

内部资料
免费交流

京内资准字0609-L0057号
北京交大印刷厂印刷

中国教育科研参考

2014年 第 23 期
总第(345)期

中国高等教育学会编

2014年12月15日

目 录

教育部关于深化高等学校科技评价改革的意见·····	(02)
上海市教育委员会关于推进上海高等学校科学研究分类评价的指导意见·····	(04)
英国人文社会科学成果评价体系及其借鉴····· 杜向民 朱燕锦 刘兰剑	(05)
全球科研评价体系的演进与发展 ——以美国、英国、澳大利亚、荷兰四国为例·····	蒋玉梅 (11)
效率和生产率方法在高校科研评价中的应用·····	胡咏梅 段鹏阳 梁文艳 (16)
关于完善我国高校科研质量评价的思考 ——基于世界发达国家的政策经验·····	陈慧颖 刘在洲 (24)
高校科研评估指标及其实证研究·····	高宏利 李作学 王 前 (27)
高等学校科研评价改革在北京上海广东率先启动·····	中国教育报 (32)

编者的话：科学研究作为高校的三大职能之一，受到了高校教师与学者的普遍关注，由此引发的“高校科研评价与管理”“人文学者与自然科学者之争”也是高等教育综合改革需要解决的问题之一。新时期，激发高校科研创新活力，提升高校科研的社会服务水平，需引进科研分类评价体制机制，规范高校科研管理模式，实现高校科研管理从硬性管理到人性化服务的转变。对不同类型高校、不同类型学科采用分类评价机制，建立多元化、多维度的评价指标体系，分类考核，科学管理，促进高校科学研究发挥应有的社会服务与引领科技发展的职责。本刊以“科研评价与管理”为选题，集中选编若干国内外相关文章，供读者参阅。

主 编：王小梅 本期执行主编：范笑仙 责任编辑：聂文静
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

教育部关于深化高等学校科技评价改革的意见

教技[2013]3号

各省、自治区、直辖市教育厅（教委），新疆生产建设兵团教育局，有关部门（单位）教育司（局），部属各高等学校：

为深入贯彻党的十八大和十八届三中全会精神，加快落实《中共中央国务院关于深化科技体制改革加快国家创新体系建设的意见》和《国家中长期教育改革和发展规划纲要（2010-2020年）》，深化高校科技评价改革，提出如下意见：

一、明确改革导向

1. 高等学校是国家创新体系的重要组成部分。科技评价不仅是高校科技管理的主要手段，而且是高等教育宏观管理和资源配置的重要依据，对高等学校办学行为具有突出的导向作用。深化高校科技评价改革，对于大力提升高校创新能力、推动高等教育综合改革、实现创新驱动发展具有十分重要的意义。

2. 高校长期以来形成的以统一、量化为特征的科技评价机制，对调动科技人员的积极性和创造性发挥了重要作用。面对全面提高质量和创新驱动发展的时代要求，科技评价中的问题日益显现：重数量轻质量、重形式轻内容、重短期轻长远的现象依然存在；评价指标单一化、评价标准量化、评价方法简单化、评价结果功利化等倾向没有得到根本扭转；分类评价实施不到位，对科技成果转化、科学普及等工作激励不足；科技支撑教学和创新人才培养的导向不够；开放评价、长效评价机制不够健全，这些问题将严重影响高校科技工作的持续健康发展。高校科技评价改革的任务十分紧迫。

3. 深化科技评价改革的指导原则是：鼓励创新，服务需求，科教结合，特色发展。注重科技创新质量和实际贡献，重点突出围绕科学前沿和现实需求催生重大成果产出的导向，建立产学研协同创新机制加快创新驱动发展的导向，推进科教结合提升人才培养质量的导向，鼓励科技人员在不同领域、不同岗位做出特色，追求卓越。

4. 深化科技评价改革的目的是，根据不同类型科技活动特点，建立导向明确、激励约束并重的分类评价标准和开放评价方法，营造潜心治学、追求真理的创新文化氛围。着力提升基础研究和前沿技术研究的原始创新能力，关键共性技术的有效供给能力，支撑高质量创新人才培养的能力，服务国家和地区经济社会发展战略需求的能力。

二、实行分类评价

5. 实施科学的分类评价。针对科技活动人员、创新团队、平台基地、科研项目等不同对象，按照基础研究、应用研究、技术转移、成果转化等不同工作的特点，分别建立涵盖科研诚信和学风、创新质量与贡献、科教结合支撑人才培养、科学传播与普及、机制创新与开放共享等内容，科学合理、各有侧重的评价标准。

6. 对主要从事创新性研究的科技活动人员实行代表性成果为重点的评价。对基础研究人员的评价以有利于潜心研究、长期积累，催生重大原创性成果为重点。对应用研究人员的评价以聚焦需求，具有自主知识产权和重大技术突破为重点。对软科学研究人员的评价以服务决策需求，形成战略研究报告，支撑思想库智囊团建设为重点。

7. 对主要从事技术转移、科技服务和科学普及的科技活动人员实行经济社会效益和实际贡献为重点的评价。鼓励高校按照国家相关规定，完善职务聘任和岗位聘用机制，面向市场和产业需求，把科技转化为现实生产力；加强协同创新，强化企业在技术创新中的主体地位；积极开展科学普及活动，提高全民科学素养；创新人才培养模式，提高创新创业人才培养质量。

8. 对从事技术支撑和服务的科技活动人员实行以服务质量与实际效果为重点的评价。鼓励高校加强职务聘任和岗位聘用的引导，提高技术支撑人员服务技能，加强自主开发仪器设备。

9. 对高校创新团队实行以解决重大科技问题

能力与合作机制为重点的整体性评价。围绕团队自身建设、代表性成果、科教结合、机制文化等方面开展评价。要认可团队科技成果所有参与者的实际贡献，同时反对无实质贡献的虚假挂名。

10. 对高校创新平台（机构、基地）实行以综合绩效和开放共享为重点的评价。围绕创新质量、服务贡献、科教结合、人才队伍、机制文化等方面开展评价。鼓励创新平台深化科教结合，注重创新成果转化、技术服务和科学普及。鼓励接受国内外访问学者。科研基础设施依照规定应该开放的一律对社会开放。

11. 对不同类别科技项目，制定科学合理的评价标准，注重创新的独特价值。基础研究项目要以原始创新性成果和创新性人才培养为评价重点，着重评价科学价值。应用研究项目要以关键技术和核心技术突破、自主知识产权成果、经济社会效益等为评价重点，着重评价目标完成情况、成果转化情况以及技术成果的突破性和带动性。产业化开发项目要以技术、产品的成熟度和市场反应为评价重点，着重评价对产业发展的实质贡献，促进科技成果资本化、产业化。

三、坚持开放评价

12. 建立开放评价机制。基础研究以同行评价为主，大力加强国际同行评价。应用研究和产业化开发应建立主要由市场决定技术创新项目和经费分配、评价成果的机制，由用户、市场和专家等相关第三方参与评价。加强开放、多元的国内外专家数据库建设和共享。充分利用信息化手段，提高科技评价工作效率和开放程度。

13. 完善公平、公正、透明的开放评价规则。完善评价答辩、公示、反馈、申诉、举报和回溯评价制度，健全随机、回避的评价专家遴选机制，健全评价专家责任和信誉制度，增强评价专家的社会责任感，保证评价工作的独立性、公正性和评价结果的科学性、客观性。

14. 建立长效评价机制，避免频繁评价。根据科技活动类型、学科特征，结合人事聘用合同、项目过程的要求，适当延长评价周期，注重评价实效。科技活动人员的评价周期原则上不少于3年，对青年科技人员实施聘期评价，创新团队和平台基地的评价周期原则上不少于5年，根据绩效情况可

减少、减免评价。加强评价结果共享，避免重复评价。

四、切实组织落实

15. 全面改革教育系统各类科技项目、人才项目和科技奖励的评审办法。按照改革导向，教育部将修改完善各类科研基地和人才项目评价体系及评价机制；提高科技奖励质量，减少数量，适当延长报奖成果的应用年限。地方教育行政部门要对本部门各类科技项目、人才项目和科技奖励做相应改革。

16. 高校承担科技评价改革的主体责任。加强以科研诚信与学风、创新质量、服务贡献和科教结合为重点的分类评价，更加强调科研成果转化应用、技术转移推广、科学普及等科技产出在评价中的作用。要改变考核评价中将科技项目与经费数量过分指标化、目标化的作法。要改变在教师评聘、收入分配中过分依赖和不合理使用论文、专利、项目和经费数量等科技指标的做法，减少科技评价结果与利益分配过度关联。

17. 在教育管理中科学合理使用科技评价结果。各级教育行政部门要优化科技指标在高校设置、资源配置、学科评估、研究生教育等管理环节中的使用方式。按照此次科技评价改革的导向，修订相应的指标和标准，更加强调创新质量和实际贡献，更加注重科教结合支撑人才培养。

18. 科技人员要弘扬求真务实的科学精神和健康向上的评价文化。要加强科学道德建设，践行诚信文化，抵制不良风气的侵蚀。要开展学术批评与学术争论，保障不同学术观点的公开表达和充分讨论，共同营造求真求实、平等争鸣、鼓励探索、宽容失败的氛围和环境。

19. 引导社会力量科学合理地使用科技指标评价高校。要增强指标的系统性和完整性，补充和完善科教结合支撑人才培养、服务与贡献、科学普及等指标。要正确处理数量、质量和投入产出比的关系，强化对特色、质量和社会贡献的考察。应避免对不同类型高校采取同一标准、依据少数定量科技指标进行简单排名与短期评价。

（原文刊载于中华人民共和国教育部官方网站，<http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/moe-784/201312/xxgk-160920.html>，2013年11月29日）

上海市教育委员会关于推进上海高等学校科学研究分类评价的指导意见

沪教委科[2014]54号

各高等学校：

为贯彻落实《教育部关于深化高等学校科技评价改革的意见》（教技[2013]3号）和《教育部办公厅关于开展高等学校科技评价改革试点的通知》（教技厅[2014]3号）精神，切实推进上海高校教育综合改革，支撑上海国际创新中心的建立，以激发创新活力为根本，以质量和实际贡献为导向，逐步建立科学合理的高等学校科学研究分类评价体系，进一步深化学校管理体制的改革，特制定本指导意见。

一、建立针对不同类型高校的分类评价体系。引导和鼓励不同类型的高校，基于学校定位和学科类型，建立科学研究的分类评价体系，使评价更加符合学校发展的需要。

二、建立针对不同类型学科的分类评价机制。引导和鼓励各高校以学科特色发展为目标，探索建立以创新质量和实际贡献为主的科研分类评价体系，科学合理地评价各个学科发展阶段，各种类型学科发展特征，促进学科健康发展。

三、建立对科研成果的多元化评价体系。针对科学研究活动的多样性，包括基础研究、应用研究、技术转移、成果转化、决策咨询、科学普及等各种形式的工作，注重研究成果的多样性，包括学术论文、专著、译著、学术工具书、艺术作品、创意设计、发明专利、研究报告、咨询报告、国家和行业标准、对经济社会发展的贡献等，推进并完善“代表性成果”的评价机制。

四、建立多维度的科研评价措施体系。针对科研工作特点和科研活动载体不同，面向重点实验室、工程中心、创新团队、个人等不同科研载体，探索建立团队考核与个人发展相结合、长周期考核与过程管理相结合、学科建设与学校顶层设计相

结合的多维度评价方法体系，积极推进“同行评议”“第三方评价”和“国际评价”等评价方法，探索开放的、多方参与的科研评价措施，建立与国外接轨的评价制度。

五、建立科研人员的分类考核体系。对基础研究人员的评价以有利于潜心研究、长期积累、催生重大原创性成果和创新人才培养为评价重点，着重评价科学价值。对应用研究人员的评价以聚焦需求，具有自主知识产权和重大技术突破为重点，着重评价成果转化情况、技术成果的突破性和带动性。对从事技术转移、科研支撑和服务人员的评价以服务质量、促进科技成果资本化、产业化等实际效果为重点，激励相关人员不断提升自身服务能力。对软科学研究人员的评价以服务决策需求，形成战略研究报告，支撑思想库智囊团建设为评价重点。对艺术创作的研究人员，以作品影响力为评价重点，改革现有重理论轻创作的评价方式。

六、以科研分类评价机制完善高校创新队伍结构。鼓励各高校根据自身发展战略，建立相应的评价机制，创新知识服务能力建设，加强工程技术人员队伍、稳定的高水平实验人员队伍、专业化的技术转移人员队伍、高水平的决策咨询队伍和专业化管理队伍等建设，促进高校创新能力的提升。

七、以科研分类评价机制支撑和推动学校人事制度改革。鼓励各高校将科研分类评价机制与人事制度改革相结合，以岗位聘任和考核为依托，以科研分类评价机制为突破，建立涵盖创新性研究、技术转移、科技服务、科学普及、技术支撑和服务的科研分类岗位评价体系，形成“公平、公开”的竞争与合作机制，促进优秀科研人才的成长。

八、加强科研分类评价对资源分配的指导作用。鼓励和支持各高等学校结合自身发展的定位、

英国人文社会科学成果评价体系及其借鉴

杜向民 朱燕锦 刘兰剑

一、引言

英国的科研评价是大学评价中最核心、最重要的部分，其目的在于评价高校的学术水平及博士、硕士研究生的教育质量。英国把评价结果与项目资金拨款相联系，把高校科研评价等级作为拨放资金的依据，其整体特点可归纳为“基于信息辅助的同行评议方法的事后评价体系”。

当前，英国的科研评价体系正在经历从RAE（The Research Assessment Exercise，研究评价实践）到REF（The Research Excellence Framework，卓越研究框架）的转变，这个转变引起了世界各国的广泛关注。总体来看，以同行评议为核心形成的RAE基本评价框架在英国已经应用了多年，但由于同行评议存在的种种缺陷，英国下决心改制为以科学计量为核心的REF框架。

20世纪80年代初，撒切尔夫人执政期间，由英国高等教育基金委员会制定科研评价制度，跨学科、范围广、种类齐全的研究项目评价活动就已经开始实施。这是基于科研项目研究成果的“实用性”（成果在政府、民间机构、企业的使用情况）、“产出”（出版数量）、“质量”和“影响力”（文献被引用的数量）等项指标进行的评价。从1986年起，由北爱尔兰、英格兰、苏格兰高等教育部门制定的RAE模型开始实施。这个模型的目的是对科研成果水平进行同行评议，其特点是评审人数扩大、同行评议使用数据分析、对事前评价的重视等。

英国全国的科研评价活动每4年开展一次。迄

今为止，应用RAE体系已经开展了6次大学科研评价。付媛媛对英国大学人文社科评价的历史演化和改革特点进行了概括分析，认为其特点有：①英国科研评价是主要以学科类别为基础的同行评议；②英国大学人文社科科研评价程序越来越公开公正；③国家通过调整经费实现大学科研方向的宏观管理和促使科研成果为国家全面发展服务；④英国大学人文社科科研评价的评价机构为社会组织，很好地监管了高校与政府间的关系；⑤英国大学人文社科科研评价注重实质性评价，由之前重视评价结果转而改变为更加重视评价的过程，主要评价人文社科科研成果对英国社会效益的全面性影响。罗侃认为，科研评价对英国科研的影响越来越大，前两次的科研评价只包括拥有高等教育地位的机构，因此能够进行评价的科研成果数量很少。但在第三次高校评价改革后，很多之前的基础科学学院、专科学校和其他中等教育单位也升级到大学的地位，使得英国高等教育机构的数量剧增，因此科研成果评价的规模也变得更为庞大。1997年的评价范围得到很大的扩张，进入评价的人员达到55000人，共收到了近3000个学校的评价成果文件，参与机构多达191所大学单位。“在2001年的评价中，评价部门接收了2598份评价材料，进入评价的人员达到50000名，参与机构达到173个大学教育单位”。国外还有很多教育人士通过实践研究得出英国科学评价的影响。例如，威尔士、苏格兰的大学教育委员会作出的《科研质量的保持》的报告表明，从1986年设立英国科学评价制度以来，大学科研成果质量

高峰高原学科建设规划，建立符合学校特色发展的分类评价标准。市教委将根据分类评价结果，在高峰高原学科建设、协同创新中心建设、高校智库建设等财政性投入专项中，优化资源配置。

九、落实教育部高等学校科技评价改革试点方案，深化教育综合改革试点工作。市教委将在教育

部下发的高等学校科技评价改革试点方案中，制定区域高校科技评价改革试点方案，研究制定相应的推进举措。

（原文刊载于上海教育，<http://www.shmec.gov.cn/html/xxgk/201410/412082014012.php>，2014年10月9日）

有所提升，高等教育中的科研气氛有了明显的积极变化，各教育单位通过评价得到了关于大学科研成果的积极回馈。亨克尔指出RAE的设立，在促进英国经济社会发展中作出了很大的贡献，通过评价，英国教育部门不仅可以更好地进行全面的科研方向调控，也使高校拥有了一定的科研自由，给予高校适当范围内的自由权。

REF（卓越研究框架）的出台与近年来文献计量的发展与应用密不可分。同时，也由于RAE在6次评价中出现的问题引起了英国高等教育及科研管理部门的高度重视，一些缺陷导致RAE体系备受争议。例如，同行评议中的主观性，倾向于著名机构，过于关注传统研究领域，而对于交叉学科和综合性学科重视不够，轻视年轻人，无法评价人文社科价值显现的长期性等。为了获取RAE中的排名，类似足球界的“学术转会制度”已然形成。因此，英国不得不下决心转向以文献计量为基础的REF体系，这个以数据分析为基础的评价体系显然更加客观。

本研究以从RAE到REF的转变为主线，从评价制度与方法两方面对比分析英国人文社科科研在学术性、行政性、效益性三方面的评价思路及方法变迁。

二、英国人文社科学术性评价制度与方法

学术性评价是指科研领域的学术研究能力的提高，促进学术交流，规范学术研究，知识的学术发展意识的提升。因此，学术性评价就是对科研成果自身的质量评价。

（一）评价制度

从20世纪80年代开始，英国科研成果评价制度主要是基于RAE评价体系。RAE体系中，评价主要从三方面展开：研究质量、研究环境、研究声誉。评价要点如下：首先是科研成果质量，即在同类科学研究中寻找出优秀的科研成果，优秀的科研成果最能体现一个科研单位或高校的研究水平，即成本产出比。科研成果质量在评价中所占比重最高，可达60%以上。其次是科研的影响力，包括科研成果对建制单位作用、其社会效益和经济效益等。研究成果影响力的具体指标需要通过实践来证明。这个要点约占25%左右。最后是科研环境，科研环境包括对本科生以上学生的学术培训、营利性企业的科

研成果合作项目数量、政策环境（政策引导的紧缺科研方向）、成果运用的广度和深度等。这个要点约占15%的比例。这三个要点是评价科研成果的基本要素，以建制单位为单元把这些材料整合交给英国高等教育部门，由评价委员会进行评价。

英国科研成果学术性评价制度有如下几个特点：第一，公开性。首先，英国教育部对科研评价材料的评审过程必须公开，评价结果公示期在三天以上（有专利保护的秘密资料除外）。其次，英国教育部必须对评价的制度和具体指标进行讲解说明。第二，一致性。英国科研成果的评价制度和办法必须与科研机构保持时间和内容的一致。科研机构可以向英国教育部门咨询文件解释。第三，可信度。英国教育部从20世纪70年代进行科研学术性评价以来，科研评价材料的评价过程的透明化、材料申报的程序和写作格式得到了英国科研机构的认可，为了保持这种认可，英国教育部保持着制度的连续性，对评价制度和方法的改进只有经过多数人的认可和权衡利弊之后才能进行改动。第四，效率性。英国学术性科研评价的经费开支要用到实处，提高效率。第五，国际化。以国际化的标准及指标来制定英国人文社科研究成果学术性评价指标，研究成果以国际化的指标进行评价，评价中需要有一定数量的外国专家参与。英国人文社科研究成果2001年学术性评价中，评价委员的组成不仅包括英国本国的565位知名教授，还包括从国外请来的245位国际著名专家。第六，中立性。英国教育部主管人员对所有课题一视同仁，做同等的支持鼓励，不能对特定课题做特殊关怀，以此保持评价的公平公正。

（二）评价方法

英国科研评价结构一般分成两级。例如，2001年的评价中就分成了15个学科领域的主要评价组与67个具体学科评价组。两级评价小组相互协作。主要评价组的任务是审查具体的评价指标和制度，最终确定被测机构的总评等级，并且发挥联系各评价组的纽带作用。具体学科评价组的主要任务是具体设定评价指标并执行，主要职责就是具体实施评价并作出评价等级建议。英国2008年的评价中收到全国1000多个教育机构的6000多个提名，最后评选其中1000人组成专家小组进行评价，他们都是在学

术界享有较高威望并且科研成果显著的专家。

英国科研评价的方法与过程具有独特之处。例如，在2001年第五次科研评价中，每一位待评价的研究者需要提交3项以上的研究成果材料，成果形式可以为发表的期刊论文、出版的书籍、录音材料等，评价委员会注重的是研究成果的质量而非数量，所以对出版物和发表的期刊没有过多数量要求，材料提交期限为3个月，英国教育部提前1年公布对所提交材料的要求标准，然后评价委员会用5个月的时间对提交上来的材料进行匿名审阅，最后共同讨论得出一份评价结果。为体现公平，每个评价人员都须写一份关于评价结果的详细标准对照，这个标准就是之前已经公布的学术性评价制度（见表1）。英国学术性评价指标最后的评价结果为5级中的一级，然后根据评级发放奖励资金。一般来说，一个科研机构的科研人员和师资力量比例越大，则这个科研机构或学校的学术综合实力就越强。例如，沃里克大学（The University of Warwick）参加英国2001年第五次科研研究评价的科研人员比例达到90%以上，其中有50%的科研人员达到4级以上，反映了其强劲的综合实力。

表1 英国科研学术性评价指标

促进学术交流	文献的引用率
	科研成果在政策的运用情况 社会经济的传播运用情况
提高学术交流水平	学术质量提高
规范学术研究	评价材料的明确清楚
	标准一致性
	内容层次性
促进知识自觉	知识的继承性
	知识的连接性

在REF中，其中一项评价内容是科研成果质量。由于人文社科研究成果和自然科学存在较大不同，REF体系指出，在使用文献计量方法的前提下，对人文社科研究成果采用和自然科学研究成果不一样的评价方法，自然科学主要以定量评价方法为主，而人文社科科研成果评价主要采取定性评价为主、定量评价为辅的评价方法。

与RAE相似，定性评价主要采取同行评议的方式进行，由专家小组对各单位进行评价。评价委员会下设四个评价学科组指导工作，根据学科不同分为36个评审单位，人文社科主要被分到第三和第四个学科组，即哲学社会科学和人文艺术科学小组，

然后由这两个学科组的相关领域专家进行评审。评价委员会主要由国内外享有较高声誉的教授参加，虽然委员会教授自己也可以制定一些指标来定量评价，但是委员会的评价意见起主要作用。英国REF科研成果的学术性评价方法的评价最终意见划分为5个质量星级（见表2）。例如，对一篇已发表的人文社科类论文进行评价，期刊的影响因子和文献引用率是定量指标，但这并不是主要的决定性因素，这样可以避免通过不正当手段提高文献引用率的现象。不过，如果论文能在影响因子较高的期刊上发表，则论文评分一般也较高。所以REF虽然主张使用定性和定量相结合的方法，但是由于人文社科的复杂性和长期性等特点，对于学术性评价方面还是以定性评价为主，文献计量方法为辅。

表2 英国科研学术性研究成果质量等级描述

四星级	质量达到世界领先的原创性，重要性和严谨性标准
三星级	质量达到国际上优秀的原创性，重要性和严谨性标准，但它达不到优秀的最高标准
二星级	质量达到在原创性，重要性和严谨性方面国际公认的标准
一星级	质量达到原创性，重要性和严谨性方面公认的全国性标准
未分类	质量落在国家认可工作的标准以下，不符合的研究本评价报告的目的

三、英国人文社科行政性评价制度与方法

以学术建制的发展为目标的学术建制单位，对本单位人员和单位的学术贡献的评价就是行政性评价。这里主要指其学术成果对本单位发展的影响，而不是学术成果对学术自身发展的影响。

（一）评价制度

建制性评价制度是行政性评价的主要特性，评价由单位组织，鼓励研究人员做出优秀成果，为本单位作贡献。英国的建制性评价主要是指建制单位质量控制制度的设立及实施。科研评价制度中荣誉性认可可是行政性评价的又一侧重点，荣誉性认可不仅是工具性展示，也是对科研人员优秀成果贡献的充分肯定，其主要内容是建制单位对科研人员的科研成果对社会经济做出的贡献所给予的荣誉和承认，英国荣誉性承认包括精神奖励和物质奖励。科研行政性评价服务于管理，其表现为对学术研究方向、研究人员的绩效和经费

拨付的管理。英国科研评价工作主要由国家教育部门管理，政府通过资金的拨付引导学术宏观研究方向，然后以课题制为根据，以科研人员的科研成果绩效来拨付科研资金。

建立绩效荣誉制度是为了鼓励做出高质量、高水平的科研成果，促进经济社会发展。绩效奖励的公平性和奖惩的力度直接关系到科研人员的积极性。英国2014即将实施的REF体系中人文社科行政性评价制度主要以延续性为主，建立的行政性评价制度有以下特点：第一，确立具有可操作性和有效性的指标体系；第二，实行科研人员责任制，对碌碌无为的科研人员降低职位职称；第三，设立阶段性考核，考核优秀的给予荣誉称号和一定科研经费奖励，对考核连续不合格的人员降职辞退。只有这样才能提高广大科研人员的科研积极性，提高他们的科研绩效。例如，英国伯明翰大学对本校科研人员做如下等级要求：第一级，好评：一个研究周期内完成申报书成果的90%以上。第二级，一般：完成项目申报书的70%以上。第三级，较差，研究成果没有达到预期目标的60%，需要大幅改进。如果连续三个研究检查周期没有达到二级以上，则降低职位并取消奖金，达到第一级的会有绩效奖金。自2000年实施这个制度以来，伯明翰大学一共提交了59项优秀科研项目，跨50个评价学科，全部达到英国教育部划定的三级以上等级项目，学校总的研究项目经费翻了一番。

（二）评价方法

“高等教育技术科学部”是英国的国家科研监督机构。这个部门通过高等教育经费发放部对科研成果进行评价，以评价结果为基础，按照评价等级给科研机构拨付资金，以此对全国的科研方向进行宏观调控，进而为国家的科研进步提供方向。科研成果的行政性评价结果与建制单位科研经费的拨付挂钩，并且这种相关性还在深入的变化发展之中。20世纪90年代第一次科研评价的结果决定了高等教育拨付部门50%以上的科研经费拨付方向。在第二次科研评价中，若项目成果评价等级为第五级，即最后一个等级，那么该项目就不能得到财政资助。进入20世纪90年代以后，修改为处于评价等级第一级和第二级的项目便得不到财政拨款（此时评价等级共有7个）。到第七次（REF）科研评价改革后

则加大了难度，将财政拨款的分水岭提升到第四级。目前，英国行政性评价方法主要是设立社会科研监督机构，借助社会力量对科研成果进行质量评价和控制，并设立奖项鼓励科研人员多出高水平成果。实施方法主要是在一定的时间段内对研究成果按照本单位的评价指标做出评判。

在REF制度中，科研成果的行政性评价方法程序大致相似，继承与发展相结合，其将行政性评价纳入了研究环境这个大因素中，扩展了行政性评价的外延，更好地推动了建制单位多出优秀研究成果。评价内容具体如下：管理方面，经费拨付数额、学术研究方向、科研人员绩效奖惩；建制单位方面，科研工资、科研人员数目、科研人员的聘用。其中，建制单位的学术研究方向和评价结果影响着经费拨付，这部分的评价中评价方法与经费拨付相结合。因为REF人文社科评价操作程序与之前相似，现以2008年英国全国科研评价的操作方法为例：第一步，设定评价分类单元（主要以人文社科下划专业为分类单元，REF将人文社科分为哲学社会科学和人文艺术科学两个大类）、设立评价人员小组（评价人员小组要与人文社科下划专业分类单元领域相一致）；第二步，申报科研成果材料（包括科研人员资料、科研成果知识产权证明）；第三步，评价人员小组评审（主要是根据科研人员提供的科研成果材料和评审小组的专业评价得出科研成果水平的总体评价）。最后评价结果是剑桥大学排名第一，牛津大学排名第二。行政性评价中的服务于管理原则主要是指以科研人员的业绩来决定经费的投入。人文社科科研水平主要资金拨付方法经过若干步骤：第一个是明确人文社科各系的分配金额总数，主要计算办法是将每个类别的科研人员总数和科研成本按照权重指标和系数进行计算得到（见表3和表4）；第二步再把计算好的科研经费拨付到各个高等教育机构。在资金的拨付中，高校得到的经费跟科研成果质量等级呈正相关。而且据表4所示，若是科研成果质量评价的等级为1、2、3级，则经费的系数权重为0，代表此项目不能拥有资金。从表中还可知道五级比四级的科研经费拨款高3倍。由此可见，英国科研评价的财政服务管理与科研成果评定等级相关性很高。

表3 英国科研单位拨款的权重指标及系数

分类	类别	权重系数
权重指标1: 科研成本	哲学社会科学	1.3
	人文艺术科学	1
权重指标2: 科研人员	研究活动积极的研究人员	1
	研究副手	0.67
	普通研究人员	0.6

表4 英国科研评价结果与拨款的权重系数

科研水平评价等级	研究质量拨付模型中的拨付权重系数
3A, 3B, 2, 1	0
4	1
5	3.18
5	4.036

四、英国人文社科效益性评价制度与方法

(一) 评价制度

人文社科的效益性主要是指经济效益和社会效益，目标是将现实效益和潜在效益相比较，以此来度量人文社科成果的价值。

英国对RAE评价体系一直存在争议。争议的核心是绩效评价指标和方式，由于评价结果和经费拨付密切相关，以及这个评价制度自身存在一些缺陷，招致有关教育部门的反对，因此新建立了REF评价制度体系。该评价制度主要针对社会经济效益制度做出了改变。REF的一个最显著变化就是把科研的效益影响分离出来作为科研评价的一个重要要素来对待，包括人文社科成果对建制单位的作用、社会效益和经济效益的评价。研究影响力评价权重占25%左右。具体制度如下：①用引用率等定量指标在一定程度上取代定性分析以评价影响力。②分为5个级别，最末一个级别效益低下，将没有科研经费资助。③科研影响评价的比重增加，占25%以上。效益评价是评价指标尽可能量化的社会、经济效益评价指标体系，并根据最终评价指标之间的计算结果为基础做出判定。

(二) 评价方法

英国对效益性评价采用的评价方式与学术性评价的定性评价方式不同，主要采用定量分析方法。社会经济效益主要指科研成果影响水平和教育质量研究。由于科研成果具有主体性、价值判断和其他特征，科研成果评价变得非常复杂，多层次的评价指标和评价结果使科学研究中的困难变大。

新的REF体系着重强调了绩效在科研评价的重

要性：“高等教育研究部门既要关注科研成果对建制单位的效益贡献，又要关注科研成果对社会经济的运用效益。”体系还对科研效率的指标做了详细的解释。因为REF是新的评价体系，对人文社科评价有部分改进，特别是效益性这一部分，英国在2014年正式实施之前以人文社科类的社会工作、公共管理、公共政策、语言文学这几个学科为试点做了尝试。英国人文社科效益性评价方法的测量方式为：①效用，有以下测量指标：毕业学生和科研人员从社会外部机构得到的私营收入、科研成果的专利数量和国内外教育机构颁发的证明；②影响力，用文献被引率表示；③报酬，主要指科研成果转化的经济效益年收入；④质量水平，按照英国高等教育的指示，科研成果的效益性由建制单位科研人员数量、科研经费资助和经济报酬比等方面构成。

五、英国科研评价制度体系与评价方法对我国的借鉴意义

英国已经形成了具有本国特色的、比较完备的人文社会科学研究成果评价体系，对于激励研究人员的积极性和创造性，提高研究效率，促进高水平成果的产出发挥了重要作用。这对我国人文社会科学成果评价具有一定的借鉴意义。

(一) 以科研成果的质量为核心对人文社科科研人员进行考核

目前，我国在人文社科科研人员聘任的过程中过分注重科研成果的数量，在人员晋升的指标中有很多数量上的硬性规定。如，论文发表的篇数，科研课题的数量、出版书的数量和获得奖状和证书的数量等。虽然在注重数量的同时也看重发表的期刊的等级，如C刊。但是期刊等级高并不一定代表论文质量就高，用于核心期刊评定的引用率和影响因子受到恶意引用以及所研究的课题属社会热点性等因素的影响。并且人文社科课题评价一般都有一年期的限制或中期检查，规定论文数量，导致科研项目组必须多出成果、快出成果。这样的硬性规定难以让科研人员真正静心仔细研究问题，发表的成果也难以真正体现科研人员的水平。教授、副教授等职称考核晋升中与科研数量直接挂钩，甚至导致有些科研人员为出版更多作品，得到更多资金而抄袭他人学术成果。

在英国2014新体系REF对人文社科的评价中，虽然科研的成果形式也可以为发表的期刊论文、出版的

书籍、录音材料等，但是评价委员会注重的是研究成果的质量而非数量。不在数量上做过多的要求，每4年评价一次，只要提交不超过3项最具代表性的研究成果就可以了，这样把短期目标变为长期目标，具有长效性，更适合对人文社科科研人员的考核评价。因此，需要综合完善科研人员考核晋升的依据，改变其考核方式，既要注重科研成果发表数量，又要注重科研质量，如增加H指数、A+指数等指标综合计算科研人员的贡献，改善我国以“简单计数”方法作为人文社科科研评价基本依据的现状。

（二）改革事前评价为事后评价

当前，我国人文社科科研项目一般采用课题制的形式，先由科研人员提交研究申报书，然后按照科研项目的级别（国家级、部级、省级、校级）来拨付资金。这样在一定程度上变成了吃大锅饭，对科研质量高的人文社科科研项目造成实质上的不公平，而且级别高的人文社科课题并不代表科研质量就高，以行政级别来拨付经费造成了课题之间位格的不平等。

英国人文社科科研人员研究成果的评价特点是评价结果与经费拨款相联系，这既使资源进行了有效配置，又促使科研人员多出优秀成果。科研评价结果被定为五级，科研成果评价将科研质量水平与经费正相关，并通过公开透明的专业评价进行，这种评价方式很大程度上促进了英国科研成果的水平和经济社会的发展。我国课题制的拨款方式实际上也可以借鉴英国的方式，将评价指标量化，资金与科研水平挂钩。并且，英国实行科研人员责任制，对碌碌无为的科研人员实行降低职位职称处理。同时，设立阶段性考核措施，加强对科研过程的管理，对考核优秀的给予荣誉称号和一定科研经费奖励，对连续考核不合格的人员降职辞退。这种重视质量并将质量等级与科研经费挂钩的科研成果考核评价方式，使考核评价更具有激励效能，能够促进科研人员多出高质量的成果。

（三）评价方式定性定量相结合

我国在人文社科成果评价时在定性与定量之间的平衡上一直难以把握。我国的科研成果评价主要是“简单化的定量评价”，远没有达到科学计量评价的水平。特别是20世纪90年代中期之后建立“简单量化”的指标评价体系，主要看核心期刊、国家期刊、三大引文索引的引用率，这样的评价方

式具有一定的主观性，不符合人文社科研究成果的特点，特别是在评价人文社科研究成果的学术性方面，并不完全适用，不具有长效性。

英国人文社科成果评价方法是定性分析与定量分析相结合，这在英国新的科研评价制度REF中较为明显，其学术性评价方面主要是采用定性评价为主，文献计量为辅的评价方式，一般采用同行专家学者评议的方式进行。同行评议专家主要是由国内外享有较高声誉的教授担任，虽然委员会教授也可以制定一些指标来定量评价，但是委员会的评价意见起主要作用。而在效益性评价中主要是定量分析方法为主，评价指标为引用率、建制单位科研人员数量、科研经费资助、经济报酬比和产出量与引文量的比例等方面构成。我国可以借鉴英国的这种人文社科评价方式，尊重人文社科研究成果的特点，反对一刀切的评价方式，在学术质量方面主要采用定性评价为主、定量评价为辅的方式，而在研究成果效益性评价方面采用定量评价方法，科学计量其投入产出比，评价其经济社会价值。

（四）人文社科科研成果评价也要重视效益性

目前，我国人文社科科研成果评价主要内容是课题最后发表的论文数量、出版书数量、在国内外的获奖情况、引用率等，这表明我国目前对人文社科科研成果评价主要停留在对科研成果发表的数量重视上，很多科研机构甚至对期刊的影响因子和论文被引情况等基本的量化指标都没有纳入考评体系。仅仅简单的“计数”，只重视投入而不重视产出，更没有结合人文社科科研成果的经济和社会效益进行评价，难以真正实现从整体上衡量研究成果的价值等级，这样就造成了目前我国人文社科研究成果与经济社会发展实际需要相脱节。而在英国的REF体系当中，针对评价只注重论文发表期刊影响力的情况做出改进，特别增加了科研成果的经济和社会效益的评价，占25%的效益性评价是其三大评价要素之一。

（杜向民，长安大学党委书记，中国人文社会科学评价中心主任，政治与行政学院教授，陕西西安 710064；朱燕锦，长安大学政治与行政学院硕士研究生，陕西西安 710064；刘兰剑，长安大学中国人文社会科学评价中心副主任，政治与行政学院副院长、教授，陕西西安 710064）

（原文刊载于《中国高教研究》2014年第10期）

全球科研评价体系的演进与发展

——以美国、英国、澳大利亚、荷兰四国为例

蒋玉梅

在知识经济时代，全球众多发达国家都旨在通过科研评价来促进本国科研实力的提升。虽然各国采取的科研评价形式不同，但这些评价都服务于一个共同的目标，即促进科研产出和质量的提升，实现国家科研实力和创新力的可持续增长。在国家层面上，科研评价可以实现科研资源的合理配置，促进重点研究领域的优先发展，催化科研成果的社会应用；在机构层面上，科研评价直接或间接地影响财政拨款、大学声誉、生源、毕业生就业，从而调控科研机构的发展方向和发展策略；在个人层面上，科研评价有利于激励科研工作者和管理者的工作热情，提升科研工作绩效。

在世界各国中，美国、英国、澳大利亚、荷兰的科研评价体系不仅历史悠久，而且各具特色。通过对这四个国家科研评价体系的比较研究，挖掘全球科研评价的演进逻辑与发展趋势，可以为我国科研评价体系的建立提供有益的经验借鉴。

一、科研评价体系的国际比较

1. 美国科研评价体系。美国国家科学研究委员会（The National Research Council，简称NRC）作为由美国国家科学院创建的民间非营利组织，在近20年里对美国高层次的科研活动进行了多次评价，并于1982、1995、2010年发布了三次美国研究型博士项目排名。其中1982和1995年的两次排名完全基于同行评价，NRC在每个学术领域中邀请一些系主任和高级学者对该学术领域的博士项目进行打分，从0（不具备博士授予资格）到5分（卓越）分为6个级别，专家评分的平均值就是各个参评单位的最终分数。然而，完全依赖同行评价的评估方式受到了外界对于评价体系科学性和公正性的质疑，鉴于此，NRC在2001-2003年期间

开发了全新的评估方法。新方法最大的特色在于其基于数据（Data-based）的特点，即通过调查和回归两种不同取径的方法来确定排名。调查法通过自下而上、显性化的途径，让教师直接在问卷中对博士项目的多个特征进行评分；回归法则是一种自上而下、推断性的评价方法，先邀请每个学科的40名教师对各博士项目进行整体评分，然后通过回归模型寻找这些评分与项目特征之间的联系，即通过回归的技术手段来推断项目各个特征的优劣。

NRC在2008-2009年对美国博士项目进行了新一轮评估，并于2010年发布了美国研究型博士项目的最新排名。评估包含三个维度：科研活动（Research Activity）、学生支持与成果（Student Support and Outcomes）、教育环境的多样性（Diversity of the Educational Environment）。其中科研活动包括教师在出版、引用、科研经费和获奖等方面的情况，出版和引用的文献计量指标来自于SCI数据库；学生支持和成果包括第一年获得全额资助的学生比例、规定时间内获得学位的学生比例、学生获得学位的时间、为学生设置的学术职位情况、毕业生就业信息的提供情况；教育环境的多样性则包括少数民族师生比例、女性师生比例、国际学生比例。最新排名的多维度结果呈现方式令人耳目一新，新排名包括调查排名、回归排名两个总体排名和科研排名、学生排名、多样性排名三个分项排名，每个项目的排名基于不同的评价者样本计算500次，取90%的置信水平。

然而2010年排名发布之后，外界评价却褒贬不一。2010年排名以数据统计为主、辅以同行评价的评估方法无疑比单纯的专家评价更科学，多维度的排名也给不同的利益关系者提供了更全面

的数据，但来自各方的质疑也不在少数。例如：有学者质疑仅从教师群体获取数据的可信性，问卷评价指标设计和指标权重分配被认为是评估委员会的一家之言，跨学科研究被认为在这种学科本位的评估中没有得到公正的评价等等。另外，NRC公布的排名区间被指责跨度过大，如密西根大学的传播学专业的回归排名在2-58之间，调查排名却在7-22之间，这令使用者很难判断项目的真实水平。

2. 英国科研评价体系。英国的科研评价由政府主导，是全球历史最悠久和最成熟的科研评价体系之一。在1986-2008年间，英国高等教育拨款委员会共开展了6次RAE (The Research Assessment Exercise) 评估。RAE排名历经20多年，延续了以学科为单位的同行评价方式。以2008年评估为例，以学科为基础设立了67个评价单元，双层结构的评价专家组包括15个主专家组和67个分专家组（每个评价单元设立一个分专家组），总共包括1000名专家。评估内容主要围绕研究成果、研究环境、声誉指标这三个方面展开，其中研究成果的评价主要基于科研人员的代表性出版物，权重至少达到50%。

RAE排名呈现出显著的精英管理特点，并取得了明显的效果：一流院校和优势学科获得了更多的科研资源，其科研质量和科研管理水平都实现了明显的提高；科研人员都力争在高质量的期刊上发表论文，科研人员的流动也更为活跃。但是，该评价也受到了评价成本过高、对跨学科评价不利以及专家构成不公等种种诟病。

为了削减开支、提高效率、提升科研的社会影响力，英国政府在2008年开发了REF (Research Excellence Framework) 以取代原来的RAE评价体系。新的评价体系有了几个重大的变化。第一，评价单元和专家大大减少。评价单元由2008年的67个减少到36个，主专家组由15个减少到4个，分专家组由67个减少到36个。为了减少由于大幅度削减评价单元带来的诸如学科分类粗放等负面影响，除专家组成员外还将聘请编外评审员来参加评估，尤其是在跨学科研究和科研的实际应用价值方面。第二，评价标准和评价内容有了较大的变化。评价内容维度由原来的4个变为3个，分别为

“成果 (Output)” “影响力 (Impact)” 和 “环境 (Environment)”，每个维度的权重都予以明确的规定。成果被赋予65%的权重，以“原创性、意义和严密性”作为其评价标准。每位科研人员被要求提供4件代表性科研成果，既可以是著作、论文、工作报告、会议论文，也可以是教材、设计与展品、媒体作品等。成果评价主要采用专家评价，同时参考 Scopus 引用数据库的文献计量数据。REF 声称考虑到文献引用率的诸多缺陷，如：论文发表的时间长短、学科背景差异、负引用 (Negative Citation)、出版语言差异等，文献计量数据仅作参考。新增的影响力维度则既备受瞩目而又饱受争议。影响力从“广度 (Reach)” 和 “意义 (Significance)” 两个方面进行衡量，强调科研在文化、经济、环境、教育、政策制定、公共服务等各个方面造成的影响。增加该维度的出发点原本是试图促进科研成果的社会应用，但在实际操作过程中遭到了重应用研究轻理论研究、重“硬”科学轻“软”科学的诟病，因而其权重由原先的25%下调到目前的20%，被评单位被要求提交一份影响力表格和案例分析报告，专家组主要根据这两份材料来评估其影响力。环境则被赋予15%的权重，以“活力 (Vitality)” 和 “可持续性 (Sustainability)” 为衡量标准，主要包括科研策略、人员（教师发展和研究生培养）、科研收入及基础设施和设备、科研合作。

3. 澳大利亚科研评价体系。澳大利亚的科研评价由政府主导，与科研经费资助挂钩。澳大利亚研究委员会 (Australian Research Council, 简称为ARC) 开展的科研评价经历了从“综合指数” (Composite Index) 到“科研质量框架” (Research Quality Framework, 简称RQF) 再到“澳大利亚科研卓越” (Excellence in Research for Australia, 简称ERA) 的转变。

始于1995年的综合指数评价以大学为独立单元进行评价，评价内容主要包括大学所获得的经费及科研产出（出版物数量及授予的硕士和博士学位数量等）。这种评价体系表面上产生的收效非常明显，大学越来越重视在Web of Science数据库收录期刊中发表论文，其科研产出量在1992-1996年间每年至少提升了8%，发表在SCI上的论文每年增

长2%。但是，澳大利亚的科研影响力却在下降，在1988到1993年间，澳大利亚文献引用率在OECD国家中从第6位下滑到第11位，学者们虽然发表的论文数量上升，但是其平均引用影响（Citation Impact）却在下降。

针对这种情况，澳大利亚政府于2006年开始筹备新的科研评价体系，即RQF。在大量收集各大学、专家及商界和教育界人士建议的基础上，RQF的评价单位由大学改为研究组，包括13个学科领域的专家组，每个组有12个成员，专家组成员包括至少3个外国专家和3个终端用户。RQF还未及实施，新一届政府于2008年宣布以ERA来取代RQF。

ERA结合文献计量数据以及专家评审内容，以学科为单位对科研质量进行评估。ERA已于2010年和2012年进行了两次评估，共41所科研机构参评。评价内容包括四个方面：（1）科研质量（Research Quality），包括出版物及会议的级别、引用分析、ERA同行评议、经同行评议的国内和国际科研收入；（2）科研数量及活动（Research Volume and Activity），包括科研产出总量、科研收入等指标；（3）科研应用（Research Application），包括科研商业收入等应用测量指标；（4）声誉（Recognition），包括在权威出版单位担任编辑的人数、参加知名学术团体的人数和全国科研奖金获得者人数等声誉测量指标。澳大利亚研究委员会根据评估结果，对具备国际竞争力的院校和学科进行重点资助。

ERA作为新生事物也遭到了诸多质疑，其中争议最多的当属ERA于2011年公布的期刊评级，把学术期刊划分为A*、A、B和C四个等级。期刊评级一公布即遭到了各方的激烈批评，并迫于压力于当年5月宣布废除。另一较大争议是科研经费到底该归入科研投入还是成果。目前ERA把科研经费数额作为科研成果指标，很多专家质疑该做法的科学性，认为很难把科研经费数额清楚地归入投入或成果，而建议用科研输入/输出率来取代。

4. 荷兰科研评价体系。荷兰科研评价体系的产生源自于“条件性资助（Conditional Funding）”体系的终结。荷兰从1983年实施“条件性资助”，对教学和科研进行资助。随着1993年该资助体

系的消亡，荷兰政府要求建立教学和科研两套评价体系。因此，荷兰大学协会（Association of Universities in the Netherlands，简称VSNU）在1993年到2003年期间开展了“科研质量评价（Quality Assessment of Research）”项目，由外部同行委员会对科研质量的四个方面进行评估：生产率、成果质量、適切性和长期生存能力。

自2003年后，荷兰大学的科研评价开始采用标准评价协议（The Standard Evaluation Protocol，简称SEP）框架，该协议由荷兰皇家艺术和科学院（KNAW）、荷兰大学协会（VSNU）、荷兰科学研究组织（NOW）联合制定。SEP评价包括外部评价和自我评价，其中外部评价每6年举行一次，由外部独立专家对研究所（research institute）进行评估，从两个层次展开：机构（院校、系科或研究学院）层次、科研小组或项目层次。SEP评价树立了两大目标：提高科研质量和代表资助部门、政府及社会进行科研工作检查。其评价的内容和形式都较为丰富，既有回顾评价，也有前景评价；既有对科研成果的评价，也有对研究前景、科研管理、科研政策的评价；既有中期审查，也有终期评价；既有常规的材料评审，也有现场审查和负责人访谈。评价包括四个维度：质量（Quality）、产出率（Productivity）、社会相关性（Societal Relevance）、活力与可行性（Vitality and Feasibility）。其中质量包括5项内容：研究质量及科学相关性、领导地位、学术声誉、人力资源和财力资源、博士研究生培养情况；产出率包含科研产出策略和科研成果，研究成果的评价基于3-5篇代表作及出版物的总数，必要时使用文献计量指标；社会相关性包括社会价值、社会影响及社会应用性；活力和可行性包括SWOT（长处、短处、机遇、威胁）分析、策略、稳健性3项内容。SEP评估以同行评价和自我评价为主，其评估结果和政府拨款之间没有直接的联系，但教育部会基于该评价结果对科研业绩较差的单位进行警告，督促其整改。通过对美、英、澳、荷四个国家科研评价体系演进逻辑和发展趋势的剖析，我们可以把这些国家最近的科研评价体系特点整理归纳如下（见表1）。

表1 美、英、澳、荷四国最新科研评价体系的特点

国家	评价体系	外部评价组织	评价目的	评价维度	评价方法	评价结果
美国	NRC	美国国家科学研究委员会	帮助大学确定发展方向及策略	科研活动、学生支持与成果、教育环境的多样性	回归分析同行评价计量指标	与拨款无关
英国	REF	英国高等教育拨款委员会	分配科研经费	成果、影响力、环境	同行评价计量指标	与拨款挂钩
澳大利亚	ERA	澳大利亚研究委员会	基于大学绩效给予资助	科研质量、科研数量及活动、科研应用、声誉指标	计量指标同行评价	与拨款挂钩
荷兰	SEP	荷兰皇家艺术和科学院、荷兰大学协会等	提高科研质量，检查和问责	质量、产出率、社会相关性、活力与可行性	同行评价自我评价计量指标	与拨款无关

二、全球科研评价体系的演进特点

遵循美、英、澳、荷四国科研评价体系的发展脉络，可以看出这四种科研评价体系在评价目的、评价内容和评价方法上各有特色，但多样性的表象之下隐藏着共同的演进逻辑和发展趋势。主要表现为如下几点：

1. 评价体系的精细化。评价体系的精细化主要体现在评价内容的多维化、评价结果呈现的动态化和评价方法的多元化三个方面。首先，评估内容从以往的单一维度朝着多重维度的方向发展，除了对绩效（产出和影响）进行评价外，对赋能（投入和过程）的评价也日益重视。在绩效方面，除了科研产出之外，科研应用和影响力在科研评价中的地位正在提高；在赋能方面，科研环境和科研策略等投入和过程指标也得到了更多的重视。其次，评价结果呈现出更大的多样性，除了传统的单维度、静态的排名或分级方式，还出现了多维度、动态的评分方式。再次，评价方法从以往的单一手段转向同行评价和计量数据相结合的多元评价手段。可以说，世界各国都在不遗余力地开发更科学的评价体系，改进评价方法。比如：美英评价体系经历了以同行评价为主到同行评价与计量指标相结合的过程，澳大利亚则经历了以简单的计量指标为主转向以专家评价结合计量指标的过程，荷兰的科研评价虽然目前仍以同行评价和自我评价为主，但正越来越多地把计量指标纳入评价过程。

2. 评价成本效益的最大化。鉴于科技竞争力的战略地位，各国对巨额的科研评价耗费并不吝惜，但却加强了对科研评价成本效益（评估成本与效益之比）的调控。早期基于规模的评价模式虽然成本较低，但是由于无法有效提升高校和科研人员的科研积极性，因而其产出的科研效益较低。当前各国普遍采用基于绩效的评价方式，从方案设计、资料提交、数据收集到专家评审等各个过程都需要耗费大量的费用，尽管如此，各国政府鉴于科技生产力的战略地位，并没有单纯削减科研评估投入，而是利用各种方式来优化评估方法，减少不必要的耗费，同时以更高的科研产出来提升科研评价的效益。比如：美国、澳大利亚和荷兰近年来正在加大投入，研制开发更为复杂的科研评价体系，而英国的科研评价体系虽然多年来一直受到耗费过高的批评，但英国政府并没有对评价投入进行实质性的削减，而仅在评价方法上做了一些微调，如减少评价单元、增加计量指标等。

3. 政府调控的明朗化。无论是在官方还是非官方组织的科研评价中，政府都扮演着至关重要的角色。在知识经济时代，各国政府深刻认识到科研实力在社会发展和国际竞争中的关键作用，通过多种显性或隐性的方式调控科研发展的方向和策略，而科研评价就是其中的一个重要手段。从各国科研评价的演进史来看，从评价目标和内容的选取到评价指标和评价单元的设计都反映出政府宏观引导的意

图。各国的科研评价体系无一例外地都把提高科研质量作为其显性目标，但在其背后均隐藏着政府对“科研卓越”的价值预设及路径规划，这种隐性预设和规划通过科研评价维度和指标的设计被反映出来。如：美国NRC把学生支持与成果、教育环境的多样性作为评价维度，折射出美国政府一贯实施的大量吸引国际优秀人才的战略；英国REF新增加的“影响力”维度反映出英国力图引导科研与社会应用更好结合的发展策略；澳大利亚和荷兰把“社会应用”作为重要维度，反映出两国对于科研社会应用的高度重视；美国虽未把“社会应用”作为评价维度，但其科研成果评价的文献指标数据主要来自于SCI数据库，同样反映出其对具有更大直接经济效益的自然和工程科学的青睐。

三、对我国建立科学科研评价体系的启示

对我国来说，尽快设计全国性的科研评价体系已成当务之急。在此过程中，国外科研评价体系的演进模式和发展经验可以为我国建立符合中国国情和国际发展趋势的科研评价体系提供参考与借鉴。

首先，确立科学的科研评价目标。科研评价应为以下三个目标服务：促进科研实力的提升、调控科研发展方向和策略、优化科研资源的配置。当前我国的科研评价在政策、理论和实践上都处于缺位的状态，不仅缺乏全国性的评价指导和调控，缺乏本土化的科研评价理论研究，也缺乏在学科基础上进行系统评价的实践。各个院校在科研评价上各自为政，呈现出无序性、盲目性。一些国家级的科研评估活动，如博硕士点的评估和国家自然科学基金项目的评审，都存在着量化指标和结果指标起决定性作用的情况。这些不科学的情况使我们在科研工作中出现了很多短视和功利性的行为，不仅不利于科学

研究的健康发展，也耗费了大量的科研资源。科学科研评价体系的建立将有利于从国家发展战略的高度来引导和调控科研发展的方向。一方面，通过设定科研评价的周期减少急功近利的科研“投机”行为。开展科研需要稳定的工作环境，适当延长科研评价周期有利于激励科研工作者投身于具有重大社会意义的中长期研究；另一方面，可以通过评价维度和评价指标的设计来引导科研朝着数量和质量双重增长、基础和应用研究比翼齐飞的良好态势健康发展。

其次，应构建多维的科研评价内容。除了科研成果这一维度之外，还应考虑把科研环境、科研应用等维度考虑进来，从而实现科研生态的整体优化，使科研环境-科研产出-科研应用形成一个良性循环过程。在设计评价维度和指标时，应把以下几组相互联系和补充的因素考虑在内：主观与客观，投入与产出、过程与结果。通过对科研活动进行完整、科学的评价监控，有效刺激科研水平的良性发展。

再次，使用多元的评价方法。目前各国最常使用的是同行评价和计量指标。同行评价虽具有权威性和专业性，但专家的主观判断容易使科研评价偏离科学性和公正性；计量指标具有客观性和科学性，成本低、耗时短，但往往不够全面和权威。因此，单一评价方法往往会带来信度和效度缺陷。我国在构建科研评价体系的时候，应综合定性和定量的评价方法，并且针对自然工程科学和人文社会科学建立多样化的评价方法体系。

（蒋玉梅，南京大学大学外语部副教授，江苏南京 210093）

（原文刊载于《国家教育行政学院学报》2013年第9期）

效率和生产率方法在高校科研评价中的应用

胡咏梅 段鹏阳 梁文艳

一、引言

作为第一生产力——科技和第一资源——人才的重要结合点，高校在国家科技发展和创新中具有越来越重要的地位并发挥着日益突出的作用。为提高高校科研水平，中国政府通过合并高校优化资源配置、增加高校科研经费投入、实施“985”和“211”工程计划等各种途径，推动高校科研能力实现跨越式发展。根据《中国科技统计数据（2011）》，高校利用占全国11.3%的科研人员，8.5%的R&D经费，承担了62.6%的基础研究活动，产生了21.6%的发明专利申请，获得了28.8%的发明专利授权，生产了64.6%的科研论文。

尽管中国高校科研水平和创新能力已有了明显增强，但应该清楚地认识到，较之世界一流大学，中国高校的科研实力仍然存在较大的差距。为了切实提高高校科研实力，在稳步增加高校科研投入的同时，还必须制定与完善科研管理体制，保证科研投入的有效配置与使用。评价是决策的依据，因此，运用科学有效的评价方法，建立健全科研评价体系，是完善高校科研管理体制的重要任务之一，同时也是增强中国高校科研实力的重要基础。从现有的高校科研评价来看，由于高校科研评价目标定位问题，现有评价方法和手段大都偏向于对产出的评价，没有重视科研活动的效率和效益问题，一定程度造成了高校科研人员的急功近利和学术浮躁现象，学科重复设置、科研资源浪费、理论性基础性研究重视不足等现象也日渐严重。

增加高校科研投入对于促进中国高校科研水平的提高非常重要，但能否充分利用科研投入、优化配置资源，对于提高科研效率和生产率同样重要。要从根本上保证高校科研生产可持续发展，必须解决现有科研评价中单一强调科研产出的问题，引入科研投入维度，引导高校科研活动走向内涵式发展道路。在科研评价中引入投入的维度，事实上是在科研评价中引入投入-产出概念，建立生产率 and

产效率分析模型促进高校加强科研生产成本核算、提高资金使用效率和投资效益，同时改善科研管理制度，最终提升高校科研生产率，促使高校科研生产持续稳定发展。

本文结构安排如下：第二部分梳理高校科研评价方法的现状和局限，第三部分介绍效率和生产率分析方法和理论，第四部分综述效率和生产率方法在高校科研评价中的实际应用，第五部分探讨高校科研效率和生产率方法在科研评估上的局限性以及值得进一步深入和扩展的空间，为后续研究提供方法论和研究设计方面的参考。

二、现有高校科研评价方法概述

正如前文所述，科研评价是基于科学、公正和可行的评价手段和方法，对机构或个人科研活动及其相关环境进行综合分析和判断，旨在提高科研管理和科研决策水平。有效的科研评价有助于推动科研活动持续健康发展，促进科研资源优化配置。从发达国家和地区的经验来看，美英等国都非常重视高校科研评价，并将评价结果与科研拨款挂钩，在科研管理体制中引入激励和约束机制，促进高校科研水平和创新能力的提升。

从国内学术研究和科研管理实践经验来看，随着国家和社会对高校科研水平重视程度以及加强公共问责的呼声不断增强，国内围绕科研活动展开的评价日益增多。概括起来，目前国内高校科研评价的主体主要可分为三类，即行政管理部门(官方)、民间组织、学术科研机构/个人；评价方法同样可分为三类，即定性评价方法、定量评价方法、综合评价方法。本部分将就国内相关评价实践中的方法及其存在的局限进行概括和总结。

(一) 定性评价方法

定性评价主要是一种以人的主观经验、价值观、思维方式等为依据的判断活动，将评价对象作概念、程度上的质的规定，然后进行分析评定以说明评价对象的性质或程度。定性评价方法主要有同

行评议法和德尔菲法。

同行评议法是指同一领域或与其邻近领域的专家采用同一种评价标准和程序对该领域的科学研究活动及其相关要素（如研究人员、研究机构、研究项目等）进行评价的一种方法。可以说，同行评议是某一或若干领域的专家采用统一评价标准，通过通信评议、会议评议、调查评议、组合评议等多种形式对相关领域的某一思想进行评价的活动，因此它是以专家判断为主的方法，其评价结果对有关部门的决策有重要的参考价值。德尔菲法的提出是为了克服一般专家讨论中存在的屈从于权威或盲目服从多数的缺陷，采用匿名发表意见的方式，专家之间不得互相讨论，不发生横向联系，通过多轮次调查专家的看法，经过反复征询、反馈、修改和归纳，最后汇总作为专家调查的结果。

目前，定性评价方法主要用于对科研项目、科学出版物、科研成果、学位与职称、科研机构的运作等的评价。定性评价方法的优点和缺点都非常明显。其优点在于：第一，启用了学术研究界那些可接受的、有专长的、受人尊敬的成员担任科研活动及其结果的鉴定人，这有利于质量控制和肩负科学责任；第二，允许本学科、专业的同事交流想法、意见、建议和反馈信息，这种交流促进了科学进步，并保持了科学交流渠道的畅通。其缺点主要在于：由于缺乏客观标准，对评价者要求非常高，由于熟人关系网、马太效应、崇尚权威等，可能会制约评价的公平、公正、合理；而且，当研究成果是应用性而不是理论性，或当研究是交叉学科或者新兴学科时，定性评价的适用范围同样将受到限制。

（二）定量评价方法

定量评价方法是依据具体的、客观的数据展开评价，它将评价对象进行数量化的分析和计算，从而判断其价值。定量评价方法能够消除主观因素在评价中的不利影响，使评价具有可比、可信的数据依据。同时，定量评价可以使一些概念精确化，加强评价的区分度，降低评价的主观性和模糊性，增加评价的说服力。从目前国内外比较常用的定量评价方法来看，可以进一步分为单指标评价方法和多指标综合评价方法。

单指标评价是指选择某一个维度展开评价。以科研论文产出数量的评价维度为例，评价者可以

利用SCI、SSCI、EI、CSSCI等论文数据库，评价特定时期科研论文发表数量；以科研论文产出质量的评价维度为例，评价者同样可以利用SCI、SSCI、EI、CSSCI等论文数据库，从特定时期科研论文引用频次、论文半衰期等文献计量学手段获得数据；最后利用统计学、计量经济学等方法进一步展开分析评价。多指标综合评价方法是指根据综合评价的目标，运用统计学模型，或者将定性研究方法与定量研究方法相结合，对评价对象的多个维度进行整体评价。纯量化的综合评价方法主要通过因子分析、典型相关分析等多元统计分析方法将单指标评价的各个维度整合成一个综合指标，以此进行排序和评价；将定性方法与定量方法相结合的综合评价方法是根据综合评价的目标，对客观事物的影响因素进行分解，以构造不同层次的评价指标体系，根据专家同行评议结果对这些指标进行赋值并确定其权重系数，最后采用层次分析方法等综合评价模型进行综合得到最终评价价值，以此进行排序和评价。

（三）高校科研评价实践中的问题

我们在对相关评价手段及其应用进行分析后发现，在高校科研评价的实践中，评价的方法和手段日益演变和完善，在实践中根据具体情况得到了许多改进。不过，高校科研评价体系的改革与完善，还应努力适应不同类型、不同层次科学研究活动的内在发展规律。特别地，由于科研活动不同于纯粹的机械运动和自然现象，它是由具有博弈行为的人所参与的特殊复杂系统。因此有效的科研管理与科研评价，往往不是“数量化”“一刀切”“投票法”等简单做法所能完成的。例如，高校在科研成果评价方面不能“一刀切”，必须区别对待。对于那些应用性研究，必须要将创造效益作为科研考评的重要指标，重在对科研成果创造经济效益以及对社会发展所做贡献进行评估；然而，对于基础性研究活动，要给予科研人员充分的研究时间，按照其项目周期来进行科研成果的考评，并且要将这些项目成果在整个科研成果评价体系中的权重突出，进而鼓励高校科研人员重视基础性理论性研究，不断提升自主创新能力。对于高校教师的科研考评不应全部量化，往往重大攻关项目或创新团队项目都是一个科研群体共同完成的，如果用“记分”方式评估每个科研人员的成果，由于难以公正地剥离

出每一项成果的个体贡献，不利于形成合作创新的良好学术氛围。利用高校内部学术委员会的“投票法”来进行科研奖励或学术职称的评选活动，投票人的“利己”动机和“中间偏好峰”的存在将会导致最优者反而被淘汰。

此外，本文发现，已有研究在讨论科研投入与产出的关系时，常将两者作为高校科研实力的并列指标，而忽略投入与产出的比例关系，即效率的问题；同时，对科研投入与产出的评价侧重不同，或仅侧重于对产出的关注，或虽两者皆有关注，但指标内容的选择随意性较大，缺乏理论依据，影响评价结果的解释，也难以获得共识性的结论。因此，需要从经济学投入产出的视角，引入效率与生产率的概念，对高校科研生产进行更加全面客观的评价。

三、效率和生产率评价：概念、方法及进展

（一）效率与生产率的概念

生产率是指生产单元所生产的产出与所需投入的比值，即“生产率=产出/投入”。生产率提高对于经济增长至关重要，在实际生产过程中，由于存在多种投入要素和多种产出要素，生产率分为全要素生产率（total factor productivity, TFP）和部分要素生产率（partial measure of productivity, 又称偏要素生产率）。全要素生产率是一种包括所有生产要素的生产率测量，定义为加总的生产产出与加总的投入要素之间的比率；部分要素生产率则是考察产出与所需某一类投入要素的比值，例如劳动生产率等。在实际研究中，如没有特殊说明，生产率测算均指全要素生产率，而生产率提高主要来源于效率提高、规模变化和技术进步。

效率是生产率提高的重要源泉。效率反映了生产单元在既定产出下使投入成本最小化的能力。1957年法雷尔（Farrell, M.J.）将效率分解为技术效率（technology efficiency, TE）和配置效率（allocation efficiency, AE）。效率则是技术效率与配置效率的乘积。技术效率指在既定投入集下获得最大产出的能力，衡量投入到物质生产过程中的资源能否得到充分有效的利用。配置效率指生产单元在给定要素价格和生产技术下以最优比例利用投入要素的能力，它涉及各种生产要素如何替代配置才符合经济原则。此后，法尔（Faere.R.）等人在

1998年提出了规模效率（scale efficiency, SE），用于测量生产率提高的另一重要源泉——规模变化，即生产单元的效率可以通过改变它们的运营规模而提高，即保持同样的投入组合但改变运作规模而获得改进。由此可见，分析生产单元的效率一般分解为技术效率、配置效率以及规模效率三个部分，它们共同构成了生产率提升的源泉。

（二）效率和生产率分析方法的进展

自从法雷尔提出分段线性包络的生产前沿估计方法发展至今，较为成熟的效率和生产率测度方法主要有四种：最小二乘法（LS）、随机前沿分析方法（Stochastic Frontier Analysis, SFA）、数据包络分析方法（Data Envelopment Analysis, DEA）、全要素生产率指数方法（Total Factor Productivity Index）。

计量经济生产模型方法是基于生产函数或成本函数来表述经济生产关系，它把产出表述为一些投入的函数（生产函数模型），或者将成本表述为投入价格与产出的函数（成本函数模型）。与随机前沿分析方法不同的是，计量经济生产模型方法只能获得平均生产函数，不能得到生产前沿；两者相同的地方是都需要设定函数形式。常见的函数形式有线性函数、柯布-道格拉斯生产函数、超越对数函数、二次函数等等。计量经济生产模型常用的参数估计方法有最小二乘法（LS）、最大似然估计（ML）以及贝叶斯估计（Bayes Estimation），其中LS和ML方法应用更为广泛。

指数是测算不同时期经济变量变化最常用的工具，其在生产率测量中最重要的作用就是对全要素生产率变化的测量，最常用的即为Malmquist TFP指数。

数据包络分析（以下简称DEA）是一种非参数方法，是在法雷尔已有研究的基础上，以单输入单输出的效率概念为基础发展起来的，评估具有多输入多输出同类型决策单元（Decision Making Units, DMU）相对有效性的效率评估方法，最早由查尔斯（Charnes）、库伯（Cooper）和罗兹（Rhodes）于1978年提出。随后，许多学者致力于DEA模型的研究，先后给出了规模收益不变的CRS模型、规模收益可变的VRS模型，以及规模收益非递增的NIRS模型和规模收益非递减的NDRS模型。DEA方法目前

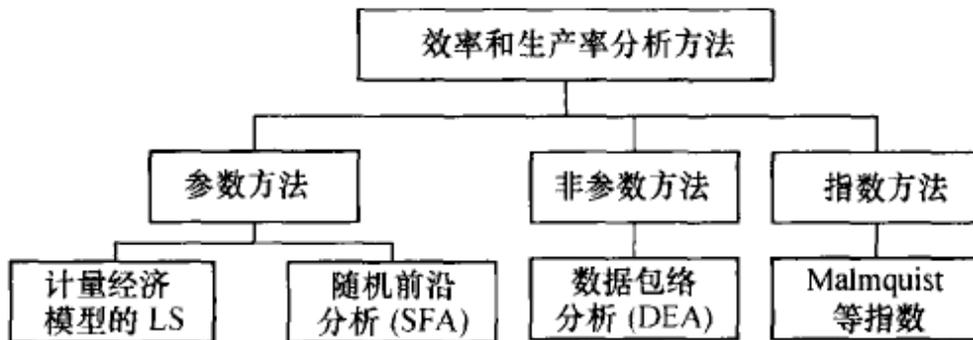


图1 效率和生产率分析方法

已经发展成为较成熟的非参数效率的方法，并被广泛地应用于管理科学领域。

随机前沿分析方法（又称随机边界分析，以下简称SFA）使用包含随机误差的经济计量方法估计生产前沿函数，从而计算效率与生产率。艾格勒（Aigner）、洛佛尔（Lovell）和施密特（Schmidt）以及穆森（Meeusen）、冯登·布劳克（van den Broeck）等学者分别独立地提出了适合的横截面数据的随机前沿生产函数模型，模型由生产函数和两个误差项组成，其中一个为零均值的随机误差项解释统计噪音，另一个为非负均值的随机误差项衡量技术无效性。随后，一些学者运用此模型进行了实证研究，并对该模型的无效率误差项分布假定做了修改，分别提出了截断正态分布（truncated normal distribution）和伽玛分布（Gamma distribution）模型。巴蒂斯（Battese）和科埃利（Coelli）提出了适合面板数据的随机前沿生产函数模型，该模型允许技术效率水平随着时间推移而发生系统变化，被称为时变无效性模型。

SFA在测量技术效率方面日益得到学者们的认可，但随机扰动与技术效率的不可观测和严格的分布假设却成为应用障碍。随着皮特（Pitt）和李（Lee）将横截面数据的最大似然估计技术推广到面板数据，巴蒂斯和科埃利运用似然比检验，能够对面板数据下的最大似然估计中的“随机前沿函数与传统生产函数的选取，随机误差和技术效率的分布假设是否稳健，技术效率是否受其他因素影响而变动”进行甄别和判断，使SFA具有更广泛的实证和应用空间。

DEA和SFA都属于前沿面分析方法，而且可

以基于这两种方法来估计全要素生产率的变化，并将其测算结果分解为技术效率变化、规模效率变化以及技术变化。法尔（Faere）、格罗斯克夫（Grosskopf）和张（Zhang）运用相对于DEA前沿面的距离来测算Malmquist全要素生产率变化指数。弗恩泰斯（Fuentes）、格里费尔-塔特杰（Grifell-Tatje）和佩雷尔曼（Perelman）、欧瑞尔（Orea）提出基于超越对数形式的随机前沿生产函数来测算Malmquist全要素生产率变化指数。

相对于DEA方法而言，SFA的优势主要体现在两点：其一，可以解释模型中的噪音项；其二，可以用来研究传统的假设检验。而其缺点则表现为：一方面，对于模型中的无效率项必须设定其分布形式；另一方面，对效率的测算要事先设定生产函数（或成本函数）的函数形式。此外，在结果的稳定性方面，SFA计算结果较为稳定，不易受异常点的影响，更适合大样本计算；而DEA则对样本的同质性要求较高，易受异常值的影响。DEA的一个优点是能直接处理多产出的情况，而SFA处理多产出则较为复杂，需将多产出合并成一个综合产出或者利用距离函数解决。

四、效率和生产率方法在高校科研评估中的应用

（一）国外高校科研效率和生产率测量的实证研究

近三十年来，越来越多国外学者尝试利用SFA、DEA方法评估高校科研效率和生产率。国外关于高校科研效率和生产率测量的研究主要以美国、英国和澳大利亚为代表。不过至今专门对高校科研效率和生产率进行评估的研究仍然较少，采用

SFA方法评估高校科研效率的研究更为少见。伊扎蒂 (Izadi) 等人的研究, 以及豪讷 (Horne) 和胡 (Hu) 的研究采用SFA方法分别估计了英国99所高校和澳大利亚33所大学的技术效率和成本效率。斯蒂文斯 (Stevens) 同样采用SFA方法评估了英国和威尔士80所高校1995~1996学年至1998~1999学年的效率, 并分析了教师和学生特征对效率的影响。卡姆克斯 (Kempkes) 和泊耳 (Pohl) 以及达巴史岩 (Daghabshyan) 采用SFA方法考察了德国和瑞典高校效率的变化, 并分析了效率的影响因素。

采用DEA方法评估高校科研效率和生产率的研究日渐增多。琼斯 (Johnes) 等采用不同指标体系, 研究了英国大学经济学系的科研效率, 结果表明DEA结果对投入或产出指标组合敏感程度较小。巴斯里 (Beasley) 分别对英国高校52所院系 (物理系和化学系) 的教学效率和科研效率进行了评估。吉奥范尼 (Giovanni) 等人采用DEA方法与文献计量学相结合的方法测度了意大利不同学科高校的科研效率, 结果表明不同学科高校之间的科研效率具有显著差异。琼斯 (Johnes) 与李雨采用DEA方法评估了中国109所高校2003年和2004年的科研生产效率, 研究表明中国高校的平均科研效率超过0.9, 通过DEA模型得到的高校两年的科研效率排名之间高度显著相关; 综合性大学的科研效率均值高于专业性大学, 沿海地区的高校科研效率均值高于

西部地区高校; 综合性大学的科研效率高这一结果给中国近年来大规模的高校合并活动提供了有利的支撑。

相对于科研效率的静态评估而言, 科研生产率的动态评估研究较为少见。琼斯采用DEA方法和距离函数方法评估了英国112所高校在1996~1997学年到2004~2005学年之间的Malmquist生产率指数。结果表明英国高校效率平均每年增加1个百分点。进一步研究发现, 高校每年有6个百分点的技术进步, 伴随着5个百分点技术效率的下降。高校部门的快速变革对高校的生产技术有着正向的影响, 但却是以技术效率的降低为代价。

(二) 国内高校科研效率和生产率测量的实证研究

随着以DEA为代表的效率测评方法传入我国并迅速普及, 国内许多学者开始采用该方法对高校科研效率和生产率进行评估。

陆根书等人在2006年前后发表了采用DEA方法评价高校科研效率的研究, 比较了不同地区教育部直属高校的科研效率, 结果表明东、中、西部地区高校科研效率依次递减, 不同地区高校科研效率和规模效率都存在波动, 变化趋势各不相同; 扩大科研规模是高校改进科研效率的主要途径。田东平、苗玉凤采用DEA方法测度我国510所高校的科研效率, 结果依旧表明高校平均技术效率呈东、

表1 国外采用DEA评估高校科研效率及生产率的主要文献

作者	研究方法	研究对象	国家	投入变量	产出变量	备注
Johnes G. & Johnes J.	DEA	36所大学的经济学系	英国	教学与科研折合人/月数; 教学与科研以及纯粹科研人员折合人月数	学术论文数; 经济学核心期刊论文数等	采用不同指标体系, DEA结果对投入或产出指标组合敏感程度较小。
Beasley	DEA	52个院系 (物理系和化学系)	英国	一般性经费支出中科研所占比例; 物质设施设备支出中科研所占比例; 科研收入	本科生数; 教学型研究生数; 研究型研究生数; 科研排名前列/中上游/中游/下游的院系科研收入	分别评估了教学与科研方面的效率。
Giovanni et al.	DEA及文献计量学	不同学科高校的科研效率	意大利	正教授数量; 副教授数量; 科研人员数量; 科研资助金额	出版物数量; 对出版物的贡献量; 科研实力	不同学科高校科研效率差异显著。
Athanassopoulos, Shale	DEA	45所大学	英国	一般学术性支出; 科研收入	毕业生数; 高等学位授予数; 加权后的科研产出	比较了两组不同投入产出指标的DEA结果。
Madden et al.	DEA	24所大学的经济系	澳大利亚	教学与科研人员数	本科毕业生数; 研究生毕业生熟数; 核心期刊论文数; 其他期刊论文数; 专著数; 其他著作数	比较了新财政资金投入前后的效率。
Avkiran	DEA	36所大学	澳大利亚	折合全时教师数; 折合全时非教师数	本科生在校生数; 研究生在校生数; 科研产出量	比较了三种不同指标组合的分析结果, 提出了应用DEA方法的建议。

中、西部地区依次递减的趋势。徐娟、李清彬和任子雄等学者以省为单位,采用DEA方法考察了我国高校科研效率,研究认为高校科研效率与经济发展水平有一定关系,但并不是强正相关,说明高校科研效率并非单纯受区域经济的影响。韩海彬、李全生以不同学科特点为分类依据,测度了我国H所高校的人文社会科学科研效率,结果表明,11所高校中有5所高校的人文社会科学科研的技术效率和规模效率运营情况都相对较好,其余6所高校的人文社会科学科研效率为非DEA有效。研究首次从学科视角评估高校人文社会科学领域的科研效率,具有一定的启发意义。胡庆江等人采用DEA方法考察了“985”工程高校的科研效率,结果表明,大部分

“985”工程院校科研效率偏低。

近些年来,一些学者开始采用基于DEA的Malmquist生产率指数方法考察高校科研效率的动态变化趋势。胡咏梅、梁文艳以2000年合并的25所高校为样本,考察了其1999~2002年间的科研生产率变化状况,结果发现技术进步是高校科研效率提高的主要来源,而合并高校的规模效应并不显著。骆卉慧从省级层面考察了我国29个省区高校科技系统的效率和生产率,研究发现,横向而言,中国各地区高校科技系统效率呈东、中、西部依次递减;纵向来看,各地区高校科技系统资源配置效率总体呈下降趋势。

从上述实证研究可以发现,首先,现有国内高

表2 国内关于高校科研效率评估的研究

作者	主要内容	样本及数据	方法	投入变量	产出变量	主要结论
陆根书,刘蕾	不同地区教育部直属高校科研效率	54所教育部直属高校,2000~2002年	DEA	高校科研人员;高校科研辅助人员;财政性科研收入;其他科研收入	R&D课题数;R&D成果应用和科技服务课题数;专著;国外学术刊物发表论文数;国内学术刊物发表论文数;成果获奖数;鉴定成果数;专利授权数;技术转让当年实际收入	46.3%的高校科研效率有待进一步提高;东中西部高校科研效率、纯技术效率依次递减;不同地区高校科研效率和规模效率都存在波动,变化趋势不同。扩大科研规模是高校改进科研效率的主要途径。
田东平,苗玉凤	高校科研效率	510所高校,2001~2003年	DEA	科技活动人员;当年科研支出经费	专著数;国外论文数;其他全国性刊物论文数;鉴定成果数;技术转让收入;专利出售收入	平均技术效率呈东中西依次递减趋势;重点高校平均技术效率高于全国平均水平10%。
徐娟	各省高校科研效率	31个省(市)的高校,2006年	DEA	教学与科研人员数;研究与发展人员数;当年科技经费内部支出;当年课题支出经费	专著数;论文数量;当年技术转让实际收入;国家级成果数	有12省市科研效率整体有效;这些省市大部分为中西部地区经济不发达省份,说明高校科研效率并非单纯受区域经济影响。
李清彬,任子雄	2002~2006年省际间高校科研效率	31个省(市)的高校,2002~2006年	DEA	教育经费;科研经费;高等学校科技人力(教学人员与科研人员总和)	专利授权数;高等学校技术转让成交额;出版专著(部);发表论文数;专利出售合同金额;科技成果国家级奖项	DEA非有效率地区占一半以上;高校科研处于规模报酬递增的地区占很大比例;经济水平与科研效率有一定关系,但不是强正相关。
胡庆江,何玮佳,柳锐	“985工程”院校科研效率	36所高校4个年份数据	DEA	科技活动人员数;拨入经费	专著;学术论文数;鉴定成果数;技术转让收入	大部分“985工程”院校科研效率偏低。
作者	主要内容	样本及数据	方法	投入变量	产出变量	主要结论
骆卉慧	高校科技系统效率	29个省区高校科技系统,2000~2004年	DEA及Malmquist指数	研究与发展全时人员;科学家和工程师人数;当年科技经费中的内部支出	直接成果(专著、论文、专利授权);成果获奖数(国家级);间接成果(科技成果转让当年实际收入)	横向来看,中国各地区大部分高校科技系统处于DEA非有效,呈东中西部依次递减趋势;纵向来看,各地区高校科技系统资源配置效率总体呈下降趋势。
胡咏梅,梁文艳	科研生产率动态变化	2000年合并的25所高校,1999~2002年	Malmquist指数	科研人员总数;科技人员中高级职称比例;科研人员人均科研经费投入;人均课题数;单个课题平均经费	科研人员人均出版专著数;人均发表国内外论文数;技术转让合同科研人员人均收入;人均获奖数;获得国家级奖励数	技术进步是整体科研效率提高的主要来源;合并高校的规模效应不显著。

校科研效率和生产率研究的对象可以归为三类：对教育部属高校科研效率的评估；对不同省区高校科研体系投入产出效率的评估；对单个学术部门或大学院系科研效率的评估。其次，就研究采用的投入产出指标体系而言，不同学者根据不同的研究偏好和数据可得性采用不同指标体系，结果也因指标体系不同而有较大差异。再者，从评估结果来看，我国高校科研效率整体状况不容乐观，呈现东、中、西效率依次递减的区域性差异；在规模效率层面，西部地区高校多处于规模递增阶段；而地区经济发展水平与高校科研效率并没有显著相关关系。

诚然采用DEA、SFA以及计量经济模型等量化方法评估高校科研生产效率和生产率已有颇多成果，在评估的精确性、评估范围的广泛性、评估的公正性等方面具有非量化评估不可替代的优势，但也要注意这些量化评估方法的局限性。第一，尚不存在没有限定条件和缺点的评估效率和生产率的量化方法。正如前文所述，无论是参数方法还是非参数的测度效率和生产率的方法，均有各自的使用条件和局限性。比如，DEA方法虽然不需要考虑投入、产出指标的相对重要性，但该方法对样本的同质性要求较高，且结果易受异常值的干扰；SFA相对于DEA，尽管在结果稳定性方面较好，但在处理多产出的生产效率问题上不具优势。第二，采用不同量化方法得到的效率排名并不总是一致的，而且没有成熟的方法或者标准在DEA和SFA两种方法之间进行选择。第三，DEA、SFA方法对投入、产出数据质量要求高。其相对效率值易受指标测量误差的影响。如果忽视了某些数据的误差，可能会导致某种异常或者反直觉的结果。而且，DEA方法不允许各指标存在缺失值。第四，DEA、SFA以及计量经济模型方法首先是在企业生产效率评估中的应用，效率反映的是生产单元在既定产出下使投入成本最小化的能力。高校科研生产组织并没有追求成本最小化的动力，多数科研组织追求的是产出质量，即产出高水平的科研成果，经济生产领域的效率评估方法是否适用于高校科研生产尚有待理论支持。

五、结论与启示

科研创新在推动当今社会进步、保持经济社会稳定健康发展方面发挥着越来越重要的作用。作

为科研活动和创新的主战场，世界各国越来越重视增加高校科研经费投入，以提高高校科研实力。政府、社会和高校自身对高校科研质量、科研绩效的关注程度日益提高，甚至将评价结果与拨款紧密挂钩。在诸多的量化评价方法中，效率与生产率方法是经济学“投入-产出”研究在高校科研评估中的应用。从以上对效率与生产率方法、理论及其实践应用的总结，本文得到以下几点结论，希望能为进一步完善科研管理体制提供有益的启发。

第一，需要丰富高校科研效率与生产率评估研究，科学对待和使用其研究结论。现有的研究聚焦于对高校科研生产技术效率、规模效率的考察，鲜有对科研生产成本效率、配置效率的考察。若能对同一研究对象进行多视角的分析，可以获得更多更有意义的结论研究。多数研究采用DEA方法考察高校科研效率和生产率状况，单一方法的使用，无法考量结论的效标关联效度，难以保证结论的有效性。与此同时，由于DEA方法属于非参数方法，对其结果的好坏并不能用通常的统计学指标加以检验，并需配套使用bootstrap、jackknife等手段，保证结果的稳定性与有效性。此外，研究者和决策者要正确看待效率评估结论。由于DEA、SFA等方法给出的均是相对效率值，对于样本的依赖性很大，而且指标数量对于效率结果也有直接影响，因而不能因为某个范围的样本效率评估情况良好，就将之推广至其他群体，应当谨慎对待和使用研究结论。

第二，需要综合使用文献计量方法、质性评估方法，并基于科研活动真实规律与生产理论构建科学的评估指标体系。从国内已有研究和实践来看，科研效率评价的投入、产出指标体系的建构通常缺乏理论基础，指标选择多限于可得性和借鉴同类研究，理论基础的欠缺使得研究结论难以获得共识，也不利于对评价结果的解释。更进一步，现有的研究在考虑高校科研产出时，评价指标体系中对于产出质量的考察大都利用文献计量法，即用发表在国外SCI或SSCI杂志上的文章数、发表期刊影响因子、文章引用率衡量被评对象科研质量。用这类指标衡量科研产出质量存在一定的局限，特别是高校整体评价中，由于学科之间在论文发表和期刊等方面的异质性存在，将这些指标纳入计量模型或者DEA、SFA模型会带来严重的估计偏误。因此，

在未来的研究和实践中，需要基于科研活动的理论与实践设计科学的评估指标体系；同时，必须同时强调量化指标与质性评价的综合使用，事实上，在国外教育研究中将量化与质性研究相结合的混合研究方法(Mixed Methods Research)已逐渐成为主流方法，被称为“第三种研究范式”。

第三，需要保证评估样本的可比性与同质性，更有利于评价结果用于科研管理。国内研究受数据获得限制，已有研究的样本大多选取教育部直属高校或重点高校，对其他类型高校科研效率的评估和考察较少，而且缺乏针对不同类型高校科研效率的分类评估，与DEA方法要求样本具有同质性相抵触，也不利于为高校的分类管理提供实质性的改善科研资源配置的建议。从国外的实践来看，效率与生产率的方法已经逐渐深入到以学院、科研团队、学科为单元，一方面，这解决了样本同质性的问题，保证了评价单元在科研生产活动在逻辑和特征上的可比性；另一方面，深入到学科、科研团队的评价，并将评价结果用于科研绩效管理中，这有助于激励科研团队内部合作、积极整合科研队伍、形成集群效应，有助于形成高水平的科研群体和产出重大科技成果。

第四，需要关注到科研效率变化的趋势，趋势的变化更具决策参考价值。关于高校科研效率和生产率的研究多以横截面数据分析为主，基于纵列数据(longitudinal data)或面板数据(panel data)考察高校科研效率和生产率的动态变化趋势的研究仍较为少见，然而，趋势性研究更具政策和实践价值。

第五，评价实践中没有重视对效率影响因素的分析，这不利于科研管理体制的改革与完善。现有研究多局限于对效率状况的评估，对影响效率和生产率的因素鲜有考察，因而在改进高校科研效率和

生产率的政策性建议方面缺乏实证依据。以参数和非参数方法为基础的量化方法是高校科研效率和生产率研究的主要手段，但仍需要借助质性研究方法对技术有效、规模有效、配置有效以及技术无效、规模无效、配置无效的部分高校进行典型个案比较研究，深入访谈高校科研管理者、教师和科研辅助人员，对高校科研评估的内、外部政策进行文本分析，剖析高校科研资源配置过程的微观机制，探究高校人事管理制度、科研管理制度、人才培养制度、资产管理制度、经费预算和决算制度等对高校科研生产过程的影响，寻求对高校科研生产效率和生产率的整体性、解释性的理解，并为高校管理者优化内部科研资源配置和提升科研生产率提供建设性意见。

综上所述，在高校科研管理中引入效率与生产率思想非常重要，但当前国内已有研究无论是在评估指标的选取、方法技术的选择，还是对时间趋势的分析上均尚有许多值得进一步深入和扩展的空间，有待后续研究在这些方面有所突破和贡献。此外，需要关注高校科研生产资源配置过程的微观机制研究，比如科研人员人力资本结构(人力资源配置)、教师教学与科研工作时间配置(时间资源配置)、教学与科研经费配置(财力资源配置)、教师个体行为选择与科研激励制度(制度资源配置)等等，通过对高校内部科研资源配置过程的细致考察，才可能揭示高校科研效率和生产率的生成机制。

(胡咏梅，北京师范大学教育经济研究所教授，北京100875；段鹏阳，北京师范大学教育经济研究所博士研究生，北京100875；梁文艳，北京师范大学教育经济研究所讲师，北京100875)

(原文刊载于《北京大学教育评论》2012年第3期)

关于完善我国高校科研质量评价的思考

——基于世界发达国家的政策经验

陈慧颖 刘在洲

2013年12月，教育部出台《关于深化高等学校科技评价改革的意见》，明确指出“科技评价不仅是高校科技管理的主要手段，而且是高等教育宏观管理和资源配置的重要依据，对高等学校办学行为具有突出的导向作用。”完善高校科研质量评价体系，做到高水准、客观评价，激发教师科研的积极性，实现科研质与量的结合，是我国高等教育提高质量的重要内容。世界主要发达国家在高校科研质量评价方面有许多经验值得我们吸取和借鉴。

一、坚持评价主体的多元化，突出民间中介评价组织作用

坚持评价主体的多元化，就是要把政府评价、社会评价和高校自主评价等评价主体结合起来，共同评价高校科研质量。评价主体的多元化可以使高校科研质量的评价更加公平、合理。许多发达国家为了提高监督效率，减少徇私舞弊、暗箱操作的可能性，都先后建立了民间中介评价组织对高校科研质量进行评价。如美国科研评价机构分为3个层次，联邦政府科技评价机构、州政府科技评价机构和研究院所及大学科技评价机构。为了评价的客观性，美国成立了美国民间科技评价机构，以美国国家科学研究委员会（The National Research Council，简称NRC）最为典型，它作为由美国国家科学院创建的民间非营利组织，是政府拿钱，委托科研机构代理评价的格局。美国联邦机构依据相关法律履行评价职能，政府委托一大批高水平、相对稳定的社会咨询机构，包括企业和营利机构开展评价。政府只是出资，具体执行就在其他部门和机构。这种出资人和执行人分离的制度保证了评价的公平性和合理性，出资人和评价执行者之间建立相互监督的机制。英国的科研评价由政府主导，是全球历史最悠久和最成熟的科研评价体系之一。为了提供更加客观的评价结果，英国政府不直接进行大学科研评

价，而是建立或委托中介机构即委托大学科研评价机构——英国高等教育基金委员会负责实施。英国科技评价呈现社会化的趋势，政府已经开始减少对科技评价的干涉，把一些具体的活动交给立场中立的中介机构。二战后德国科研机构的主导模式是以著名的马夫学会、弗劳恩霍夫学会、亥姆霍兹联合会、莱布尼茨科学联合会等为代表的非营利科研机构模式，它们具有公共组织的性质，但并不隶属于联邦政府的任何部门，在法律上是以协会注册的独立社团法人，通过与政府签订合同来确定双方之间的权利义务。我国目前高等教育的评价基本是政府行政主管部门的行为，采取的是自上而下的行政检查，评价手段通常是通过上级压力来推动科研质量的提高，缺乏对科研质量内在的自我约束机制和社会力量及专门的评价机构以及专业评价人员的第三方监督。我国行政评价有利于高效组织评价，推动评价工作开展，形成了我国评价工作独有的模式。但是在评价的过程中，政府是评价主体，所进行的评价带有强烈的行政行为性质和较强的权威性评估，不利于评估的公平公正性，不利于打破我国科研质量评价的窘境，也不利于高校活力的增强。因此我国政府应尽快鼓励建立一些颇具权威的大学评价中介机构，委托这些中介机构对大学的科研等进行定期评价。

二、注重评价成员的广泛性

评价人员来自不同的国家、不同的机构，有来自教育界、企业、政府和工商界等人士，从而可从多角度、全方位地进行评价。如英国在2001年“科研评估机制”（RAE2001）的评价专家中，除了465名来自英国本国各大学的教授、学者外，还有345名的海外顾问，其中包括哈佛大学、斯坦福大学、普林斯顿大学和麻省理工学院等世界一流大学的教授，还包括瑞典、瑞士、法国、澳大利亚和

新西兰等国的大学教授；日本在对社会关注的研究课题中，评价人员中还吸收反映公共意见的人士参加；荷兰为了保证公正性，委员会成员主要由外国专家担任；德国马普学会的科学咨询委员会成员主要来自高校和科研机构的高层科学家，其中超过97%来自马普学会外部，超过75%来自国外。我国的评价人员基本上都是政府派来的人员，其他领域的人员和外国的评价人员基本上没有，这不利于我国高校科研质量评价的全面性、公正性。

三、坚持评价标准的发展性

坚持评价标准的发展性，通过重点拟定某些能够促进高校科研质量发展的评价指标以及加大某些具有促进高校科研质量发展的指标权重，促使高校科研质量不断发展。发达国家开展科研评估的指导思想非常明确，即通过发挥评价的导向作用，促进大学、院系、研究所及科研人员的科研质量的提高，产生更多世界一流的科研成果。他们根据各自对评价标准的重视程度制定了发展性的评价标准。

第一，有些国家在建立评价指标体系时，对成果质量指标特别重视。如英国RAE不要求科研成果的数量，只要求科研人员提供4份有代表性的科研成果，注重考察成果的质量而不是数量，既能够反映出研究者的水平又能够减轻评估团的工作负担；荷兰除了要求科研人员提供出版物列表，还要求提供5份关键出版物及其质量和声誉的指标。我国高校科研成果评价长期以来存在重数量轻质量、重形式轻实质、重短期轻长远等问题，这种现行的科研评价使有些学者放弃了科学研究的本义，把学术的内在精神求索转变成乏味的简单操作，致使科学研究片面追求数量和速度，缺乏创新思维低水平重复，导致我国科研质量很低。2011年11月，教育部颁布了《关于进一步改进高等学校哲学社会科学研究的意见》，确立了“质量第一的评价导向”，一切都应该以提高科研成果的质量来进行，正确运用评价导向，回归科研本身价值，正确处理“量”与“质”的关系，建立以“质”为主，兼顾数量的合理的科研评价制度，加强高校科研工作的薄弱环节，推动学校的科研发展。明确科研成果可以供全球共享的思想理念，争取我国的科研成果在国际上处于领先地位，提高国家综合实力。

第二，还有些国家很重视论文和科研项目的

原创性和质量。正如日本的研究大学强化促进事业委员会在2013年明确指出论文指标中要求创造国际性的高质量论文，英国在其高等学校科研水平评价指标中也将科研项目的原创性、严密性作为评价标准，在英国高等学校科研水平评价的等级和标准中也以“原创性、精确性、重大意义”作为评价的标准（见表1）。我国的科研质量评价指标主要包括：科研项目、科研成果、学术交流、成果奖励和成果转化。其中科研项目包括项目申报、立项和项目完成；科研成果包括科研论文、著作与教材、职务专利和新产品；学术交流包括学术会议和学术报告；成果奖励包括自然科学成果奖励和教学、社会成果奖励。从资料分析中发现，我国对科研论文和科研项目的权重赋予了较高的分值，而对于专利和成果转化的权重很低。科研项目获得多少及质量的高低是显示学校科研力量的重要标尺，一些教师单纯为了申报课题，一旦成功不顾课题完成质量。发表SCI文章及核心期刊的文章数量、科技成果与利益挂钩，导致部分单位部门和个人过分地强调产出，造成“天下文章一大抄”的局面，高质量论文数量少，垃圾文章充塞眼球，学术泡沫化现象严重的状况。

表1 2008年英国高等学校科研水平评价的等级和标准

等级	质量概况描述
4*	在原创性、精确性、重大意义方面达到世界领先水平
6*	在原创性、精确性、重大意义方面达到较高的国际水平，几次于世界领先水平
5*	在原创性、精确性、重大意义方面达到国际认可的水平
1*	在原创性、精确性、重大意义方面达到国内认可的水平
无级别	科研质量尚未达到国内认可水平或不符合该评估对科研的界定

第三，有些国家很重视专利的申请与成果转化，比如日本的研究大学强化促进事业委员会在2013年提出的研究成果向社会的回报这一指标中的要求就是强调科技成果的转化，也正是因为日本重视专利及科技成果的转化使日本走上了科技强国的道路。我国长期以来实行科技成果鉴定及其奖励制度和专利法的激励制度，即科技管理双轨制，这种

体制使得各种奖励政策向科研成果偏移，从而使高校科研质量指标体系的设计也向科研成果偏移，而对专利等知识产权以及对科研成果转化为生产力的重视程度不够，对专利等知识产权的权重设置很小，这使得我国投入产出的智力成果大部分以科研成果而非专利形式保持其技术形态，导致我国的科技成果只停留在知识资产的形态上，没有转化为国家知识产权的产权形态。据统计，我国高校每年通过鉴定的科技成果达1万项左右，科技成果转化不足15%，我国高校科研呈现出“成果多、转化少、推广难”的局面，造成了我国大量科技产权的流失，科技成果只限于学术上的认可，不利于我国科技的发展。因此，我国应该把专利纳入评价指标体系并给予重视，不能只停留在科研成果鉴定、论文发表的数量、SCI等收录情况、科研成果奖励等形态上，还要赋予专利、知识产权相当的权重。

四、坚持评价过程的公开化

高校科研质量评价要追求评价前、评价中和评价后的公开化，但评价过程的公开化最让公众信服，只有过程的公平才能保证结果的公平，国外发达国家非常重视评价过程的公开，英国在开展高校科研评价之前，评价机构要公布指导高校提交评价材料的详细说明书，并且各个评价小组必须事先向社会公布评价标准和方法，接受社会各界的质询和监督。评价结果中的每一项得分都予以公布。因此英国的高校科研评价从评价单元的确定到学科专家组的任命，从评价组工作程序的制定到评价结果的公布，各个环节的信息，都可以直接从RAE或英国高等教育基金会的主页上查阅得到，允许各界人士对此进行公开和充分的讨论。日本把包括评价过程有关的全部内容和结果公布于众，1998年，日本科学技术会议对各省厅的评价工作状况做的调查显示：大部分科研机构都公布了评价结果，约有2/3的大学公开发表了结果，在数量上呈递增趋势。评价结果基本上采用互联网发布和向新闻媒体发布的形式。澳大利亚的RQF科研评价的过程和结果都以事实为依据，并对科研评价涉及的所有相关人员公开透明，也包括评价的标准和手段。

从现实看，我国高校科研评价缺乏完整的公示，很多公示仅仅是一个分数，造成了公众不信任，怀疑其公正性和客观性。因此，无论是对人的

评价，还是对项目、机构、基地的评价，评审工作越公开透明，被评审对象有更多的知情权，评价工作越容易得到被评审对象的理解、支持与配合。我们应该从正式评价之前就做到透明化，把评价人员、评价过程和评价结果公布于众，做到透明化，接受社会的监督，这样才可以增加评价的可信度。

五、坚持评价方法的“质量结合”

所谓坚持评价方法的“质量结合”，就是把质性评价方法和量化评价方法结合起来。具体来说，在评价高校科研时，应坚持将同行评议法、成果计量法和文献计量法等评价方法相结合。科研质量评价是依据一套比较完整的技术标准和规范对科技系统及其相关活动的作用、影响和效果进行的科学、独立、客观、公正的评价。对于科研质量评价方法的研究一直是各国进行科技评估理论研究的重点。目前较多国家采用的是同行评议法和以定性分析为基础，以定量分析为手段，定性与定量相结合的方法，即在专家评议的基础上，结合科学计量学、经济计量学方法。如美国在评价科研质量时，往往采用同行评议法这样的质性评价方法和文献计量法这样的量化评价方法相结合。美国国家科学研究委员会（NRC）是以同行评议法为主的，这种方法常常被科学界用来判断工作程序的正确性，确认结果的可靠性以及对有限资源分配的公认性。同行评议是某一或若干领域的专家采用同一种评议标准，共同对涉及相关领域的某一事项进行评价的活动，即通过专家评分的平均值就是各个参评单位的最终分数，其评价结果对有关部门的决策具有重要的参考价值。Lutz Bornmann认为同行评议并非十分完美，仅仅依靠同行评议并不能完全说明科研质量，这种评价方式是主观判断，会受到评价人的个人好恶、远近亲疏的影响，其公正性和客观性受到质疑。鉴于同行评价存在着一定的不足，美国采用了调查法、回归分析法、文献计量法等定量的评价方法做补充。NRC在2001-2003年期间开发了全新的评估方法，它最大的特点在于其基于数据（Data-based）的特点，即通过调查法和回归分析法两种不同途径的方法来评价。

文献计量学最大的优势能够定量，尤其是引文计量法是定量评价基础研究成果的重要手段，具有简单性、可操作性、准确性、客观性等特点，这也是行

高校科研评估指标及其实证研究

高宏利 李作学 王 前

《国家中长期人才发展规划纲要（2010-2020年）》在人才重大政策中的关于实施有利于科技人员潜心研究和创新政策中指出，“改进科技评价和奖励方式，完善以创新和质量为导向的科研评价办法，克服考核过于频繁、过度量化的倾向”。因此，开展公正、科学、合理的科研评估活动，不仅可以为政府制定科研发展决策提供必要的信息和依据，同时引入竞争机制的评估活动也对科学研究质量的提高、科研人员研究水平的提升、资源的优化配置、科研领域的特色发展具有现实意义。

一、高校科研评估指标模型的建立

根据高校科研评估体系构建的原则，并借鉴已有对高校科研评估的结构研究，本文提出了高校科研评估体系结构维度模型。高校科研评估体系结构维度模型，需要先对大量国内外有关高校科研评估体系的研究文献，然后采用专家访谈的形式，对这些维度元素进一步整理而确定下来。

内利用引证数据来进行各项评价的基本前提。同时引文计量法已能为决策科学化、民主化提供较为客观的定量参考意见，因此能弥补同行评议的不足。

如同美国一样，荷兰在评价科研质量时，兼有同行评议法和文献计量法。评估小组对高校科研从四个方面进行评价、打分，撰写报告，指出问题和建议。除此之外，评估委员会委托有关机构进行相关的文献计量分析，这样可以保证评价的公正性和客观性。

英国的高校科研质量评价方法具有同行评估、评价方法定性与定量结合等特点，同行评价的评价组成员由来自不同机构的资深评价人员组成，他们不代表某一特定团体和某些可能的受益者，仅代表个人进行客观的评价。此外学科组还和海外专家进行商讨，以确保评价是否达到国际先进水准。定性与定量结合，突出定性评价，这样就避免了评价中重视数量而忽视质量的倾向性，坚持成果质量第一的原则。

整理后的高校科研评估体系结构维度元素有：科研队伍、科研平台、科研成果、科研收入和学术声誉。从高校科研评估体系结构维度模型关系图中（图1），可以看出科研队伍居于模型的中心，科研平台是高校科研的基础，科研成果、科研收入和学术声誉分别位于三角形顶点处，支撑高等院校科学研究的地位和声誉，受到科研平台的支撑和科研人员队伍的影响。

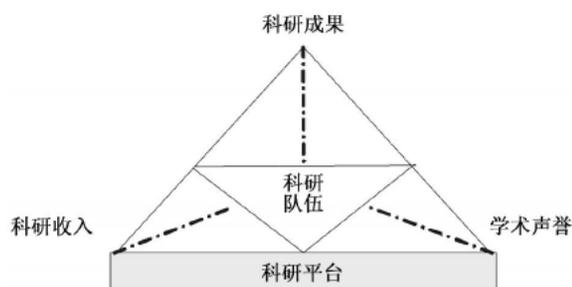


图1 高校科研评估结构模型

就目前我国高校在科研评价方法上主要有同行评议、成果计数和文献计量等方法，还远未形成一套有效的方法。事实上，由于评价对象具有复杂性，科研质量作为一种科学的评价方法不可能适用于所有的科技活动，每一种评估方法只能适用于特定的环境、特定的评估对象和特定的评估目标，并各有利弊。因此，对于科技评估的理论和方法，要不断地研究、发展和创新，要针对不同用户需求的差异性和项目的多样性、技术指标的复杂性，探索科研质量评价的一般规律，使评价工作既缜密可靠又具备良好的适用性。

（陈慧颖，长江大学大学发展研究院硕士研究生，湖北荆州 434025；通讯作者：刘在洲，武汉纺织大学党委副书记，长江大学大学发展研究院教授，湖北武汉 430200）

（原文刊载于《中国高教研究》2014年第10期）

高校科研评估体系是评估高校科学研究的能力和水平,表现为高校科学研究元素的系列指标。本研究基于大量国内外有关高校科研评估的研究文献,特别是参考英国的高校科研评估的研究,总结归纳出高校科研评估体系的要素,然后采用与专家访谈的形式对这些要素进行调整、归类、删补,最后抽象出19个具体要素。这些要素涵盖了高校科研的科研队伍、科研平台、科研成果、科研收入和学术声誉5个维度。

二、高校科研评估指标的统计分析

本研究采用问卷调查的方式,请被调查者对这19个反映高校科学研究的能力和水平的具体相关要素在一般大学和重点大学上的重要性进行评分,用以判断高校科学研究的能力在不同高校的作用和不同工作岗位上的结构差异。

问卷调查采用李克特5级评分方法,分值从小到大代表着其重要程度从低到高,如数字1代表完全不重要,数字5代表非常重要。在正式调查问卷发放之前,小范围试用了问卷10份,请调查者问卷中的问题进行提问,以便于保证问卷的有效性。根据被调查者的反馈,就填表说明、问卷形式进行了重新设计,强调如何填写问卷,争取在正式调查过程中得到被调查者的积极配合。总的说来,此次共发出调查问卷124份,回收有效问卷87份,回收率为70.16%。回收样本结构见表1。

表1 高校科研能力要素重要性调查回收有效问卷样本结构

分类		人数	比率
性别	男	49	56.3%
	女	38	43.7%
工作性质	一般大学人员	56	64.4%
	重点大学人员	31	35.6%
工作年限	工作5年以下	19	21.8%
	工作5-10年	28	32.2%
	工作10-15年	19	21.8%
	工作15年以上	31	35.6%

(一) 科研队伍

高校科研队伍是以基础科学研究及科学技术研究与开发为内容,由为数不多的且技能互补的、愿意为共同的科研目的、科研目标和工作方法而相互承担责任的科研人员组成。科研队伍是高校科研创

新能力的核心和主要力量。因此它的地位和作用处于模型的中心地位。

根据高校科研机构的特点和评估指标体系建立的原则,本研究在调查问卷中设计了一个开放性的问题,即“您认为作为一所科研能力较强的大学,其能力表现在哪些方面?”从回收的问卷来看,被调查者就这一问题进行了不同程度的回答,我们选择了科学研究工作经验在15年以上的31份问卷的回答进行整理,结果如表2所示。由于是开放性问题,所以表中根据语义对不同回答进行了汇总,根据被调查者所属重点大学和一般大学性质,列出了回答频率在10次以上的因素。

表2 高校科研能力表现调查表(N=31)

重点高校科研能力表现	频数	一般高校科研能力表现	频数
科研人员水平	28	科研队伍	26
领军人物	25	论文和著作的影响力	24
学术成果	24	博士和教授人数	24
博士人数	24	科研平台	22
科研基地和平台	22	研究经费	21
研究生人数和研究团队	21	研究项目数量	21
研究经费	20	科研工作在校的受重视程度	20
研究项目	20	科研氛围和自由度	17
学科竞争力	19	硬件条件	16
学术环境	18	学生素质	10
优势学科	18		
学术受尊重的程度	17		
硬性条件	13		
学生素质	10		

根据表中的结果,两类不同工作类型高校的工作人员都认为高校有较强科研能力主要表现在科研队伍水平和能力、学术成果、科研平台、研究经费和项目、学术氛围和环境、硬性条件和学生素质等方面。而且科研队伍人员都排在了第一位,由此也看出高校科研队伍在评价高校科研能力和水平中的重要作用,而且也验证了前面高校科研结构模型中科研队伍在高校科研中的核心作用。当然科研人员队伍中评估还可以进一步整理归纳,例如被调查者提到的博士人员数量、领军人物、教授的数量和研究生人数都可以归纳到科研人员队伍中。这样,科研队伍所表现出来的具体因素有职称结构、学位结

构、层次结构和研究生人数4个方面。表3是被调查对象对4个方面的重要性评分结果。

表3 科研队伍元素调查表

元素	一般大学	重点大学
	中值	中值
职称结构	4.3563	4.5402
学位结构	4.4828	4.5172
层次结构	4.4253	4.5517
研究生队伍	4.2299	4.7241

从表3调查结果可以看出，一般大学和重点大学的教师工作性质虽有区别，但是他们在高校科研能力中科研队伍的表现的中值都超过了4分以上。职称结构、学位结构和层次结构在两类院校中的认可程度都较高，分值分别为4.5402、4.3563，4.5172、4.4828和4.5517、4.4253，这说明两类院校都非常重视科研人员的职称结构、学位结构和层次结构，也说明了职称结构、学位结构和层次结构在保证高校科学研究顺利进行中的重要作用。

(二) 科研平台

科学研究过程中需要在一定的实验条件下进行，依靠一定的实验设备和仪器方可进行。例如高校的重点实验室、工程训练中心等等都是研究者和学生进行科研训练的必要条件。由此科研平台的内涵可概括为：从事研究活动的场所、支撑科研活动的实验设备和仪器、负责管理与运行的技术支撑队伍、赖以维持的经费和其他资源，以及将这些资源有效结合、高效运行的政策体制保障。科研平台建设是一所高校教学与科研资源长期积累的结果，反映了该校此学科的人才培养和学术水平以及承接重大重点项目的能力并对可持续发展产生综合性和久远的影响。

调查结果表明，科研平台这一维度对两类院校都很重要，但是得分差别很大。本研究的科研平台主要从国家级科研平台和省级科研平台两个方面来进行比较的。重点大学这两个方面的评分分别为4.2759、4.5632，而一般大学这两个方面的评分分别是3.4713、4.1839。从评分上看，国家级科研平台均值相差较大，而省级科研平台均值相差相对较小。各个高校非常重视科研平台的建设，重点大学更加重视国家级的科研平台建设和水平。

(三) 科研成果

高校是基础研究的主力军，基础研究的成果主要是著作和论文，专利是体现科技工作的重要指标。而科研奖励是代表研究水平的高低。通过对文献的检索和实际调研，发现具有较强科研成果的科研机构表现在论文、著作、专利和科研奖励四个方面上比较突出。

调查显示，科研成果对于重点大学和一般大学的重要性存在重大差别。在四个要素的重要性评分上，重点大学普遍比一般大学的评分高。这说明重点大学的科学研究在工作中取得高的科研绩效，必须比一般大学的科学研究具备更强的科学研究成果，表现在论文、著作、专利、科研奖励四个方面的实力较强。比较结果见图2所示。

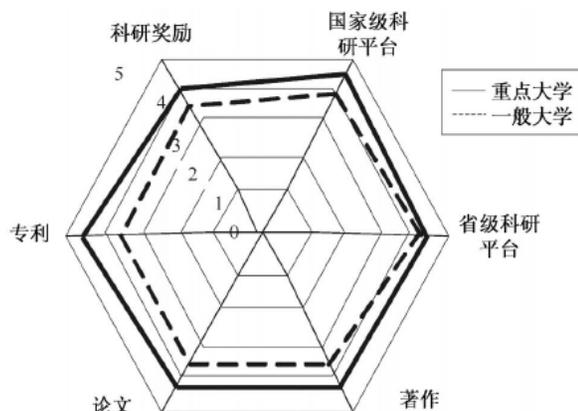


图2 科研成果和科研平台重要程度比较图

(四) 科研收入

科研经费是各类科技活动的基础和前提条件，反映了科研机构的科技工作规模和科技实力。在实际调研中，重点大学和一般大学都非常注重研究经费和研究项目在科研中的重要作用。本研究通过项目类别和资助规模两个因素来评估一个高校的科研收入这个因素。从调查结果来看，科研收入和科研项目两个方面对重点大学和一般大学的重要性方面有着很大区别，一般大学在这两个方面的评分分别为4.0460、3.5862，而重点大学的评分为4.3563和4.4828。说明重点大学的科研收入和科研项目普遍要比一般大学重要。

(五) 学术声誉

学术声誉是指一所大学以科研水平为核心的办学实力，在社会上尤其是学术界影响的深度与广

度。它虽然是一个主观指标（一般是通过广泛的问卷调查得出的），但几乎是世界著名大学学术指标中最主要的标准。它本质地、准确地反映着一所大学的地位和影响。综观当今大学评价指标体系的构成，尽管学术声誉所占权重各有差异，但一个基本事实是学术声誉起着越来越重要的作用，成为大学评价体系建构的基点。

根据已有的研究和调查结果，高校的学术声誉主要表现有学术竞争力和学术环境两个方面，学术竞争力可以通过师资力量、优势学科、学术成果、学术受尊重的程度四个方面来评估；而学术环境可以通过硬性条件、学术氛围、学生素质三个方面来评价。实际调研结果发现，高校的学术声誉几个方面指标总体上讲重点大学的得分都要高于一般大学。具体而言，优势学科、学术受尊重的程度在一般大学和重点大学比较接近，其他几个方面重点大学要比一般大学得分高很多。这可以从图3显示出来。这说明了一般大学总体上不如重点大学在学术声誉的表现，但是一般大学也有其自身的优势，比如一般大学在某些学科具有自己的研究特色，进而在某些学科上的学术受尊重的程度也比较高。

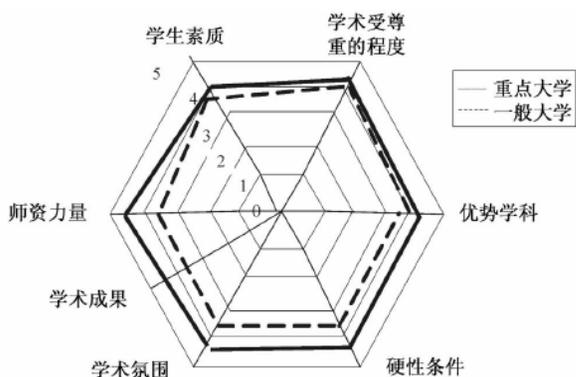


图3 学术声誉重要程度比较图

三、高校科研评估指标结构差异的统计假设检验

据调查问卷的评分均值的差异直接来判断一般大学和重点大学在科研评价的结构上差异。然而，这些评分均值不一定能够真实地反映两者之间的实质性差异。为了判断两类院校在科研评估结构上的差异，本文进一步采用统计假设检验的方法，根据上面调查问卷的数据，检验在高校科研评估结构的各个维度上，表现高校科研评估结构的各个要素对

于一般大学和重点大学重要性评分在5%的置信水平上是否存在显著差异。问卷分析采用独立样本T检验的方法。

问卷分析首先就19个高校科研评估结构要素对一般大学和重点大学的重要性进行对比，提出19个假设。如：

假设1：学位结构对一般大学和重点大学的重要性不存在显著性差异；

假设2：职称结构对一般大学和重点大学的重要性不存在显著性差异；

为了简单起见，可以把这19个假设归纳为一个包括所有假设的总假设。

假设：各维度高校科研评估要素对一般大学和重点大学的重要性不存在显著性差异。

根据统计检验结果，来分析这些要素对一般大学和重点大学重要性的差异。下面就各个维度的具体要素进行分析。

（一）科研队伍的统计结果

统计结果表明，学历结构和层次结构的Levene显著性水平都大于0.05，说明数据服从方差齐性，应该选择齐方差假设下的T检验显著性水平分别为0.725，0.258，都大于0.05，接受原假设，说明这两个要素对两类院校无显著差异。而职称结构的Levene的检验显著性水平虽然大于0.05，说明数据服从方差齐性，但是它们的齐方差假设下的T检验显著性水平是0.76，大于0.05，说明这个要素对两类重要院校无显著差异。由此看来，重点大学和一般大学都非常重视科研的职称结构、学历结构和层次结构，这也可以解释现在一般大学也都积极地引进人才的原因。而研究生队伍的Levene的检验显著性水平小于0.05，说明数据不服从方差齐性，应该选择非齐方差假设下的T检验显著性水平，也小于0.05，拒绝原假设。这个要素拒绝原假设，说明这个要素对两类院校来讲具有差异性。重点大学在研究生的数量上要比一般大学要多，因此研究生队伍当然要比一般大学的重要性上强得多，尤其是一些研究性大学，研究生招生的数量可能要和本科学生相当。但从均值评分上来看，两类院校对这个要素的重要性评分都超过了4分，这可以看出这个要素对两类院校都是非常重要的，一般院校也积极申请授予研究生学位的权利，不断提高学校的科研能力

和水平。

（二）科研平台的统计结果

科研平台有国家级和省级科研平台。统计结果显示，这个维度的要素的T统计量的显著性概率都小于0.05，所以拒绝原假设，即这些要素对于重点大学和一般大学的重要性之间存在显著性差异。以辽宁省高校的省级重点实验室为例，大连理工大学有13个，东北大学有11个，大连海事大学3个，辽宁大学2个，沈阳航空航天大学1个。由此看来，不同类型的大学在科研平台建设上具有明显的差别。

（三）科研成果的统计结果

统计结果表明，科研奖励这个要素对两类院校的重要性不存在显著差异，而在论文、著作、专利三个要素对两类院校的重要性存在显著性差异。以辽宁省高校中的辽宁大学和沈阳航空航天大学为例，2007年辽宁省教育统计年鉴上的数据显示，省级科技成果奖辽宁大学3项，沈阳航空航天大学也是3项，这说明一般院校在某些科研奖励上也会有所突破。结合均值的分值来看，重点大学和一般大学都非常注重知识的创造能力，在科研成果诸如论文、著作和专利等方面具有奖励导向政策，在这方面重点大学奖励额度要比一般大学要大，再加上老师评职称等一些因素，因此使得重点大学的论文、著作和专利的平均得分要比一般院校的得分要高很多。

（四）科研收入的统计结果

在5%的显著性水平上，人均科研经费这个要素的Levene显著性水平大于0.05，说明数据服从方差齐性，应该选择齐方差假设下的T检验显著性水平分别为0.024，小于0.05，拒绝原假设，说明这个要素对两类院校有显著差异。人均科研项目这个要素的T检验显著性水平也小于0.05，所以拒绝原假设，这个要素对重点大学和一般院校的重要性存在显著性差异。从均值的差异程度来看，这两个要素对重点大学的重要性远远大于一般院校的重要性。2007年辽宁省教育统计年鉴上的数据显示，大连理工大学的科研项目数为2128项，大连海事大学为1322项，沈阳航空航天大学为406项。大连理工大学科技课题当年拨入经费51315.9万元，大连海事大学科技课题当年拨入经费13318.7万元，沈阳航空航天大学科技课题当年拨入经费5201.4万元。

（五）学术声誉的统计结果

在5%的显著性水平上，学术声誉的七个因素的T检验显著性水平都小于0.05，所以拒绝原假设，这说明几个要素对两类院校的重要性存在显著性差异。从均值的差异程度来看，重点大学学术声誉的得分都要比一般大学要高，这只是说明一个总体的情况，并不意味着重点大学所有个人的学术声誉要比一般大学的学术声誉高。造成这种差异的原因与人们平常认识分不开的。对于重点大学而言，某些专业或领域有着较高的学术声誉，可能会代替人们对该学校学术声誉总体情况的认识。

通过上面的实证研究可以发现，重点院校和一般院校在科研评价的要素上存在很大的差异。除了职称结构、学历结构、层次结构、科研奖励要素不存在显著性差异外，研究生队伍、国家级平台、省级平台、论文、著作、专利、人均科研经费、人均科研项目、师资力量、优势学科、学术成果、学术受尊重的程度、硬性条件、学术氛围、学生素质要素，重点大学都比一般大学有着显著性的差异。这些重要程度的差异实质上反映了两类院校科研水平的差异。

四、结论

重点大学和一般大学的科研整体实力相差悬殊，但是一般大学在某些特色专业或领域的科研也有自己的优势。这样的研究成果对高校科研管理和高校科研评价的实践有着重要的意义。在科研评价上，首先可以根据不同类型的学校在评估方法上应该有所区别对待。由此想到的是在基础科学和应用科学上，评估工作也应该有所区分。还有在社会科学和自然科学的学科评价上，在制定评估指标上应该有所区别，例如英国、德国的科研评估既有面向自然科学、工程科学的评估体系，也有面向人文社会科学的评估体系。因此可根据科研评估的目标的要求，开发和制定不同类型的科研评估体系。这种分层次和分类别的指导思想对我国的高校科研评价的进行具有重要的指导意义。

（高宏利，大连理工大学科技管理与科技伦理研究中心，辽宁 大连 110624；李作学，沈阳航空航天大学经济与管理学院，辽宁 沈阳 110136；王前，大连理工大学科技管理与科技伦理研究中心，辽宁 大连 110624）

（原文刊载于《教育科学》2011年第2期）

高等学校科研评价改革在北京上海广东率先启动

中国教育报

近日，高校科技评价改革试点单位上海市教委、北京市教委和广东省教育厅相继研究或出台政策，加快建立科学合理的高校科研分类评价体系，以质量和实际贡献为导向，通过多元化的评价机制，促进高校管理体制和教育综合改革。

上海市教委发布《关于推进上海高等学校科学研究分类评价的指导意见》，今后上海高校的科研能力评价，将不再仅依据发表论文的数量来一刀切，而是将逐步建立科研人员的分类考核体系和科研成果的多元化评价体系，根据创新质量和实际贡献进行评价。并将积极推进“同行评议”“第三方评价”和“国际评价”等评价方法，探索开放的、多方参与的科研评价措施，建立与国外接轨的评价制度。

根据北京市教委的统一部署，北京市属高校将迎来一轮科研评价机制改革，有望改变“重科研轻教学”的现象，破解“没项目——不能晋职——被转岗”的难题，以科研质量为导向，防止片面地将学术项目、学术成果与职称评定、职务晋升等挂钩，按照学术发展和科学研究的规律，逐步建立一套有效促进和形成优良学风的评价体系。

广东省教育厅以广东高校“创新强校工程”和“四重”（重点学科、重点人才、重点平台、重大科研项目）建设为契机，结合全省教育综合改革试点工作，加快推进高校科技评价改革和体制机制创新，多次组织专家召开研讨会，正在制订整体方

案，研究省内试点高校，探索重点实验室、工程技术研究中心、“珠江学者”等基地、人才评估改革方案。

教育部将探索制定高校科技评价活动规程，系统梳理和规范科技评价的环节和程序，从组织程序上保证评价的质量和开放公正。同时，加强改革实践中的检查指导和跟踪调研，及时了解各地、各校试点进展情况，收集挖掘在改革实践中涌现的新思路、新办法、新举措，评估试点成效。

据介绍，自2013年年底以来，教育部印发了《关于深化高等学校科技评价改革的意见》，着力解决高校科技评价改革“世界观”的问题，回答“为什么做科研”“科研为什么而做”的问题；教育部科技委发布了《高等学校科技分类评价指标体系及评价要点》，提供科技评价工作的“方法论”，回答“怎么评价”“评价什么”的难题；教育部办公厅发出《关于开展高等学校科技评价改革试点的通知》，要求典型地区、典型高校针对明确问题，先行先试，加快探索实践经验，强化引领示范。教育部就高校科技评价改革工作3次发文，基本形成了完整的工作体系，要求地方和高校统一思想，抓紧抓好政策落实，同时在《中国教育报》《中国高等教育》上推出专题，正确引导舆论，浓厚改革氛围，赢得普遍支持。

（记者 宗河）

（原文刊载于《中国教育报》2014年10月21日第1版）