

# 中国公共政策评估研究的方法论取向:走向实证主义

## 和经纬

**[摘要]** 中国公共政策评估实践尚不成熟,而且缺乏严密的方法论指导。在借鉴先进国家政策评估理论方法的过程中,应该确立起中国公共政策评估研究的实证主义取向。借鉴吸收国外主流的政策评估方法,使本土化的公共政策评估研究和实践能够经得起方法论的检验。主流政策评估方法的核心就是社会实验,通过分离政策效果的方式对政策进行评估。在资料收集处理过程中也有许多技术,要合理选择使用。政策评估的社会实验方法主要源自西方的方法论传统,在中国亦有其适用性。

**[关键词]** 实证主义; 社会实验; 方法论; 中国; 公共政策评估

**[中图分类号]** C93-03 D601 **[文献标识码]** A **[文章编号]** 1006-0863(2008)09-0118-07

### 一、研究意义

虽然公共政策的过程理论被作为一种过于模块化的僵硬模型已经广受质疑,但无论是作为政策链条中的一个环节,还是作为一个独立的活动,公共政策的评估都具有非常重要的意义,这是毋庸置疑的。在政策学习(Policy Learning)、政策调整乃至补充合法性上,政策评估都无法被取代。<sup>[1]</sup>

目前中国国内对公共政策评估的研究多数停留在政策评估的重要性和必要性、评估主体、评估的价值取向和伦理等方面,而对更为重要的评估方法却研究甚少。用国内学者李静芳的话说就是“目前从事官方评估工作的人士绝大部分都非‘内行人’,即使从事政策研究的专家学者大多数都是研究社会科学出身的,很少有人兼备自然科学和社会科学知识,其知识结构和思维方式比较单一,难以兼有思辨和公理化思维方式的长处。因此,他们在进行政策评估时不可避免地倾向于用价值判断代替事实分析,用定性分析取代定量结论。”<sup>[2]</sup>这个问题主要可能有三个方面的原因:

首要的原因在于中国特殊的国情以及目前的公共政策过程尚未给政策分析师和政策研究专家提供“用武之地”。朱仁显就将中国公共政策评估实践的问题归结为:“很多人把政策评估视为可有可无的事。即使进行了政策评估,也缺乏科学态度和方法,如,以评估作为炫耀工作绩效、沽名钓誉的手段;借评估使效果不佳、绩效不良的政策合理化;以评估来掩饰决策的失败或错失;以评估来攻击与自己偏好不一的政策;以形式取代实质;以先入为主的方式进行评估等。”<sup>[3]</sup>

第二个方面的原因在于,中国目前是否存在具有多学科背景和复合型分析技术,并能用社会科学的话语进行研究写作的“政策分析师”,本身还是一个问题。在国外,类似兰德公司、布鲁金斯学会、阿斯本研究所这样的第三方研究机构以及散落于欧美各大学的研究人员构成了政策分析师的主体,他们也是政策评估方法更新的主要源泉。<sup>[4]</sup>

第三个原因与公共政策学科的发展趋势有关。目前国际上主流的公共政策研究者主要由两个群体的学者组成,传统政治学家和具有很强的定量分析背景并受过专业训练的经济学家。前者中比较著名的包括 Frank Fischer(美国罗格斯大学教授)、Guy Peters(美国匹兹堡大学教授)、Michael Howlett(加拿大西蒙·弗雷泽大学教授)以及 M. Ramesh(新加坡国立大学教授)等,他们思辨与论辩的研究特质在主流的公共政策研究中虽然仍有很大影响,但主要局限在政策研究(Policy Studies)领域,而在专业化的政策分析(Policy Analysis)实践中,整个景象则主要是经济学主导,学理探究和思辨在这里基本没有市场。

具体的政策评估研究需要非常严密的方法论作保障,用国际通行的质量标准来检视中国的一些政策评估研究,在方法论上都或多或少存在瑕疵,有的甚至经不起基本的方法论推敲。马骏教授对中国公共行政学研究缺陷的总结其实很多也存在于政策评估的研究中。“许多似乎是实证研究取向的研究其实并没有运用现代社会科学研究方法。许多研究根本没有一个明确的研究问题,没有文献评估,没有深入地研究因果机制,没有构建出具有说服力的理论,更没有进行理论检验,或者理论观点没有经验事实支

\* 本文系 2008 年 3 月由中国行政管理杂志社和南开大学周恩来政府管理学院联合举办的“中国公共政策过程的理论与实践”研讨会优秀论文。

持。许多非实证研究取向的研究也并没有遵循诠释研究和批判研究的研究方法。”<sup>[15]</sup> 贡杰也认为，中国的公共政策评估还远未形成一套成熟、完善的方法体系，这在很大程度上制约了政策评估的科学化程度和规范化水平。<sup>[6]</sup> 就中国的研究而言，目前能够找到的中文文献中，对公共政策评估方法贡献比较大的往往是经济学、数学、系统工程等领域学者。但是，在公共政策（尤其是政策评估）学科智识的增长、积累以及方法论的验证、更新方面，国内不同学科之间的交流明显不足，使得很多已经得到检验的政策分析技术还未进入中国公共政策研究的“工具箱”和大学的教科书，因而造成了前文中李静芳总结的中国公共政策评估方法存在的诸多问题。本文试图在厘清政策评估方法核心问题并总结国外主流途径的基础上，结合中国的实际，提出适用于中国并经得起方法论检验的评估方法。

## 二、文献回顾

按照德国学者 Wolfram 的观点，政策评估作为一种分析工具，首要的任务就是为评价政策绩效提供政策过程和结果的信息。同时，政策评估又是政策循环中一个重要的阶段，它需要将这样的信息回馈给政策制定过程。<sup>[7]</sup> Wolfram 在他最新的文章中提出，“经典”的政策评估（classical evaluation）就是要分析政策的成果和失败，以及政策如何影响着现实世界。在他看来，政策评估本质上就是通过揭示因果机制来反映其有效性。<sup>[8]</sup> 荷兰学者 Bovens 等人将这种政策评估的取向称为“理性主义的政策评估”（rationalistic policy evaluation）。他们总结这种理性主义的评估哲学植根于实证主义，要把事实和价值截然分开，提供对公共政策绩效的“绝对客观认识”。<sup>[9]</sup> 以色列学者 Naclmias 的定义最能代表“理性主义政策评估”的基本前提。他认为，政策评估是对政策和公共项目是否实现预期目标的客观性、系统性、经验性检验。<sup>[10]</sup>

Howlett 和 Ramesh 指出，公共政策评估的研究多年来一直被所谓的“技术学派”所主导，他们把政策评估看成是价值中立的技术性活动。<sup>[11]</sup> 对于这些学者来说，公共政策评估的本质就是要通过经验研究来检验某项政策是否实现了预期的政策目标，如果没有，问题何在？如何解决？但是，有相当一部分学者对此表示怀疑，芝加哥大学的 Coleman 教授早在 1975 年就写道，“在未来的研究中，我不相信有一种综合性的研究方法能够对公共政策的效果做出评价。”<sup>[12]</sup> 世界银行的学者 Rist 也认为，无论评估的科学方法多么严密，都不应该是判断政策评估效果的标准。他认为，政策评估应该关注特定政策是否促进了全社会的治理，而不是评估方法本身。<sup>[13]</sup> 丹麦学者 Albaek 说的更为直接，他认为，那种把政策评估作为客观性和价值中立研究的观点，不过仍然是建立在社会科学定量实证主义（quantitative positivism）的基础上，这种观点在本质上还是把公共政策看成是理性决策的结果。<sup>[14]</sup>

在前文提到的公共政策研究者中的政治学家中，大部分都赞同类似观点。Anderson Kerr Manzer 等学者认为，公共政策本来就是高度政治性的，想要确定一个完全客观

的标准作为评估的参照几乎不可能。而且，不同评估者对评估的标准也会有自己的认知，实证主义的政策评估在现实世界中很难行得通。<sup>[15][16][17]</sup> Howlett & Ramesh 和 Ingram & Mann 也认为，政策目标往往是不同利益集团妥协的产物，同样的评估结果在不同群体中会有完全不同的解读，具有很强的政治性。他们就此认为实证主义政策评估以成本收益分析和风险评估为基础，借助经验数据的定量化模式不应是政策评估的主流方法。<sup>[18][19]</sup>

## 三、实证主义的方法论取向

有趣的是，目前能够找到的相关文献中持以上观点的几乎都是政治学家，而其结论也无外乎是说政治因素很重要，就算用实证主义假设检验的定量方法做出了分析，也会由于利益立场多元而无法得到解释，凡此种种。而被论辩学派（the argumentative turn）批驳的那些以经济学背景为主的政策分析学者却不事争论，而“默默无闻”地在 Journal of Policy Analysis and Management American Political Science Review American Economics Review 这类的主流刊物上一篇接一篇地发表他们的经验研究论文。应该认识到，第二部分文献回顾主要学者的研究都是立足于美国两党民主政治和利益集团博弈的假设之上，其观点有其很强的情境属性（contextuality）。国外经典文献中存在这样的“分歧”，对于目前需要学习借鉴国外先进理论方法的中国政策评估研究而言，要发展出一套符合社会科学基本质量标准并经得起方法论检验的评估方法，就必须回答公共政策的产出和影响是否客观可测这个重大的方法论问题。

发生在西方（主要是美国）学界关于政策评估方法的争论很容易让人感到困惑。对于仍旧处在介绍、消化西方国家理论方法阶段的中国公共政策研究而言，本土化的政策评估究竟应该走哪条路线？是循着影响日隆的实证主义走一条事实与价值分离的定量化道路，还是与批判实证主义的学者一起分析利益冲突和政治影响？我们需要做出选择。当然，有人可能会说“百花齐放、百家争鸣”，但是在公共政策的研究中，总要有一种主导的范式，在很多根本性的问题上，总要给出一个相对肯定的答案。直接回答上面的问题存在很大的难度，甚至要上升到认识论的高度。那么，我们不妨尝试着把这个问题放之于中国现代化进程和政治行政民主化的背景之下来分析。毕竟，学习借鉴西方理论的最终目的还是为我所用，通过知识的积累指导和推动中国的公共政策实践。基于这样的认识，我们必须厘清上面提出的重大方法论问题。

公共政策的产出和影响是否客观、可测，这是实证主义评估方法能否站住脚的理论前提。将其放诸于已有文献的语境下，其实就是问：实证主义的政策评估和政治学家们鼓吹的政治利益分析和事实价值结合是不是相互冲突矛盾？笔者认为，在先进国家的公共政策研究中，理论的发展有一定的历史脉络。在西方先进国家大多数重要的公共政策问题上，其实是先有实证方法的政策评估，后有对实证主义的合理质疑，两者并非水火不容；相反，通过这种学术的质疑问难，深化了对公共政策的理解。受篇幅所限，这里仅举一

例,是关于经济增长和社会不公的关系。传统的经济学理论和个人的直觉告诉我们,收入分配不均对于经济增长有负面效应。<sup>[20]</sup>在公共政策(主要是税收政策)的实证研究中,很多经验分析也证明了这一结论。但麻省理工学院的经济学教授 Kristin Forbes 则通过更加有信度的数据库以及更加严谨的固定样本分析,重新进行回归方程的参数估计并且顺利通过敏感度检验,证明初始条件下的收入分配不公对于后续的经济增长具有正面作用。<sup>[21]</sup>这样一项颠覆性的研究几乎是对过去美国收入再分配政策的否定,而且也与常人的经验常识相悖。

Forbes 及后续的评估很自然地招来了多方面的反对。然而他的研究具有很坚实的方法论基础,反对的声浪并非针对评估方法本身。对于社会低收入阶层而言,这样一个评估结果不但可能使他们的切身利益受到更大损害,在感情上也是无法接受的。就具有特殊利益的政府而言,更加害怕类似的政策评估带来难以预料的负面影响。这印证了 Howlett 和 Ramesh 所说的,公共政策很难有一个确定的目标,一个逻辑严谨方法得当的评估可能会受到很多利益的影响而陷于超乎技术层面的争论。可以看出,反对实证主义政策评估的理由并不是因为它的评估方法错误或者是评估技术不足,而是背后的政治意涵(Political implications)过于复杂,很难在公共政策的众多利益相关者中间形成关于政策结果的一致结论。因此这并不是否定实证主义政策评估方法的合理性,而可以看作是在实证评估的经验结论基础上进行的公共对话。在美国,仅仅一项弱势族群儿童的学龄前教育项目 Head Start 就在联邦和地方层面进行了数十次评估(主要是定量途径)。之后对评估实证主义的质疑已经不是 Head Start 是否真正提高了这些孩子的智力和认知水平,而成为是否要保护少数族群和弱势群体的政治对话,整整持续了四十年。<sup>[22][23]</sup>

而中国的情况则不同,除了一些事关国计或者影响特别重大的公共政策有比较严肃的评估之外,一般性的政策或者由地方政府主导的公共政策就很少有过于严谨的评估,这和欧美先进国家完全不同。对于中国而言,现在最为缺少的就是用实实在在的经验研究为公共政策的制定、执行、调整和终结提供有力的支撑;而在美国和其它先进国家,这样的研究称其为汗牛充栋,当不过分。中国治淮十年耗费 600 亿,不同地方、不同部门在认定淮河水到底有没有变清的问题上都说法不一,除了地方利益和部门利益作祟,政策评估本身也存在问题。<sup>[24]</sup>那种认为“西方国家很多政府兴师动众搞政策评估有浪费纳税人钱之虞,不适合我们发展中国家”的论调,当可止于淮河治理的例子了。

通过以上分析,我们可以理解为什么经典文献中出现的“争论”其实并不是实证评估方法是不是正确的争论,而是实证的评估结论能否被完全应用而作为政策调整或者终结根据的争论。美国学者 Gray 和 Jenkins 就指出,理性、科学的政策评估有时无法得到应用,问题并不出在评估本身,它在深层次上反映的是政治和组织的低效以及自利化的倾向。<sup>[25]</sup>也就是说,文献中的“分歧”其实不是实证主义是否

应该主导政策评估的问题。为了厘清这个疑问,美国学者 Guba 和 Lincoln 提出了一个较为符合人们经验认知的框架。他们认为,政策评估其实有四个阶段:第一个是技术阶段(technical),就是用客观的标准和严谨的方法论来评估政策在技术层面的达成程度;第二个阶段是描述性阶段(descriptive),即以第一阶段得到的经验结果为基础,描述政策的达成程度;第三个阶段是做出判断(judgment)及以前两个阶段的结果来判断公共政策是否实现了了许诺的政策目标;第四个阶段是价值多元(value pluralism)。在这个层次,众多利益相关者的利益诉求和公共政策的伦理价值才被纳入考察的范围(表 1)。<sup>[26]</sup>Guba 和 Lincoln 的框架对于厘清经典文献中的“争论”并论证中国公共政策评估研究取向应该走实证主义路线具有非常好的启示意义。

表 1 政策评估的四个阶段

时序	政策评估阶段	主要任务	方法导向	基本特征
↓	技术阶段	客观分析政策是否达到预设的技术目标	实证研究方法,主要是准实验研究(Quasi-experiments)	价值中立 客观严谨
	描述阶段	根据预设政策目标来描述公共政策的成绩和问题		
	判断阶段	对政策目标是否达成的总体判断	比照政策目标和前两阶段的分析	
	价值多元阶段	考虑多元群体的利益诉求,分析政治因素	价值多元主义	价值导向,不寻求绝对中立

资料来源:根据 Guba and Lincoln(1987),作者绘制。

这个框架首先把与政策评估有关的过程识别为四个前后相继的阶段,这基本符合人们的经验认识。通过这个有用的框架我们就不难理解,其实经典文献中关于政策评估方法的“分歧”主要是发生在评估的第四个阶段。也就是说,考虑多元价值、政治影响和利益诉求仅仅是在政策评估的最后阶段才发生作用的,完全没有否定这之前的政策评估过程应该严谨地遵循实证主义的评估取向。

用 Guba 和 Lincoln 的阶段论来看中国的政策评估,我们就可以发现其中的差距。中国公共政策评估的研究往往在没有走完第一个技术性环节的时候就开始考虑“价值多元”(利益团体、政府形象、社会稳定、GDP 党政关系、长官意志等等)。这大概是我们的政策评估“缺乏说服力和科学性”的重要原因。可以说,中国的政策评估研究在很多方面都缺乏方法论的有力支撑,实证研究能力较为薄弱。负杰就认为,目前的中国公共政策评估“学科基础薄弱,缺乏对其他学科分析方法的借鉴和整合,缺乏具有自身学科特色的方法论体系;政策评估的价值标准和技术标准的筛选和处理不够规范;在评估的具体方法方面,尚普遍存在‘两多两少’的问题,即经验推理多,实证研究少;定性分析多,定量评估少。此外,在政策评估前后比较分析时采用的样本数目较少,代表性不足,有时甚至缺乏必要的实证检验,使评估的可信度与效果大打折扣。”<sup>[27]</sup>因此,对于中国的公共政

策评估理论和实践而言,远没有到批判“经济学帝国主义”的阶段,而应积极学习经济学和社会学的实证主义思维方式和研究方法,首先遵循基本的方法论质量标准,增强评估研究的说明力和解释力。

总体而言,笔者认为,从主流政策评估方法的趋势和中国研究者在评估方法上的落后局面综合来看,大力学习、应用实证主义以假设检验、经验数据、定量分析为主要内容的政策评估方法,应该是未来中国公共政策评估方法的努力方向。

四、实证主义政策评估方法的主要思想:社会实验

公共政策的本质就是对社会生活的一种干预(intervention),那么政策评估也就是对这种干预效果的评价。在交叉重叠又有多元客体的政策系统中,要评估某项公共政策的唯一途径就是分离政策干预的效果,然后对其进行分析。在一个简化政策系统模型中,关系如图1:

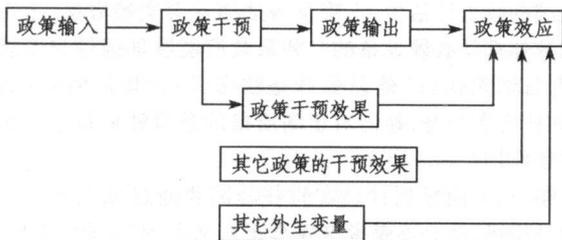


图1 分离政策干预效果的图示

资料来源:Wu Ramesh and Frizep 2007

我们所要做的就是将政策效应中的政策干预效果与其它政策的干预效果和其它外生变量引起的政策效应相分

离,单独加以考察。Weiss特别指出,在分离政策干预效果时,一定要廓清因果机制,即确定政策效应中的对应部分确实是由政策干预直接导致的,而不是相反。<sup>[28]</sup>

从政策评估的研究设计角度来看,目前主流的方法是社会实验法(social experimentation)。它的基本思想就是将政策客体以完全的随机原则分配到实验组(treatment group)和对照组(control group),然后对实验组施加政策干预,设定的时间段后观察两组的效应(这是实验设计的基本思路,但是在实践中由于多方面的限制,很多情况下无法使用对照组)。如果两组同时发生同向的变化,无论强度是否相同,都可以严重质疑政策干预是否产生实际效果;如果实验组发生变化,而对照组几乎接近于静止,那么两者的差异便是成功分离出的政策干预效应。如果实验组和对照组发生异向变化且强度显著,则需要进一步的分析。Greenburg和Shrode在整理美国公共政策评估的历史后发现,从1960年代至1990年代,美国各级政府就先后进行了90次之多的社会实验,横跨医疗政策、教育政策、劳工政策、住房政策等领域<sup>[29]</sup>,耗资超过十亿美元<sup>[30]</sup>,证明社会实验的适用性极广,在合适的政策问题上,中国的公共政策评估也可以尝试采用。

(一)政策评估的实验研究设计

一般而言,在政策评估的社会实验法中,分离政策干预效果的常用实验设计有四种,其中按照是否有时间向度(time dimension),有前测—后测的实验设计又有时间序列的变形(见表2)。在下表中,X代表政策干预,O代表实验设计中每个点的观测值。从理论上讲,它们的效度(validity)随着实验设计的逐渐完整而自上而下递增。

表2 政策评估常用的实验设计

实验设计	形式	分离政策干预效果	效度
简单后测实验	X-O	O	↓
非等值后测实验	$\frac{X-O_{1a}}{O_{1b}}$	$O_{1a}-O_{1b}$	
简单前测—后测实验	无时间序列实验: $O_1-X-O_2$	$O_2-O_1$	
	时间序列实验: $O_1-O_2-O_3-X-O_4-O_5-O_6$	$\overline{\Delta O_4} \sim \overline{O_6} - \overline{\Delta O_1} \sim \overline{O_3}$	
非等值对照组 前测—后测实验	无时间序列实验: $\frac{O_{A1}-X-O_{B1}}{O_{A2}-O_{B2}}$	$(O_{B1}-O_{A1})-(O_{A2}-O_{B2})$	
	时间序列实验: $O_{A1}-O_{A2}-O_{B1}-X-O_{A4}-O_{A5}-O_{A6}$ $O_{B1}-O_{B2}-O_{B3}-O_{B4}-O_{B5}-O_{B6}$	$(\overline{\Delta O_{A4}} \sim \overline{O_{A6}} - \overline{\Delta O_{A1}} \sim \overline{O_{A3}}) - (\overline{\Delta O_{B4}} \sim \overline{O_{B6}} - \overline{\Delta O_{B1}} \sim \overline{O_{B3}})$	

资料来源:作者绘制。

之所以我们说上表是政策评估中的常用实验设计,是因为理论上讲社会实验有多达至少14种模型,这在Cook和Campbell的著作《Quasi-experimentation: Design and analysis Issues for Field Settings》一书中已经做过详细介绍。但是现实的政策评估很少像这样抽象化地进行,政治支持、资

金、配合和研究对象本身的特性都对实验设计造成很大限制。<sup>[27]</sup>有的模型太过复杂,有的因客观条件很难满足苛刻的实验要求而无法使用。在几乎所有的实验研究方法教科书中,都认为简单后测实验是效度最弱的一种,但很多情况下现实情况并不允许更为复杂的实验设计。第一,社会实验因为需要政策受体(subjects)的随机分配(randomization),为

为了保证样本,一般需要数量很大的受体。当他们是数据或者物体时,要收集一定数量还有可能;但是在很多情况下政策受体是人,问题就变得非常复杂,即便保证了随机分配,也会有其它很多潜在的问题可能威胁到研究效度,进而造成评估结果失真。第二,很多政策的关键实时信息并没有得到保存(在中国这往往是常态),而政策本身又不可逆,这就使得前测—后测实验设计很难进行。第三,进行社会实验设计很多时候需要充裕的时间和足够的物质保证,因而对试验设计的选择造成限制。因此总的来说,在进行实验设计选择时固然优化的设计具有更高的效度,但是在各种约束下,评估者往往不能自由选择,而只能寻求次优。但这并不意味着理论上次优的实验设计产出的研究效度就一定低于更优的实验设计,也就是说简单的后测实验所得到的评估结果未必就劣于非等值后测实验。这取决于很多因素,如样本容量、测量手段、建构效度、评估者的能力等等。

### (二) 实验研究的资料收集和分析

作为政策评估社会实验法核心的分离政策效果其实只是一个虚拟的过程,本质上要通过选择变量、构建指针、收集数据来实现。按照 Wu Ramesh 和 Fritzen 等人的观点,第一手的资料收集(Primary data)方法主要有问卷调查、个案研究、关键人物访谈、关注群体讨论、小区群体访谈和直接观察。<sup>[32]</sup>根据笔者对文献中政策评估实验分析法的总结,大约有 70% 其实都并未使用第一手的数据收集,或者只是二手资料(如统计年报、部门档案、备忘录等)的补充。资金和时间约束固然是其中的主要原因,但文献中的这些评估之所以能够进行,其实很大程度上得益于发达国家先进的信息系统和严格的档案制度,笔者将在后文中举例阐明。

在大多数情况下,数据的分析首先需要使用一般描述性统计分析,展现政策干预前后主要观测指针和变量的数量、分布、相关性等的变化。在需要探求因果关系以及进行测量的时候,则需要借助多元回归、因子分析等多变量的统计分析。在很多时候,还需要通过回归模型进行预测,用以评估政策的长期效应。

### (三) 社会实验评估方法在中国的使用

应当承认,社会科学中绝大多数的研究方法都是具有普世性的,没有国界。从文献中可以发现,大规模的政策评估实验方法已经在墨西哥、菲律宾、泰国等国得到采用。这种主流方法在中国如何使用,笔者认为有三个问题需要注意:

第一,社会实验并非必须通过庞大的人为力量组织设计,模拟政策的运行,这是必须明确的问题。事实上在国外政策评估的研究中,这样的例子也是很少的。对于中国的公共政策研究者而言,如果要对一项政策进行评估,并非必须申请政府的评估项目或者是研究机构的课题,而可以根据自己的研究领域和研究兴趣对一项政策进行评估,这是完全可行的。很多政策问题其实是一种自然实验(natural experiment),个人不需要也无法施加外力进行操纵,而只需要观察并收集数据。国外学者进行的很多政策评估其实就是收集政策过程中有信度的数据。如果可及的数据涵盖了

政策实施前后两个阶段,那就已经构造出一个事实上的实验前测和后测,通过对数据的分析即可完成初步的政策评估。当然,这需要研究者有比较扎实的计量经济学功底。

第二,接着上一个问题,之所以国外研究者在不需要人为进行实验操作控制的情况下就可以完成初步的政策评估,一个重要的前提就是有充分的数据。这就回到了前文中说到的政策信息系统和档案制度的问题,这里举例说明。1970年代美国为了提高总统和联邦议会选举的投票率,实施了投票日报名(election day registration)政策。要判断这项政策到底是否确实提高了投票率,进行新的社会实验显然不可行。美国学者 King 和 Rodney 于是收集 9 个有代表性的州在过去十年中的投票率数据,采用有对照组的时间序列模型,通过回归分析证明这项政策对于提高投票率并没有显著影响。<sup>[33]</sup>这样一个时间跨度如此之长的政策,要对其评估,各州完整翔实的数据库数据起到了关键性作用。在我国的政策评估中,不管是否采用社会实验方法,丰富可信的数据都是必须具备的。如果政府能够加强信息系统建设,并且切实执行“信息公开是常态、不公开是例外”的原则,向研究者公开,相信对于高质量的政策评估将起到很大的推动作用。

第三,上面提到社会实验并非需要通过庞大的人为力量设计操纵,但并不是说这样真正意义上的“实验”不重要。事实上,国外经验证明,在很多重大公共政策问题上,通过以计量技术为主的非实验法进行政策评估存在一定的局限,主要是无法控制随机偏差问题(randomized bias),而对于社会实验来说,这是致命的威胁。<sup>[34]</sup>要解决这个问题,提高评估研究的效度,最好的办法就是进行实实在在的社会实验,并且样本容量越大越好。但是很显然,这样大规模的社会实验非常复杂且耗资巨大。如要实行,需要来自高层的强力支持,否则几乎没有可能。在这个方面,墨西哥用社会实验方法做了可能是目前为止全世界规模最大的政策评估。墨西哥政府 2002 年开始试点全民医疗保险( Seguro Popular de Salud),最初在五个州试点,一年后推展到二十个州。为了在全面施行之前对这项政策进行评估,墨西哥政府邀请了哈佛大学定量社会科学研究所以来为全民医保政策进行评估。评估者在高层的大力支持下,在全国范围内选择了五个州进行社会实验评估。<sup>[35]</sup>可见,只要能够获得政府的全力支持,大规模、有效度的政策评估是完全可能的。李树桥在总结当代中国公共政策的特色时曾经指出,“试点”是中国公共政策评估的重要形式。<sup>[36]</sup>在中等收入的大国进行以社会实验为主的大规模政策评估已经被墨西哥的经验证明可行,事实上在我国医疗卫生政策方面已经有不错的经验(如新型农村合作医疗的试点),科学严谨的评估方法如果能和政府的强烈政治意愿相结合,相信一定可以显著提高中国公共政策的评估质量,从而在决策的科学化、民主化中起到至关重要的作用。

### (四) 实验法与非实验法的政策评估

在科学的政策评估方法中,历来就有实验与非实验(non-experimental)两种方法,两者各有特点。非实验法又

称计量法,植根于计量经济学。它的基本思想是,假设政策受体和非受体(如未受某项政策影响的人群)在关键性指向上有充足的纪录数据可以支撑计量运算,那么不需经过社会实验,也可以通过计量经济学的手段获得政策受体发生(或不发生)预期变化的机率。但是与严谨的社会实验法相比,计量法存在一个很难逾越的障碍,即:它很难分离政策效果与政策受体自身发展所带来的效应,因而给评估政策效果带来潜在的偏差。受篇幅所限,这里不予展开。另外,计量法的政策评估需要非常完整的与政策执行过程有关的数据资料和训练有素的分析人员,这在实践中经常是不具备的,也在一定程度上影响了其适用性。关于社会实验法与计量法的比较,可以参见 James Hedgeman和 Jeffrey Smith等学者的论文,国内学者余向荣等也有阐述。

## 五、结论

政策评估是公共政策活动中非常重要的环节,但是由于政治行政体制、学科发展水平、人才力量、多学科整合程度等原因的限制,中国公共政策评估的发展还较为落后,不仅鲜有系统、成熟的评估实践,在评估方法论上更有很多问题,现有的很多评估没有科学的方法论指导,或者方法论存在瑕疵,这些都大大限制了学科知识的增长和方法的更新。要促进中国公共政策评估方法的进步,必须借鉴先进国家的研究成果。这个过程中首先要正确理解国外经典文献的情境意涵及其适用性,走出字面翻译、生搬硬套的老路,选择性地吸收,并结合学科发展的背景进行理解。

虽然很多学者对实证主义研究提出批判,但不可否认的是,实证主义的政策评估方法仍旧是目前的主流。经典文献中呈现的一些“冲突”之处并非否定实证主义的价值。相反,恰恰证明它应该是公共政策评估的主导方法。结合现在中国公共政策评估方法匮乏的实际,实证的评估方法应该是未来努力的主要方向。

在实证的政策评估方法中,社会实验法是最重要的研究设计,能够较为有信度地将政策干预的效果从整体政策效应中分离,从而进行分析,这也是主流公共政策评估的主要逻辑。中国很多“公共政策评估”存在方法论的缺陷或者完全经不起方法论的检验,问题大多出在这里。虽然理论上讲社会实验法有很多模型,但由于主客观的多重限制,实际研究中经常使用的只有四种方法。它们各有特点,也各有适用性,要根据公共政策的性质和其它客观条件选择使用。

分离政策效果只是一个虚拟的过程,本质上是通过选择变量、构建指针、收集数据来实现的。在大多数情况下,数据的分析首先需要使用一般描述性统计分析。在需要探究因果关系以及测量变量之间关系的时候,则需要借助多元回归、因子分析等多变量的统计分析。在很多时候,还需要通过回归模型进行预测,用以评估政策的长期效应。

文章的最后提出了在中国使用社会实验法进行政策评估的可行性及可能遇到的问题。本文的限制在于受篇幅所限,无法深入结合中国公共政策评估的特点,在社会实验法各种模型的适用、原则、步骤上做出探索和阐释。这些留待

以后完成。行

## [参考文献]

[References]

- [1] [11] [18] M. Howlett & M. Ramesh *Studying Public Policy: Policy Cycles and Policy Subsystems* (2nd Edition), Oxford: Oxford University Press, 2003.
- [2] 李静芳, 当前我国地方公共政策评估现状与对策[J], 江西行政学院学报, 2001(4).
- [2] Li Jingfang *Status Quo and Recommendations of China's Public Policy Evaluation at the Local Level* *Journal of Jiangxi Institute of Public Administration* 2001(4).
- [3] 朱仁显, 发展我国政策评估的若干构想[J], 理论探讨, 1999(4).
- [3] Zhu Renxian *Several Proposals for the Development of China's Policy Evaluation: Theoretical Exploration* 1999(4).
- [4] 余章宝, 作为非政府组织的美国智囊库与公共政策[J], 厦门大学学报, 2007(3).
- [4] Yu Zhanbao *America's NGO Think-tanks and Public Policy* *Journal of Xiamen University* 2007(3).
- [5] 马骏, 中国公共行政学研究的反思: 面对问题的勇气[J], 中山大学学报, 2006(3).
- [5] Ma Jun *Reflections of China's Public Administration Research: Courage to the Problems* *Journal of Sun Yat-sen University* 2006(3).
- [6] [27] 贞杰, 论现代社会条件下的政府政策评估[J], 江苏行政学院学报, 2005(4).
- [6] [27] Yun Jie *Public Policy Evaluation in the Contemporary Society* *Journal of Jiangsu Institute of Public Administration* 2005(4).
- [7] H. Wollmann *Evaluation in Public Sector Reform* *Cheltenham: Edward Elgar* 2003.
- [8] H. Wollmann *Policy Evaluation and Evaluation Research* in F. Fischer, G. Müller and M. Sidney (eds), *Handbook of Public Policy Analysis: Theory, Politics and Methods* Boca Raton: CRC Press, 2007.
- [9] M. Bovens et al. *The Politics of Policy Evaluation* in M. Moran, M. Rein and R. Goodin (eds), *The Oxford Handbook of Public Policy* Oxford: Oxford University Press, 2006.
- [10] D. Nachmias *Public Policy Evaluation: Approaches and Methods* N.Y.: St Martin's Press, 1979.
- [12] J. S. Coleman *Problems of Conceptualization and Measurement in Studying Policy Impact* in K. Dolbeare (eds), *Public Policy Evaluation* Beverly Hills: California: Sage, 1975.
- [13] R. C. Rist *Policy Evaluation: Linking Theory to Prac-*

- tion London: Edward Elgar, 1995.
- [14] E. A. Bæk. Policy Evaluation: Design and Utilization. *Knowledge in Society* 2(4), 1989—1990.
- [15] C. W. Anderson. The Place of Principals in Policy Analysis. *American Political Science Review* 73(3), 1979.
- [16] D. H. Kerr. The Logic of Policy and Successful Policies. *Policy Sciences* 7(3), 1976.
- [17] D. Manzer. Policy Rationality and Policy Analysis: The Problem of the Choice of Criteria for Decision-making. In Q. P. Dwivedi, ed., *Public Policy and Administrative Studies*. Guelph: University of Guelph, 1984.
- [19] H. M. Ingram & D. E. Mann, eds. *Why Policies Succeed or Fail*. Beverly Hills, California: Sage, 1980.
- [20] S. Kuznetz. Economic Growth and Income Inequality. In M. Seligson et al., eds., *Development and Underdevelopment: the Political Economy of Inequality*. Boulder, Colo.: L. Rienner, 1993.
- [21] K. J. Forbes. A Reassessment of the Relationship between Inequality and Growth. *The American Economic Review* 90(4), 2000.
- [22] J. Currie. How Should We Interpret the Evidence about Head Start? *Journal of Policy Analysis and Management* 26(6), 2007.
- [23] W. Williams & J. Evans. The Politics of Evaluation: the Case of Head Start. *Annals of American Academy of Political and Social Science* 385(9).
- [24] 谢望礼. 治淮十年付诸东流的反思 [J]. *江西财经大学学报*, 2004(5).
- [24] Xie Wangli. Why the Ten-year Efforts in Water-control of Huai River Eventually Turned to be Vain? *Journal of Jiangxi University of Finance*, 2004(5).
- [25] A. Gray & B. Jenkins. Policy Evaluation in a Time of Fiscal Stress: Some Reflections from British Experience. *Knowledge in Society* 2(4), 1989—1990.
- [26] E. Guba & Y. Lincoln. The Countenances of Fourth Generation Evaluation: Description, Judgment, and Negotiation. In D. Palumbo (eds.), *The Politics of Program Evaluation*. Beverly Hills, California: Sage, 1987.
- [28] C. Weiss. Nothing as Practical as Good Theory: Exploring Theory-based Evaluation for Comprehensive Community Initiatives for Children and Families. In J. Connell, A. Kubisch, L. Schorr and C. Weiss (eds.), *New Approaches to Evaluating Community Initiatives: Concepts, Methods and Contexts*. Washington, D. C.: Aspen Institute, 1995.
- [29] D. Greenberg & M. Shroder. *Digest of the Social Experiments*. Madison, Wisconsin: Institute for Research on Poverty, University of Wisconsin, 1991.
- [30] [31] G. Burtless. The Case for Randomized Field Trials in Economic and Policy Research. *Journal of Economic Perspectives* 9(2), 1995.
- [32] X. Wu, M. Ramesh & S. Fritzen. *Integrated Policy-Making for Sustainable Development: A Guide*. Lee Kuan Yew School of Public Policy, National University of Singapore, 2007.
- [33] J. D. King & A. W. Rodney. Impact of Election Day Registration on Voter Turnout: a Quasi-experimental Analysis. *Policy Studies Review* 14(3/4), 1995—1996.
- [34] J. J. Heckman & J. A. Smith. Assessing the Case for Social Experiments. *Journal of Economic Perspectives* 9(2), 1995.
- [35] G. King et al. A “Politically Robust” Experimental Design for Public Policy Evaluation: with Application to the Mexican Universal Health Insurance Program. *Journal of Policy Analysis and Management* 26(3), 2007.
- [36] 李树桥. 当代中国公共政策的特色 [J]. *中国行政管理*, 2000(12).
- [36] Li Shuqiao. The Characteristics of Public Policy in Contemporary China. *Chinese Public Administration* 2000(12).
- (作者单位: 新加坡国立大学李光耀公共政策学院博士生)

(责任编辑 方晋)

## Methodological Orientation for China's Public Policy Evaluation: Towards Positivism

He Jingwei

[Abstract] Not being built on a solid methodological ground, China's public policy evaluation is still far from maturation. A positivism orientation should be set up in light of embracing alien experiences. Mainstream methodology in policy evaluation should be introduced to China on a localized basis. This helps to improve the quality of domestic research and make it convincing and methodologically sound. Basically, the heart of mainstream policy evaluation methods involves social experimentation in which evaluation is processes by isolating net policy effects. There are a series of data analyzing techniques that should be utilized on a case-by-case basis.

[Key words] positivism; social experimentation; methodology; China; public policy evaluation

[Author] He Jingwei is a doctoral candidate at Lee Kuan Yew School of Public Policy, National University of Singapore