

公共组织第三方评估何以可为？ 一项双案例比较研究

阎波

内容提要:文章通过对国家自然科学基金国际评估和全国高等学校一级学科评估双案例的比较分析,探讨了第三方评估方案要素对评估效果的影响。研究发现,评估定位、评估对象、组织机构和评估主体与第三方评估的独立性之间具有紧密关联,评估内容、信息收集、评估周期和评估结果的运用则关系到其科学性。机制设计不当的第三方评估会诱发机会主义行为。研究对迷信量化管理和第三方评估的“神话”提出了反思,并为改进公共组织第三方评估提供了启示。

关键词: 第三方评估 绩效评估 信息公开 问责 案例研究

中图分类号: D630.9

文献标识码: A

文章编号: 1003-3947(2018)03-0125-11

一、引言

自改革开放以来,第三方机构在公共治理中开始扮演越来越重要的角色(陈潭,2017),第三方评估也随之悄然兴起。人们通常认为,作为评估对象的公共组织是“第一方”,它所服务的公共对象是“第二方”,与这两方没有直接利益关系的即“第三方”;第三方评估就是由这些与评估对象具有弱关系的利益相关者来担任评估活动的组织机构或评估主体的价值判断活动(吴建南、阎波,2006)。随着国家治理能力的日益现代化和公民意识的逐渐觉醒,我国教育、科技、公益服务等领域的第三方评估日益崭露头角(廖秀健,2018;李艳、林秀玉,2016;马佳铮,2016;邵邦、范明超,2018),近年来更是在中央政府的鼓励下蔚然成风。仅在2015年就有民政部等多个中央部委和单位以及北京、湖北、广东、贵州等各地地方政府启动了第三方评估,教育部等中央政府部门还专门逐级下发文件,提出“要主动委托第三方开展全面、深入、客观的评估”。不仅如此,李克强总理在2015年的国务院常务会议上屡次提及重点政策落实情况的第三方评估,特别强调“决不能让评估报告‘束之高阁’”,客观上对第三方评估实践起到了显著的推动作用。

长期以来,学术界对第三方评估推崇有加,将其视为化解“市场与政府双失灵”的重要机制(赵环、严骏夫、徐选国,2014),认为独立的第三方评估能有效规避公共组织既当“运动员”又当“裁判员”带来的弊端(徐双敏、陈尉,2013),甚至能够解决“执行难”这样的痼疾(潘溪,2017)和遏制绩效目标走样(郑方辉、邓霖、林婧庭,2016),社会舆论和政府宣传更是对第三方评估给予一片溢美之词。但是,第三方评估可能仅仅是“看上去很美”,它并非包治百病的灵丹妙药,人们往往忽视了其作用发挥的机理而想当然地夸大“独立第三方”的正效应。事实上,学术界对第三方评估的“评估”仍处于探索阶段(负杰,

作者简介: 阎波,西安交通大学公共政策与管理学院副教授、博士生导师。

基金项目: 国家自然科学基金项目“问责机制何以奏效?面向公共部门政策执行的实证研究”(项目编号:71473190);全国学校共青团研究课题重点课题“学校共青团工作第三方评价机制与载体研究”(项目编号:2015ZD06)

致谢: 感谢中国人民大学公共管理学院马亮博士、中央党史和文献研究院闫健博士的修改意见。本文文责自负。

2015) ,对公共组织第三方评估的“冷思考”以及对其效果影响因素的探讨也相当欠缺(马亮、于文轩, 2013; Yu & Ma 2016)。为此,本文通过双案例比较来探讨公共组织的第三方评估何以有效及其关键影响因素。案例研究的优点之一在于其分析的深度(Gerring 2004),尤其适用于对某种现象的具体表现进行细致的经验描述(Eisenhardt and Graebner 2007),故本文采用案例研究方法对第三方评估效果的影响因素进行分析。

二、研究述评

在西方发达国家,第三方评估实践由来已久,但它通常被视为绩效评估的一种形式或公共部门管理的一个环节,专门针对第三方评估的研究并不多见。一般认为,现代意义上的公共部门绩效评估起源于美国,其历史可以追溯到19世纪和20世纪之交的“进步时代”(Progressive Era),而自20世纪70年代的“新公共管理运动”开始,绩效评估得以在美国等发达国家流行(Newcomer & Brass 2016)。绩效评估的出现具有深刻的社会和历史背景——当社会问题和公众诉求日益增加而公共部门传统的管理和治理工具效用不彰时,人们就试图采用绩效评估这样一种看上去相对新颖、科学和专业的手段来促使公共部门改善其作为,以应对全球化和信息化所带来的挑战(Abma & Noordegraaf 2003)。强调“独立第三方”的评估活动则恰好体现了这种美好的期望,其本质是一种试图借助外部专业力量形成“鲶鱼效应”的绩效问责机制。这就意味着,实施独立第三方评估的压力通常来自于组织外部,而非组织内部改进管理的需要。在行政问责机制失灵、公共问责机制缺失或公共组织与公共对象之间难以达成共识的前提下,凭借第三方机构的独立性和专业性来推动绩效评估,往往是一种兼顾成本与效果的折衷策略,因此在世界范围内日益风靡。

在实践中,“谁对评估负责”历来是绩效评估方案设计的一个关键问题(Talbot 2010)。有的公共组织依赖于内部专家进行绩效评估,有的则选择将绩效分析的工作外包给外部咨询专家(Carnochan et al. 2014)。国际评估专家罗西(Rossi)及其合作者认为,评估主体都应该是接受过专业训练并拥有基础知识的专家,而独立评估是指那些评估主体在制定评估方案、实施评估和发布评估结果方面承担主要责任的活动;评估活动的发起机构需明确规定评估的目的和性质,并委托评估主体开展具体工作,同时评估主体也可就评估内容与利益相关者进行磋商(Rossi et al., 2003)。美国联邦政府问责办公室(GAO)则认为,是否选择公共组织外部的独立第三方机构(如咨询公司、研究机构或政府的独立监察机构)实施绩效评估,主要取决于是否具备所需的外部专家和资源以及评估主体的独立性对于该项评估的可靠性有多大,而选择恰当的评估方案和获取可靠的数据信息是确保独立性的关键因素。它们指出,依赖于外部专家的评议和依赖于指标统计的科学计量(Bibliometric)是两种主流的评估方式(GAO 2012)。

但是,公共部门绩效评估被称为“世界性难题”,由专家参与的第三方评估也需要面对充满不确定性与目标模糊的公共治理环境。大量研究和实践表明,尽管绩效体制(Performance Regime)已经成为一个普遍现象,但绩效评估未必能够强化问责(Jones & Bouckaert 2017)。由于绩效评估实践中的多种利益相关者(例如部长、其它高级官员、技术专家、普通公务人员、大众媒体和公众等)在参与意愿、认知领会和时间范围等方面有着明显的不同之处,容易诱发互动博弈过程中机会主义行为的出现,即使是采用精巧的绩效指标和专家评估,也仍然难以避免绩效数据背离实际绩效的“绩效悖论(Performance Paradox)”(Pollitt & Dan 2013)。为了减少这种影响,美国联邦政府授权预算管理办公室使用项目评估定级工具(Program Assessment Rating Tool, PART)对联邦政府资助的项目进行第三方评估,对联邦政府项目给出简明的评估结论,但依然招致了评估对象和外界对其评分过于简单而缺少原因分析、评估结论有失主观、自上而下的“一刀切”式控制等批评意见(Moynihan 2008)。事实上,每个绩效管理体系都需要根据情境进行调适,尤其是在将绩效评估与绩效预算挂钩的情况下(Newcomer & Caudle 2011)。当绩效难以监测且与评估结果挂钩的财政激励与惩罚过强时,第三方评估依然难以避免评估对象弄虚作假等非预期后果(Girth 2017)。

当前, 第三方评估已经逐渐成为评价研究和绩效管理研究等领域的一个常见议题, 但遗憾的是国内对这一现象的探讨仍停留在探索阶段。吴建南等学者较早关注了由学术专家、企业和社会代表等参与的第三方评估实践(吴建南、庄秋爽 2004; 吴建南、阎波 2004), 一些学者主动参与到地方公共管理的第三方评估实践当中并对问题和经验进行了归纳, 或对第三方评估的实践做了述评(包国宪、张志栋, 2008; 郑方辉 2008; So 2014), 也有学者针对特定领域第三方评估的内涵、评估主体、评估程序等进行了分析(王春业、邓盈 2017; 陆汉文、梁爱有 2017; 寿志勤等 2010)。总体上看, 在当代中国公共组织绩效评估实践中, 学术专家等具有一定独立性和专业性的群体在绩效评估中的作用仍相当有限(Walker & Wu 2010; Jing et al. 2015; Yang & Wu 2013)。尽管如此, 第三方评估的价值和作用得到了许多学者的推崇(傅军 2013; 徐双敏 2011; 郑方辉、尚虎平 2016; 廖秀健 2018), 关于其独立性、机制设计及发展方向的讨论也日渐增多(程燕林 2017; 潘旦、向德彩 2013; 负杰 2015; 张旭、李会军、郭菊娥, 2016)。

其中, 一些对中国第三方评估实践的分析开始涌现。有学者对两项在中国影响较大、研究对象和内容近似的第三方政府公共服务绩效评价项目进行了比较和评价, 对评估结果在独立性、信度、效度、全面性、可理解性、功能性等方面的表现做出分析和评估, 指出地方政府应谨慎对待第三方评估、综合利用多种来源和渠道的评价结果(马亮、于文轩 2013)。相关的案例研究发现即使是由国外组织发起的第三方评估也面临着独立性、信度和效度等方面的挑战(Yu & Ma 2015、2016)。尽管如此, 国内外学术界对中国第三方评估实践的研究仍有待深入, 亟需更多丰富的案例研究和深入分析。为此, 本研究选取了国家自然科学基金国际评估和全国高等学校一级学科评估两个具有广泛社会影响的典型案例, 在对每个案例进行结构化描述的基础上开展比较分析, 归纳其政策启示。

三、案例描述

(一) 国家自然科学基金国际评估

自改革开放以来, 国家对基础研究的投入大幅增长, 其绩效状况日益引发了社会关注, 国家领导人和有关部委也多次提出要高度重视政府问责和绩效评估, 负责资助和管理基础研究项目的国家自然科学基金委(简称“基金委”)也需要回应各方诉求, 开展绩效评估遂日益成为共识。2008年初, 基金委六届一次全委会报告中明确提出“将适时启动科学基金资助和管理绩效国际评估工作”, 新中国成立以来最大规模的科技财政支出绩效评估随之被提上日程。经过一年多的专家团队研究和反复沟通, 在考察了美国、日本等发达国家国际评估模式之后, 基金委和财政部决定采取一种充分依靠第三方机构和评估专家的国际评估, 并于2010年初正式启动实施。此次评估旨在从历史发展和国际的视角, 独立、全面地评估国家自然科学基金自成立以来25年的战略定位、资助与管理绩效, 在总结和分析其成功经验和不足以及所面临的挑战基础上, 为改进科学基金的管理绩效提供对策, 并为明确科学基金在国家创新体系中的战略定位提出建议。

如图1所示, 基金委既是评估对象, 也是评估的委托方之一。由基金委与财政部双方领导组成的领导小组负责全面指导评估工作, 对相关重大事宜作出决策, 下设的评估工作小组及其办公室负责具体组织协调。它们授权国家科技评估中心(事业单位独立法人)负责整个评估活动的组织实施, 在委托方配合下负责准备评估所需证据材料。作为独立评估主体的国际评估专家委员会由来自六个国家的13位专家组成, 他们均在科技政策和科学基金资助与管理领域具有相当高的国际声望。评估专家从战略定位、绩效、国际合作、资助工具、管理等五个方面对科学基金资助与管理绩效进行了综合性、回溯性的评估。他们采取了“国内准备+国际评估”的循证评估模式, 即在前期外部研讨基础上设计总体评估方案和准备评估所需的证据材料, 进而提交给第三方的国际评估专家委员会, 由专家结合国内调研和国际比较进行评估(吴建南、马亮、郑永和 2012)。此次评估开展了大量调研、案例分析和大规模的文献计量分析, 调研工作涵盖了科学基金的主要相关主体, 包括对连续三年以上申请未获资助人员也进行了调查

和面访^①。

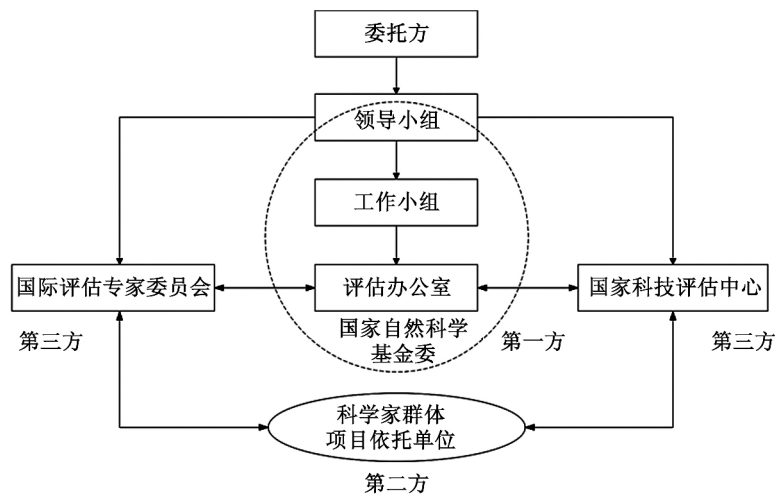


图1: 国际评估中的行动者关系

此次评估的结果引起了国内外的广泛关注,也对科学基金资助与管理产生了影响。根据评估专家建议,基金委从2012年开始设立“优秀青年科学基金项目”,为有科研潜力的青年研究人员设立了新的项目类型,并采取了提高面上项目资助强度、设立“重大非共识项目”等措施。国际评估的顺利实施得到了国务院领导、国内舆论和国际科学界的高度肯定^②,不仅相关报道见诸于人民网、《科技日报》和美国《科学》杂志等,也促使国务院加强了对各类科技计划的综合绩效评估,在《关于深化中央财政科技计划(专项、基金等)管理改革的方案》中明确“对科技计划(专项、基金等)的绩效评估通过公开竞争等方式择优委托第三方机构开展,评估结果作为中央财政予以支持的重要依据”。该评估之所以成功的一个重要原因在于,委托方从一开始就意识到绩效评估是“世界性难题”,而对基础研究绩效的评估更是难上加难,故而能够慎之又慎^③,不仅预先委托绩效评估专家组织多家专业机构设计国际评估方案,而且反复考察评估模式的適切性,并在国际评估中坚持独立评估的原则。

(二) 全国高等学校一级学科评估

改革开放以来随着高等教育的不断发展,对不同高校同一学科之间进行比较和评价的全国高等学校一级学科评估(简称“学科评估”)开始在诸多高等教育评估中日渐突出,成为最受国内高校关注和认可的第三方评估活动(袁本涛、李锋亮,2016)。2000年,在高等教育主管部门支持下,根据多位高校研究生院院长建议,学科评估拉开帷幕(林梦泉、姜辉、任超,2010)。经过一年多的研究和筹备,首轮学科评估试点工作于2002年4月启动,迄今已经完成四轮评估,评估对象数量逐轮大幅增加。从首轮评估旨在提高研究生培养和学位授予质量,到第三轮学科评估所强调的“服务大局、服务高校、服务社会”,再到第四轮又增加了“服务国际”,历次学科评估的定位不断变化。总体上看,学科评估越来越突出“向上负责”,将评估升格为落实国家教育发展改革战略、推动“双一流”大学建设的一项工具,同时也越来越强调扩大其影响力,向国内外展示中国教育发展改革的成就和学科建设成果。这些变化反映出发起者越来越希望以评估为依据重新配置教育投入资源,从而推动学科建设者强化自身创新能力和国际化竞争意识的绩效问责诉求。

学科评估从一开始就是在以教育部为代表的国家教育管理部门主导下进行的活动,其组织机构的

^① <http://scitech.people.com.cn/GB/15335456.html>

^② http://digitalpaper.stdaily.com/http_www.kjrb.com/kjrb/html/2015-06/02/content_305198.htm?div=-1

^③ <http://news.sciencenet.cn/htmlnews/2010/6/233583.shtm>

独立性历来饱受质疑。学科评估组织机构的前身是 1994 年在北京理工大学成立的评估所,伴随着高等教育体制改革的不断深化,最终于 2003 年成为挂靠在清华大学、具有独立法人资质的学位中心。尽管学位中心自称“一直以第三方方式独立开展评估工作”,且评估对象均系“自愿申请参加”,但许多学者不以为然,认为不仅学位中心与政府部门并未“脱钩”、二者利益关系紧密,而且参评高校与政府部门之间的资源依赖关系也使评估具有强制性(陈学飞等 2016)。而党的十八届三中全会明确提出“管办评分离”和 2015 年国务院《统筹推进世界一流大学和一流学科建设总体方案》强调“积极采用第三方评价”,也间接说明当前教育领域有待进一步提升第三方机构的独立性。不过,前两轮学科评估对外发布的均是《关于开展学科评估工作的通知》,而第四轮学科评估则已改为《全国第四轮学科评估的邀请函》,文件标题从“通知”到“邀请函”的改变的确反映出其行政色彩已略有淡化,而第三方评估色彩则更加鲜明。

表 1: 历次学科评估参评学科、单位数

轮次	批次	年度	涉及一级学科数	参评单位数	参评学科数
第一轮	第一批	2002	12	89	309
	第二批	2003	42	157	620
	第三批	2004	26	131	437
第二轮	第一批	2006	31	193	1067
	第二批	2008	50	248	1302
第三轮		2012	95	391	4235
第四轮		2017	95	513	7449

数据来源:中国学位与研究生教育信息网。

基于指标统计的排名是学科评估最主要的特征之一。尽管学科评估采用了主观评价与客观评价相结合的方式,但为了便于操作而采用了许多显性的量化指标,倚重根据指标体系量化测算的结果进行排序,使评估排名在很大程度上取决于指标数据。评估所需的客观数据信息由各评估对象提供给学位中心核查确认,这给信息公开程度参差不齐的评估对象增添了很大负担,也提供了操纵和包装数据的可能。由于评估结果是教育主管部门和各高校制定学科建设规划、经费投入与资源分配的重要参考,由此诱发了评估对象的大量机会主义行为。为了获取潜在的资源和维护学校声誉,一些学校甚至公开打出了“举全校之力决胜学科评估”的标语,抽调骨干人员参与应对,采取拼凑和取消学科专业、攒“学术 GDP”、卷入“挖人大战”等方式来争取获得好的评估名次。这使得有关学科评估的社会争议近年来愈演愈烈,第四轮学科评估中“A 类期刊名单”甫一发布旋即被教育部叫停的“熔断”现象即为一个明证。与一些发达国家相比,我国学科评估结果的使用与组织资源配置的关联更加紧密,产生了一系列负面效应(蒋林浩、沈文钦、陈洪捷、黄俊平 2014)。

四、比较与启示

大多数学术研究认为第三方评估存在多种不同模式(徐双敏 2011; 潘旦、向德彩 2013; 赵环、严骏夫、徐选国 2014; 殷明 2015),但它们大都停留在政策评估或项目评估的宏大理论框架上,对具体模式设计和分析语焉不详,或没有准确区分评估主体与组织机构的差别,更缺乏对不同评估模式效果的考察。实际上,公共组织绩效评估实践通常具有多元化的目的(Behn, 2003)和一系列关键议题(Hatry, 2013; Boland & Fowler 2000),包括评估定位、评估对象、组织机构、评估主体、评估内容、信息收集、评估周期、结果运用等影响评估活动效果的要素(吴建南 2009),对这些要素的不同组合就形成了不同的评估模式。在实践中,设计评估方案的“黄金标准”,就是根据特定的评估问题和预期的结果来选择适当的评估模式(Patton 2011)。然而,作为针对特定组织的综合性绩效评估活动,传统的项目评估或政策评估理论(Owen & Rogers, 1999)很难反映国际评估或学科评估这种复杂活动的特征,因而难以为分析和

选择评估模式提供有效的技术和智力支持。为此,本研究通过对上述双案例中评估方案的要素进行逐一比较,揭示两项评估在评估方案要素方面的诸多差异,如表2所示。

表2:国际评估与学科评估的比较

	国际评估	学科评估
评估定位	独立全面评估、促进绩效提升、明确战略定位	服务大局、服务高校、服务社会、服务国际
评估对象	国家自然科学基金委	高等学校及科研机构
组织机构	国家科技评估中心	学位与研究生教育发展中心
评估主体	国际评估专家委员会	国内专家、学位中心及其他利益相关者
评估内容	资助与管理整体绩效(战略定位、资助绩效、管理绩效、对中国科技体系的影响)	一级学科建设(师资队伍与资源、人才培养质量、科学研究水平、社会服务与学科声誉)
信息收集	国内准备+国际评估(循证评估)	数据报送+主观评价(指标统计排名)
评估周期	长周期(一次性)	短周期
结果运用	国家自然科学基金委改进资助与管理,国家加强对各类科技计划综合绩效评价	各高校学科规划与建设改进,中央/省级教育行政主管部门重新分配资源

首先从评估定位来看,二者都具有多元化的评估目的,但学科评估通常被舆论认为带有一定的行政色彩。除均强调促进评估对象的绩效改进外,二者还以不同方式回应了国家战略。国际评估更强调“提出一系列前瞻性的发展思路,支持财政部和基金委重新审视基金委在中国国家创新体系中的战略定位”,突出的是战略分析和调整,将评估的关键作用定位于出思路、给建议;学科评估则更强调“深入贯彻落实”《国家中长期教育改革和发展规划纲要(2010~2020年)》《统筹推进世界一流大学和一流学科建设总体方案》和“全面落实”国家研究生教育综合改革精神,突出的是战略落实和执行,将评估的关键作用定位于促使评估对象服从国家战略,带有更为明显的自上而下行政色彩。同时,学科评估更强调“服务”评估的各利益相关者,国际评估则关注的是“提出重要评估发现、经验教训和建议”以及“提出一系列前瞻性的发展思路”,也说明后者具有更加超然的评估定位。其次从评估对象、组织机构和评估主体的关系来看,学科评估较之国际评估更加复杂。第一,学科评估的名义对象是各参评单位的一级学科,实际对象是数百所全国高等学校及少数科研机构,其复杂性远远大于仅有单一评估对象国际评估。第二,尽管国家科技评估中心与学位与研究生教育发展中心均为独立法人机构,但后者的关键支持者是教育部,本身又是其直属事业单位,其独立性很容易受到质疑。第三,国际评估中的评估主体是由十余位国际顶级科学家组成的委员会,组织关系较为简明,且委托方主动公开承诺对此次评估“积极支持但不干预”,而学科评估主体则包括国内专家、学位中心、学生与用人单位和行业企业专家等,构成相当复杂,其中很难排除一些评估主体与评估对象之间的利益关联,从而弱化了评估主体的独立性。

再者从评估内容、信息收集和评估周期来看,二者反映出截然不同的评估理念。国际评估是一次性的长周期评估,在信息获取方面采取了“国内准备+国际评估”的组织模式,在推导评估结论时创造性地采用了循证评估方式;学科评估则是周期性的评估,采用了较为常见的指标统计和量化排名方式。从表面上看,学科评估的方式似乎更加简便、可控,但是用指标统计的量化方式来测量难以简单量化的评估内容,反而会因“绩效测不准”原理的存在和技术手段的不完备而出现事与愿违的后果(阎波、高小平,2013;渠敬东、周飞舟、应星,2009)。而国际评估所采用的循证评估是近十余年来备受推崇的评估方式,即根据反事实(Counterfactual)的理论观点,通过提供经得起推敲和考验的证据来得出评估结论(Forss & Bandstein, 2008; Heinrich, 2007),其关注焦点在于事实证据而非量化计分,更易于呈现丰富的绩效信息。

最后从评估结果运用来看,二者都对评估对象产生了重要的影响,但影响的方式不同,评估对象的相应行为倾向也不同。第一,国家自然科学基金委采纳国际评估结果,采取了提高面上项目资助额度、设立优秀青年基金项目等一系列新举措,进一步增强了信息公开程度;而学科评估结果向社会公开后,

许多高校领导都感受到了沉重的压力,由于相信学科评估结果将与教育部对高校的投入挂钩,各高校纷纷将学科建设列为工作重点,开始了新一轮的“学科建设竞赛”。第二,国际评估结果不涉及排名和直接的资源分配,评估对象只需要根据评估主体所提出的发现和建议改进工作,压力适度且主要集中于组织内部;与国际评估相比,学科评估对评估对象的激励作用相对更加显著,但无论是学科评估的支持者和委托方、组织机构还是评估主体又都难以约束评估对象的自利行为,造成大量的利益冲突和信任缺失,而这正是学科评估引发诸多争议的一个重要原因。

通过上述分析,我们可以更好地理解中国第三方评估效果的影响因素。这两项第三方评估实践具有很多相似特征,也存在明显差异。国际评估是在科研管理领域由第三方机构组织、少数独立第三方专家实施的一次性评估,而学科评估是在教育管理领域由第三方机构组织、多类第三方专家和利益相关者实施的周期性评估,二者都成为各自领域中第三方评估的典型案列。由于二者的背景及所要解决的问题不同,简单地评判孰优孰劣显然过于表面化。然而本文的案例比较表明,国际评估与学科评估分别代表着组成要素截然不同的第三方评估模式,其效果也迥然不同,其中的某些规律性现象尤其值得深思。

其一是第三方评估的独立性。在公共组织自我评估缺乏公信力、服务对象评估缺乏信度和效度的困境下,独立第三方评估被视为保证评估结果公正的重要途径,与“第一方”和“第二方”具有弱利益关系且专业的第三方机构就成为相对最佳的评估主体或组织机构(吴建南、阎波,2006)。然而,不同“第三方”的独立性存在差异。在评估定位方面,学科评估有着相对更加明显的行政干预色彩;在评估的关键行动者方面,国际评估中的评估主体构成相对简单且与评估对象的利益相关性较弱,评估组织机构与评估主体、评估对象的关系也相对更弱。在更深层次上看,对国际评估与学科评估起到重要支持作用的不同中央部门在机构性质和职能、组织规模、历史渊源、领导特征、组织文化等方面都存在着明显差异,很难排除这些差异与第三方评估独立性之间的关联。而确保独立性的根本,是使评估的方案、活动、模式、方法和评估结论避免不必要的偏差或影响,而不是看评估的组织机构具备何种身份(Newcomer & Brass, 2016)。换言之,独立法人、事业单位、专业研究机构等“标签”既不能等价于较强的独立性,更不能等同于高水平的评估信度、效度和公信力。国际评估中的“第三方”并不因其存在委托关系而弱化其独立性;相反,学科评估并不存在明显的委托方,但这也并不意味着评估主体的独立性就高。

其二是第三方评估的科学性。在传统观念中,独立第三方机构具有较强的专业性和权威性,其评估结果具有更强的科学性(徐双敏、翟胡,2012)。实际上,绩效评估是一项组织政治活动,评估内容及其信息收集的选择体现出组织机构及其背后的委托方或支持者对于“绩效”的期望以及对于如何获得高绩效的理解,其评估周期、手段和结果运用的差异映射出组织机构管理思维的差异(吴建南,2009)。例如,国际评估是充分倚重评估专家主观评价的循证评估,而学科评估采用的是以多元指标统计来整合多源信息的量化排名,“专家导向”和“指标导向”哪个更科学?要回答该问题,应该反思对科学基金资助与管理绩效和学科建设绩效本身的理解,科学地分析客观指标计量结果与专家主观评价结果存在差异的原因,以及哪一个能够更真实地反映委托方或支持者的预期。事实上,无论是充满风险、目标模糊和结果不确定性的基础研究还是充满探索性、难以标准化和直接测量的学科建设都具有特定的自然规律和生态效应,盲从于专家和迷信量化管理都与科学精神背道而驰,不应为了实现短期目标而一味追求简便易行的方式。

其三是第三方评估的局限性。尽管第三方评估能够弥补传统绩效评估的不足,但管理者不应过度夸大和迷信其有效性。实际上,评估只是绩效管理的一个关键环节,不能期望用评估来解决所有的管理难题(吴建南、阎波,2004),更不宜将给出复杂问题解决之道的希望寄托于第三方。评估可以实现监管督促的功能,但无论是对基础研究还是学科建设的长期绩效提升都很难产生直接作用,对第三方评估赋予过高的期望显然不恰当。更重要的是,在第三方评估实践中,由于信息不对称和强激励机制的存在,承受绩效问责压力的评估对象依然会表现出逆向选择和道德风险等机会主义行为,以追求短期收益的最大化,机制设计不当的评估甚至会加剧这些行为产生的负面影响。例如,由于预期学科评估结果与潜

在资源分配相挂钩,许多参评高校竭尽所能地应对第三方评估,不惜以“关停并转”非优势学科等方式来博取更理想的评估得分和排名,营造学术“大跃进”的繁荣假象,对正常的学科建设反而产生了不必要的干扰。事实上,如何解读和运用绩效信息往往比如何评估绩效更为重要,尤其是在具有治理偏好的管理决策者面对多元绩效目标的情境下(Christensen et al., 2018)。

总体上看,现实中公共管理者对第三方评估独立性、科学性与局限性的理解,反映出其特定的管理哲学。对公共事务和组织活动的管理,其核心仍然在于对个体动机的理解和尊重,激发人的主观能动性,这是使组织可持续发展和绩效不断改进的根本所在。这种“以人为本”的理念,与传统公共管理实践中强调完成目标任务的“以事为本”导向形成了鲜明对比。从表面上看,强调明确目标、量化指标、按部就班的绩效测量程序、精巧的加权统计和分类排名体现出公共组织管理的规则显性化和规范性,其背后隐含着韦伯式科层组织一以贯之的“以事为本”逻辑(渠敬东、周飞舟、应星 2009;周雪光 2015),但是这些显性规则和规范化管理并不能杜绝组织环境下关键行动者滥用其自由裁量权,以及与此有关的变通、串谋、违规等机会主义行为。尽管人们总是试图寻找更理想的评估指标,但指标导向的绩效评估总会使公共组织不可避免地付出代价,例如绩效短视、测量偏差、弄虚作假、僵化等,形成各种“管理主义悖论”(Johnsen 2005; Hood 2007)。因此,第三方评估也需要“祛魅”,以避免沦为“换汤不换药”的控制工具或新的“政绩工程”。

五、结 论

本文通过对国家自然科学基金国际评估和全国高等学校一级学科评估双案例的比较分析,展示了两项典型的第三方评估在背景、方案和效果上的差异,进而比较了二者在评估方案各要素方面的不同特征。之所以第三方评估受到社会舆论青睐并在公共管理领域流行,最重要的一个原因就是它具有相对较强的独立性和科学性。本项研究发现,在中国,评估定位、评估对象、组织机构和评估主体与第三方评估的独立性之间具有紧密关联,而评估内容、信息收集、评估周期和评估结果运用则关系到第三方评估本身的科学性,同时第三方评估也具有局限性。要使第三方评估发挥预期作用,就需要理性分析第三方评估实践的构成要素,在深入研究的基础上形成符合第三方评估背景要求的评估方案。在学术界普遍缺乏对公共组织第三方评估的冷静思考和深入分析的背景下,本研究力图弥补现有研究之不足,并期望为公共组织的第三方评估实践提供一些政策启示。

第一,深化对评估对象的理解和研究是设计第三方评估方案的起点。基础研究很难评估,学科建设也同样不易,类似的越是见效周期长、难度大、不确定性强的组织活动,越不易进行绩效评估。对于不确定性较高和目标较模糊的组织活动来说,在一定评估周期内,达到外部第三方所预期的指标要求未必等同于高绩效;反之,未能达到相关指标要求也并不必然意味着低绩效,此种情形下采用绝对化的指标量化管理很可能会取得适得其反的效果。

第二,在不具备高质量信息公开的条件下,应慎用第三方评估。第三方评估应以信息公开为前提,缺乏高质量的公开信息,会令评估活动的关键行动者疲于应付。如果缺乏可靠的数据信息,绩效评估很难实现促进学习、改进和问责等目的,而绩效信息和数据的可靠性也与评估的独立性息息相关。

第三,为确保第三方评估的独立性,应对第三方评估中关键行动者之间的关系进行合理设计。绩效评估活动中存在着不同的“第三方”,对组织机构和评估主体的选择需要格外谨慎和灵活。所谓“独立第三方”的关键并不在于其身份,而是要综合考察它的组织特征和背景因素对于评估结论的影响,尤其是要避免评估主体沦为“伪第三方”。

第四,当涉及评估对象众多、影响面广时,应谨慎将评估结果与资源配置相挂钩。面对数量较多、情况各异的评估对象,旨在试图通过将评估结果与资源配置挂钩来实现对评估对象的管理控制,很难避免激励约束不相容现象的产生。在第三方评估方案或结论存在争议的情况下,过度激励可能会加剧彼此主体间的利益冲突和信任缺失,这本身也与第三方评估的独立性存在矛盾之处。从长远来看,应将绩效

评估与绩效预算等其它改革举措纳入统一的管理框架,避免单一举措的盲动和冒进。

第五,在第三方评估的模式设计方面,应树立“以人为本”的理念,避免陷入“以事为本”的困境。作为一项问责机制设计,采取何种模式开展第三方评估,体现的是利益相关者对组织绩效的期望以及对于如何获得较高绩效的潜在认识。一味追求看似精巧的指标设计和量化管理,实际上很可能过度简化了组织绩效的形成机理,使第三方评估沦为巩固科层组织“以事为本”逻辑的技术工具。相反,现代意义上的绩效评估和问责机制更强调人而非技术的作用,不仅应充分尊重评估主体的价值判断,也应寻求对评估对象的正确引导和有效激励,树立“以人为本”的绩效管理理念。

参考文献:

- 包国宪、张志栋 2008: “我国第三方政府绩效评价组织的自律实现问题探析”,《中国行政管理》2008,1: 49—51。
- 陈潭 2017: “第三方治理:理论范式与实践逻辑”,《政治学研究》2017,1:90—98。
- 陈学飞、叶祝弟、王英杰等 2016: “中国式学科评估:问题与出路”,《探索与争鸣》2016,9: 59—74。
- 程燕林 2017: “如何保证第三方评估的独立性”,《中国科技论坛》2017,7: 14—18。
- 傅军 2013: “政府绩效评估的推进思路——基于六省市绩效评估的经验”,《理论探索》2013,3: 87—90。
- 蒋林浩、沈文钦、陈洪捷、黄俊平 2014: “学科评估的方法、指标体系及其政策影响:美英中三国的比较研究”,《高等教育研究》2014,35: 92—101。
- 李艳、林秀玉 2016: “国家治理体系创新与治理能力提升的新探索——广东省人大委托第三方评财政绩效的现实思考”,《公共管理学报》2016,3: 147—151。
- 廖秀健 2018: “‘对抗式’重大决策社会稳定风险评估模式构建”,《中国行政管理》2018,1:58—63。
- 林梦泉、姜辉、任超 2010: “学科评估发展与改革探究”,《中国高等教育》2010,21:43—44。
- 陆汉文、梁爱有 2017: “第三方评估与贫困问题的民主治理”,《中国农业大学学报(社会科学版)》2017,5: 103—110。
- 马佳铮 2016: “政府绩效第三方评估模式的实践探索与优化路径——以中国(上海)自贸区为例”,《上海行政学院学报》2016,4:17—25。
- 马亮、于文轩 2013: “第三方公共服务绩效评价的评价:一项比较案例研究”,《南京社会科学》2013,5:55—63。
- 潘旦、向德彩 2013: “社会组织第三方评估机制建设研究”,《华东理工大学学报(社会科学版)》2013,28: 16—22。
- 潘溪 2017: “第三方评估:适用于执行难问题的解决机制及其限度”,《理论月刊》2017,2: 56—60。
- 渠敬东、周飞舟、应星 2009: “从总体支配到技术治理——基于中国30年改革经验的社会学分析”,《中国社会科学》2009,6: 104—127。
- 邵邦、范明超 2018: “政府科技项目第三方评价的影响因素与发展对策”,《中国科技论坛》2018,2:10—14。
- 寿志勤、陈鑫、邹彦林等 2010: “安徽实验:第三方政府网站绩效评估机制研究”,《中国科技论坛》2010,2: 114—119。
- 王春业、邓盈 2017: “重要立法事项第三方评估机制研究”,《中南大学学报(社会科学版)》2017,6: 27—33。
- 吴建南 2009: “公共部门绩效评估:理论与实践”,《中国科学基金》2009,3: 149—154。
- 吴建南、马亮、郑永和 2012: “基于循证设计的科学基金绩效国际评估研究”,《科研管理》2012,33: 137—145。
- 吴建南、阎波 2004: “政府绩效:理论诠释、实践分析与行动策略”,《西安交通大学学报(社会科学版)》2004,24: 31—40。
- 2006: “谁是‘最佳’的价值判断者:区县政府绩效评价机制的利益相关主体分析”,《管理评论》2006,18: 46—53;58。
- 吴建南、庄秋爽 2004: “‘自下而上’评价政府绩效探索:‘公民评议政府’的得失分析”,《理论与改革》2004,5: 69—71。
- 徐双敏 2011: “政府绩效管理中的‘第三方评估’模式及其完善”,《中国行政管理》,2011,1: 28—32。
- 徐双敏、陈尉 2013: “‘第三方’评估政府绩效的制度环境分析”,《学习与实践》2013,9: 22—27。
- 徐双敏、翟胡 2012: “国外地方政府委托评估模式比较研究”,《学习与实践》2012,8: 72—79。
- 阎波、高小平 2013: “政府绩效管理创新中的‘样本点’”,《中国行政管理》2013,10: 21—26。

- 殷明 2015: “第三方评估: 走出价格市场化改革的困境——基于对石油和电力行业的观察”, 《学术界》, 2015 2: 210—218。
- 贡杰 2015: “中国地方政府绩效评估: 研究与应用”, 《政治学研究》, 2015 6: 76—86。
- 袁本涛、李锋亮 2016: “对我国学科评估发展的调查与分析”, 《高等教育研究》, 2016 3: 28—33。
- 张旭、李会军、郭菊娥 2016: “经济体制改革试点第三方评估理论基础与未来发展”, 《经济体制改革》, 2016 1: 5—13。
- 赵环、严骏夫、徐选国 2014: “政府购买社会服务的逻辑起点与第三方评估机制创新”, 《华东理工大学学报(社会科学版)》, 2014 3: 1—6。
- 郑方辉 2008: “第三方评价地方政府整体绩效的实证研究——以广东省市、县两级政府为例”, 《中国行政管理》, 2008 5: 13—17。
- 郑方辉、邓霖、林婧庭 2016: “补助性财政政策绩效目标为什么会走样? ——基于广东三项省级财政专项资金绩效第三方评价”, 《公共管理学报》, 2016 3: 122—134 + 59。
- 郑方辉、尚虎平 2016: “中国法治政府建设进程中的政府绩效评价”, 《中国社会科学》, 2016 1: 117—139。
- 周雪光 2015: “项目制: 一个‘控制权’理论视角”, 《开放时代》, 2015 2: 82—102。
- Abma, Tineke A. , and Noordegraaf, M. , 2003. “Public Managers Amidst Ambiguity: Towards a Typology of Evaluative Practices in Public Management.” *Evaluation*. 9: 285 – 306.
- Behn, Robert D. , 2003. “Why Measure Performance? Different Purposes Require Different Measures.” *Public Administration Review*. 63: 586 – 606.
- Boland, T. , and Fowler A. , 2000. “A Systems Perspective of Performance Management in Public Sector Organisations.” *International Journal of Public Sector Management*. 13: 417 – 46.
- Carnochan, S. , Samples, M. , Myers, M. , and Austin, M. J. , 2014. “Performance Measurement Challenges in Nonprofit Human Service Organizations.” *Nonprofit and Voluntary Sector Quarterly*. 43: 1014 – 32.
- Christensen, J. , Dahlmann, C. M. , Mathiasen, A. H. , Moynihan, D. P. , and Petersen, N. B. G. , 2018. “How Do Elected Officials Evaluate Performance? Goal Preferences, Governance Preferences, and the Process of Goal Reprioritization.” *Journal of Public Administration Research and Theory*. 28: 197 – 211.
- Eisenhardt, K. M. , and Graebner, M. E. , 2007. “Theory Building from Cases: Opportunities and Challenges.” *Academy of Management Journal*. 50: 25 – 32.
- Forss, K. , and Bandstein, S. , 2008. “Evidence – based Evaluation of Development Cooperation: Possible? Feasible? Desirable?” *IDS Bulletin*. 39: 82 – 89.
- GAO, 2012. “Designing Evaluations: 2012 Revision.” In *Applied Research and Methods*, edited by Kingsbury, N. R. , United States Government Accountability Office.
- Gerring, J. , 2004. “What is a Case Study and What is it Good for?” *American Political Science Review*. 98: 341 – 54.
- Girth, A. M. , 2017. “Incentives in Third – Party Governance: Management Practices and Accountability Implications.” *Public Administration Review*. 77: 433 – 44.
- Hatry, H. P. , 2013. “Sorting the Relationships among Performance Measurement, Program Evaluation, and Performance Management.” *New Directions for Evaluation*. 137: 19 – 32.
- Heinrich, C. J. , 2007. “Evidence – Based Policy and Performance Management.” *The American Review of Public Administration*. 37: 255 – 77.
- Hood, C. , 2007. “Public Service Management by Numbers: Why Does it Vary? Where Has it Come From? What Are the Gaps and the Puzzles?” *Public Money and Management*. 27: 95 – 102.
- Jing, Y. , Cui, Y. , and Li, D. , 2015. “The Politics of Performance Measurement in China.” *Policy and Society*. 34: 49 – 61.
- Johnsen, Å. , 2005. “What Does 25 Years of Experience Tell Us About the State of Performance Measurement in Public Policy and Management?” *Public Money and Management*. 25: 9 – 17.
- Jones, S. and Bouckaert, G. , 2017. “Performance for Accountability: The Case of Joint Agreements in a Federal System.” *Public Performance & Management Review*. 40: 765 – 96.

- Moynihan , D. P. ,2008. *The Dynamics of Performance Management: Constructing Information and Reform*. Washington D. C. : Georgetown University Press.
- Newcomer , K. , and Brass , C. T. ,2016. “Forging a Strategic and Comprehensive Approach to Evaluation Within Public and Nonprofit Organizations: Integrating Measurement and Analytics Within Evaluation. ” *American Journal of Evaluation*. 37: 80 –99.
- Newcomer , K. and Caudle , S. ,2011. “Public Performance Management Systems: Embedding Practices for Improved Success. ” *Public Performance & Management Review*. 35: 108 –32.
- Owen , J. M. , and Rogers , P. ,1999. *Program Evaluation: Forms and Approaches*. CA: Sage.
- Patton , M. Q. ,2011. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. N. Y. : Guilford Press.
- Pollitt , C. , and Dan , S. ,2013. “Searching for Impacts in Performance – oriented Management Reforms: A Review of the European Literature. ” *Public Performance and Management Review*. 37 (1) :7 –32.
- Rossi , P. H. , Freeman , H. E. , and Lipsey , M. W. ,2003. *Evaluation: A Systematic Approach*. L. A. : Sage Publications.
- So , B. W. Y. ,2014. “Civic Engagement in the Performance Evaluation of the Public Sector in China: Building Horizontal Accountability to Enhance Vertical Accountability. ” *Public Management Review*. 16: 341 –57.
- Talbot , C. ,2010. *Theories of Performance: Organizational and Service Improvement in the Public Domain*. Oxford: Oxford University Press.
- Walker , R. , and Wu , J. ,2010. “Future Prospects for Performance Management in Chinese City Governments. ” *Administration and Society*. 42: 34 –55.
- Yang , Y. , and Wu , J. ,2013. “Are the ‘Bigger Fish’ Caught? China’s Experience of Engaging Citizens in Performance Measurement System. ” *Public Administration Quarterly*. 39: 1 –33.
- Yu , W. , and Ma , L. ,2015. “External Government Performance Evaluation in China: A Case Study of the ‘Lien Service – Oriented Government Project’ . ” *Public Money and Management*. 35(6) : 431 –37.
- 2016. “External Government Performance Evaluation in China: Evaluating the Evaluations. ” *Public Performance and Management Review*. 39: 144 –71.

How Does Third – Party Evaluation Work? A Comparative Case Study from China

Yan Bo

(School of Public Policy and Administration , Xi’an Jiaotong University , Xi’an)

Abstract: This paper investigates how the third – party evaluation practices affect evaluation results by means of a dual – case structured comparison of evaluation schemes , including the International Evaluation of the National Science Foundation of China and the National Disciplinary Evaluation of Chinese Universities and Colleges. The findings show that the independency of the third – party evaluation depends on the purpose as well as the characteristics of the key actor involved in that , whereas the scientific nature of that relies on the contents , information gathering , and time horizon. In addition , an inappropriate mechanism of the third – party evaluation might trigger opportunistic behaviors. It implies that it is essential to rethink the myth of quantitative management and disenchant the third – party evaluation. Theoretical and policy implications are also discussed.

Key words: Third – party Evaluation; Performance Evaluation; Information Disclosure; Accountability; Case Study

(责任编辑: 闫 健)