

文章编号:1000-2995(2012)06-009-0137

基于循证设计的科学基金绩效国际评估研究

吴建南¹, 马 亮², 郑永和³

- (1. 西安交通大学 公共政策与管理学院, 陕西 西安 710049;
2. 西安交通大学 管理学院, 陕西 西安 710049;
3. 国家自然科学基金委员会 计划局, 北京 100085)

摘要:针对科学基金国际评估总体方案设计面临的方法论难题,本文尝试构建了循证评估设计方法,对其原理与应用进行了阐述,并通过科学基金国际评估总体方案设计展示了它的可行性。通过提取评估方案的关键构成要素并寻找支撑其内容的证据来源,构建了评估要素—证据矩阵并将之用于评估方案设计。研究表明,循证评估设计方法能为科学基金国际评估方案设计提供很好的支持。

关键词:国家自然科学基金;资助与管理绩效;国际评估;循证评估设计

中图分类号: N945.16

文献标识码: A

1 引言

作为我国基础研究领域的主要资助渠道,成立于1986年的国家自然科学基金委员会(以下简称基金委)在我国科学界享有盛誉。2008年初,基金委提出国家自然科学基金(以下简称科学基金)资助与管理绩效国际评估设想,对科学基金成立20多年资助与管理活动的整体绩效进行国际评估^[1]。尽管世界主要国家的科学资助机构都已开展了绩效评估或国际评估,但科学基金国际评估仍然面临严峻挑战^[2]。什么构成了科学合理的评估方案?评估方案设计面临着哪些难题?如何为评估方案设计提供令人信服的证据支持?这些都是此次评估亟待解决的关键问题。

文献检索显示,迄今关于科技评估的研究主要集中在科技计划和项目层面,对科学资助机构整体评估的研究还不多见^[3]。目前各国科学资助机构都在探索绩效评估,但针对这一问题的研究却还远远不够。欧美一些国家开始认识到这一问题的重要性,如欧洲科学基金会(ESF)自2007年以来四次召开“资助计划与研究计划评估”成员组织论坛。美国国家科学基金会(NSF)也在2008年召开了“科学政策的科学”八国集团研讨会,并首次邀请中国代表参加^[4]。虽然国际评估的意义已经得到了国内科学界的认可,但目前还局限于少数组织,如基金委中德科学中心^[5]、上海交通大学^[6]、中国科学院上海生命科学研究院^[7]等。

除了研究不足和缺乏经验外,科学基金国际

收稿日期:2009-12-07;修回日期:2010-06-11。

基金项目:国家自然科学基金委主任基金项目(J0910011,科学基金资助与管理绩效国际评估研究项目,2009-04至2010-03);国家自然科学基金项目(J0910012,国家自然科学基金资助与管理绩效评估总体方案研究,2009-04至2010-03);国家自然科学基金项目(70873092,目标责任考核、财政管理与组织创新——面向中国地方政府绩效改进的实证研究,2009-01至2011-12);西安交通大学腾飞人才特聘教授项目(2008-06至今)。

作者简介:吴建南(1970-),男,陕西西安人,西安交通大学公共政策与管理学院副院长、教授,研究方向:政府创新与绩效管理。

马 亮(1984-),男,江苏铜山人,西安交通管理学院博士生,研究方向:绩效管理、组织创新与科技评估。

郑永和(1964-),男,辽宁丹东人,国家自然科学基金委员会计划局副局长,研究方向:科技政策与管理。

评估还面临一些具体难题。就组织情境来说,评估涵盖 1986 年以来的 20 多年,恰与中国改革开放进程交叠,巨大社会变迁增加了评估的复杂性,并使绩效信息追溯变得困难。就其组织特征来说,科学基金的资助对象规模庞大、内容复杂,基础研究本身的复杂性也使评估难度增加。就评估本身来说,绩效是一个多维构念,绩效的多维性使其评价非常困难^[8]。综上所述,常规性的评估模式很难用于科学基金国际评估,因此需要从评估方法论入手探讨此次评估的出路。

科技绩效评估是一项专业性较强的活动,需要对科技本身有充分的理解。但与其他许多管理领域一样,绩效评估又是在日常普遍活动,绩效评估的专业性要求和应用普遍性导致研究和实践的脱节。尽管研究人员提出了许多复杂的评估理论和技术,但实践人员在评估中却很少采用^[9]。实践人员更多地是采用一种“集体学习”的方式,通过各方利益相关者的讨论,形成对绩效评估的共识,并据此制定评估方案^[10]。更有甚者,只是简单地“拍脑袋”、“定调子”,凭个人主观经验决定或照搬照抄其他评估方案,简单粗放,往往有违评估的需要^[11]。

绩效评估实践与研究之间存在的巨大鸿沟使人们反思二者关系,并催生了循证实践的产生^[12]。“证据运动”和循证实践在过去十余年得到了较快的发展,并成为各个领域争相应用的思维模式和行动范式。证据在词典中的意思是“表明某种信念或某个命题是否正确或有效的可以获得的事实或信息”。循证实践强调实践人员不仅需要依靠个人的工作经验,而且需要参考最新的研究成果,将研究融入实践,使研究指导实践。循证实践首先在循证医学中应用^[13],随后扩展到循证政策^[14]和循证管理^[15]。

循证评估是基于证据的评估,即评估结论的得出应该建立在经得起推敲和考验的证据的基础之上^[11]。绩效评估可以分为设计和实施两个阶段,关于循证评估的文献主要突出在评估实施中收集各种证据来支持评估结论,却很少探讨评估设计的证据支持。然而,如果评估方案没有证据支持,则根据该评估方案实施的评估也将无法得到证据的有力支撑。因此,循证评估设计指评估方案的设计也需要像评估实施一样,建立在可靠

证据的基础之上。

尽管对于如何设计评估方案,实践界与学术界进行了大量探索,但对于如何在设计过程中加以论证仍然还没有足够研究。有鉴于科学基金国际评估所面临的诸多难题,本文提出了基于证据的评估设计——循证评估设计,并论证了它的可行性及未来前景。这种方法不仅有助于解决科学基金国际评估方案设计的现实问题,而且对于推进科技绩效评估方法论的研究与实践也具有一定的理论意义。

2 研究设计—循证评估设计

评估方案的证据有哪些?它们可以为评估方案设计提供什么支持?评估方案的证据反映了评估委托方对绩效评估的需求以及评估活动的特点,能够为评估方案设计提供合法性、科学性和可行性的保障。综合来看,评估方案设计的证据来源包括评估委托方和利益相关者的评估需求、已有评估实践的例证、评估文献中对评估理论与方法论的研究成果、试评估与风险控制、评估方案设计人员的经验等,具体分为如下几个方面(如图 1 所示)。

(1) 评估委托方和关键利益相关者的支持。能否获得评估委托方的理解和支持,满足关键利益相关者的评估需求,是评估方案获取合法性中的关键^[16]。识别绩效评估的主要目的和关键议题,了解关键利益相关者的态度,并根据其建议和意见进行修改和完善,直至获得多数人的共识和支持,能为评估方案设计提供合法性依据^[17]。在征求意见时采取行动研究和行动学习的方式,既推动了评估方案设计的进程,也培养了设计和实施评估方案的实践人员,为评估方案的执行提供条件^[18]。

(2) 已有评估实践。绩效评估是“经验科学”,切忌“闭门造车”。如果针对对象 A 的绩效评估已经在类似的对象 B 中进行了实践且取得了成功,尽管二者在情境和特征方面存在一些差异,但却有着广阔的参考空间^[19]。而且这种相互学习也有助于创新的扩散,避免重复劳动,并使各种评估实践在保持多样性的同时实现融合。从实践中来看,通过考察学习、标杆管理等方式实现不

同评估案例的交流,也有利于绩效评估朝着“最佳实践”的方向迈进,并促进绩效评估的规范化和标准化。

(3)评估文献。从循证实践的角度而言,学术界所积累的大量研究文献是实践人员需要重点利用的资源,也是循证实践最初所倡导的思想。长期以来,学术研究与管理实践的脱节及由此形成的学术严谨性与管理关联性之间的巨大鸿沟,一直是管理研究的难题^[20]。循证实践正是对这种研究与实践相互割裂的现象的强烈响应,期望促进学术研究在管理实践中的应用^[15]。在科技评估领域,研究人员探索和验证的新理论、模型、方法和技术等,都是绩效评估需要吸收和利用的有力证据。这些研究成果的转移、转化和应用,既有助于学术研究发挥最大的功效,也为管理实践解决了许多现实问题。在评估方案设计中引入研究文献,可以避免评估设计的盲目性,增强评估设计的专业性和科学性。

(4)试评估与风险控制。评估方案设计本质上是一个复杂的过程,而非简单的线性模式,它需要以过程管理的理念进行,涉及不断的反复和试错^[21]。有必要对评估方案中的一些成熟设想进行小规模试验,对其可行性、成本和效果等进行初步估计,为评估方案获得证据支持。与此同时,在评估方案的设计过程中,需要有意识地思考评估风险的潜在来源及防范措施和控制手段,尽可能将评估风险消灭在萌芽阶段。

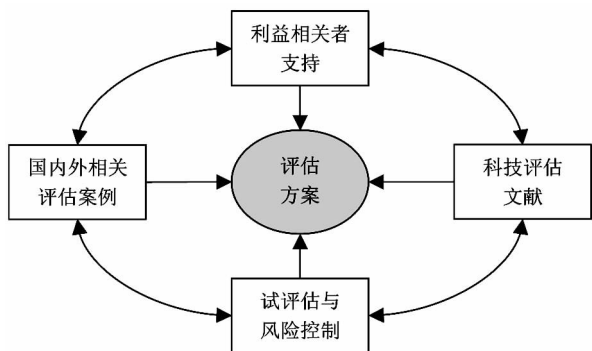


图1 循证评估设计框架

Figure 1 The framework of evidence-based evaluation design

除了上述证据来源以外,评估方案设计人员的经验也是设计的重要基础。评估经验主要是无

法显性化的隐性知识,是设计人员日积月累对绩效评估的领悟和理解,通常只可意会不可言传,但却是对上述证据的转化与整合,同样可以作为评估方案的有力支撑。如果上述证据都能得到贯彻落实,那么评估方案设计人员的调整空间和自由裁量权将不会非常大,设计的评估方案也不会有太大差异,而据此执行的评估也不会产生南辕北辙的结果。通过对评估证据的梳理,可以考察评估方案主要反映了哪些证据来源的影响,以及它们投射在评估方案上的“影子”,进而明确评估方案的倾向性和可行性。

3 结果与讨论

3.1 证据来源

结合循证评估设计的基本框架,我们构建了科学基金资助与管理绩效国际评估方案设计的证据来源,具体包括如下4个方面。

(1)评估委托方对评估的需求与目的,科学基金领导层的评估意图和科学基金的实际情况,是佐证评估方案是否合理的关键依据。基金委主任通过面向媒体的访谈和会议讲话等渠道传达了对国际评估的需求和期望实现的目的^[1],是评估方案设计的重要内容。针对科学基金,通过档案资料和深度访谈了解了其职能特点和数据基础,理解其资助与管理活动的实质和内涵,并掌握科学基金管理人员对国际评估的认识,从而使评估方案可以操作执行。最后,在评估方案设计的后期,多方面征求包括基金委领导层与管理人员、国际评估专家等在内的利益相关者的意见和建议,对评估方案进行不断完善。

(2)其他国家科学资助机构绩效评估的实践经验。因为评估方案设计的经验性非常强,针对类似对象或相关领域的评估实践是最有益的参考依据。许多国家已围绕科学资助机构绩效评估进行了探索,为科学基金国际评估提供了借鉴^[2]。2000年前后,针对科学资助机构的绩效评估成为国际科学界的一个热点,许多国家都实施了周期性绩效评估或国际评估,其评估报告和方法论成为科学基金的重要参考。结合这些评估案例的科学性及其与科学基金国际评估的相关性,我们参考了7个案例(具体见表1)。通过考察报告、评

估报告和私人沟通等形式,我们总结了上述评估案例的特点,并据此凝练了国际评估的原则和模

式,为科学基金国际评估提供了基础条件。

表 1 科学基金国际评估设计的证据来源

Table 1 The sources of evidence for international evaluation design of NSFC

证据标签	证据说明
基金委主任	基金主任陈宜瑜院士的讲话及接受新闻媒体的谈话资料
基金委访谈	2009 年 5 月对 10 余名基金委管理人员的深度访谈资料、2009 年 10 月对基金委科学部和职能局室的座谈会资料
DFG 2000	德意志研究联合会 (Deutsche Forschungsgemeinschaft, DFG)
RCN 2001	挪威研究理事会 (Research Council of Norway, RCN)
JSPS 2002	日本学术振兴会 (Japan Society for the Promotion of Science, JSPS)
FWF 2003	奥地利科学基金 (Austrian Science Fund, FWF)
SGCRP 2005	中德科学中心 (Sino - German Center for Research Promotion, SGCRP)
NSF 2008	美国国家科学基金会 (National Science Foundation, NSF)
FP6 2009	欧盟第六框架计划 (the Sixth Framework Programmes, FP6)
评估文献	科技评估研究文献,主要是理论和方法论层面的
试评估	科学基金战略定位的试评估方案设计和发现科学基金研究类、人才类和环境条件类项目的试评估方案设计和发现
风险控制	科学基金评估风险控制课题组的理论和案例研究发现

(3)科技领域绩效评估的研究文献。关于评估模式、框架、指标设计和信息收集方法等方面的研究文献都是支持评估方案的证据。尽管针对科学资助机构整体评估的文献不多,但面向具体研究项目和资助计划的评估研究却较多,如基础研究评估、影响评估、逻辑模型、科学计量学等,它们构成了科学基金国际评估的重要方法参考库。此外,针对评估目的、评估框架与关键议题、评估结果使用、评估预算等具体问题,已有文献也进行了广泛的研究,为此次评估提供了参考依据。

(4)试评估和风险控制。在评估方案设计的过程中,针对未来评估中必然涉及的一些对象和内容进行试点评估,衡量评估设计是否可行、数据能否收集、质量能否保障、成本有多高等,都为评估方案设计提供了可靠依据。与此同时,开展大量的理论研究和调研,对绩效评估中经常出现的风险予以事先识别和防范,并在评估设计中有所体现,可以避免评估走向歧途。

上述评估方案的证据来源构成了评估设计的重要依据,也成为论证评估方案的主要内容。通过对有限的重要证据的梳理和分析,能够充分支

持科学基金国际评估的设计需要。

3.2 评估要素

在上述评估证据的有力支撑下,我们需要明确科学基金国际评估方案的构成要素及其结构,也即国际评估方案及其各个部分“长什么样”。一般而言,评估方案需要回答的关键问题包括:为什么进行评估?谁发起评估?谁负责组织实施评估?评估哪些方面的内容?这些方面的关系是什么?如何收集信息支持评估?评估可以分为哪些阶段?评估结果是什么?如何使用评估结果?评估需要花费多少时间和经费?等等。上述问题的回答将形成绩效评估的若干关键要素,它们之间的关系构成了评估方案^[11]。结合上述评估证据,特别是公共部门绩效评估研究文献和国外科学资助机构绩效评估报告,我们认为科学基金评估方案包括如表 2 所示的评估要素。

3.3 评估要素—证据矩阵

针对每个评估要素,可以结合各类评估证据进行设计,并协调各个评估要素之间的相互关系并形成评估方案。评估要素与评估证据并非一一对应的关系:对于某个评估要素,可能需要不止一

个评估证据,需要设计人员权衡和选择;某个证据可能为若干评估要素提供佐证,形成多证据支持。因此,循证评估设计并非完全“迷信”证据,而是对各种证据进行甄别、评价和比较,选取并应用最适合的证据。

通过评估证据与评估要素的交互作用,产生

评估要素—证据矩阵(如表3所示)。矩阵左列是评估方案的各个组成要素,上栏是各类评估证据来源;该矩阵纵横交错,既说明了每个评估要素的证据来源,也显示了不同证据所支持的评估要素。

表2 科学基金国际评估的要素框架

Table 2 The framework of essentials for international evaluation of NSFC

评估要素	说明与解释
评估背景	评估发起的背景信息,说明评估的基础条件已经成熟
评估委托方组织者	评估的发起者、资助者组织实施评估活动的主体
评估对象评估目的	评估对象及其可能影响评估的主要特征评估期望实现的目标或达到的效果
评估框架	阐释各个评估内容之间关系与结构的理论模型,能够解释绩效的形成与高低
评估方法信息来源	评估需要的信息类型、来源以及获取这些信息的方法
评估主体	执行评估的主体,对评估的过程与结果负责
评估程序评估时间	评估过程需要经历的各个具体阶段及所花费的时间
绩效报告	评估的中间产品和最终产品,是评估信息的有效组合
评估结果使用	评估之后对评估产生的绩效信息的利用,藉此实现评估目的
经费预算	完成评估任务需要支付的财务资源

表3 科学基金国际评估方案设计的要素—证据矩阵

Table 3 The essential – evidence matrix for international evaluation design of NSFC

评估证据 评估要素	DFG	RCN	JSPS	FWF	中德 中心	NSF	FP6	科学 基金 访谈	科学 基金 委主任	评估 文献	试评估	风险 控制	证据 支持
评估背景								■	■				2
评估委托方 评估组织者			■		■				■	■		■	5
评估对象评估目的								■	■	■			3
评估框架	■	■	■	■		■		■	■	■	■	■	10
评估方法信息来源	■	■	■	■	■	■	■	■		■	■	■	11
评估主体	■	■	■	■	■		■		■	■		■	9
评估程序评估时间	■		■		■				■	■	■	■	6
评估结果使用	■		■					■	■	■		■	6
经费预算		■								■	■		3
绩效报告			■			■				■			3
提供证据	5	4	7	3	5	3	2	5	7	7	4	5	

(1) 评估背景。

要点:1993 年中共中央办公厅在温家宝同志的领导下对科学基金进行了专项调研;2007 年财政部对科技计划(基金)绩效评估的文件要求;全球化与信息化的背景下国际评估成为发展趋势。

证据来源:基金委工作人员访谈;其他科学资助机构国际评估的背景介绍(如 JSPS)。

(2) 评估委托方与组织者。

要点:科学基金资金来源的主管部门——财政部(特别是其教科文司)和基金委作为此次国际评估的委托方;基金委领导层、计划局、政策局、国际合作局、财务局、办公室等单位组成,通过国际评估领导小组的形式组织开展国际评估。

证据来源:科学基金陈主任的访谈;关于绩效评估中的委托—代理关系和促进评估操作化的文献^[16];风险控制的研究结果。

(3) 评估对象与评估目的。

要点:评估目的可以按照时间维度分为以下三个方面:总结过去、评估现状、展望未来。

证据来源:公共部门绩效评估文献中的评估目的分类^[22];与基金委工作人员的访谈,他们认为评估首先需要实现提出问题和改进绩效的目的;其他科学资助机构(如中德科学中心)的案例。

(4) 评估框架。

要点:此次评估采取“国内准备+国际评估”的模式。基于“目标—管理—效果”评估框架,对科学基金战略目标的相关性、资助与管理的适当性以及效率性、效果性与充分性进行评估,据此形成若干评估内容,对应于需要回答的评估问题,并提供支撑性信息以及信息来源和收集方式。

证据来源:逻辑模型和科技评估的通用术语^[23];其他机构(如 RCN 和 NSF)的实践;访谈中获悉的关键信息;试评估对评估内容的探索。

(5) 评估方法与信息来源。

要点:国内评估的具体方法包括:文本分析、问卷调查、访谈与座谈会、科学计量学、案例研究等。国际评估的具体做法包括:通讯评估、现场评估、会议评估。评估的信息来源包括:科学基金和公开数据库已有的数据来源;通过评估过程产生的数据。

证据来源:已有评估文献对数据来源的研

究^[24];其他多数科学资助机构评估中采用的方法和数据;试评估所采用的评估方法和信息来源;风险控制的研究结果。

(6) 评估主体。

要点:评估主体分为国内评估专家组和国际评估专家组。通过招标邀请国内外专家,建立若干个课题组完成评估的具体实施。国际评估专家以 10 位左右为宜。构成包括发达国家资助机构前领导、国际著名科学家、国内科技管理专家。

证据来源:评估文献对评估主体的利益相关性和知识拥有度的权衡^[8];其他国际评估案例的建议(如 JSPS、DFG 等);风险控制的研究结果。

(7) 评估程序与时间。

要点:评估时间以 1 年半为宜,国内评估大概需要 1 年,国际评估大约需要半年。

证据来源:来自试评估的工作量估计;其他国际评估案例的建议(如 JSPS、DFG 等);试评估所耗费的时间估计;评估风险的防范,避免时间不足。

(8) 绩效报告。

要点:充分考虑国际评估专家的信息需求,结合科学基金的绩效信息予以有机组合,并体现其风格。

证据来源:叙事分析和“绩效故事”方面的文献^[25];其他科学资助机构的范例(如 JSPS、NSF 等)。

(9) 评估结果使用。

要点:包括向上级报告和结果公开、绩效改进及供有关部门参考使用。

证据来源:评估结果使用的相关文献^[26];与科学基金工作人员的访谈;其他科学资助机构的范例(如 JSPS、DFG 等);风险控制的研究结果。

(10) 经费预算。

要点:综合国内准备与国际评估的情况,确定经费预算为 1000 万元。

证据来源:Horn 的评估预算设计方法^[27];试评估的经费估算;财政部文件《中央财政教科文事业专项资金项目委托评估评审及付费管理暂行办法》(财教[2004]110 号)规定对于委托评估额在 10 亿元以上部分的适用费率是 1%,以科学基金 2009 年的预算为准计算评估经费;RCN 国际评估招标价为 700 万元。

统计分析显示, JSPS、基金委主任访谈、评估文献所支持的评估要素最多, 均为7个。一些案例(如欧盟框架计划)只对少数评估要素提供了证据(2个)。就评估要素而言, 评估方法和信息来源、评估框架获得的证据来源最多, 分别是11和10个; 其次是评估主体的证据来源, 共9个; 评估背景获得的证据来源最少, 仅为2个; 而经费预算和绩效报告的证据支持也不多, 均只有3个。评估要素获得的证据数量从一个侧面说明了它的稳健性和确凿性; 而评估证据能够支持的评估要素数量也体现了其效用性。但这些数字仅供参考, 不能表明证据本身的解释力有多高, 也不能说明评估要素得到多大程度的支持。

为了验证评估方案的合理性, 我们面向评估专家发放调查问卷。结果显示, 专家对评估方案的许多要素都认同, 仅对评估框架和绩效报告等少数内容存在异议, 表明循证评估设计的科学可行性^[28]。我们还面向基金委工作人员召开座谈会征集意见和建议, 与会人员都谈到了国际评估的难度和挑战, 对评估方案的进展程度表示肯定并提出了许多建议, 构成了评估方案设计的新证据链。这些努力既是循证评估设计中不可或缺的构成部分, 也是对循证评估设计的一种肯定和支持。

4 结论

科学基金资助与管理绩效国际评估既有别于计划或项目评估, 也不同于常规性的绩效评估, 它对既有评估设计方法提出的挑战使我们不得不反思科技绩效评估的方法论问题。我们发现循证评估设计是一个值得利用和推广的框架, 对其原理与应用进行了阐述, 并通过科学基金国际评估总体方案设计展示了它的可行性。基于循证评估设计思路, 本文构建评估证据—要素矩阵, 探讨评估方案的构成要素与可资利用的证据来源, 以期促进评估设计的科学化 and 规范化。

本文发现, 评估方案设计的证据来源包括: 评估委托方和利益相关者的需求和支持、相关评估“最佳实践”的例证、评估文献中对评估理论与方法论的研究成果、试评估与风险控制等。科学基金国际评估的构成要素包括: 评估背景、评估委托

方与组织者、评估对象与目的、评估框架、评估方法与信息来源、评估主体、评估程序与时间、评估结果使用、经费预算、绩效报告等。结合证据来源, 本文对这些评估要素进行了初步设计。研究表明, 通过上述渠道设计的评估方案能够得到足够证据的支持, 这种设计与论证的过程也为追溯每个评估要素的来源与逻辑提供了依据, 并有利于得到评估委托方和关键利益相关者的认可。

任何实践领域都可能都存在着寻找特定的证据以支持其合法性和权威性的行动, 因此循证实践是人类的一种普遍行为, 只不过人们可能没有意识到他们是在寻找证据罢了。但从更为系统的角度来看, 循证实践实际上是一种“范式转换”, 将这种人人都可能体验的朦朦胧胧的潜意识显性化, 从而使人们从认知走向行动, 真正将知识转化为行动和生产力^[29]。对于科学和医学等“硬学科”来说, 循证实践的可行性更强, 因为科学共同体对科学范式和概念体系都达成了非常高的共识。而对于管理学等“软学科”而言, 概念杂多、体系混乱、情境多样, 循证实践在应用时还面临许多难题^[12, 15]。本文对循证评估设计进行了探索, 未来还需要围绕其理论、方法和应用等展开深入系统研究。

致谢: 感谢国家自然科学基金委主任基金项目课题组成员在历次研讨会上提出的意见和建议。依行惯例, 文责自负。

参考文献:

- [1] Chen Chen(陈晨). Chen Yi-yu: The Nation must ensure all disciplines of basic research being supported(陈宜瑜: 国家必须保障所有学科的基础研究都能得到支持)[N]. Science Times(科学时报), 2008-6-16.
- [2] Wu Jian-nan(吴建南), Ma Liang(马亮), Zheng Yong-he(郑永和). The Framework, Issues, and Methods of International Evaluation of National Science Funding Institutions: A Cross-national Comparative Multi-case Study(科学基金国际评估的框架、内容与方法—基于多案例的跨国比较研究)[J]. Studies in Science of Science(科学学研究), 2010, 28(5): 704-712.
- [3] Zhao Xue-wen(赵学文), Gong Xu(龚旭). Theory and Practice of Science Research Performance Evaluation(科学研究绩效评估的理论与实践)[M]. Beijing(北京): Higher Education Publisher(高等教育出版社), 2007.

- [4] Wu Jian - nan(吴建南). International Practice of Performance Evaluation of S&T Program (Foundation): A Report from G - 8 Research Assessment 2008 Workshop(科技计划(基金)绩效评估的国际实践—参加八国集团研究评估2008年会议的报告)[J]. Science Foundation in China(中国科学基金), 2009, 23 (4): 240 - 243.
- [5] Han Jian - guo(韩建国), Chen Le - sheng(陈乐生), Zhu Dong - hua(朱东华). International Evaluation Practices of Science Foundation — Evaluation of Sino - German Center for Research Promotion(科学基金国际评估的实践—中德科学中心评估工作)[J]. Science Foundation in China(中国科学基金), 2009, 23 (3): 173 - 176.
- [6] Zhang Jie(张杰). Thoughts on Mid and Long term International Evaluation of Colleagues and Departments(院系中长期国际评估的思考)[N]. Guangming Daily(光明日报), 2007 - 10 - 10.
- [7] Zhang Xue - quan(张学全). The First International Evaluation of Institutes in China(中国科研所将首次进行国际评估)[N]. CCTV International(央视国际), 2004 - 1 - 15.
- [8] Wu Jian - nan(吴建南). Public Sector Performance Evaluation: Theory and Practice(公共部门绩效评估:理论与与实践)[J]. Science Foundation in China(中国科学基金), 2009, 23 (3): 149 - 154.
- [9] Coplin WD, Merget AE, Bourdeaux C. The Professional Researcher as Change Agent in the Government Performance Movement[J]. Public Administration Review, 2002, 62 (6): 699 - 711.
- [10] Wu Jian - nan(吴建南), Yan Bo(阎波). The Path Choices of Local Government Performance Evaluation System—An Analysis of Fujian Province(地方政府绩效评估体系的路径选择—福建的分析)[J]. Chinese Public Administration(中国行政管理), 2008, (2): 25 - 29.
- [11] Forss K, Bandstein S. Evidence - based Evaluation of Development Cooperation: Possible? Feasible? Desirable? [J]. IDS Bulletin, 2008, 39 (1): 82 - 89.
- [12] Solesbury W. Evidence Based Policy: Whence it Came and Where it's Going[R]. London: Centre for Evidence Based Policy and Practice, 2001.
- [13] Sackett DL, Rosenberg WMC, Gray JAM, et al. Evidence based medicine: what it is and what it isn't[J]. British Medical Journal, 1996, 312 (7023): 71 - 72.
- [14] Sanderson I. Evaluation, Policy Learning and Evidence - Based Policy Making[J]. Public Administration, 2002, 80 (1): 1 - 22.
- [15] Pfeffer J, Sutton RI. Evidence - Based Management[J]. Harvard Business Review, 2006, 84 (1): 63 - 74.
- [16] Roessner D. Quantitative and qualitative methods and measures in the evaluation of research[J]. Research Evaluation, 2000, 9 (2): 125 - 132.
- [17] Stufflebeam D. Program Evaluations Metaevaluation Checklist [R]. Michigan: The Evaluation Center, Western Michigan University, 1999.
- [18] Dewar B, Sharp C. Using evidence: how action learning can support individual and organisational learning through action research[J]. Educational Action Research, 2006, 14: 219 - 237.
- [19] Hartley J, Benington J. Copy and Paste, or Graft and Transplant? Knowledge Sharing Through Inter - Organizational Networks[J]. Public Money & Management, 2006, 26 (2): 101 - 108.
- [20] Fincham R, Clark T. Introduction: Can We Bridge the Rigour - Relevance Gap? [J]. Journal of Management Studies, 2009, 46 (3): 510 - 515.
- [21] Neely A, Mills J, Platts K, et al. Performance measurement system design: developing and testing a process - based approach[J]. International Journal of Operations & Production Management, 2000, 20 (9/10): 1119 - 1145.
- [22] Behn RD. Why Measure Performance? Different Purposes Require Different Measures[J]. Public Administration Review, 2003, 63 (5): 586 - 606.
- [23] McDavid JC, Hawthorn LRL. Key Concepts and Issues in Program Evaluation and Performance Measurement[C]. Program Evaluation and Performance Measurement: An Introduction to Practice. Sage Publications, Inc, 2006: 1 - 38.
- [24] Arnold E, Balázs K. Methods in the evaluation of publicly funded basic research[R]. London: Technopolis, 1998.
- [25] Wu Jian - nan(吴建南), Ma Liang(马亮), Zheng Yong - he(郑永和). How do International Science Foundations Report Their Performance?: A Narrative Analysis on the Performance Report of Japan Society for the Promotion of Science(科学基金国际评估如何报告绩效:关于日本学术振兴会绩效报告的叙事分析)[J]. Science of Science and S&T Management(科学学与科学技术管理), 2009(12): 55 - 59.
- [26] Nagarajan N, Vanheukelen M. Evaluating EU Expenditure Programmes: a Guide[M]. Luxembourg: Office for Official Publications of the European Communities, 1997.
- [27] Horn J. A Checklist for Developing and Evaluating Evaluation Budgets[R]. The Evaluation Center, Western Michigan University, 2001.
- [28] Ma Liang(马亮), Yang Yu - qian(杨宇谦). Promoting International Cooperation and Enhancing NSFC Performance Evaluation—A Report on the International Conference of S&T Performance Management and Research Methods(加强国内

外合作交流 推进科学基金绩效评估—“科技绩效管理与研究方法国际会议”综述[J]. Science Foundation in China(中国科学基金), 2009, 23 (6): 359 – 362.

[29] Nutley S, Walter I, Davies HTO. From Knowing to Doing: A Framework for Understanding the Evidence – into – Practice Agenda[J]. Evaluation, 2003, 9 (2): 125 – 148.

International evaluation of NSFC's performance by evidence – based design

Wu Jiannan¹, Ma Liang², Zheng Yonghe³

1. School of Public Policy and Administration, Xi'an Jiaotong University, Xi'an 710049, China;
2. School of Management, Xi'an Jiaotong University, Xi'an 710049, China;
3. Bureau of Planning, National Natural Science Foundation of China, Beijing 100085, China)

Abstract: The Natural Science Fund of China (NSFC) has kicked off an international evaluation program since 2008 in order to assess its global funding and management performance establishing in 1986, which is faced with grave challenges. An evidence – based evaluation design approach is proposed to deal with the challenges for designing the overall international evaluation system of NSFC. The essentials of the international evaluation system are identified, appropriate evidences are searched to support them, and the evaluation essential – evidence matrix, which is applied to design the international evaluation system of NSFC, is constructed. The study reveals that the evidence – based evaluation design approach provides the sufficient support for designing the international evaluation system of NSFC, and could be widely used in other science and technology performance evaluation areas.

Key words: NSFC; funding and management performance; international evaluation; evidence – based evaluation design