

中国科学院研究所评价浅议

李晓轩* 石兵

(中国科学院评估研究中心 北京 100080)

摘要 中国科学院研究所评价,已经过近 10 年的实践与探索,在知识创新工程试点工作全面推进阶段开始之际,围绕研究所评价是否有必要继续下去、不同性质研究所是否可以评价以及应该如何评价等问题,仍然存在一些不同的看法。就这三个方面的问题进行了简要分析。

关键词 中国科学院,研究所评价

一 中国科学院研究所评价简介

中国科学院研究所评价始于 1994 年,由院科技政策局和科技政策与管理科学研究所共同成立的院情研究室承担当时的评价工作,每年在院属研究所采集其产出及发展方面的定量数据,按统一体系计算出评价结果。因该评价结果报告的封面为蓝色而称其为蓝皮书评价。蓝皮书评价受到了中国科学院乃至中国科学界的广泛关注。蓝皮书历年的评价结果成为研究所定位以及知识创新工程试点基地型研究所遴选的重要参考依据。1998 年以后,为配合研究所分类定位,蓝皮书评价又增加了基础研究、高技术研究与发展以及资源环境与可持续发展三类研究所评价体系。1999 年,研究所分类定位工作完成,蓝皮书评价只进行分类评价,单一体系评价终止。2002 年,进入知识创新工程试点序列基地型研究所的遴选工作基本完成。至此,历时 8 年的蓝皮书评价,完成了其历史使命。

1999 年,中国科学院开始对进入知识创新工程试点序列的研究所进行试点评价,试点评价单位为 8 个。同年,成立了中国科学院评估研究中心。2000 年,试点评价单位扩大到 36 个,2001 年为 59 个。这 3 年的试点评价体系主要由两部分构成,即目标完成度评价和“三性”贡献评价。因试点评价结果报告的封面为白色,又称为白皮书评价。目标完成度评价主要评价研究所的战略规划(即创新目标任务书)的完成情况,包括科技目标完成度评价和管理目标完成度评价,科技目标完成度由同行专家评价;“三性”贡献评价,是指在基础性、战略性与前瞻

性三个方面做出的重要贡献,涉及重大科技任务的承担、高质量科学论文、重要国际会议邀请报告、重大社会经济效益、人才培养重要成果以及重要的科技奖励等定量指标。“三性”指标标准高,蓝皮书评价所涉及的全部定量数据中只有少数属于“三性”贡献。为推动研究所创新文化工作,从 2000 年起,在白皮书评价中增加创新文化建设评价。2002 年,随着知识创新工程试点工作进入全面推进阶段,研究所评价体系又做了重大调整。蓝、白两个评价体系合并。历时 3 年的白皮书评价,也完成了其历史任务。

在调整 2002 年研究所评价体系时,按照“质重于量与分类评价”两个原则,分别增加了定性的重大创新贡献评价,以及针对基础研究、高技术发展与研究、资源环境与可持续发展研究、产业化四个方面的分类评价。

二 研究所评价是否要继续下去

我院实施知识创新工程试点工作以来,整体科研环境有了相当大的改善,科研人员工作的积极性极大地提高,大家把这一大好形势看作是我国科学界的又一个春天。在这种情况下,不少人希望院层面最好不要继续评价,少一些扰动,让科研人员安安心心做学问,做出成绩回报国家和社会。如何看待这种观点呢?

首先,从社会发展角度来看,现代社会越来越成为一个评价的社会。在企业界,著名的 GE 前总裁杰克·韦尔奇不仅以其“数一数二”的评价标准关闭、出售了 GE 的众多企业,导致 10 余万员工离开 GE,同时,韦尔奇还强硬地坚持 C 类会议,每年要有 10%

* 中国科学院评价研究中心副主任,副研究员

收稿日期:2003 年 3 月 5 日

的员工被评为 C 类而淘汰出 GE。在韦尔奇的理念里,评价就是追求卓越的手段。韦尔奇在 GE 的变革带来了 GE 近 20 年的高速增长, 他的巨大成功,使他成为各国企业管理者竞相效仿的对象。从政府管理的角度看,新的公共管理将绩效管理置于最为突出的地位。美国 1993 年通过的政府绩效和结果法案(GPRA)就是最为典型的例子。GPRA 要求联邦机构制定 5 年的战略规划,并提供每年量化的实施目标,即年度绩效规划,以便对照年度绩效规划进行评价,形成年度绩效评价报告,并将绩效评价结果与预算的批准过程结合起来。

其次,科技领域也不是绩效评价的禁区。科技评价大体经过了三个时期。数百年前,随着以实验和数学为基础的近代自然科学的确立以及科学研究机构的专业化,科技评价就被用来作为保证科学研究质量的重要手段。由于当时的科学研究活动尚处于自由发展时期,其科技评价活动也主要是科学家个人的活动。这一阶段可以称为自由评价时期。“二战”之后,科研经费的分配、研究方向的遴选、研究机构的调整等,成为西方主要国家科技政策的重要内容。这些相关政策的实施,广泛借助了科技评价手段。另外,自 20 世纪 60 年代起,经济合作与发展组织(OECD)的一些国家就开始尝试对政府支持的科学研究活动的投入、产出、成果和影响力等方面进行评价。这些评价大体是科技组织自发进行的,可以称为自发式管理评价时期。

近十多年来,各国政府都将科学技术作为提高国家经济实力、赢得军事竞争优势的重要手段,科学研究活动不再仅仅是科学家个人及其团体的兴趣性活动,科学研究的国家战略导向越来越突出。各国政府对科技活动的评价更加重视,科技绩效评价越来越被广泛地、制度化地开展起来。同时,随着信息技术的日益普及和广泛应用,使得大量数据的采集、统计、处理与建模越来越简便易行。在许多发达国家,绩效评价已进入国家和政府层面的法制化、规范化时期。

在美国,GPRA 生效后,研究机构与其它联邦机构一样要接受绩效评价。对于应用研究,绩效评价的推行没有引起争论。但能否在基础科学研究领域推行,则引起了美国科技界的广泛争论。对 GPRA

的担心与抵制主要有:(1) 重大科学突破的产生时机和发展方向无法预测;(2) 没有什么定量方法可以真正衡量基础研究的质量;(3)GPRA 会妨碍科学研究的自由探索。为此,美国科学、工程与公共政策委员会(COSEPUP)从基础科学评价的理论与方法方面进行了专门研究,提出相应的政策和建议,以保证 GPRA 在基础科学领域的实施。

日本在 20 世纪 90 年代连续颁布了几项法律法规,以保证科技评价得以制度化、规范化地开展起来。1995 年日本政府颁布《科学技术基本法》,以法律的形式确定了科技评价的地位;1996 年在《科学技术基本计划》中,提出要实行公正严格的评价制度;1997 年,又通过《国家研究开发评价实施办法指南》,提出了科技评价的基本框架,包括评价的目的、原则、基本程序与方法以及不同类型科技活动的评价特点与内容等。

中国科学院的研究所评价目前还不是来自国家法律的要求,而是属于自发行为。从以上科技评价的三阶段来看,这种研究所评价还处于科技评价的第二个时期,即自发式管理评价时期。其发展的方向应该是法制化、规范化时期,而不是反过来,再回到仅仅由科学家个人主导的自由评价时期。

三 不同的研究所是否可以比较

从蓝皮书评价到白皮书评价,研究所大排队的评价模式没有改变。是否可以按大排队的模式来评价科学院众多的研究所? 这些研究所各有各的特色、方向和基础,在谈到是否能够相互比较这个问题时,有人不客气地说这是牛和马的比较。言外之意,风马牛不相及也。有不少人持有这种观点。

牛和马比较的深层的哲学道理,需要而且值得继续探讨。但是,在实践中,我们经常不得不将牛和马相比较。比如,在企业里,要比较研发人员和销售人员的绩效,要比较研发人员和销售人员的职位价值以确定其工资水平。这是因为外部的信息难以获得,而且外部环境与企业自身的差异太大,难以做到牛和牛比,马和马拼。因此,在企业管理的实践与理论研究中,不断推出新的绩效评价模式,不断推出新的工资级别的职位评估方法,以求化解这种矛盾。除了前面提到的不得不比较的原因以外,这些评价模式与方法的基本假设是,不同员工对组

织的贡献是可以通过共同的指标进行比较的。

在白皮书评价中,目标完成度评价不存在不可比的问题。“三性”贡献指标主要反映不同研究领域的贡献,如高水平论文反映基础研究的贡献,授权发明专利反映高技术领域的贡献。不同领域可以有不同贡献,并不是要以同一指标相互比较。同时,评价体系经过统计分析调节后,基础、资环与高技术三类研究所在“三性”贡献上得分均值相等。因此,对于白皮书而言,评价的是牛的力气与马的速度。在 2002 年研究所评价中,增加了分类评价。这在一定程度上更加弱化了大排队方式的矛盾。另一方面,按院领导的部署,院评价中心正在进行研究所国际评价的研究工作。随着全球化的发展,国际科技界科技产出的共享信息将会更充分;同时,随着我国科技竞争力的增长,我院研究所国际地位的提 高,直接与国际上同类的著名研究机构进行比较将成为可能。

四 研究所评价方法中的问题

不论是企业管理,公共行政管理,还是科研管理,对绩效的追求与重视,往往使绩效评价成为众目关注的焦点。因此,绩效评价总是最容易发现缺陷与不足的地方。在高校里对教授的评价,有人称为是“农民计工分的方法”。中国科学院研究所评价一直是“众矢之的”。不过,来自研究所的批评与指正,已成为推动研究所评价方法不断完善和发展的重要动力。现在的研究所评价体系与方法依然存在很多问题,需要不断改进与完善。以下,我们选择同行评议、文献计量与经济计量三种典型的方法来谈一谈。

1 同行评议

同行评议是科学共同体公认的科学评价的最重要的方法,在科学的自由发展时期就有了同行评议。同行评议方法的进步在于,由专家代替外行拥有了对学术问题的判断决策权,从而使科学活动能够在一定程度上独立于政治和宗教。但是,自从它产生的时候起,一边伴随着赞誉的同时,一边也招致了各种各样的批评。批评主要集中在对它的有效性和公正性的怀疑。对其有效性的怀疑牵涉到真理标准问题,即同行评议能否真正有效地判断和评价科学事物,这种方法能否达到对评价客体的真理性

认识;对其公正性的怀疑则涉及到价值标准问题,即同行评议是否会因为受一部分人操纵而导致对另一部分人的不公平。此外,同行评议还存在评审成本较高,容易导致权威至上、马太效应以及近亲发展等弊端。

在中国科学院研究所评价中,白皮书评价一开始就引入了同行评议,请同行专家评价研究所科技目标完成情况。在 2002 年研究所评价体系中,同行评议涉及研究所科技目标完成情况评价、重大创新贡献评价以及创新文化建设评价等。与以往相比,在 2002 年评价体系中,定性的同行评议发挥了更大的作用。前 3 年白皮书评价的研究分析显示,同行评议结果总体比较好,但仍存在一些问题。主要表现在两个方面,一方面与评价专家有关,如评价分数过于集中在高分区,区分度不够大;个别专家打人情分,评价结果有失公允。另一方面,与创新任务书设定有关,不少研究所任务书过于宏观抽象,缺少定量与定性的参照标准,专家难以准确评价。为此,在中国科学院研究所评价中改进和完善同行评价办法包括,其一,加强专家库建设,逐步建立专家信誉档案。在合适领域如基础研究领域,引入国际专家评价。在国家自然科学基金委项目评审中,尝试引入海外专家评价,称作“掺沙子”,效果不错。其二,与任务书有关的问题,可能有两个方面的解决方案,一是对不确定性强的基础研究,弱化针对具体目标完成情况的评价,强化学术水平与学术影响的评价;二是对可预见性强的应用研究领域或国家战略导向的研究领域,增加研究所任务书的操作性描述,使之更适合评价与监督。

2 文献计量法

简单地说,文献计量法就是指出版物的数量和质量,质量用引证的次数来表示。在文献计量法中,最有名的有:SCI (Science Citation Index)、ISTP (Index to SciTech Proceedings)、ISR (Index to Science Review)、EI(The Engineering Index Annual)等引文库,我院也建立了自己的引文库 CSCD (Chinese Science Citation Index)。以 SCI 等引文库为基础的文献计量法是一种定量方法,但出版物在被录用以及引用时又是同行评议的结果,是一种大范围、无规定边界的同行评议的结果。伴随着计算

机技术以及 Internet 技术的快速发展,文献计量法在科技评价中的应用越来越普遍,成为国家、地区以及大的科研机构之间科研能力比较的重要指标。但是,文献计量法通常要采用大样本才具有统计学意义,对于小规模研究机构与个人在一段短时间内的绩效评价是没有意义的。其次,文献计量法还有一些不足之处,比如,一种刊物引文率高并不代表其发表的所有文章引文率都一样高,存在“搭车”现象;一篇论文引用率高并不是其自身科学质量的充分证据,有些引证甚至是用来提醒防止错误的。此外,科学家若具有非常专门化的研究兴趣引用率通常也会比较低。近年来,科技界针对 SCI 开展了大量讨论。赞成的一方认为,以 SCI 为评价标准,能够鼓励我国科研人员发表高水平学术论文,推动“中国科学家走向世界”,同时,能够避免同行评议中出现的不正之风;反对的一方认为,由于 SCI 方法本身是有缺陷的,以 SCI 作为主要的甚至是惟一的评价标准,将鼓励科研人员只追求数量,而不是踏踏实实做科研工作、力争做出重大创新成果,从而助长了科学研究中的浮躁风气和功利性趋向。这些争论恰好针对了文献计量法的两个不同的方面:积极的一面与消极的一面。

文献计量法也是中国科学院研究所评价中的重要方法。以 2001 年白皮书的三性贡献评价为例,院评估中心共设定了 341 种分属各学科领域的高影响因子的刊物,主要包括院文献情报中心金碧辉教授关于 SCI 刊物分区的第一区及第二区的部分刊物。在“三性”贡献中,论文得分占全部得分的 15% 左右。作为国立研究机构,以 SCI 刊物中高水平的部分作为研究所评价标准应该是与科学院定位相一致的,国家在基础研究领域要求科学院做出原始性创新贡献。这在一定程度上可以防止对 SCI 论文数量的过分追求。从这几年的实践来看,最大的问题在于如何做到既坚持高标准刊物导向,又符合不同研究所的特点。由于学科原因,有的研究所几乎无法在以上选定的刊物中发表论文,他们自己领域内即便是最好的刊物也可能没有进入 SCI 分区的一区或二区。这反映了 SCI 本身的缺陷。如何克服 SCI 的缺陷,反映科学院全部研究所高水平的科研工作,应该是中国科学院研究所评价指标与标准研究中

考虑的重要内容。

与刊物选择相关的一个指标是重要国际学术会议。与高水平论文一样,重要国际学术会议的邀请报告也是反映学术贡献与学术地位的重要指标,2001 年白皮书评价中选出了 97 种重要国际学术会议作为评价标准。与刊物不一样,重要国际学术会议选取的困难主要不在于学科领域不同,而在于如何确定会议的重要标准。目前院评估中心所尝试采用的国际会议选择标准还需要进一步研究。另外一个相关的问题是,研究所在所内研究人员评价中机械套用院对所的评价标准,将论文数量分解到人头。这是对文献计量方法的误用,导致部分研究人员对研究所评价的抵触。

3 经济计量方法

经济计量方法主要有成本效益法和生产功能分析法两种。其中,成本效益法主要考虑成本、效益、净现值和回报率。这一方法可以用来评价产业方面的应用性研究,但是难以评价基础研究及资源环境和可持续发展研究。因为基础研究不确定性大,其成本、效益难以测定,而许多资环类研究则是没有经济回报的。在我院研究所评价中,与成本效益法有关的指标主要是成果转让合同经费及所有者权益等。到目前为止,还没有考虑投入成本。

生产功能法是一种比成本效益法更直接的经济计量方法,主要通过产品、资本、劳动力以及研究中的经费支出来估计研究工作对企业附加价值的边际贡献。这种方法旨在直接计算研究的经济价值。但是,由于研究工作和附加价值之间的关系往往是非线性的与间接的,同时,又受到技术、生产及市场等诸多因素影响,因而很难准确计算。同时,计算所用的数据需要来自企业和社会,涉及企业和社会是否愿意提供数据以及其科学管理与信息化的水平,搜集数据存在相当的难度。即便在美国,这种方法主要还限于少数研究者以及咨询公司使用,很少正式用于对政府资助的研究活动进行评价。该法在科学院研究所评价中暂时还没有用到。目前,院产业局正在积极研究该指标,不过,只有数据支撑的问题得以解决,生产功能法才有望真正应用于研究所评价之中。