

国外公立科研机构绩效评估方法研究

苏秋月，张义芳

(中国科学技术信息研究所，北京 100038)

摘要：公立科研机构是国家科技创新体系的重要主体，科学的绩效评估有助于公立科研机构的良性发展和组织效率的提高，而评估方法的选择是实现客观评估的基础。本文在分析国外科研机构绩效评估研究与实践的基础上，对公立科研机构绩效评估最常用的四种方法进行了归纳分析及比较，以期为我国全面实施公立科研机构绩效评估提供适用的方法借鉴。

关键词：公立科研机构；科研绩效；绩效评估；评估方法

中图分类号：F272.5 **文献标识码：**A **DOI：**10.3772/j.issn.1009-8623.2016.05.010

公立科研机构的绩效是指科研机构有效使用其资源并提供与该机构使命目标相一致的科研产出和服务的水平和能力。客观而全面地评价科研机构的科研绩效，是一个需要认真对待的科研管理问题。然而，由于科研机构的多样性，科研活动的复杂性、不确定性以及科研成果的不易衡量性，对公立科研机构的绩效评估难以做到全面而完整，而开发利用适用的评估方法能在一定程度上确保评估结果的信度和效度，使得评估结果易于为受评机构所接受。

本文在分析国外相关研究与实践的基础上，概括出科研机构绩效评估方法的选择原则、方法分类，以及四种常用评估方法的基本原理、优缺点及适用范围，以期为我国全面实施公立科研机构绩效评估制度提供方法借鉴。

1 评估方法的选择原则

工欲善其事，必先利其器。评估方法的信度和效度是实施科研机构绩效评估的基本议题。国外研究表明，科研机构绩效评估方法一般应遵循以下几个原则：

(1) 评估方法应适合评估目标和可获取的资源。如果不考虑这两个因素，理想的评估方法就无

从谈起。因此，在选择评估方法之前，应充分了解科研机构的活动性质及可资利用的评估条件，避免理论评估方法与实际应用脱节。

(2) 评估方法应具有效度。评估方法通常可被认为是多种效度类型的折中，不仅包括内部效度（因果关系），还包括统计结论效度（统计假设的真实性）、外部效度（可概括性）和结构效度（所测度对象的适合程度）。

(3) 评估方法必须可重复运用。科研机构的绩效评估不仅关注近期直接“产出”（Outputs），如发表论文数、出版专著数、发明专利数等，还应关注科研机构在学术以及经济社会诸方面的中期乃至长期的综合性“成效”（Outcomes），如人才培养与流动、技术创新、经济回报、对社会的整体贡献、国际合作等。因此，为获取更为完整的评估，通常建议在未来合适的时间将之前评估中所使用的方法重复运用到新的评估活动中。

(4) 评估方法应关注最终结果而非具体实现方式。例如：对科研机构科研项目进行评估时，最棘手的一个问题是明晰项目开展方式和项目最终结果之间的界限。很多评估活动的失败都是由于未能专注于最终结果，或未能将有价值的成果与这些成

第一作者简介：苏秋月（1993—），女，在读硕士研究生，主要研究方向为科技战略与科技政策。

收稿日期：2016-04-02

果的具体实现方式划分开来。

(5) 评估方法应具有可比性。很多评估方法对成效的衡量在不同的项目间或科研机构间不具有可比性。在评估中使用可比标准时,决策者能够更有效地对不同意见进行权衡。

2 评估方法的分类

公立科研机构的绩效评估可繁可简。各国在科研机构评估的起始阶段,一般试行采取简单易行的评估方式和方法,随着经验的积累和条件的完善,再逐步提高和深化。如表1所示,科研机构绩效评估方法同其他科技评估一样,大体分为三类^[1]:

(1) 定性评估方法,即由专家评估组按照一

定的标准对科研机构的绩效给出定性评价,如同行和专家评估、实地考察等,这是迄今为止使用最广泛的评估方法,在科研机构绩效评估中占据主导地位;

(2) 定量评估方法,即通过文献计量/科学计量、经济分析等方法,以客观的统计数据为基础,考察科研机构的科研产出和社会经济影响,评估指标主要包括论文、专著、专利、经济收益、社会效益等;

(3) 综合评估方法,即将定性与定量的评估方法有机结合,评估公立科研机构的绩效,最常见的是以专家评估为主、辅以文献计量和经济分析的方法。

表1 三类评估方法

方法分类	方法描述	主要代表性方法
定性评估方法	基于专家知识和经验的评估方法	同行评议、专家评估、德尔菲法、实地考察等
定量评估方法	基于统计数据的客观评估方法	文献计量法/科学计量法、经济计量
综合评估方法	定性与定量相结合的评估方法	结合计量方法的专家评估

过去一段时间,出于客观性和显见性的需要,一些国家的政府对其科研机构的绩效评估曾力求可衡量和量化,但实践表明,对科研绩效完全诉诸于量化是不现实的,因为客观量化指标和方法无法完整呈现科研机构的科研绩效,结合专家评审才是一个比较适合科研机构绩效评估的方法^[2]。因此,目前各国公立科研机构的绩效评估多采取定性和定量相结合的综合评估方法,以求全面、系统、客观地评估科研机构的绩效。

3 典型评估方法的特点分析

3.1 同行和专家评估

在科研机构绩效评估中,同行和专家评估一直是主要的方法。同行和专家评估是指由某一领域或若干部门领域的专家按照既定的评估原则和评估标准,运用科学的方法所进行的专业化评价活动^[3]。从国际上看,基础研究类科研机构的绩效评估采用同行评估,而对于应用技术开发类机构,仅仅依靠同行评估显然不够,需要扩大评估专家的范围,邀

请研发用户等多元化利益相关者(比如研究的上下游单位或产业界专家)参与评估,即本文所说的专家评估。

同行和专家评估作为评价科研机构绩效的一种重要方法,在欧美等国家和地区的科研机构评价评估中已被广泛采用,其主要有两大优势^[4]:首先,科技界对它广泛认可。这种方法的广泛应用基于以下理念,即只有在某个领域名列前茅的内行专家才能有效评价科研机构同领域的工作。因为这些专家具有评估研发成果和未来应用前景的能力,他们的观点和建议能够为被评价科研机构接受和尊重;其次,同行和专家不仅可以评估科研机构的科研质量,还可以评估科研机构的科研管理、人才培养和科研环境情况,进而形成对科研机构的改进意见和建议,使科研机构未来的发展更趋完善、更富成效,而这正是体现科研机构绩效评估的一个重要目的。

同行和专家评估的一个不足是评估过程给科研机构和评估专家所带来的成本和负担。组织实施同行和专家评估的开支可能很大。筛选和招集专家、

组织参观科研机构实验室并与科研人员深入面谈等需要投入时间、资源和精力。同行和专家评估的另一个主要问题是,它以专家个人的主观判断为基础,因而不可避免地会带有评议偏见。这种偏见可能有多种根源,如评估专家的个人偏好、既往的研究经历和经验,以及存在利益冲突情况下可能的非客观性、非公正性问题。

然而,上述问题并没有影响同行和专家评估作为一种主要的方法在科研机构绩效评估中发挥重要作用,因为迄今为止还没有一套可行的方法能够完全代替同行和专家评估的作用。例如,德国马普学会为了提高科研质量和改进科研管理,通过外聘的科学顾问委员会定期地对下属研究所进行绩效评估。科学顾问委员会的成员由国内外知名科学家组成,不仅针对马普学会研究成果质量提出评价意见,也提供对学会组织结构和资源配置方面的建议,以在确保研究水平的同时保证研究的经济效率^[5]。英国的研究卓越框架(REF)则力求在更广泛的背景下采用专家评估,在评估科研学术价值的同时,还对科研影响力和科研环境进行评估,通过吸收用户专家代表参加评审,强化对科研影响力的评估。当然,为消除影响同行和专家评估公正性和客观性的因素,减少大规模的专家评估造成的负担,各国科研机构评估中常常补充采用其他评估方法,如下文的文献计量法。

3.2 文献计量法

科研产出是反映一个科研机构科研实力的重要方面。而文献计量法通过评估科研机构的论文、专著、专利等科研产出的数量和传播情况来考察科研机构的科研成果产出。文献计量法主要从以下方面来提供分析成果:

(1) 数量指标。以论文、报告或专利数量来衡量特定研究领域的科研成果,实际上也就是衡量科研机构科研产出的规模。

(2) 知识转移量指标。它依据的原理是论文引用过程能反映科学界内部的知识交流,并能间接衡量科研工作质量。

(3) 引用分析。反映科研机构科研产出效果和影响的一个或多个方面。

文献计量法是科研机构绩效评估的常用方法之一,具有独特的优势:

(1) 由于使用数据库统计数据,文献计量法具有客观、量化、易于比较等特点,能够对科研机构的科研产出能力等有较为直观的评价,可以大大减轻同行和专家评估对科研机构和科研人员造成的行政管理负担。在英国,已出现这样的趋势:将文献计量法融入范围更广的科研评估,并且在生物医学研究领域,利用文献计量分析法,依据该领域中引用率最高的论文来查找该领域最杰出的科研组织。

(2) 文献计量法可以生成衡量跨学科研究工作的合作情况及其性质和范围的量化指标。它能处理多种评估对象,包括研究产出、合作关系以及知识传播的模式与强度。

不过,文献计量法用于反映抽象概念(比如科研工作的“质量”)的能力在理论上仍然未能让人完全信服。更为人诟病的是,基于数量的指标导向极易导致科研机构和科研人员选择那些短周期、稳妥的渐进性研究,而规避那些长周期、高风险的先驱性研究。

文献计量法还有其他一些局限性,比如综合的文献计量数据库对不同学科期刊的覆盖面差异会导致计量结果的偏差(适合本学科的数据库是合理的数据来源);文献计量只把论文、专利等作为科研产出,而忽视其他的产出(如人才培养)和长期结果;论文和专利引用具有时滞性;论文和专利均不具备均一的价值;文献计量法无法衡量科研成果未来的潜在应用前景等^[6]。

英国2014年实施的REF^[7]评估实践表明,文献计量法不足以代替同行和专家评估,但在一定范围内(至少在一些学科中),论文的引用信息可以用作同行和专家评估的补充。在各国的科研机构绩效评估中,文献计量法均是作为同行和专家评估的辅助工具和参考,而不是单独使用。越来越多的研究表明,文献计量分析可以对同行和专家评估起到有益的互补作用,对于大规模同行和专家评估来说尤其如此,因为这类评估工作需要分析的科研成果相当多,复杂程度也很高,所以一定形式的量化佐证不无裨益。

3.3 经济分析方法

政府资助公立应用型科研机构的目的之一是获取经济效益,对科研机构开展有效的经济绩效评估是政府了解科研投入效果和拟定投入政策的重要依

据。经济分析通过评估科研活动的投入、产出、效率和效益来反映科研机构的绩效，既包括科研成本效益比和经济影响评估，也可以包括科研活动对经济和社会贡献的测度。经济分析多用于应用开发性强的公立科研机构的绩效评估。应用型科研机构技术产出的多样性和复杂性，给科研机构绩效的测度带来很大的困难和挑战。经济分析则是以科研以外的视角，描述研发活动对经济和社会产生影响的过程和由此产生的经济效果。

经济分析在科研机构的科研管理中发挥着重要作用。首先，经济分析为科研机构内部科研经费调整和政府宏观科技预算分配提供经济依据。若没有论述充分的明确依据，无论是科研机构内部还是国家科技财政的分配都会缺失公正性，难以服众。其次，对科研机构效率的持续关注促使人们开展更多的回顾性影响评估，以了解科研机构的整体绩效和内部科研活动的绩效，并将评估结果作为下次资源分配的依据^[8]。

经济分析有几种不同的类型，在科研机构绩效评估中，常用的是成本效益分析和经济影响评估。成本效益分析将科研机构的科研活动取得的可计量效益与该活动的实施成本联系在一起。该方法的主要优势在于，它是一种易于理解的绩效评估工具，主要劣势是其实施需要使用大量资源，并且大多数时候，很多可确定的效益难以计量。经济影响分析

则需要一个多学科视角，如技术评估、微观经济分析和公司财务等。在科研机构绩效评估活动中开展经济影响评估，主要是针对机构的研发项目或技术，利用一些经济学指标和统计学方法，将科研成果与经济变量（就业趋势、生产率、市场份额等）联系在一起。研发项目或技术产出对经济的价值，有的短时间能够表现出来，如研发合作、吸引风险资本，有的则需要较长时间才能看出来，如生产率的提高、对国家经济的影响（见图1^[9]）。因此，研发机构的经济影响评估必须是多层次的，且由图1可以看出，理想的经济影响研究分析是在一项研究后至少3年、最多10年的期间。

经济分析有着不同的形式，具体采取哪种形式主要取决于需要分析或比较的成本和效益的货币化程度。每种类型的经济评价都有着不同的目的，这决定了它的使用条件。鉴于可以采取不同的经济视角，经济评价并没有一种单一的方法。

3.4 案例研究

案例研究是经验性实证研究的一种特殊方式。在科研机构评估中，案例研究通常只起某些方面的例证作用，通过描述科研机构科研项目的成败案例来为评价科研机构的科研绩效良莠提供佐证。

案例研究的主要优点是它处理复杂问题的灵活性。它能够明确地阐述某一现象出现的方式和过程以及出现的原因，深入而全面地表现一项科研工作

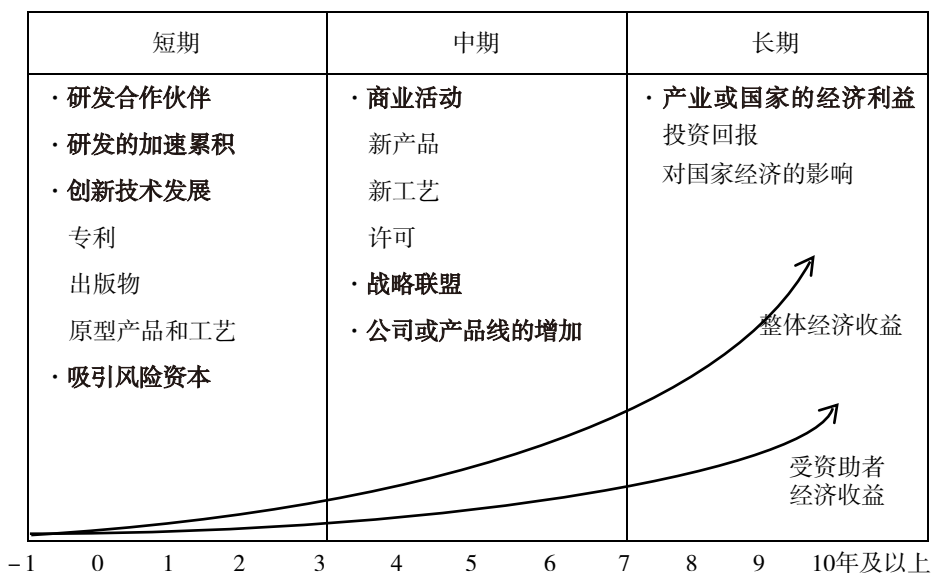


图1 科研机构项目或技术的价值整合框架

的细节问题和所涉情境,在科研机构绩效评估中,案例研究的灵活性足以反映多样化的科研影响(包括意外的影响),可以探究各种类型的科研产出和结果并用于多种目的的科研绩效评估。英国 REF 就采用了案例描述+影响力模板的方法来考察大学科研工作学术界以外产生的社会影响和经济影响^[10],因为针对科研影响评估,很难拟定一组广泛适用于所有科研成果的统一的指标。REF 选择采用案例研究这种方法,是因为其具有足够的灵活性来评估科研活动可能发生的各种影响。美国国家标准与技术研究院(NIST)的实验室评估中曾选择光学检测器项目作为代表案例,评估 NIST 校准设施给其直接用户带来的效益,以证明 NIST 工作的社会价值。

案例研究的主要局限性是它的研究成果不易具备概括性,因而通常用在需要举例说明的时候。在科研机构绩效评估中,案例研究常受到的一个质疑是,为什么能从一个案例或几个案例的结果得出科研机构科研绩效的概括性结论。虽然实际评估中通过慎重选择一些案例组成规模足够的案例组合,能在一定程度上解决这一问题,但在对案例结果进行概括并将其推到更广的范围时,还是需要谨慎。案例研究的另一种局限性是它固有的主观成分,一些研究者存在一种为了得出某个结论而把数据牵强附会的倾向。尽管案例研究可以采用事实型数据,但它的结果很大程度上来自研究者对案例的解读与推论,而且结果也很难检验。案例研究的另一个不足之处在于它的成本较高。通过案例研究评估科研机构绩效时,需要花费大量时间来采集、分析数据,势必影响评估工作的总体进度。而且,相对其他方法而言,案例研究的运行成本较高。

4 结论

本文在分析国外科研机构绩效评估研究与实践的基础上,选取同行和专家评估、文献计量方法、经济分析方法、案例研究四种典型的评估方法进行归纳比较分析。从国际经验来看,科研机构绩效评估使用最多且被普遍认可的方法是同行和专家评审,但对于不同的科研机构,同行和专家评审的过程以及专家的选择又不尽相同,且主观性强,难以客观、全面、公正地把握科研机构的绩效;文献计

量法是以统计的方式,利用文献的特征(如论文的数量与分布、被引次数、引文网络等),来反映科研成果的影响和绩效^[11],相对客观、公正,但用于反映抽象概念(比如科研工作的“质量”)的效果不是很好;经济分析在应用开发型的科研机构绩效评估中应用的比较多,主要考察科研机构科研活动的成本效益和经济影响,但评估面临的挑战和难度较大;案例研究能够处理评估中许多复杂的问题,但它只能针对某一特例和特定情景,分析的结论不具有概括性。

综上所述,任何一种方法都有具体的适用范围和相对的局限性。绩效要依靠多维建构,单一的方法不能完整而全面地测度公立科研机构的绩效。因此,在具体的评估实践中,需要依据评估的要求和目的,考虑各种评估方法的配合使用,以提高绩效评估的效度和信度。另外,绩效评估方法研究应注重可应用性。对比国内外研究,国外在进行理论研究的同时重视应用和对实践的指导,而国内的方法研究往往局限于理论层面,较少考虑是否能真正落实到科研机构的绩效评估实践中。这是我国科研机构绩效评估方法研究中的一个突出问题。

我国的科研机构绩效评估实践刚刚起步,评估方法的运用还不是很科学和完善。因此,从系统的观点出发,以定性定量结合、长中短期结合、内部评价与外部专家评价结合、科技进步与效益结合为原则,建立科研机构综合评估的方法框架体系,仍是我国未来研究的重点^[12]。■

参考文献:

- [1] 李晓轩,汪凌勇.国际科技评估的理论与实践[J].科技成果纵横,2003(5):15-17.
- [2] 张义芳.政府科研机构绩效评估及相关制度分析[J].科技进步与对策,2006,23(12):5-7.
- [3] 孔红梅,刘天星,段婧.同行评议初探[J].生态环境学报,2010,19(4):1004-1008.
- [4] 文庭孝.国立科研机构实力比较:理论、方法与实证[M].湖南:湘潭大学出版社,2010:47.
- [5] 孟强.德国评估科研机构的标准、程序和方法[J].全球科技经济瞭望,2004(5):26-28.
- [6] Susan Guthrie, Watu Wamae, Stephanie Diepeveen, et al. A

- Guide to Research Evaluation Frameworks and Tools[R]. America: RAND Corporation. 2013.
- [7] HEFCE. REF policy development[EB/OL]. (2016-4-11) [2016-4-20]. <http://www.hefce.ac.uk/rsrch/REFreview/evaluation/>.
- [8] 乔治·泰奇 . 政府研发的经济影响评估 [M]. 盖红波, 黄军英, 程如烟, 等, 译 . 北京: 科学技术文献出版社, 2005: 7-12.
- [9] 周建中 . 美国国家标准与技术研究院绩效评估的实践、方法及启示 [J]. 中国科技论坛, 2009 (1) : 135-139.
- [10] HEFCE. Response to Lord Stern' s review of the Research Excellence Framework[EB/OL].(2016-4-11) [2016-4-21].http://www.hefce.ac.uk/media/HEFCE,2014/Content/Research/Review,of,REF/Stern_review_response.pdf.
- [11] 赵学文, 龚旭 . 科学研究绩效评估的理论与实践 [M]. 北京: 高等教育出版社, 2007: 96-103.
- [12] 戴丽英, 李海燕, 赵镇 . 我国重点实验室绩效评估体系和评价方法 [J]. 中国医院管理, 2011, 31 (1) : 60-62.

Research on Performance Evaluation Methods of Public Research Institutions in Main Countries

SU Qiu-yue, ZHANG Yi-fang

(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: The public research institutions are an important part of national science and technology innovation system, and a scientific performance evaluation method will contribute to the healthy development of the public research institution and organizational efficiency improvement. The choice of evaluation methods is the foundation of the objective performance evaluation. Based on the analysis of evaluation on public research institutions in main countries, this paper explores four types of performance evaluation methods which were most frequently used in public research institutions, in the hope of providing reference of methods to the implementation of performance evaluation in China's public research institutions.

Key words: public research institutions; scientific research performance; performance evaluation; evaluation method