

·基金纵横·

# 关于对美国国家科学基金会基础研究 绩效评价若干方法的思考

沈新尹

(国家自然科学基金委员会,北京 100085)

科学研究成果的评价、项目计划乃至资助机构的绩效评价一直是科学界关注的热点。从管理层面的角度考虑,所关心的是科学投资对科学本身、以及对社会、经济和文化发展的回报——直接的或间接的,近期的或长期的,显现的或潜在的。对基础研究而言,其绩效评估比其他任务导向的项目计划更复杂,因其产出或影响往往是长远的、潜在的或间接的,不易为人们所认识。如何正确地评价基础研究的绩效是一项科学性、政策性很强的工作。评价指标系统建立不当,过于强调某一方面或忽视另一方面,容易带来负面的影响,造成误导,从而损害基础研究本身。因此,如何从基础研究的规律性、产出的特殊性出发,对可能体现基础研究绩效的各项指标进行客观、全面的分析,以建立一整套科学、合理的综合指标体系,是我们应该认真加以研究的。

我委在科学基金运作的长期实践中,对资助项目的成果评价方式和指标进行了积极的探索,从结题项目的评价分级到与项目资助挂钩,从分析国家自然科学基金、科技进步奖获奖项目与科学基金资助项目的联系到论文发表奖励的尝试性政策措施,体现了对资助项目研究成果、对科学基金投入回报的充分重视。尚不足的是,我们的评价工作目前大多局限于研究项目的层次上,从总体上,在深层次上系统地科学基金投入、某项大型项目计划的实施对学科发展、对社会、经济、文化发展影响和成效的评价不够,或者往往容易简单地采用统计数字加成果范例的评价模式,停留在一般意义上的工作或成果总结上,与国际上目前广泛开展的绩效评价工作在认识上和实践上存在一定的差距。

美国国家科学基金会(NSF)根据美国政府绩效法(GPRA)的法定要求,在每年向国会递交的经费预

算及年度绩效报告中,均包括有关 R&D 活动和管理层面绩效评估结果以及下年度绩效目标的内容。在 NSF 的许多有关文件和报告中,绩效目标的设立和评估是一项重要的内容。但对于基础研究,如何建立科学、合理的绩效评价指标体系,科学地开展绩效评估工作,对于已有 50 年工作实践的 NSF,也是一个需要经历多年不断摸索的过程。下面介绍的是 NSF 在评估基础研究绩效时所考虑的一些指标、要素和方法,以及在运用它们时应注意的一些原则意见和对我们的启示,对我委预算报告编制、成果管理、开展绩效评价等有关工作有一定的参考意义。

## 1 美国政府绩效法中关于项目计划产出与成效的概念

美国政府绩效法(GPRA)要求各联邦机构在每年向国会提交的年度预算和绩效报告中,提供有关项目计划产出(Outputs)和成效(Outcomes)的信息。GPRA 将项目计划实施所产生的可直接观察到的产物,如出版物、毕业生等称之为产出,而将其中期和长期的效果,如生产知识、改善国民健康和加强国家安全等称之为成效。与产出和成效相对应的是投入(Inputs),如研究人员的知识、使用的设备、仪器、设施以及其他支撑条件等。

在白宫管理和预算办公室(OMB)关于执行 GPRA 的指导意见中讨论了项目计划实施产生的影响问题。有些影响是长远才能显现、直接或间接的。这种影响可能是预期的或是预期之外的、正面的或负面的。

从国家科技管理部门的层面,就绩效评价明确提出科学的政策性指导意见,不仅对绩效评价标准本身的确定,而且对科学基金的总体运作,包括资助

本文于 2001 年 5 月 24 日收到。

项目评价标准、资助政策以及项目选题导向等产生重要的影响。对基础研究的绩效评价,既重视直接产出的一面,又看到其长远、间接、潜在的一面,将为科学家进行创新性、开拓性的研究工作提供更为宽松的自由探索氛围和更大的发展空间。

此外,在进行绩效评价时,采用什么作为产出和成效的指标还取决于资助机构的任务和目标,以及是从某项具体计划考虑还是从国家目标考虑。因此,对于科学基金组织而言,其任务和职责或某一项目计划的目标可以用于衍生出各自的绩效评价指标。反之,在进行绩效评价时,则必须对照和紧密结合科学基金的任务和职责以及项目计划的预定目标。

## 2 常用的研究成果评价要素、指标和方法及在基础研究绩效评价中的应用

NSF认为,目前研究成果评价通常采用的要素、指标或方法往往局限于研究活动层次,仅着眼于项目计划产出与成效的一小部分,未能与GPRA的绩效评价概念明晰对应,从而尚不适合直接用于机构层次的绩效评价,因而建议,联邦资助机构在考虑如何设计最有效的评价方法时,现有的一些研究成果评估方法只能作为一个参考起点。

下面是一些通常采用的研究成果评估要素、指标或方法:

### 2.1 有关统计数字

#### (1) 论文发表数

NSF认为,论文发表本身意味着科学研究的发现向公众领域的转移,在经同行评议的杂志上发表的论文表示了对其所含信息的正面科学评价。虽然当论文发表数与其他一系列指标相结合并进行分析时,可提供有用的信息,但其单独使用或在缺少有关绩效的其他方面的足够信息时,得到的将是一种不完整的或不准确的描述。例如不同学科论文发表率的不同可能只反映其发表论文习性倾向的差别、最小可发表单元的定义和合作模式等的差别,而不是真正知识产出率的差别。此外,仅仅采用论文发表数作为绩效指标可能影响论文发表的模式和发表率——刺激人们热衷于产生更多的论文而不是发现新的知识。

作为一种直接可观察的产出,论文数因其易于检索和统计,在我委的成果、绩效评价工作中,通常被视为主要的评价指标。需要清醒地看到,仅仅论文数本身只能对成果提供“不完整的或不准确的描

述”,这将有助于我们注意克服一种简单化的倾向,即重数量、重形式上被SCI等权威检索系统收录而忽视对论文内在质量(学术水平和科学价值)的认识。

#### (2) 论文被引用数

对于大尺度的科学评价,论文被引用数是不为人们所注意的。NSF认为,与论文发表数一样,在进行绩效评价时,对论文被引用数的使用 and 解释也应该谨慎。有些情况下,高的被引用数可能意味着负面的评价(如有争议的冷聚变结果),小团体之间或螺旋循环式的引用也说明不了科学本身。在不同的学科领域间引用率也不同。在许多领域,实验工作的引用率通常比理论工作高。有时方法学的论文常常获得极高的引用率,结果使引用率过低估计科学认识上的进展,过高估计纯粹的实验活动。此外,有些以著书为主要产出形式的领域(如一些社会科学),被引用数是一种不公正的价值评价,因为SCI仅仅收录杂志的参考文献。

论文的科学价值并不与被引用数成正比。论文被引用数比论文数更容易提供“不完整的或不准确的”甚至虚假的描述。因此对论文被引用数更应重视其内在的科学意义而不是形式上的数量。我们在实际工作中通常仅注意要求剔除“自引”,但对于广义的“自引”——合作者、师生间、小团体间的无实质科学意义的“友情互引”以及一般意义上的“罗列式引用”仍然等而视之,对我们真正需要了解的引用者对论文科学价值的评价更缺乏足够的分析。从引文分析可以获得有关论文科学价值的丰富信息,但这需要进行细致的分析,不是简单的列举被引用数就能回答的。

#### (3) 专利数

NSF认为,专利数、新的器件、计算机程序和其他发明的数量并不能说明一项科研项目是否在知识前沿进行世界水平的研究。但一些任务导向的机构可以据此了解项目计划与机构任务间的联系。如果专利数等用来评价基础研究项目,应该与其他信息结合在一起,作为更完整、更详尽的评价系统的一部分。在利用专利数时应充分认识其局限性。尤其是,对基础研究而言,其专利数和发明活动的其他指标一般是低的。

专利数等数量少,其年度变化是不稳定的。在一个短短的项目评价报告中将发明活动的度量指标纳入为数不多的综合性指标中,对基础研究将是一种危险的做法。对这些数量小变化大的指标,可以

采用滚动的3年平均值来加以处理。

根据国家赋予我委“资助自然科学基础研究和部分应用研究”的职责,在采用专利数作为成果、绩效评价指标时更应谨慎。应该根据学科的不同性质、不同项目计划的具体目标确定是否将专利数列为评价指标或给予不同的权重。

## 2.2 用同行评议方法进行价值评议

NSF认为,正是因为缺乏度量手段,目前往往采用同行评议进行价值评议,以提供有关绩效的重要信息。这类评议通常着眼于项目计划层次。因为各个有关机构正在按照GPRA的要求发展各自的评价方法,专家评议的方式如何设计和组织尚需进一步探讨。评价小组可以集中对一个机构的主要项目计划或一组相关的项目计划进行评价(覆盖期5年左右)。

NSF还认为,科学评价的可靠性和有效性完全取决于它们是如何组织的以及评价人的组成。一个评价小组必须具有与被评项目计划相称的能力,其成员必须是受人敬重和客观的,如他不会因为评价小组的结论可能影响今后对自己有关研究工作的资助而影响他的评价意见。只有在这种情况下,评价意见才是可靠的。为确信评价小组成员判断的客观性,原则上应主要考虑来自那些未受被评项目计划资助或介入遴选资助项目过程的研究人员的意见。

一个评价小组应该对被评项目计划是否在科学知识的前沿开展工作进行评价。此外,项目官员还可以尝试对项目计划对提高其他有关潜在能力的贡献进行评价,如对保持高素质科学研究人员就业的贡献、保证设备、仪器的运转以支持前沿科学研究的贡献等。项目计划的实施对向某项特定任务提供知识基础的贡献以及对实现国家目标的贡献也可列在其中。

此外,有意向使用项目成果的用户可以提供项目成果与应用的相关性和重要性的有关信息。他们展望的前景也可适当提及。对任务导向的机构,有意向的用户是预期使用项目成果的工业、农业界等用户,而非任务导向的机构,如NSF,用户是指在该项目研究产生影响的领域工作的研究人员。

NSF关于同行评议的上述原则性意见主要集中于三个方面,即评价应与资助机构的主要任务目标紧密结合、注重项目计划是否在科学知识的前沿开展工作以及评价工作的可靠性和公正性。这些原则性意见对于我委在“十五”期间加强成果和绩效评价工作具有一定的参考意义。

## 2.3 国际地位

保持科学知识前沿的领先地位是美国投资科学的战略要素。就各个研究机构而言,评价准则是该机构是否在科学知识的前沿进行研究。从国家科学技术理事会(NSTC)或国家的角度考虑,则需要了解美国在国际上所处地位的信息。NSF认为,虽然已有一些数据和方法对各个国家的研究活动及其总体产出进行国际比较,但国际研究比较的方法仍是不成熟的。NSF特别强调,国际地位的评估并不仅着眼于单纯的数字排序,对领先地位的兴趣注意在于要使美国的研究和教育计划在与其他国家的科学家的竞争和合作中进行。

基础研究国际化的特点和趋势要求我们对我国各类学科基础研究的国际水平和地位有一个清晰的认识和分析。尽管国际上的比较研究方法尚不成熟,但我们在这方面的系统分析和研究更为薄弱,尤其是缺乏系统、完整的背景资料的长期积累。我们应在充分利用国际上已有数据(如NSF定期编辑出版的《科学技术指标》、世界银行的《世界发展报告》中的有关内容等)和成熟方法的基础上加强这方面的工作。知己知彼,才能使我国的基础研究真正置身于世界科技发展前沿的国际舞台上。

## 2.4 关于对其他政府目标的贡献

NSF充分重视一个项目计划实施对政府其他目标可能的贡献。这种贡献是与项目绩效相关的,不管它们是否是项目的预定目标之一,都可以纳入项目绩效报告中。对这类贡献的度量一直在进行尝试,如一项研究项目实施结果作出的其他贡献,包括人力资源和基础设施的发展、不同学科间和不同部门间合作关系的建立、在科研活动或非正式科学教育活动中参与的大学生数目等。

研究活动的产出有些可以从发表的报告中获得,有些可以从项目负责人的项目总结报告中获得。如果数据是从个别的项目负责人和项目官员获得的,必须在机构的层次上对其进行整合并纳入报告。同时也需要指出,并不是所有的项目计划都需要对实现一个机构的所有目标有贡献。明确这一点在管理上是有利的,这将有助于保证研究人员有进行创新研究所需的灵活性。

不仅从有关直接产出的统计数字,而且从对国家其他目标的贡献和社会影响的侧面评价成果和绩效,为我们从更广泛、更长远的角度来考虑和分析基础研究的绩效提供空间。这对于我委一些不以论文为主要产出的项目计划尤为重要。当然,论及“对国

家其他目标的贡献和社会影响”也需要提供确切的根据和充分的背景资料并进行实事求是的科学分析。此外,在论及“对国家其他目标的贡献和社会影响”时,应将各个有关项目的贡献整合、提升到科学基金投入和项目计划的整体层次,各个有关项目的贡献应是在整体层次框架下的有机表述。

## 2.5 经济学家发展的其他评价方法

NSF介绍了一些由经济学家发展的其他评价方法。有些方法试图评价科学研究带来的利益或回报,一般都是直接或间接地将科学研究产生的知识与在实际应用中运用这些科学知识而最终产生的利益联系在一起。但目前已有的经济学方法和数据只足以评价基础科学重大成效和影响的一个侧面。在联邦政府众多的各类项目计划中,经济学的方法也许只适合评价一些任务导向的项目计划,至多适合于评价不直接面向特定应用的项目计划。在方法和数据允许的情况下,经济学的方法可以用来度量研究活动带来的收益的大小和意义。

“评价科学研究带来的利益或回报”的经济学方法曾用于对美国国立卫生院(NIH)的生物医学研究和国家标准技术局(NIST)的计量学研究带来的经济影响进行研究评价。鉴于经济学方法只适合评价一些任务导向的、或至多适合于评价不直接面向特定应用的项目计划,对我委而言,经济学的方法也许仅适合对某些个别的项目计划进行尝试。

## 2.6 关于对未来利益的贡献

决策者和政策制定者有时需要知道,如果对某方面的一系列研究进行投资,未来的期望是什么? NSF认为,现在还没有一种方法能度量科学研究给未来带来的利益。其部分原因至少是因为未来的模式和科学研究影响的进程是未知的。因此,并不试图去发展某种度量对未来利益贡献的手段。

就基础研究而言,与对已实施完成的项目计划带来的回报进行评价不同,由于“未来的模式和科学研究影响的进程是未知的”,预测尚未实施的项目计划将给未来带来的回报是不现实的。正如我委关于“十五”计划纲要的报告中指出的,“基础研究具有很强的探索性和不确定性,一项研究工作能否成功,什么时候、在哪个方面取得成功往往是不可预计的。一些重大科学发现是由某些意外现象所激发而产生,这是难以准确地计划、规划的”,充分认识基础研究的这一发展规律,将有助于科学基金管理部门在研究制定项目资助计划和学科发展规划时更好地把握科学自身的发展规律和国家需求的关系。

## 2.7 在科学发展的长期进程中进行跟踪评价

NSF建议,绩效报告不必也不应仅依靠定量手段。例如年度绩效报告可以阐明在一个滚动的历史时期(例如在过去的20年中)朝向目标的进展情况;可以例举一些突出的或典型的研究成果;也可以设计一个描述性的案例,说明通过科学研究,知识的增长最终导致实际的应用,从而对国家目标作出贡献。

对基础研究的绩效,不仅看到其直接产出,而且将其放在科学发展的历史长河中、放在科学研究-技术-实际应用的进程中进行跟踪评价,有利于揭示基础研究长远、潜在的影响力。对我委实施时间较长的一些项目计划,如高技术新概念、新构思探索项目、数学天元基金等,也许可以进行这方面的评价尝试。

以上介绍的NSF关于基础研究绩效评价的若干方法,对每一种个别的评价方法既肯定其适用性,又指出其局限性;对基础研究的绩效,不仅从直接产出、从量化的统计数据,而且从更广泛、更长远的角度,在科学发展的长期进程中进行跟踪评价,其总体思路是值得借鉴的。

# NSF'S PRINCIPAL CONSIDERATION ON PERFORMANCE EVALUATION OF BASIC RESEARCH

Shen Xinyin

(National Natural Science Foundation of China, Beijing 100085)