关于科技评价和科技评价改革的讨论

陈兆莹

国家科技评估中心 北京 100081

E-mail: chenzhaoying@ncste.org

2017-09-12 收稿, 2018-01-31 修回, 2018-02-02 接受, 2018-02-06 网络版发表

摘要 目前, 科技评价已成为我国科技界备受关注的问题, 科技评价的改革也被提到落实国家创新驱动发展战略的高度。然而, 什么是科技评价? 科技评价要遵循哪些基本准则? 如何理解评价的独立性? 对于这些关于科技评价和科技评价改革的基本问题, 目前各方面的认识和理解还存在一定的差异。本文提出, 为了推动科技评价的改革取得实质性进展, 从落实国家创新驱动发展战略的需求和当前科技界反映最强烈的问题出发, 可以将公共R&D (research and development) 的绩效评价和学术评价中的价值导向问题作为科技评价改革的突破口。并针对公共R&D绩效评价的三个基本原则(独立性、可信性和有用性)进行了讨论。简要介绍了我国公共R&D绩效评价的成功案例——“国家自然科学基金资助与管理绩效国际评估”, 重点分析该评估如何体现这三个基本原则。试图针对科技评价和科技评价改革的一些基本的、有争议问题的进行讨论, 希望有助于促进科技管理者、科学家和评价专家之间的对话, 并为推动我国科技评价的改革提供参考。

关键词 科技评价, 科技评估, 公共 R&D, 绩效评价, 公共管理

目前, 科技评价(科技评估)已经成为我国科技界备受关注的问题。无论是高层决策者、科技管理者、科学家还是专业评价人员都一致认为, 为落实国家创新驱动发展战略, 我国的科技评价系统亟待改革。然而, 什么是科技评价? 科技评价的对象和范围如何界定? 科技评价要遵循哪些基本准则? 改革我国科技评价系统的突破口在哪里? 对于这些关于科技评价和科技评价改革的问题, 目前各方面的认识和理解还存在明显的差异。因而, 促进科技管理者、科学家和评价专家之间的对话, 对科技评价的基本问题形成共识, 对于我国科技评价系统的建设和科技评价的改革是十分重要的。

1 科技评价和科技评价的改革

1.1 科技评价在我国是一个备受关注但尚未完全取得共识的问题

在政府颁发相关文件中, 科技评价的对象主要

包括中央或地方财政资金资助的科学技术计划、项目、机构、人员和成果, 旨在强化政府科技投入的问责机制, 改进管理决策, 提高研发活动的整体绩效。不同文件的提法虽有所不同, 但科技评价的基本属性被定为是政府科技管理的手段和工具。

从科学共同体的观点来看, 科技评价的基本属性可理解为学术评价, 认为科技评价是科学共同体内部自治的一项基本制度, 是衡量科技创新活动及其主体水平和价值的主要依据。除此以外, 在我国关于科技评价的讨论中, 还涉及到立项目评审、成果鉴定、职称评定、院士遴选、科技奖励、大学排名以及为技术交易服务的技术价值评价等不同类型的评价活动。

尽管在我国关于科技评价的文献中, 经常可以看到关于“国际科技评价”、“美国科技评价”、“英国科技评价”的介绍, 但实际在国际上很难找到与我们所说的科技评价相对应的概念术语。在与国际评价

引用格式: 陈兆莹. 关于科技评价和科技评价改革的讨论. 科学通报, 2018, 63: 611–617

Chen Z Y. Reflections on science and technology evaluation in China (in Chinese). Chin Sci Bull, 2018, 63: 611–617, doi: 10.1360/N972017-00976

专家的交流中,经常遇到的问题之一就是如何解释我国科技评价的内涵.2008年本人应邀参加了经济合作与发展组织(OECD)科技政策委员会组织的关于科技领域的评价系统的研究,并以“中国科技评价系统的演变”为题目提交了研究报告的初稿.报告的主要内容是介绍中国国家科技计划的评价.OECD专家的反馈意见表示,“科技评价”这个概念还没有一个规范的定义或公认的解释.从中国专家报告的内容来看,应该属于公共R&D评价.我们的报告终稿的题目定为“公共R&D评价在中国的进展——重点考察国家科技计划的评价”(Government-Funded R&D Evaluation in China—with Special Reference to the Evaluation of National R&D Programs, OECD DSTI_STP_TIP (2008) 11-National Case).同时还注意到,在美国评价协会(AEA)的评价活动分类中,没有科技评价这个类别.有关公共R&D活动的评价,归入“政府评价”这个类别.大学的评价属于“高等教育评价”.

可以看到,科技评价涉及的内容和范围非常宽泛,目前还很难对科技评价给出一个准确的、各方面都认可的定义.科技评价在我国是一个备受关注但尚未完全取得共识的问题.如何促进科技管理者、科学家和评价专家之间的对话,对科技评价的基本问题形成共识,是我国科技评价系统建设的一项重要任务.

1.2 关于科技评价改革的突破口

当前,我国科技界已将科技评价的改革提到落实国家创新驱动发展战略的高度.最近,全国政协教科文卫体委员会调研组就“优化创新环境改革科技评价体系”进行了专题调研(《人民政协报》,2017-05-26).面对科技评价这样一个范围宽泛、对象复杂的问题,选择改革的重点和突破口十分重要.为了推动科技评价的改革取得实质性进展,我们可以从落实国家创新驱动发展战略的需求和当前科技界反映最强烈的问题出发,选择改革的重点和突破口.本文提出,公共R&D的绩效评价和学术评价中的价值导向可以考虑作为科技评价改革的突破口.

近年来,中国研发经费投入总量呈不断上升趋势,先后超过英国、法国、德国和日本,成为仅次于美国的世界第二大科技经费投入大国,2012~2016年均增长达11.1%(来源:国家统计局, http://www.stats.gov.cn/tjsj/sjjd/201707/t20170727_1517417.html).然而,

目前我国公共R&D绩效评价的制度和评价能力却落后于研发经费的增长,与我国科技的国际地位也不相称.建立有效的公共R&D的绩效评价系统已经成为落实国家创新驱动发展战略的迫切需求.

目前我国科技界对学术评价中不正确的价值导向反映非常强烈.不少资深科学家指出,我国当前科技评价最大的问题就是对科研成果、科技人员的不恰当的量化评价,并与利益挂钩^[1].特别是以SCI作为唯一评价标准的弊端日益突出,一味追求期刊的影响因子,不求解决重大科技问题.由科学共同体发挥主导作用,纠正学术评价中错误的价值导向,应该成为科技评价改革的重要突破口之一.本文对此不做进一步讨论.

2 公共R&D绩效评价

如前所述,随着我国政府研发支出的规模逐年增大,公共科技资源的配置和使用效果备受政府、科技界和公众的关注.从大科学时代的特点来看,公共R&D活动,即由国家主导的旨在解决基础科学、共性技术和其他体现国家意志的科学研究活动,已经成为政策研究和绩效评价的重要对象.

目前,国内已有一些文献涉及到公共R&D的绩效评价,多数针对评价对象、评价指标和评价方法等内容^[2].但是对于公共R&D的绩效评价的一些基本的问题还缺乏深入讨论.例如:关于公共R&D绩效评价的基本原则,如何理解评价的独立性,第三方评价和评价独立性的关系等问题,无论在政府部门还是在专家学者之间,还存在明显的认识差异.

2.1 公共R&D绩效评价的基本原则

2009~2013年,联合国开发计划署(UNDP)组织了三次大型的国际会议(NEC),每次都邀请来自100多个国家的政府官员和评价专家参会,围绕国家公共政策和计划评价进行深入讨论,并分享各国的经验.这个系列会议形成的重要共识之一是将独立性、可信性和有用性作为国家公共政策和计划评价的基本原则和国家评价能力建设的基础^[3].这三个基本原则对于我国公共R&D绩效评价也是适用的,值得研究和借鉴.

根据联合国评价小组(United Nations Evaluation Group, UNEG)编制的《评价规范和标准》^[4]和相关资料,关于独立性、可信性和有用性的基本概念解释如下:

(i) 独立性(Independence). 评价的独立性可从制度和行为两方面来理解。

① 制度/组织上的独立性要求评价职能独立于管理决策,以保证评价免受干预。评价报告可以直接提交给适当的、与评价对象有关的管理决策层。评价人员不得直接参与政策制定、项目管理以及对于评价对象的其他管理工作。② 行为层面的独立性要求评价者有充分自主权开展评价活动,可以得到评价所需信息,能够自主地表达他们的评价意见,能够对评价内容、范围、方法和结论独立地做出最终决策而不要受管理者干预。

(ii) 可信性(Credibility). 评价必须是可信的。独立性是保证评价可信的必要条件之一,但并不能确保评价的可信性。

影响评价可信性的关键因素还包括:① 评价者/评价机构的能力和信誉,可信性要求评价人员遵守评价的道德规范,并体现专业能力;② 评价活动的过程和方法,要求评价过程公开透明,评价方法严谨规范,评价组织实施模式和评价方法的选择合理;③ 评价报告的质量,包括:评价报告要说明使用的评价方法以及方法的局限性,评价报告必须不带偏见地阐述事实、结论与建议,反映不同意见者的观点,不同渠道的信息要相互验证。

(iii) 有用性(Utility). 评价的有用性体现在对管理决策的问责和改进两方面。

评价结果对改进管理决策的具体作用可以包括:① 评价的建议得到采纳,直接用于管理和决策的改进;② 评价活动促进利益相关方的对话和沟通,有助于形成共识;③ 参与评价活动的组织或个人在评价过程中获取知识,提高能力;④ 可能有些评价结论或建议已由其他方式提出过,但独立、可信的评价可以使决策更加令人信服。

联合国开发计划署评价办公室用图1的“评价三角形”形象地表达三个基本原则之间的关系。独立性位于三角形的底部,表明独立性是评价的基础,评价的独立性可增强其可信性,独立性和可信性为评价的有用提供了重要的支持。同时也强调独立性不是目的而是评价的保障条件,是为了产生可信、有用的评价结果。三个咬合的齿轮表示这三个原则互相关联。

2.2 关于评价独立性的讨论

多年来,评价的独立性一直是国际评价领域的

图1 (网络版彩色)评价的三个基本原则以及它们的相互关系。来源:联合国开发计划署评价办公室

Figure 1 (Color online) Three principles of evaluation and the relationship between them. Source: Independent Evaluation Office (IEO), United Nations Development Programme

一个热点议题。在我国关于科技评价的讨论或争论中,也经常涉及到评价的独立性问题。本文针对我国科技界关于评价独立性的一些有争议的问题进行讨论,并简述对这些争议问题的观点。

(i) “第三方评价”是否等同于独立评价?对于政府公共支出的评价,国内在提到“第三方评价”时,并没有给出严格的定义。多数情况下,第三方评价是指与政府管理部门没有行政隶属关系的“第三方评价机构”实施的评价。从前面介绍的关于评价独立性的两个层面来看,“第三方评价”在制度/组织层面与评价的独立性相关,即评价机构在组织上独立于管理决策,但不能保证行为层面的独立性。从这个意义来说,“第三方评价”不能等同于独立评价。国际上讨论公共政策、政府计划的评价问题时,也出现过“第三方立场”、“第三方视角”等提法,但很少用“第三方评价”的概念,通常采用独立评价这个国际上普遍采用的、具有规范解释的概念。所以本文认为,独立评价比“第三方评价”更准确,也便于国际交流。

(ii) 评价的独立性是绝对的还是相对的概念?一种观点认为,评价的独立性是绝对的概念,非黑即白,一个评价活动要么是独立的,要么不是独立的。通过分析不同国家、不同组织的评价系统,可以清楚地发现,独立性是一个相对的概念,不可能达到绝对的独立性。不同的制度安排,不同的政策背景和不同的发展阶段可能需要不同程度的独立性。例如:世界银行独立评价局(IEG)和联合国开发计划署独立评价办公室(UNDP IEO)是在国际上非常有影响力的两个独立评价机构。世界银行独立评价局的主任由董事会任命,他有权直接向董事会报告评价的结果和建

议,无需获得世界银行行长的批准。UNDP独立评价办公室主任由UNDP署长任命,董事会通过,评价结果要通过署长向董事会报告。可见,这两个评价机构关于评价独立性的制度安排有所不同,独立性的程度也有区别。

(iii) 为保证评价的独立性,管理者是否要与评价者和评价活动隔离? 独立性有时被解释为在评价者和管理者保持一条明确的界限,并认为保持距离越远越好。然而,对于公共领域的评价,管理者的参与不一定危害评价的独立性,相反,评价活动完全和管理者隔绝反而会损害证据的搜集,影响评价的质量。因而,管理者在保证评价的独立性前提下,为评价活动提供所需要的协调和支持,是公共领域的评价成功的重要经验。本文在后面的案例中将进一步讨论。

(iv) 内部评价和独立性是对立的吗? 在我国关于科技评价的讨论中,内部评价往往被认为与独立评价是对立的,也就是说内部评价不具备独立性。国际经验表明,公共领域的评价系统往往包括内部评价和独立评价两个组成部分。以世界银行的项目绩效评价为例,评价系统包括项目经理为责任主体的项目自评价(即内部评价),以及由独立评价局(IEG)负责的项目绩效评价(独立评价)。世界银行的项目绩效评价的成功经验在很大程度上归结为将项目内部评价和独立评价很好地结合起来。

美国国家纳米技术计划的评价可以看作是内部评价和外部评价(独立评价)相互支持的成功案例。纳米技术计划的评价采用内部评价和外部评价并行的模式。内部评价由计划组织管理机构自行开展,总统科技顾问委员会(PCAST)成员组建评价委员会,每两年进行一次,报告以评价委员会主席的名义提交给时任总统。外部评价由美国国家研究理事会(NRC)主持,每三年进行一次,评价结果提交国家纳米顾问组,商务、科学和交通参议院委员会及众议院科学代表委员会^[5]。二者内容上互为补充,在时间上交替开展,取得满意效果。美国国家纳米技术计划从2001年启动以来,跨越了克林顿、小布什和奥巴马这三届总统的任期,计划评价一直坚持内部评价和外部评价并行的实施模式。

3 公共R&D绩效评价的成功案例: 国家自然科学基金资助与管理绩效国际评估

国家自然科学基金资助与管理绩效国际评估是

在国际国内产生了重要影响的公共R&D绩效评估^[6]。该评估于2010~2011年实施,其目的是对科学基金25年的资助与管理的问责和改进提供一个基于证据的支持,并从全球视野重新审视科学基金在中国国家创新体系中的战略定位,提高资助与管理的绩效。本文简要分析该评估是如何体现独立性、可信性和有用性这三个基本原则。

3.1 如何保证评估的独立性?

该评估采取“国内准备+国际评估”的模式。国内准备是指国家自然科学基金委员会(以下简称“基金委”)和财政部联合委托独立评估机构(国家科技评估中心)全面负责评估的国内准备,完成评估证据报告(《科学基金资助与管理绩效国际评估综合证据报告》,2011, <http://www.nsf.gov.cn/publish/portal0/tab112>);国际评估是指聘请国内外战略科学家组成评估专家委员会,以国内准备的证据材料为基础,结合国内调研和国际比较,独立得出评估结论,完成评估报告(《科学基金资助与管理绩效国际评估报告》,2011, <http://www.nsf.gov.cn/publish/portal0/tab112>)。在国家自然科学基金资助与管理绩效国际评估实施期间,我担任国家科技评估中心的副主任,并作为该评估的国内准备部分的主持人,负责评估活动的设计和 implementation 过程的技术支持。

该评估从制度和行为两方面充分尊重和保障了评估的独立性。在制度和组织层面,财政部和基金委从评估开始就做出了明确的规定,要求有关管理部门不得干涉评估活动的实施。国家评估中心作为外部专业评估机构,不受利益相关方的干预和影响。国际评估专家委员会成员与基金委没有利益关系;在行为层面,评估中心在实施过程中根据需要获取相关数据,自主地确定面访、座谈和问卷调查的对象,独立编写综合证据报告。国际评估专家委员会在国家科技评估中心的专业支持下独立开展工作,评估中心与国际评估专家委员会直接沟通与互动。

本人作为国内准备的主持人向国际评估专家委员会汇报国内准备的评估证据报告时,国际专家委员会主席Richard Zare(斯坦福大学化学系教授,美国国家科学理事会前主席,美国国家科学院院士和中国科学院外籍院士)首先要求回答几个关于评估独立性的问题(主要是行为层面的),例如:在你们准备评估证据的过程中,是否能够获取评估需要的数据信

息,基金委是否配合?在你们选择面访、座谈和问卷调查对象时,是否受到限制,例如无法接触某些调查对象?你们编写证据报告时,信息的选择和证据的分析是否受到基金委及其他方面的影响和干预?

需要说明的是,评估的独立性并不意味着管理者完全与评估活动隔离。基金委作为被评对象,能够按照评估要求,全面收集并提供历史资料、典型案例、项目申请及资助数据、人力资源、财务收支等内部管理信息,但对评估过程和评估结论不加干预。

3.2 如何提高评估的可信性?

该评估采用“从关键问题出发、基于证据”的方法。评估设计了反映基础研究特点和公共R&D绩效评价需求的关键评估问题,针对关键问题,提供相关证据。评估中心根据关键评估问题将收集的证据进行整理,对同一关键评估问题下不同来源的证据进行了交叉核实,最终确定回答关键评估问题所需的证据,并完成了综合证据报告。本次评估结果的核心内容由两部分构成:①综合证据报告提供针对关键评估问题的各类评估证据;②评估报告提供基于证据的评估发现、评估结论及建议。

对于科学基金资助与管理的评估,是否客观、全面地反映了各方面的意见,尤其是不同意见,对评估的可信性影响很大。该评估的问卷调查和面访对象除了包括各类受资助人员,还包括了未能获得资助的群体,特别关注了一个特殊的群体,即三年连续申请科学基金项目,但一直没有获得资助的科研人员。这个特殊的群体的声音往往比较弱,其意见难以得到充分的重视。为了在评估中客观地反映他们的意见,该评估对2008~2010年连续三年申请但都未获得资助的7838名科研人员进行了问卷调查,回收有效问卷1371份。召开各类座谈会、面访51场,涉及272人,其中42人属于上述连续三年申请但都未获得资助科研人员。对这类人员的调查,除了征询他们对科学基金项目管理、同行评议等共性问题的意见以外,特别询问以下问题:你认为你的申请连续被拒绝的主要原因是什么?你认为基金委给出的申请失败的理由是否公正合理?是否可接受?如果你收到了对你申请的评审意见,你认为这些意见对改进是否有帮助?

3.3 如何促进评估的应用?

国际经验表明,影响评估结果使用的主要因素

来自两方面:评估活动本身和评估的用户。来自评估活动本身的因素包括评估的时机,评估对于管理决策需求的针对性,评估的质量及评估过程中各方面的参与等。来自用户方面的因素包括用户对于评估的需求及应用评估结果的意愿,评估结果的反馈、使用与扩散等后续活动安排及其他管理制度方面的条件。

该评估在上述两方面都具有一定的有利因素。评估来自管理决策的内在需求,是基金委自己提出了进行独立评估的需求。基金委和财政部作为该评估的主要用户和委托者,从一开始就明确了要通过评估发现问题,改进管理决策,并对政府科技投入的绩效评估进行探索。基金委在保证评估的独立性前提下,为评估活动提供所需要的协调和支持,促进了评估的应用,这一经验也得到国际评估界的肯定。

由于评估的独立性和可信性得到保证,使得评估建议对改进科学基金的管理、优化资助结构和管理模式发挥了重要作用。例如:评估报告高度肯定了受益面最广的面上项目,并提出应坚定不移地把自由申请项目作为资助重点,始终保持面上项目作为基金项目的主体,扩大其资助强度和延长资助周期。基金委采纳了评估建议,将面上项目资助强度提高至80万元,资助周期从3年延长至4年。此项措施实施以来,取得了很好的效果。不仅扩大了受益群体,体现了鼓励科学家自由探索、培养创新能力的基本面,同时,相对稳定的资助,也有利于科学家减少申报项目花费的时间和精力。

4 结语

2017年6月,第599次香山科学会议汇集了我40名关注科技评价的资深科学家、科技管理和科技评价专家,“以加强科技评估,助力创新驱动发展”为主题进行深入的对话、交流和研讨。香山科学会议作为我国顶层的“科学前沿论坛”,这是第一次将科技评价列为学术会议讨论。受到香山科学会议上专家学者思想碰撞的启发,结合作者本人多年来关于国内外公共领域评价的研究和实践,本文试图针对科技评价和科技评价改革的一些基本的、有争议问题的进行讨论,希望对于促进科技管理者、科学家和评价专家之间的关于科技评价的对话有所裨益。

参考文献

- 1 Dai T, Li X X. Analysis on problems and reform perspectives of China's S&T evaluation (in Chinese). Policy Manag Res, 2013, 28: 750–755 [代涛, 李晓轩. 我国科技评价的问题分析与改革思路. 政策与管理研究, 2013, 28: 750–755]
- 2 Guo J. Reviewed on public R&D performance evaluation research (in Chinese). Sci Technol Manag Res, 2014, 1: 52–57 [郭嘉. 公共R&D绩效评价相关研究综述. 科技管理研究, 2014, 1: 52–57]
- 3 María B. Solutions related to challenges of independence, credibility and use of evaluation. In: Proceedings from the Third International Conference on National Evaluation Capacities, 2013
- 4 United Nations Evaluation Group. UNEG Norms and Standards for Evaluation. New York: UNEG, 2016
- 5 Committee to Review the National Nanotechnology Initiative of National Research Council. Triennial Review of National Nanotechnology Initiative. Washington DC: The National Academies Press, 2013
- 6 Zhen Y H, Chen Z Y. The roles of users in enhancing utility of evaluation. In: Proceedings from the Second International Conference on National Evaluation Capacities, 2011, Johannesburg, South Africa. Publication by UNDP Evaluation Office, 2012, 55–62

Summary for “关于科技评价和科技评价改革的讨论”

Reflections on science and technology evaluation in China

Zhaoying Chen

National Center for Science and Technology Evaluation (NCSTE), Beijing 100081, China
E-mail: chenzyhaoying@ncste.org

Currently, science and technology evaluation(S&T evaluation) has become a major concern in China. However, what is “S&T Evaluation”? What does the term “S&T Evaluation” mean internationally? What are the top priorities for the establishment of S&T evaluation system in China? For the issues above mentioned, the Chinese government officials, scientists and evaluators have not yet come to an agreement. The main idea this article discusses is that, given the significant expansion of public R&D funding in China, the need for understanding performance of the public R&D funding has become increasingly necessary and urgent. As a critical element in China’s innovation systems, public R&D performance evaluation seems to react more slowly than other aspects though it should be one of the top priorities for S&T evaluation in China.

This article focuses on three interconnected aspects for public R&D performance evaluation: independence, credibility and utility, which are intrinsically linked and support the establishment of national R&D evaluation systems. The author attempts to open a discussion on the current debates about independence of evaluation, including the criteria to assess the independence of evaluation functions, different levels and types of independence for different situations and self-evaluations vs. independent evaluation. The article provides a successful case of public R&D performance evaluation in China, the *International Evaluation on the Funding and Management Performance of the National Natural Science Foundation of China*. The case is the most influential R&D evaluation up to now, and represents good practice from China. The article focuses on the approaches or solutions that have been used in the case to ensure the independence, credibility and utility of evaluation. The facts and viewpoints presented in this case analysis are based on the author’ direct involvement in the real evaluation practice of the case.

S&T evaluation, public R&D, performance evaluation, independence of evaluation

doi: 10.1360/N972017-00976