增加了合作，促进了资源分配，为人力资源决策提供了信息等方面。最终，在科研绩效方面的改进使得科研投资回报率增加，提高了科研的社会投资回报率。小规模的社会投资回报率会产生显著的财政收益。例如，社会投资回报率每增加一个点数（例如从5%-6%），则每年会产生477亿美元的GDP增长量。第二是推进了大学的计划、战略和运作。表现在推进了合作和管理，促进了战略规划以及提高了认知和提升了能力等方面。正是由于这些好处，提高合作、管理和战略规划之后，会带来相应的成本节约，而且可以增

加高校的收入。除此之外，还可以吸引更多的国际学生到澳大利亚学习。第三是增加了责任心与透明度。表现在增加了与澳大利亚大学有关的责任心、透明度和监控程度，而且明确了多种多样的计划、战略、评价和政策。

六、结论与启示

澳大利亚的科研评价体系经历了一场重大的改革，这场变革整体上改变了澳大利亚科研成果数量增加而质量下降的问题，不但提高了科研成果质量，而且促进了产业界和高校的合作，产生了明显的效果。综合来看，澳大利亚的科研评价体系具有以下特点：第一，这场科研评价改革是由政府发起、推动并实施的一次重大改革。第二，对人文社会科学和其他学科采用了不同的分类评价方法，人文社会科学采用了科学计量和同行评议相结合的方式进行评价，而其他学科主要采用了科学计量等方法进行评价。第三，摒弃了以简单数量评价为特征的评价方法，建立了一套以质量为核心的科研评价体系，明显促进了科研质量的提高。

澳大利亚科研评价改革运动对中国科研评价的启示如下：

第一，以简单计数评价为特征的科研评价方法必然会导致低劣的科研成果质量。当前中国大多数单位的科研评价方法都是采用简单计数的办法，即发表了多少篇什么样的论文，有什么样级别的课题，有多少科研经费等，这种简单计数也是一种“定量评价”，但和真正采用科学计量方法进行的定量科研评价存在着巨大的差别。当前中国的科研成果也存在着数量多而质量不高的问题，许多人把原因归结于定量评价，甚至一些人要求取消定量评价，这里存在着一个误区，就是混淆了简单计数评价与科学计量方法，需要摒弃简单计数办法，而真正的定量的科学计量方法需要进一步强化和普及。这种简单计数方法在澳大利亚导致了低劣的科研成果质量，在中国也不例外，因此，简单计数评价必然会导致大量质量不高的科研成果出现，而真正的科学计量方法在科研成果质量甄别上是非常有效的，合理地运用科学计量方法进行科研评价，可以在一定程度上提升科研成果的质量。

第二，建立具有权威性的科研成果评价体系。当前，以科研评价为中心的中国大学评价由民间机构进行。这些民间机构的评价由于缺乏质量控制机制，其评价结果往往受到质疑，公信力不高。而且，由于其评价的目的并不是为了提升科研成果的质量，因此评价结果对科研质量的提升作用推动不大。因此，需要建立目标在于鼓励卓越科研评价体系，该体系应该由政府主导实施，借鉴澳大利亚、英国、加拿大等国的做法，采用一套行之有效的办法控制评价的客观性，树立公信力，推动中国科研水平的提高。

第三，人文社科和其他学科的评价方法应有适当的区别。人文社会科学科研成果由于具有影响力显现的长期性、对思想影响的潜在性等特点，统一使用科学计量等方法进行评价存在着一定缺陷，应当采用科学计量与同行评议并重的办法，对人文社科科研成果进行评价，而对其他学科，则应采用以科学计量为主的评价方法进行评价。

第四，以质量为核心的科研评价体系可以提高中国科研成果的水平。当前中国科研评价体系的显著特征是简单计数评价，因此导致了和当年澳大利亚非常相似的情况出现，即科研成果的数量激增，但影响力却微不足道。因此，需要尽快建立以成果质量评价为核心的科研评价体系，尽快扭转中国的科研风气，使中国从科研大国转向科研强国，从根本上改变中国创新能力不足的问题。

第五，按照世界标准进行科研评价。如果要达到世界一流的科研水平，则需要以世界标准对科研成果进行评判。中国科研评价活动较少有国外专家参与，各民间机构的科研评价标准也大相径庭，结果差异很大，因此，需要组织专家，研究设计世界水平的科研评价标准与指标，广泛吸纳国际专家参与中国的科研评价，推动中国科研水平向世界水平迈进。

（刘兰剑，长安大学政治与行政学院教授，陕西西安 710064；杜向民，长安大学中国人文社会科学研究评价中心教师，陕西西安 710064）

（原文刊载于《科学学与科学技术管理》2015年第12期）

日本大学教师学术评价的理念、内容及方法

韩小娇　高　军

一、日本大学教师学术评价的发展及理念

（一）日本大学教师学术评价迅速发展的原因

近年来，日本大学教师个人业绩评价越来越受到重视。据文部省统计，2000年实施大学教师业绩评价的学校共有103所，占总数的16%，到2008年实施教师评价的大学已升至341所，占总数的46%。大学教师业绩评价能在短期内迅速发展起来的主要原因可以归结为以下两点：一是进入20世纪90年代以后，日本经济长期处于低迷状态，各国间的科技之争愈演愈烈。为刺激经济增长、提升本国在全球中的地位，日本提出“科学技术重振日本”的教育目标。政府加大在大学科研经费投入的同时，对学术研究成果的期待也与日俱增。教师是大学最重要的人力资源，教师的科研能力与学术水平将直接影响大学培养人才的质量。2003年，日本冈山大学率先实施教师个人评价制度，在日本国内引起很大反响，因为此前的教师评价一直包含于大学自我评价当中，从未在真正意义上单独实施评价。此后，众多日本大学也相继制定、实施大学教师个人业绩评价制度，教师业绩评价不再是大学自我评价的从属内容，定期实施大学教师业绩评价已成为大学日常管理工作中的重要任务。二是大学教师任期制法案的通过促进了教师业绩评价的迅速发展。虽然任期制被设置为选择性制度，也就是由各大学自行决定是否实施任期制以及怎样实施，但是不可否认的是大学教师任期制的建立使传统的大学教师终身雇佣制度受到一定的冲击。据统计，任期制法律颁布后的五年，也就是2002年，有30%的大学引进了任期制，但以任期制形式所招聘的教师仅占大学教师整体的4%左右。可以看出，大学教师业绩评价多是作为一种激励教师学术研究的考核制度而存在，对大学招聘教师的影响微乎其微。日本大学教师业绩评价主要涵盖教师学术评价、教师教学评价、教师社会贡献评价和教师组织运营评价四个方面，教师学术评价是与其他三项评价相互独立且较为重要的一个评价内容。大学教师学术评价有时也被包含在大学整体评价中或由校方向专门机构提出学术评价申请，由日本第三方评价机构进行评价。

（二）日本大学教师学术评价的理念

在实施学术评价时，专家学者对学术评价所持有的评价理念和认识态度会对评价产生重要影响。2004年日本科学技术·学术审议会在一篇有关学术评价报告中建议，“在实施学术评价时，以积极发展的观点看待学术评价，通过鼓励和建议的方式提高学术活力是学术评价制度所要达到的根本目标。如果评价形式过于统一、僵化，会导致学术科研萎缩，并且会削弱研究者对需要长期进行研究、具有高难度的课题的挑战热情。因此，必须按照各个专业领域的实际情况，建立多元化学术评价制度，及时公布评价结果。”对于日本学术评价理念的更深层理解，可以从以下四点出发：首先，学术的独创性决定了其学术含量，实施学术评价时必须用长远的眼光、站在文化传承的角度，注重科技创新是学术评价的根本。其次，学术评价在关注前沿领域的同时，也要注意推动处于起步阶段的科研项目，帮助其长期开展。再次，学术评价还必须注重培养年轻学术工作者的科研热情，学术评价不能仅局限于事后对研究成果的评价，还应对目前投身于科研工作中的研究者的积极性、发展潜力做出适当评价，以达到促进学术活力的目的。最后，大学作为公共机构使用公共财政进行科学研究时，大学有义务向公众说明科研经费的使用情况，学术评价结果应向公众公布。

二、日本大学教师学术评价制度的内容、方法及应用

（一）学术评价的内容

大学教师学术评价的内容根据每所学校的评价理念、专业领域的差异而略有不同，学术评价内容

大致可以分为以下几点：（1）学术发表情况的统计与公布。统计有关学术著作的出版数量，学术书籍的编撰情况，学术书籍的翻译与评论，学术论文的发表数量，影响因子的累计数值，在国际会议和国内会议上所发表的论文数，受邀在国际会议和国内会议上发表演讲的次数。（2）学术发明。统计专利的申请次数，获得专利权的次数。（3）获奖情况。统计国家颁发的奖章数量，获得学术会议或学术团体的表彰次数，其他有关学术方面的获奖情况。（4）学术研究资金的获得情况。统计教师学术科研资金的获得情况，所受委托的科研项目的数量，参与共同研究的次数，获得捐赠资金的次数，获得其他特别科研项目的次数。（5）受邀成为特聘研究员。统计成为国内外研究机构特聘研究员的次数，成为日本学术振兴会特别研究员的次数，获其他研究机构邀请的次数。

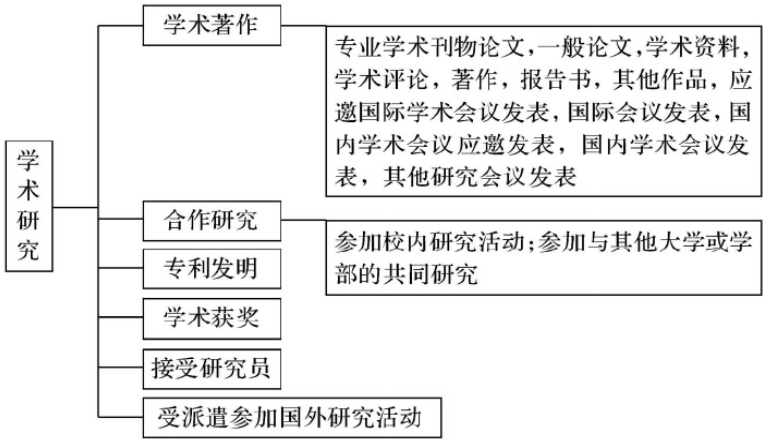

图 1　日本德岛大学综合科学部学术审查指标体系

（二）学术评价的方法

目前，日本大学教师学术评价方法可以分为以下三种：

第一种是记录研究成果和任课数量，按一定方法进行计算，以数值的形式来表示，这种方法称之为综合分数计算法。日本德岛大学在对综合科学部的教师实施个人业绩审查时，运用的就是此种评价方法。2003年12月，日本德岛大学综合科学部以“教师自我检查与评价专门委员会”的名义提出教师个人业绩审查方案，具体评价项目分为教学活动、学术研究、学校管理和社会贡献4个部分。按评价项目类型设置不同评价指标和指标小组，并为各类活动设定不同系数，教学系数为2，学术系数为1.6，管理系数为0.2，贡献系数为0.2。学术研究类活动设置了6组18项指标，如图1所示。评价时对教师学术的各项具体指标进行数字评分，分值范围是1-5分，以便进行量化评价，计算出的平均得分就是教师个人学术研究项目的最后得分。但是，大学一般

会对教师进行教学活动、学术研究、学校管理和社会贡献四方面的综合评价。如果要对教师进行综合评价，就必须求出上述四项指标的综合分数，然后分别乘以对应项目的系数就可得出教师综合评价分数。

第二种是统计教师个人业绩，按照优劣程度分为A、B、C、D四个等级进行判断，这种方法称之为业绩等级判定法。名古屋工业大学在对教师评价时所采用的方法就是等级判定法，学校事先设定了22个指标项，每一项指标又分别有不同等级，教师首先按照统一标准将教师个人学术情况录入到评价数据库当中。例如，其中一项指标是国内会议论文发表情况，如果是6次以上输入a；4-5次输入b；2-3次输入c；仅有一次输入d；没有则输入e。值得特别注意的是，如果本人发表论文次数是4-5次，本应输入b，但是本人所提交的学术论文学术性较强，获得同行的一致认可，可以将此事写在个人记述栏中，申请填写a类指标。设置灵活、有弹性的评价指标有效地确保了学术评价的公正性。此后进行数值换算，a换算为4分，b为3分，c为2分，d为1分，e为0分。如此将22项指标进行换算后，计算出教师个人平均得分。教师按照学校设置的标准，对号入座确定出自己的具体等级，如表1所示，s为卓越、a为优秀、b为良好、c为不足、d为不合格。

表 1　名古屋大学教师评价等级表

评价	学术研究		
	教授	准教授	助教
D	0.00 以上	0.00 以上	0.00 以上
C	1.00 以上	0.90 以上	0.75 以上
B	1.30 以上	1.25 以上	1.00 以上
A	1.90 以上	1.60 以上	1.40 以上
S	2.20 以上	1.90 以上	1.70 以上

第三种是按照教师在年初所制定的目标与全年完成的情况进行对比评价，这种方法被称之为目标管理型方法。这种评价方法是由教师按个人情况自行设定的，是一种以自我评价为主、同行评价为辅的评价方法。在采用目标管理型方法实施评价时，容易出现个别教师夸大个人能力和研究成果的倾向，无法客观地进行自我评价，不但导致评价结果的失效，也会影响教师之间的信赖感和合作关系。因此，目标管理型方法的使用受到很多局限。通过上述举例我们可以看出，前两种评价是对完成情况本身进行评价，每位受评教师都有统一的评价标准。目标管理型评价则是通过设定评价基准，参照受评教师个人全年取得的学术成果进行评价，每个受评教师的评价标准不尽相同。

在评价过程中这三种方法并不是独立的，例如运用综合分数计算法的同时可以将其作为基本信息运用到目标管理型评价当中。综合分数计算法使用频率最高，目标管理型评价方法使用频率最低，各大学按照学校和受评教师的具体情况，采用不同的教师学术评价方法，日本国立、公立、私立大学采用各种评价方法的比例如图2所示。

（三）第三方学术评价

日本长期奉行大学自治、学术自由，抗拒外来的学术评价，这导致日本学阀色彩浓重、高校近亲繁殖现象严重。并且，日本是一个惯行终身雇佣制的国家，大学教师流动性差、竞争压力不大，导致教师学术研究热情不高。虽然，日本政府积极推动大学实施自我评价，然而，由于缺乏监督力量，学校自我评价结果不尽如人意，教师夸大自我评价结果的弊端也随之暴露。为提高大学教育质量、促进大学的学术活力、保证大学评价的公正透明性，在实施大学自评七年之后，日本又以法律条款的形式规定所有大学必须定期接受第三方评价机构的审查。2004年4月，由文部大臣授予评价权力的“大学评价·学位授予机构”正式成立，自此，教师学术评价由过去的大学自评转变为学校与第三方机构共同评价的模式。2009年7月，日本首都大学向“大学评价·学位授予机构”提交申请，希望评价机构能对本校的系统设计学院的五个专业（机械电子学、信息通讯系统专业、航空宇宙系统工程专

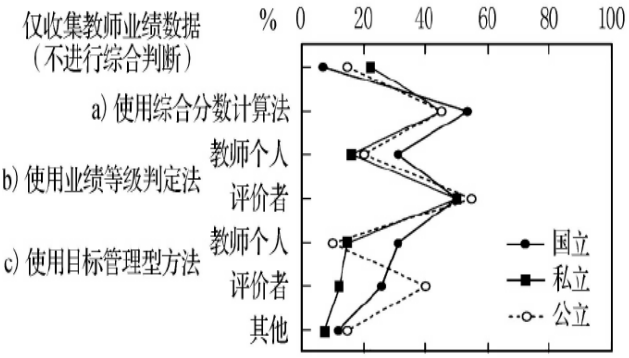

图2　日本各大学所采用三种教师评价方法的比例

业、经营系统设计和工业制造）教师的学术研究成果进行外部评价。机构对学校提交的教师学术研究成果资料进行审查，如表2所示。5名评价委员对研究组织、研究成果、研究资金、社会贡献和学术交流等项目，根据实际情况，按照特别优秀、优秀、普通、不足和较差五个等级分别给出5-1分，每项评价的最后得分是5名委员给出的平均分数。在此基础上，评价委员会对每项评价内容加以分析，告知校方分数的由来，指出未能获得优秀评价的项目的不足之处，并为校方提出今后的改革方向和方法。最后，为保证评价结果的公开有效性，机构会将学术评价结果上传至评价机构的官方网站。

（四）学术评价结果的应用

评价结果的应用主要体现在以下几个方面：薪酬与奖金的分发，教师的升迁和聘用合同的续签，科研经费的分配，学校内部的表彰。对于学术评价结果较差的教师，为确保教师学术研究时间和精力，会适当减少教师业务工作量，对于个别教师还会实施特别指导。目前，大学只在少部分岗位聘请任期制教师，评价结果将决定他们能否可以连任，然而任期制教师只占少数，所以评价结果更多地被用于激励教师的学术热情，评价结果与职位提升、奖励补助等方面相联系。学校性质不同，学术评价结果的应用情况各有差异，如图3所示。

三、日本大学教师学术评价制度的特点

（一）注重学术活力，强化教师自我学术评价意识

日本实施大学教师学术评价制度，其主要原因在于日本属于终身雇佣制国家，人事制度过于保守，致使许多大学教师安于现状，缺乏竞争力和危机意识使得大学教师对于学术科研产生倦怠心理。

表2　日本首都大学教师申请学术评价的统计资料（单位：人 / 次）

项目			2007 年		2008 年		2009 年	
			一年总数	每人平均数	一年总数	每人平均数	一年总数	每人平均数
全体人数（提交者人数）			82(79)		88(87)		91(89)	
研究成果		论文	171	2.34	213	2.70	157	1.96
		著作及评论	34	0.43	68	0.78	60	0.67
		国际会议演讲	271	3.34	323	3.71	250	2.81
		国内会议演讲	594	7.52	613	7.05	373	4.19
		作品	38	2.71	47	2.76	42	2.33
		邀请学术讲演	32	0.41	39	0.45	19	0.21
		与学术相关的获奖	33	0.42	21	0.24	7	0.08
		参加国际会议	129	1.63	148	1.70	113	1.27
学术交流	共同研究	国内	99	1.25	129	1.48	129	1.45
		国外	14	0.18	23	0.26	22	0.25
	研究交流	国内外研究者接纳	8	0.10	5	0.06	10	0.11
		海外派遣	3	0.04	6	0.07	3	0.03
		招聘客座教授研究员	7	0.09	10	0.11	12	0.13
研究资金获得	科研费	新课题申请	55	0.07	64	0.74	77	0.87
		采纳・继续	26	0.33	30	0.34	38	0.43
		总金额（千日元）	62,400	789.87	72,447	833	99 957	1 123
	外部资金	共同	28	0.35	28	0.32	33	0.37
		受委托	8	0.10	15	0.17	14	0.16
		捐赠	20	0.25	28	0.32	25	0.28
		公开征集	19	0.24	17	0.20	19	0.21
		外部资金总数	91	1.15	104	1.20	99	1.11
		总金额（千日元）	158,586	2,007	184,122	2,116	213 795	2 402

大学教师学术评价制度的引进，并非是为教师增加外部压力，而是希望通过学术评价激发教师的创造性与主动性，从而提高学术活力和学术产量。日本学术评价理念倾向于大学教师通过自我评价，不断提高自身研究能力，外部评价是促使教师对学术自发追求的内化方式。学术评价结果一方面成为人事升迁、薪酬发放的依据，另一方面是能及时了解教师的学术研究情况，以此为依据，为他们提供建议

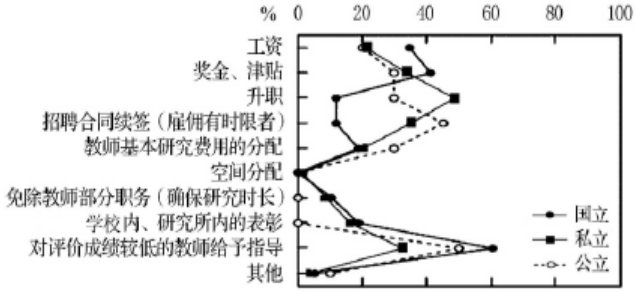

图3　日本大学教师学术评价结果的应用率

和帮助。日本建立以促进教师再发展为根本目的的学术评价制度，行之有效地促进了大学教师的学术活力。

（二）注重学术评价结果的有效性，加强教师的话语权

日本大学在实施教师学术评价前，会设立专门的学术评价委员会。学术评价分为几轮进行。如名古屋工业大学对教师进行学术评价时，由董事会、教师评价委员会、领域专家组和评价小组共同承担。各组各尽其责，评价小组主要负责核对各教师所提交的个人学术研究结果资料的真实性，并在此基础上进行公正评价。领域专家组主要负责对评价小组所呈交的教师评价结果，以第三方评价立场重新进行审核，确认评价结果的正确性，保证学术评价结果毫无偏差。教师评价委员会主要负责监督学术评价过程的实施，以确保评价过程中每个环节的

公正。董事会对教师最终的综合评价结果进行确认。学校设立专门机构对教师业绩实施评价，各组织部门责任划分明确，评价组织与监察组织并存，保证了教师学术评价结果的有效性。并且，大学在实施学术评价的过程中，非常重视教师参与评价的话语权，不是仅将教师作为受评对象看待，当教师对评价内容有不同意见时，可以在个人记述栏中提出个人意见。对评价结果有异议时，可以及时通过学校设立的重审通道向学术评议委员会提出再审诉求，充分体现出教师学术自由的权力。

（三）创办第三方评价机构，加强学术评价的客观性和透明度

日本为确保学术评价的客观、公正性，第三方评价机构在学校自我评价的基础上进行监督和审查。第三方评价机构的创办，使得日本学术评价制度从单一化评价向多元化评价方向发展，从自我评价向外部评价进行转变。专业化的外部学术评价组织的成立，尽可能避免了教师自我评价过于夸大或管理层教师动用手中权力干涉学术评价结果的情况的发生。不仅如此，评价机构会及时将结果公布于网上，这不但能保证学术评价过程的公正性和结果的透明性，也能促进各大学教师之间的相互交流，更是评价机构在学术科研方面对社会负责的一种表现。

四、日本大学教师学术评价的积极作用与消极影响

在科技创新的立国战略背景下，日本极其重视大学的学术研究功能的发挥。国家投入专项资金，设立“Top30”、“COE计划”“GP”等项目，这些项目无不强调着大学的学术研究功能。通常情况下，政府在划拨科研经费与学术研究竞争性资金时，并不按照大学整体进行划拨而是按照学术研究课题的数量以及课题的学术含量，实施倾斜性的科研经费拨付政策。此时，大学教师学术评价结果在划拨科研经费方面可以起导向作用。教师学术评价制度能刺激学术研究的活力。以往日本大学教师是终身聘任制，学术研究仅靠教师的自主性和本人对科研的热情，学术研究活动曾出现了停滞。实施教师学术评价制度后，虽然刺激效果不是非常明显，但研究数量确实在稳步上升。日本科学技术·学术审议会在2012年2月14日公布的《有关大学改革学术研究方面的主要观点及现状分析》报告统计称，日本1998-2000年间论文发表数量年均62457篇，2008年－2010年间论文发表数量上升为年均71149篇。

虽然日本近年论文发表和科研成果数量有所提高，大学教师学术评价制度在刺激学术活性化方面起到了一定的作用，但同时也显现出一些负面影响。据调查，2002年大学教师在教学方面每年平均花费的时间，人文社会科学专业的教师年均时间为675小时，占总业务时间的26%，自然科学专业的教师在教学上所用时间为627小时，占总业务时间的21%，多数高校教师认为学术研究比教学更重要。大学教师学术研究成果的多少不但影响所在学校科研成果的数量，也关系到学校在社会上的声誉和影响力，甚至直接影响教师本人获取科研经费的数额，关系到教师本人科研项目的进一步发展。所以许多日本大学教师在学术科研方面因投入过多精力而忽视了教学工作，大学最基本的教育功能被削弱，导致高校人才培养质量有所下降。部分大学教师为提高自身短期业绩，片面追求学术研究的数量而忽视研究质量，这种急功近利的学术研究态度对大学的学术活动发展，尤其是长期、基础性的学术研究活动产生了负面影响。

总之，教师学术评价是日本大学教师个人业绩评价的中心环节之一，但由于学术研究具有滞后性和潜隐性的特点，不少大学通过统计教师学术成果的数量、论文发表的期刊的等级使学术研究量化以便能准确进行评估，然而学术研究评价不能等同于学术业绩评价，学术业绩评价与学术能力之间也不能直接画等号。日本实施大学教师学术评价时，除了应注重激发教师学术研究活力，也要尽可能避免由于实施学术评价所引发的教师重学术轻教学情况的产生，这是今后日本学术评价改革时所应注意的方面。

（韩小娇，沈阳师范大学教育科学学院硕士研究生，辽宁沈阳　110034；高　军，沈阳师范大学教育科学学院副教授，辽宁沈阳　110034）

（原文刊载于《高教发展与评估》2013年第2期）

韩国高校学术评价制度体系的借鉴与反思

王晓杰　金佳律　付瑞萱

科学、完善、合理的高校学术评价制度体系对于一个国家高校学术研究事业的健康发展具有十分重要的意义。他山之石,可以攻玉。学术研究工作是高校的主要职能和核心任务之一,学术国际化是高等教育国际化的重要组成部分,其发展情况也是体现高校学术地位的重要标志。韩国与中国在地理上是近邻，在文化上互相影响。认真研究韩国高校的学术评价制度体系，对于健全与完善中国高校学术评价制度体系具有十分重要的借鉴意义。

一、韩国高校学术评价制度体系的一般现状分析

（一）韩国高校学术评价制度的变迁历程

在韩国的高校中，浦项工科大学最早开始实施学术评价制度。该校于1986年示范性地实施了学术评价，并规定学术研究成果评价是教师业务评价中的主要内容之一。同时，以此为开端，关于学术评价的探讨也正式开始。20世纪90年代中期，随着韩国高校自主评价制度的实施，学术评价成为大学评估项目之一。这样的评估项目也逐渐在所有四年制大学中开始推行，并一直延续至今。

目前，韩国高校普遍将学术评价作为大学教师业务评价的重要部分。教师的业务评价尤其是科研评价的具体规定由各高校自主制定。对于详细的评价标准、评价项目等内容，每个高校的规定也不尽相同。但是20世纪90年代以后，大部分高校将教师的工作范围划分为教育学生、科学研究、社会服务三个领域，从这三个方面来评价教师的业务能力与水平。教育、研究、服务三个领域的评价分数所占的比重，由高校的不同定位以及教师个人发展所侧重的领域所决定。比如，在研究型大学，教育、研究、服务的评价分数比例为30:60:10；教学型大学（如：教育大学）的评价比例设置为60:30:10。同时，高校也根据教师个人侧重领域的不同如选择教学型方向或者研究型方向而采用不同的评价比例。

表1　韩国国立综合大学教师评价类型及评价项目分数

大学	个人评价类型选择	教育	研究	服务	产学合作
A大学	无	40	50	10	
B大学	无	无限	无限	无限	
C大学	标准型	40	50	10	
	教育型	60	30	10	
	研究型	30	60	10	
	服务型	40	40	20	
E大学	无	50	50	20	20
F大学	无	20	60	20	
G大学	标准型	50	40	10	
	教育领域	60	30	10	
	研究、产学合作	30	60	10	
	服务领域	30	30	40	
H大学	一般	45	45	10	
	教育型	60	30	10	
	研究型	30	60	10	
I大学	工业学院:教育型/研究型**	50	40	10	
	社会学院:无	50	40	10	
J大学	一般教授	45	45	10	
	产学合作教授	30	30	10	30
K大学	无	50	40	10	
L大学	法科学院	40	40	20	
	生活科学学院	35	45	20	
	自然科学学院/数理科学部	40	40	20	
M大学	无	无限	无限	无限	
N大学	无	无限	无限	30分	
O大学	无	无限	无限	无限	
P大学	无	部分上线	部分上线	50	
Q大学	无	无限	无限	无限	
R大学	1类型	70	20	10	
	2类型	60	30	10	
	3类型	50	40	10	
	4类型	40	40	20	
	5类型	40	50	10	
	6类型	30	60	10	
	7类型	20	70	10	

如果没有特别的侧重方向，一般教育和研究所占的比例就基本相同。在大部分高校，教育和研究合计的评价比重占80%-90%，其中教育和研究评价比重各占一半，或研究评价所占的比重更高。

表1展现了2010年韩国部分国立综合大学的教师评价类型及评价项目分数分配情况。如表1所示，有些高校可由教师本人决定自己的评价方向，也有些学校根本没有这样的制度。由教师本人自选评价方向的制度，各个学校的规定也各不相同。大多数学校可以选择教学、研究、服务三个评价领域之一。但也有像J大学根据教师身份特征决定评价方向的情况，或者像I、L大学根据学院划分评价方向的情况。如上所述，存在多种评价教师业务的方式，因此研究领域的评价比重也十分多样化。如表1中阴影部分所示，研究领域的评价比例根据学校定位不同及教师评价方向的不同，最低占20%，最高达到70%。

私立大学中除了少部分学校外，大部分学校并未实施教师业务评价中的方向划分。只有极少的私立学校实施了评价划分制度，如K大学专职讲师和助理教授选择“研究型”评价，副教授可以在“研究型”、“教育型”、“研究·教育·服务综合型”中任选一个方向；J大学根据专业及方向实施评价。

评价项目分数指不同高校不同项目之间的比例或原分数的上线分数。

教育型：“教育+服务”60%, 研究40%；研究型：“教育+服务”40%, 研究60%。

（二）韩国高校学术评价的外部制度环境

对韩国高校教师学术评价产生最重要影响的制度是大学评估制度。1982年韩国大学评估制度开始实施，到2007年大学评估制度有了强有力的法律依据，并变成了体系化的制度。在原有的《高等教育法》中新设了关于大学评估的条款（第11条第2项），随着此项法令的推行，韩国全国高校都开始实施大学评估。大学评估的主要目的是通过提高高校发展的质量和水平来提升韩国大学的国际竞争力。这也是韩国政府针对2005年UNESCO和OECD共同发表的《跨国高等教育质量保障方针》（UNESCO/OECD Guidelines on Quality Provision in Cross-Border Higher Education）的应对策略。依据韩国《高等教育法》第11条第2项规定，根据教育部长官令的内容，高校需要自行评价检查该校的教育研究、组织管理、设施设备等情况，并公示自评结果。或委托教育部许可的评估认证机构评估或认证大学管理和教育水平，并公示相关结果。如接受外部评估认证机构评估的大学，可以免除自行评估。但即使接受外部评价认可的大学，也需要提交大学的自主评估材料。因此，不管何种情况，大学的自主评估资料都十分重要。根据《高等教育机构的自行评估规定》，现行的高校评估每两年实施一次。接受外部认证机构评估及获得认证的大学会得到较高的社会认可，所以目前各高校更倾向于获得外部评估认证机构的评估。大学评估结果通过媒体、认证机构网页、相关大学网页，详细地公示内容。这样的公示制度产生的效用直接关系到高校的社会声誉及国际影响力。比如：直接影响到国内学生及家长择校意愿；外国留学生也通过这样的信息判断哪所大学的质量较高。同时，除了社会影响之外，评估结果也会直接影响到大学的外部支持力度。《高等教育法》第11条第2项第4点指出“政府对大学行政或财政上的支援，可以参照评估或认证结果。”大学要想确保政府的相关支持，就必须重视评估结果。因此，对于评估的过程及结果，各个高校之间的竞争也十分激烈。对高校教师的学术评价是大学评估内容中的一部分，为了获得优秀的大学评估结果，教师学术研究成果的数量和质量就成为了重要的影响因素。政府并没有直接干预高校教师学术评价，而是由各个高校自行实施学术评价政策，因而各高校学术评价的具体规定和结果存在一定的差异。

近来，韩国高校学术评价制度中最具争议的问题便是“国立大学教师成果比例薪资制”。虽然此制度只以国立大学教师为适用对象，但这样的制度也会影响到整个高等教育界，所以备受瞩目。虽然该制度已经实施，但是不断有教师提出异议，因此该制度也在不断地修改完善之中。“国立大学教师成果比例薪资制”是根据韩国政府《公务员报酬规定》的修订，将之前以定岗年薪为主要形式的国立大学教师报酬体系转变为成果比例薪资制。此成果比例薪资制是指按一定周期评估教师的教育研究、

社会奉献以及其他业绩成果，从而决定教师个人每年收入总额的制度。根据该制度规定，国立大学新入职教师从2011年开始，在职教师中非终身教授(non-tenure-track)从2013年开始，终身教授(tenure-track)从2015年开始执行成果比例薪资制。该制度将教师的成果和业绩分为四个等级，排位前20%的教师可获得平均成果年薪的1.5——2倍薪资，排位后10%的教师年薪冻结，只能够得到较低待遇。不仅如此，随着定岗年薪制度的取缔，成果薪资的一部分附加到下一年度的基本工资中。因此，今后教师之间的收入差距也会越来越大。例如，四年连续获得最高等级与最低等级收入的教师，薪资总额可达到1000万元韩币（约57000元人民币）以上的差距。但政府由于财政上的制约，不能采用针对优秀教师的附加预算方式（plus sum），反而采用削减最低等级教师薪资，补充优秀教师薪资的归零预算方式（zerosum）。即，获得优秀评价的教师获得最低等级同事工资的方式。这样的方式会激化高校教师之间的竞争，诱发不和谐因素的增加，因此教师群体强烈反对这一政策。虽然目前在韩国一小部分私立大学已经开始实施这种成果比例薪资制度，但是如果在收入水平相对较低、职业稳定性相对较高的国立大学全面实施这样的制度，就会不断加剧韩国高校内部的激烈竞争氛围。如前所述，教师学术评价中科研成果占据重要比重，这样的制度因素必然对高校教师的学术研究活动造成沉重的心理压力。

表2　国立综合大学国际学术刊物及学术交流的加权值

	国内学术期刊		国际学术期刊			学术交流	
	可检索	不可检索	可检索	著名	不可检索	国内	国外
A大学	100	70	200		70	30	50
B大学	15	5	20		8	2	5
C大学	30(人文社会) 20(自然学科)	30 20	30		30 20	5	10
E大学	150	50	300		100	30	100*
F大学	30	5	50		20	3	5
G大学	150	50	300		150	30	100
H大学	100	20	200	1000**	150	20	40
I大学	30(工科) 35(社会学科)	15	60 70/80		30 30		
J大学	100	50	200	400	150	15	30
K大学	20		30		20	4	6
L大学	各专业不同						
M大学	120	70	200		150	10	30
N大学	20	5	40	50	15	3	6
O大学	100(人文学科) 70(自然学科)	50 50	150 100		100 70	10 10	15 15
P大学	150	60	300		100	30	80
Q大学	15 10(理工学科)	10 5	15	20	7	3	5
R大学	150	70	200		100	30	60

二、韩国高校学术评价制度体系的内涵分析

大部分韩国高校的学术评价主要针对学术论文、学术著作、发明专利、项目经费、学术奖项等内容。对高校教师学术成果进行数量评价与质量评价，大多数高校遵循的标准是依据论文发表的学术刊物等级、学术发表为国内级别还是国外级别等标准。例如，在国内发表的论文质量评价主要是根据是否能在韩国学术期刊检索（Korea Citation Index：KCI）上登载的学术期刊上发表，以此来评定论文质量水平的优劣。论文在国际学术期刊发表

的情况，根据是否发表于SCI、SSCI等学术期刊，以此来区别评定。学术交流成果也要根据国内学术研讨会和国际学术研讨会而区别评定。学术著作也要根据国内、国外出版分别计算分数。根据相关调查显示，大部分高校对于国外刊发的期刊比国内刊发的期刊、国际学术会议的交流比国内学术会议的交流赋予更高的分数。国际可检索期刊根据期刊级别可达到国内可检索期刊分数的130%-200%。表2为国立综合大学国际学术刊物及学术交流的加权值赋予现状。

如表2所示在C大学人文学科的评价中，国内可检索期刊与国际可检索期刊的评价分数没有区别。Q大学也只有理工学科的国际检索刊物存在加权值。H大学一般国际检索期刊的加权值为国内检索期刊的2倍，《Nature》或《Science》刊物等国际顶尖学术期刊的加权值为10倍。另外，针对教师个人最低科研成果水平的限定，各高校有不同规定。虽然有一些高校不存在最低加权值，但大部分高校以1年、2年、3年为单位设定最低加权值。表3为韩国J大学（私立大学）以2年为单位设定最低标准研究成果的案例。依据研究方向的种类、终身任职期限以及专业领域规定不同的最低加权值。但是J大学不仅设定了最低标准，还设定了满分标准。由于满分标准要求较高，这就给将要晋升职称的教师造成了较大压力。

在国立大学，专业学术著作与论文的对比加权值因学校而异，一般为100-250%。根据学校的不同，专业译著一般为论文加权值的50-150%。由此可见，对于译著的评价比重一般与论文相似或低于论文。对于政府机构和公共机构的研究报告，只有一部分学校赋予科研成果分数。

韩国高校学术评价结果的应用主要是对于高校教师的业务评价尤其是科研评价，其作为大多数高校教师升职、返聘、续约、终身保障等人事考核的基础材料。同时也作为获得研究补贴、研究经费、优秀教师奖励、学术休假等物质或非物质奖励的评定依据。同时也是实施年薪制度大学用以确定教师年薪标准的基本资料。

三、韩国高校学术评价制度体系的特征分析

（一）主要表现特征

当前韩国高校实施的学术评价制度的主要特征为十分注重定量评价，数量化取向较为突出。无论国立、公立大学，还是私立大学基本都采用将论文发表、著作出版等学术活动进行数量化计算的评价方式。定性评价则退到次要的位置，比如，只评价论文是否发表在相关检索期刊上，或者是否在国外发表。

韩国高校学术评价制度体系的第二个显著特征是高校教师之间围绕着学术评价而形成的竞争越来越激烈。20世纪90年代中期正式开始实施的大学评价制度，到21世纪初期以后开始逐渐强化，高校之间的竞争日益激烈，从而对于教师学术评价的压力也逐渐增加。而且如前所述，由于国立大学实施成果比例薪资制度，这也加剧了高校教师内部的激烈竞争氛围。

韩国高校学术评价体现出来这样的特征，与其高校谋求生存和发展的社会环境息息相关。目前，韩国国内大学入学人数呈现出逐年递减的趋势。这样的现象是由过去几十年低出生率导致的人口结构变化所造成的，是不可避免的问题。韩国大学可容纳人数超过大学报考人数的现象以后还会出现。客观环境因素导致韩国的高等教育不得不进行结构性调整，即，市场中会出现被淘汰掉的大学。韩国的国立大学已经从几年前就开始了院校之间的合并调整。高校想要在这样的环境之中生存下来，只有提高自身竞争力，接受学生与家长的选择。只有确保具有较高教育水平、科研水平的教师，为学生提供优质的服务，具备良好教学设施的高校，才能在激烈的国内学生大学入学率竞争中占据优势。韩国改善高校生计的另外一个方法便是大力招收海外留学生。从竞争的角度而言，吸引海外留学生是应对国内生源减少的有效方案。吸引海外留学生是现在很多地方私立大学选择的生存策略。但是在这样的情况之下，也只有提升学校的学术水平，达到或者超过国际先进标准才能实现长远的发展。因此，客观环境因素促使政府管理部门以及各个高校都十分关注提升学术研究工作质量，其中作为提升竞争力的核心要素的教师学术水平至关重要。

（二）关于韩国高校学术评价制度合理性与否的相关争议

当前，大力提升高校的学术研究水平已经成为韩国各个高校普遍认同的命题。但是在具体实施的过程中，往往会引发高校之间过度的竞争。现在韩国绝大多数高等学校都形成了以定量评价为主的学术评价方式。对此，韩国高等教育界存在着许多不同的看法，争议主要表现在两个方面：

第一，高校对教师进行学术评价的必要性。

第二，学术研究定性评价的相关问题。

除了以上争议之外，有人还提出国外检索期刊登载的学术论文获得较高评价分数的方式会对使用母语的学术研究者造成消极影响。由于意识到评价分数高低的因素，很多高校教师的学术成果并没有发表在国内学术期刊上，而是发表到国外学术期刊上，这并不利于学术成果的实际应用。虽然这些观点有一定的道理，但是当前在韩国高校全面推行同行评议等质量评价手段尚存在诸多现实困难。韩国大学的组织文化与欧美不同，实施真正意义上的质量评价还存在一定的困难。在这样的现实环境之中，确保高校学术评价具有相对公正性的现实方案只能是研究成果的计量评价和检索期刊的定性评价之间的相互结合。

四、借鉴与反思

韩国高校的学术评价制度于1986年开始实施，随着20世纪后期大学综合评估制度的正式实施，学术评价制度开始逐渐形成规模体系。学术成果的评价方法包括定量评价以及是否登载于检索期刊的定性评价，即采用了综合式的评价方法，但其数量化导向的特征仍然十分明显。此种情形与我国高等学校学术评价制度体系现阶段的特征十分类似，这也是东亚国家高等学校在进行学术评价时共同需要研究解决的问题。

许多韩国高校教师由于学术成果评价而产生的压力较大。学术评价作为教师成果评价的三个标准之一，其中教育评价由于授课时间的限制，很难区分差异；服务评价由于在综合评价中所占比重较低，也很难区分差异。因此，只能通过学术评价区分教师之间的差距。同时，随着高校之间的竞争日益激烈以及最近国立大学实施的成果比例薪资制等制度的影响，导致高校教师的科研压力逐渐增加。随着知识经济时代的到来，提高高校学术研究工作的质量和水平是世界性的发展趋势，这已经成为共识。但是在过度竞争的环境中，就容易形成以计量评价为主的高校学术评价方式。这背离了学术评价原本的积极功效，并没有实现“促进教师发展”的初衷。不仅影响了高校健康向上学术氛围的形成，也损害了真正意义上的学术自由，是对学术共同体精神的异化。我国高校学术评价制度体系的改革必须贯彻以人为本的发展理念。人的全面自由发展是任何社会制度体系建构所必须遵循的终极价值取向，人是发展的目的而不是手段，这必须成为我国高等学校学术评价制度体系建设与改革的基本原则。

目前，韩国高等教育界正努力尝试吸收来自多方面不同的声音，对于韩国高等学校学术环境建设和学术评价方式方法完善的相关研究也日渐增多。以这样的努力为基础，相信以后韩国高等学校学术评价制度的改革既能反映出教师本人的发展诉求，也能满足社会发展需要，并不断向着更加科学合理的方向发展。

（王晓杰，吉林师范大学经济学院教授，吉林四平　136000；金佳律，吉林师范大学经济学院特聘教授、韩国梨花女子大学教授，吉林四平136000；付瑞萱，吉林师范大学外国语学院教师，吉林四平　136000 ）

（原文刊载于《吉林师范大学学报（人文社会科学版）》2014年第4期）