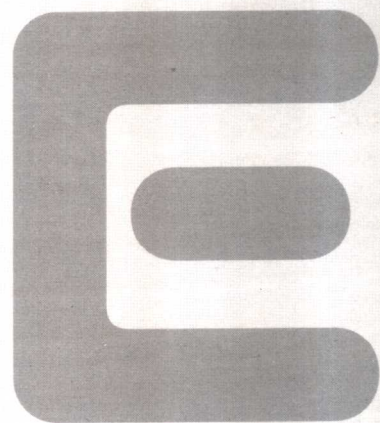


万卷方法 | 社会测量与评估方法译丛

EVALUATION
A SYSTEMATIC
APPROACH

第7版



评估：方法与技术

彼得·罗希
马克·李普希
霍华德·弗里曼

著

邱泽奇 王旭辉 刘月 等

译



重庆大学出版社

<http://www.cqup.com.cn>

万卷方法 | 社会测量与评估方法译丛

EVALUATION
A SYSTEMATIC
APPROACH

第7版

评估：方法与技术

彼得·罗希
马克·李普希
霍华德·弗里曼

著

邱泽奇 王旭辉 刘月 等

译

重庆大学出版社

Authorized translation from the English language edition, entitled EVALUATION: A SYSTEMATIC APPROACH, 7th edition by Peter H. Rossi, Mark W. Lipsey, Howard E. Freeman, published by Sage Publications, Inc., Copyright © 2004 by Sage Publications, Inc.

All rights reserved, No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher. CHINESE SIMPLIFIED language edition published by CHONGQING UNIVERSITY PRESS, Copyright © 2005 by Chongqing University Press.

评估:方法与技术。原书英文版由 Sage 出版公司出版。原书版权属 Sage 出版公司。

本书简体中文版专有出版权由 Sage 出版公司授予重庆大学出版社,未经出版者书面许可,不得以任何形式复制。

版贸渝核字(2006)第 37 号

图书在版编目(CIP)数据

评估:方法与技术/(美)罗希(Rossi, P. H.),
(美)李普希(Lipsey, M. W.), (美)弗里曼(Freeman,
H. E.)著;邱泽奇等译. —重庆:重庆大学出版社,
2007.4

(万卷方法·社会评估与测量方法译丛)

书名原文:Evaluation: A Systematic Approach

ISBN 978-7-5624-3994-3

I. 评… II. ①罗…②李…③弗…④邱… III. 评估—方
法—教材 IV. C93-03

中国版本图书馆 CIP 数据核字(2007)第 027060 号

评估:方法与技术(第 7 版)

彼得·罗希 马克·李普希 霍华德·弗里曼 著

邱泽奇 王旭辉 刘 月 等译

责任编辑:雷少波 林 萍 版式设计:雷少波

责任校对:邹 忌 责任印制:张 策

*

重庆大学出版社出版发行

出版人:张鸽盛

社址:重庆市沙坪坝正街 174 号重庆大学(A 区)内

邮编:400030

电话:(023) 65102378 65105781

传真:(023) 65103686 65105565

网址: <http://www.cqup.com.cn>

邮箱: fxk@cqup.com.cn (市场营销部)

全国新华书店经销

重庆科情印务有限公司印刷

*

开本:787×1092 1/16 印张:19.5 字数:403千 插页:16 开 2 页

2007 年 4 月第 1 版 2007 年 4 月第 1 次印刷

印数:1—3 000

ISBN 978-7-5624-3994-3 定价:49.00 元

本书如有印刷、装订等质量问题,本社负责调换

版权所有,请勿擅自翻印和用本书

制作各类出版物及配套用书,违者必究

前言 中文版

我非常荣幸地知道,我们这本书是第一本被翻译成中文的评估教材。中国是一个伟大且强大的国家,正如我们在书中解释过的一样,我希望能够借助评估研究的思路和方法促进中国社会更快地发展。

评估不能代替政治过程。政策是由政治官员和立法人员制定的。但是,政策的实施却有很多方式。评估的任务就是了解项目的形成,即形成政策的特定形式。系统地评估,可以说明哪些项目更可能达到政策目标,同时也说明哪些项目对社会、人民和社会制度而言成本更小。

评估也被应用在社会科学中。美国的评估工作者来自几乎所有的社会科学领域——经济学、心理学、社会学、政治学以及人类学。评估所使用的方法和理论也来自社会科学的各个领域。除了学科背景的差异以外,所有的评估工作者都有着共同的目标——通过自己的职业实践来改善社会环境。

我们期待着通过评估工作者的努力,中国的评估实践能够为本书未来的美国版本增添最好的评估实践案例。

罗希
李普希

前 作 言 者

本书第7版吸纳了一些新的研究材料,并对前一版中的某些内容进行了大幅修订。本版的修正涉及:广泛地调整了产出测量(outcome measurement)和督导(monitoring)的内容;更完善地说明了项目影响评估的设计和筹划;更加全面地阐述了评估研究中的重要统计问题;此外,还对事后分析作出了更详尽的描述。我们相信,这些调整将会使本书的相关内容更贴近评估研究领域的前沿。

然而,本书的目标和核心主题始终没变——那就是为对社会项目的设计、实施和利用进行评估的各种研究活动提供一个引导。我们试图把握评估研究的全局,分析社会项目评估的设计、实施、绩效和效率。尽管做出了很多修订,但是本书的目标仍然是为准备从事评估职业或者需要了解评估领域的人们,提供有关评估研究的基本知识和技巧,并分享在长期评估实践中所积累起来的集体经验。本书的读者对象包括:学生、实际工作者、社会项目负责人、社会评论者,以及任何其他成员,只要他对意在改善社会状况的社会项目的成败测量感兴趣。

我们相信,通过这本书的阅读,您将可以获得一些必要的知识以理解和评判评估研究。应该说明的是,尽管我们对评估研究程序进行了一定的描述,并对相应的细节进行了说明,甚至提及了先前研究者们已有的研究成果和讨论;不过,本书并不想成为一本仅仅介绍评估步骤的教材。根本上讲,具体项目设计和操作中的研究经验,是这本书所不能带给广大读者的。我们鼓励所有想进入评估研究领域的读者,通过亲自实践来获取工作经验。

20世纪70年代,当本书第1版出版发行时,评估还未成为评定社会项目的一种完善方法。现在的情况则大不相同。21世纪,评估研究在全世界范围内获得了发展,评估的研究程序已经牢固地进入了世界范围内每个政府层次的日常活动中,也进入了非政府组织的运作中,甚至进入了社会议题的讨论中,成为不可缺少的一个组成部分。媒体几乎从未停止过向公众发布某些社会项目评估的结果。我们相信,评估研究对于社会政策的制定与改善,会起到十分重要的作用。成为一名评估者,将会担负起一个富有刺激性的职业角色,有机会通过运用专业技能和人际关系技巧,参与社会状况的改善。

我们将本书的第7版献给丹尼尔·帕特里克·莫怡尼安(Daniel Patrick Moynihan)。他已于最近逝世。半个世纪以来,他有着令人惊奇的个人经历,既保持着在学术领域的核心位置(哈佛大学),也在联邦政府机构中(肯尼迪和约翰逊总统执政时,他担任劳工部长助理)身居要职,还在尼克松政府中担任城市问题方面的白宫顾问,并两次当选纽约市的联邦参议员。同时,他还出版了数本有关联邦政府社会政策

和决策的有影响力的著作。莫怡尼安在参议院的工作,明显提高了参议院商议社会政策的理性程度。尽管经历丰富,但是在他从事的每份工作中,改善社会政策都是其关心的核心问题。另外,他在社会政策尤其是评估研究领域,是一名出色而坚定的鼓吹者(我们可以在第1章的专栏1—A中看到一个例子)。无论是在直接推进或者间接支持联邦政府的评估活动方面,还是在推进改善我们社会生活状况的其他项目中,他都扮演了十分关键和积极的角色。

彼得·罗希(P. H. R.)

马克·李普希(M. W. L.)

目 录

1 项目、政策和评估	1
什么是评估研究	2
评估研究简史	5
项目评估的特征	12
实践中的评估研究	16
谁能从事评估工作	20
小结	21
基本概念	22
2 准备评估	23
评估方案必须包括哪些方面	24
评估方案应该考虑哪些问题	25
评估者与项目各方关系的性质	35
评估问题和评估方法	39
小结	45
基本概念	46
3 确定议题和设定问题	47
获得好的评估问题的条件是什么	48
确定评估应当回答的具体问题	57
核查评估问题和确定问题的优先秩序	66
小结	68
基本概念	69
4 需求评估	70
评估者在诊断社会状况和服务需求中的角色	71
界定社会问题	74
将问题具体化:时间、地点和范围	75
界定和识别干预对象	82

描述目标人群	85
描述服务需求的特征	87
小结	92
基本概念	92
5 项目理论的表达与评估	94
可评估性评价	96
描述项目理论	98
构造项目理论	105
评价项目理论	108
项目理论评估的后果和影响	117
小结	118
基本概念	119
6 督导项目的过程和绩效	120
什么是项目过程的评估和督导	121
项目过程督导的各种观点	127
服务利用的督导	129
组织功能的督导	135
项目过程督导资料的分析	139
小结	140
基本概念	141
7 项目产出的测量和督导	142
项目产出	143
识别项目产出	146
测量项目产出	149
督导项目产出	156
小结	161
基本概念	161
8 项目影响评估——随机实地实验	162
开展影响评估的时间选择	163
影响评估的关键概念	164
随机实地实验	166
随机实验的局限性	179
小结	181
基本概念	182
9 项目影响评估——备选设计	183
项目效果估计的偏差	184
准实验影响评估	189
在影响评估中运用准实验方法的注意事项	204

小结	206
基本概念	206
10 探明、解释和分析项目效果	208
项目效果的大小	209
探明项目效果	212
估计项目效果的实际意义	218
分析项目效果的差异性	221
事后分析的角色	225
小结	228
基本概念	229
11 效率测量	231
效率分析的重要概念	233
成本—收益分析	239
成本—绩效分析	254
小结	256
基本概念	257
12 评估的社会背景	258
评估的社会生态学	261
评估职业	274
评估标准、准则和伦理	281
评估结果的利用	285
尾声:评估事业的前景	290
小结	290
基本概念	291
参考文献	292

项目、政策和评估

1

本章介绍项目评估。笼统地讲,评估就意味着确定某些预期目标的价值或者将特定价值赋予到某些目标之上。在本书中,我们将评估这一概念限定在项目评估或者评估研究的范畴,项目评估是一种社会科学活动,涉及搜集、分析、解释和沟通有关旨在改善社会环境的社会项目的实施和绩效。评估有各种实用目的:帮助决定项目是否应该继续、改善、扩展或缩减;评估新项目的用途和创意;提高项目管理和指导的绩效;满足项目各方的要求。评估也有助于丰富实际的、方法论性的社会科学知识。

要理解现代背景下的评估活动,需要懂得一些评估的历史、评估特有的概念和目的、评估关注的问题和操作中所面临的内在张力与挑战。项目评估意味着把社会研究方法运用于分析一定政治和组织环境下的社会干预,而合理的评估将有助于对社会问题的评判,使项目的设计、实施、影响和绩效满足干预的需要。单个的评估研究以及许多类似研究的知识积累,能够为旨在改善人类环境的社会行动提供实质性的帮助。

自古以来,许多有组织的努力都致力于描述、理解和改善人类环境的缺陷。这本书承袭了对社会问题进行科学研究的传统——一种渴望改善我们的物质和社会环境质量、通过系统地创造和运用知识,提高个体和集体福利的传统。尽管**项目评估**(Program evaluation)和**评估研究**(Evaluation research)是新近创造的词汇,但却不是刚刚才有的活动。评估活动可以追溯到科学的初创期。科隆巴赫及其同事(Cronbach,1980)指出,三个世纪以前,霍布斯(Thomas Hobbes)及其同时代追随者曾努力用数量方法评估社会环境,探寻死亡率、发病率和社会解体的原因。

即使是社会实验,作为当代评估研究面临的**最大技术挑战**,也不是最近的发明。最早的“社会实验”发生在18世纪初期。那时,一位英国船长观察到,在地中海地区的诸多国家船只上工作的海员很少患坏血症;他同时注意到,柑桔类果实是这些海员的口粮之一。因此,他把自己的船员分为两半,一半人食用柠檬,另一半人则沿袭原有的饮食。实验表明,食用柠檬有助于防止坏血病。这位船长也许不知道他在评估一项示范性项目,当然,也不可能获得细致的“项目理论”(后面我们将讨论这个术语)。按照现在的医学知识可知,患坏血症是因为缺乏维生素C,而柠檬富含维生素C。总之,干预发生了作用。由此,英国海员都被迫食用柑桔类果实,这就是柠檬茶流行的由来。巧合的是,人们花了大约50年的时间才使船长的“社会项目”得到广泛利用。不过,直到现在,传播和接受评估的发现,仍然不是一件易事。

什么是评估研究

在不同时期,为了获得某些期望的结果,政策制定者、资助机构、规划者、项目管理者、纳税人或项目对象,都需要区分值得实施的和不值得实施的社会项目(Social program)^①,启动新项目和改善既有项目,从而达到特定的预期目标。为了做到这一点,他们必须获得下面一些问题的答案:

- 问题的特质和范围是什么? 问题出在哪里? 影响到谁? 影响了多少人? 如何影响?
- 什么样的问题或结果能够说明新的、扩展的或修订的社会项目的必要性?
- 可行的、能对问题产生明显改善作用的干预是什么?
- 干预的对象是什么?
- 特定干预是否落实到了目标群体?
- 干预活动实施得好吗? 提供了原定的服务吗?
- 干预对于实现预期目标或利益是否有效?
- 与绩效和收益比较,项目的成本是否恰当?

^① 这是本书的一个核心概念,已经在本章的基本概念一栏中列出。

即使是局部性的、具体的项目,获得上述问题的答案也是必要的。譬如小镇的职业培训、小学新数学教学计划、社区精神健康出诊。对全国性的、全州性的项目如健康、社会福利以及教育改革而言,也是必要的。提供上述问题的答案,是项目评估工作者的责任。具体地说,评估研究者(评估者)运用社会研究方法,研究、评价,并帮助改善社会项目的所有重要方面,包括社会问题诊断、概念化与设计、实施与管理、产出,及其效率(专栏1—A表现了一位活跃的参议员对项目绩效评估证据的重视)。

专栏1—1 政策制定老手要看评估结果

尽管不断有行政官员说服我们,要增加这个或那个社会项目的经费,但我们仍然在尽最大的努力(可以说,现在取得了很大的成功)[减少赤字]……在这些琐碎中,我印象最深的是“家庭维持”项目。家庭维持项目是另一类援助项目(已经有了许多了),属于社会服务的一块,并由某个附属委员会管理。该项目为期5年,耗资9.3亿美元,1994财政年度的启动经费为6000万美元。在过去的30年里,我看到了家庭与社会的分离;现在,每个新小组的成员都告知我,只要再加一个项目……让我冒昧地把我1993年7月28日写给泰森(Laura D' Andrea Tyson)博士(后任经济咨询委员会主席)的信作为有关“家庭维持”项目的档案:

亲爱的泰森博士:

您也许记得,上星期四在您参加民主政策委员会会议时,我和您谈到了总统的家庭维持项目。您指出,他非常支持这个项目。我向您保证,我也支持这个项目,只是我希望看到一些证据,来说明项目会产生效果。对此,您明确表示已经有资料可以证明,但出于好奇,我曾要求两个佐证。

次日,我便收到了您的工作人员格赖德(Sharon Glied)发来的传真,提供了一些佐证和一篇关于“结果评估”文章,文章似乎是华盛顿社会政策研究中心的法诺(Frank Farrow)和芝加哥大学霍尔(Chapin Hall)中心的理查曼(Harold Richman)写的。文章非常直率地写道:“就家庭维持服务而言,尚缺乏能影响整个国家经费安排的实在证据。”

也就是昨天,霍尔中心又发布了“伊利诺斯家庭优先防护项目评估:总报告”。这是自1987年伊利诺斯家庭维持法案后,对家庭优先安置项目的一项大型研究。研究“旨在考察家庭优先安置项目的效果和其他影响,譬如随后的儿童虐待”。家庭优先安置项目的工作人员提供了大约4500个案例资料和服务特点资料:大约1600个家庭参与了项目的随机实验。研究结果非常清楚。

总体上,家庭优先安置项目略微提高了一点安置率(当综合所有实验资料时)。但是,如果考虑到具体案例和区域因素,这个效果就消失了。换句话说,这一项目既没有负面效果,也没有正面效果。

这个结果并不新鲜,因为在1992年,罗希(Peter Rossi)在“评估家庭维持项目”的结论中已经说过了。今天的评估“并没有提供充分的证据以说明家庭维持计划是否有效果”。我是否可以对您说,这些发现并没有令人惊奇的地方?从20世纪60年代中期起,这样的评估一再重复,甚至持续不断。但很少具体说明项目有正面效果或负

面效果。而20世纪70年代的负债税实验,似乎还增加了家庭解体的比例。

这种“与印象相反”的发现,最早出现在20世纪60年代。格里雷和罗希(Greeley and Rossi)以及我的部分工作,还有科尔曼(J. S. Coleman)的工作,都说明了这一点。直到今天,我仍然不能确定,我们面对的只是方法的构造,还是规模更大的和更棘手的社会项目现实。任何一种情况都印证了罗希1978年提出的铁律,即“如果说过去几十年评估活动有什么经验规律的话,那就是,社会项目测量效果的期望值等于零”。

我之所以给您写这封长信,是因为我认为,有一件事情非常重要。在过去的6个月里,克林顿政府的不少人不断试图说服我,让我知道一些社会政策已经获得了极大的效果,但是在我看来,那些社会政策没有任何效果。我认为这样的事很危险。因为政策的效果并不稳定,甚至某些人的看法从根本上错误。因此,要确切地知道什么样的项目的确没有效果。意识上的自信很容易演变为固执,进而忽视一切事实。

这个时候(甚至这一代),政治保守派的最大优势在于他们对复杂思想的开放态度。而自由派的弹性则在于抵制复杂思想。在过去的12年里,我曾经极力去改变;现在看来,这些状况不仅没有被改变,反而得到了强化。如果是这样的话,自由主义的复活将只会昙花一现,并不会产生任何结果。

丹尼尔·帕特里克·莫怡尼安 (Daniel Patrick Moynihan) 参议员 敬上

资料来源:D. P. Moynihan, *Miles to Go: A Personal History of Social Policy* (Cambridge, MA: Harvard University Press, 1996), pp. 47-49.

尽管本书强调的是对社会项目(特别是人类服务项目)的评估,但是项目评估并不局限于此。美国审计总署(GAO)项目评估方法部的工作,就很好地说明了项目评估的广泛性。在这个部门的发展史上,它评估了军用设备的获得与测试、饮用水的质量控制、重要高速公路的维护、使用激素促进菜牛生长,以及其他各种有组织的和非服务性的活动。

的确,本书提供的技术基本上适用于各种有组织的社会活动所提出的绩效问题。举例而言,评估手段已经被运用于大众传媒和广告业的发展与市场开发中。商业和工业企业对选择、培训、晋升职员和组织劳动力的程序需要评估;政治人物则通过评估选民诉求来发展竞选策略;人们要对消费品性能、耐用性和安全性进行测试;公共和私有部门的管理者对职员、财务和组织的人事实践也要进行评价。当然,还有很多类似例子可以说明评估手段的广泛运用。

五花八门的评估之间的区别,就在于各种努力的特质和目标。本书强调的是评估各种旨在改善人类环境的项目,而不是评估那些旨在增加利润或扩大影响和权力的项目。之所以这样,是考虑把评估的活跃领域和实践的需要结合起来,进而限制本书的篇幅。这一点可以从我们所使用的概念中得到印证,比如评估、项目评估和评估研究这三个概念,它们是可以互换的。

为使项目评估更贴近现实和实际,下面提供了一些社会项目评估的案例,其中评估的主办方既有地方、州、联邦政府机构、国际组织、私营基金会和慈善机

构,也有其他非营利机构、营利机构和企业。

- 在美国一些主要城市,某一私营基金会提供了一笔启动费,用于在低收入社区建立社区医疗中心。目的是为当地居民提供流动医疗服务,以免他们花大笔的公共支出让医生出诊或进急救室。进一步的目的是,让社区居民获得就医机会,这样,既减少了就医时间,也降低了就医费用。评估结果表明,与医院门诊比较,有些医疗中心的确在成本—绩效方面具有优势。
- 在纽约市,部分对学校提供经费资助的人创立了一个私人资助项目,用于帮助在较差公立学校就读1—3年级的穷人孩子。符合条件的家庭可以获得奖学金资助,用于支付在任何一所私立学校就读3年的学费。在14 000份申请中,项目方随机选择1 500人进行资助。评估小组利用了这一选择模型,认为这是一项随机实验,并试图比较获得奖学金进入私立学校念书的结果与没有获得奖学金仍然在公立学校念书的结果。
- 在过去10年中,联邦政府允许州政府修改福利项目,而由此对项目对象产生的影响和需要的成本,则需要评估。一些州提高了职业训练的要求,另一些州要求在一定时间内见到效果,少数州则禁止为新生婴儿提高福利。评估结果表明,这些政策可以降低福利费用和提高就业率。为此,联邦政府1996年通过的福利改革法案(个人责任和工作机会协调法案)吸纳了项目的许多特征。
- 全世界足有2/3的农村儿童营养不良或严重营养不良,这对他们的健康和身心成长构成了严重影响。中非某些地区通过提供补充营养,展现了改善儿童身心健康的巨大潜力。项目方每天向孕妇、处于哺乳期的母亲、12岁以前的儿童提供高蛋白、高热量的食物补充。尽管这样的营养补充对身体成长的效果非常明显,但在对心智提高的影响上却表现一般。
- 为了提高工人的满意程度和生产质量,一家大型制造公司对职工进行重组,成立了一些独立的工作小组。在小组内部,工人自己指派任务,向管理部门提供产量建议,根据产量和质量改良情况,投票决定奖金的发放。项目评估所得到的资料显示:这种的做法减少了旷工率、人员调整率,并因此提高了劳动生产率。

这些短小的例子说明,系统评估工作涉及多样化的社会干预项目。但是,以上所有这些都涉及一项特别的评估活动:项目产出评估。正如后面将要讨论的,评估也可以专注于项目的需要、设计、操作、服务或者效率。

评估研究简史

尽管评估的历史可以追溯至17世纪,但是,系统的评估研究则是现代社会的产物,评估研究的发展是20世纪以来的事情。项目评估所使用的社会研究方

法,与方法本身的发展和改进,以及意识形态、政治和民主的变迁是相伴随的。

作为社会科学活动的评估研究

系统的社会项目评估最早出现在教育和公共健康领域。在第一次世界大战以前,最有意义的努力就是扫盲、职业培训以及为降低死亡率和流行病发病率所实施的公共健康项目。20世纪30年代,各学科的社会科学家开始致力于用严格的研究方法评估社会项目,由此,使得系统的评估活动变得越来越频繁(Freeman, 1977)。譬如,列文(Lewin)开创性地对“行为研究”的研究、李普特(Lippitt)和怀特(White)对民主和集权领导的研究,都是影响广泛的评估研究;类似的例子还有,涉及劳动生产率的著名的西部电器实验发现了所谓的“霍桑效应”(参见Bernstein and Freeman, 1975,进一步的讨论参见Bulmer, 1982; Cronbach et al., 1980;不同的历史视角看法参见Madaus and Stufflebeam, 1989)。

从这些开创性研究开始,应用社会研究得到了加速发展,特别是第二次世界大战中,应用社会研究的贡献尤其明显。斯托福(Stouffer)及其同事和美国军队一起,获得了用于监测士兵士气的方法、评估人格的策略以及宣传技术。这些研究发现被美国战争情报机构用来监测军人的士气(Stouffer et al., 1949)。还有很多规模较小的研究,如评估价格控制和媒体攻势在改变美国人饮食习惯方面的功效。在英国和全世界其他地方的社会科学领域,也有类似的努力。

评估研究的繁荣期

第二次世界大战以后,为了满足城市发展和房产、技术及文化教育、职业培训、预防疾病等方面的需要,出现了无数由联邦政府或私人资助的发展项目。同一时期,还有大量由联邦政府机构或者私人基金支持的国际项目致力于家庭计划、健康和营养以及农村发展。由于花费巨大,项目各方理所当然要知道“结果如何”。

20世纪50年代末期,项目评估研究变得很流行,社会科学家们忙于从事防止过失、精神心理治疗和精神药物治疗、公共住宅、教育活动、社区组织创建以及无数其他类型项目的评估。这样的评估研究不仅发生在美国、欧洲和其他工业化国家,也发生在不太发达的国家。渐渐地,亚洲的家庭计划、拉丁美洲的营养与健康、非洲的农业与社区发展等类型的项目,都成为了评估研究的重要内容(Freeman, Rossi, and Wright, 1980; Levine et al., 1981)。同时,社会研究方法的发展和相关知识的普及,包括抽样调查、高级统计方法,以及不断增长的基金资助和行政管理知识,使得大规模、多层面的评估研究成为可能。

在20世纪60年代,评估研究论著的数量急剧增长。比如,海伊斯(Hayes, 1959)阐述了评估研究在欠发达国家的发展,萨奇曼(Suchman, 1967)对评估研究方法本身进行了回顾,坎贝尔(Campbell, 1969)举例说明了社会实验方法。在美国,引起人们对评估研究极大兴趣的,是由约翰逊(Lyndon Johnson)总统在联邦范围内发起的有关贫困的争论。到20世纪60年代后期,用华尔街的话说,评估研究已经变成了一个成长中的产业。

在 20 世纪 70 年代早期,评估研究已经成为了社会科学界的一个重要学术领域。各类书籍纷纷出笼,包括第一本教材(Weiss, 1972)、对评估研究方法论的批评(Bernstein & Freeman, 1975)以及涉及评估研究组织与机构局限的讨论(Reicken & Boruch, 1974)。1976 年,《评估评论》(*Evaluation Review*)创刊,并逐步成为评估工作者广泛阅读的刊物。后来,又相继出现了不少期刊,到今天为止,评估方面的期刊已达十多种。在这一时期,许多学术和实践会议的热门话题就是评估研究。也是在这一时期,业界还成立了专门的评估研究者协会(专栏 1—B 列出了评估研究的主要期刊)。到 1980 年,科隆巴赫及其同事已经在说,“评估研究已经成为了美国社会科学中最有活力的前沿阵地”(Cronbach et al., 1980: 12-13)。

评估研究的发展也导致了评估研究本质的变迁。在早期,致力于评估研究的主要是社会研究者;但是到后来,评估研究的使用者(消费者)对这个领域也产生了重要的影响。现在,评估研究的持续发展则主要缘于决策者、项目计划者和行政管理人士的推动,他们利用评估研究的成果,并认为评估研究的成果值得信任。当然,评估研究的发展还得益于公众和项目对象的支持。尽管评估研究不能制造新闻,但是,评估研究的发现却总会牵动项目主办方、决策者、有见地的公民以及利益受到项目直接或间接牵连的人,影响范围十分广泛。

专栏 1—B 评估研究的主要期刊和专业组织

主要致力于项目和政策评估的期刊

- Evaluation Review: A Journal of Applied Social Research (Sage Publications)
- American Journal of Evaluation (JAI Press) (formerly Evaluation Practice, before 1998)
- New Directions for Evaluation (Jossey-Bass)
- Evaluation: The International Journal of Theory, Research, and Practice (Sage Publications)
- Evaluation and Program Planning (Pergamon)
- Journal of Policy Analysis and Management (John Wiley)
- Canadian Journal of Program Evaluation (University of Calgary Press)
- Evaluation Journal of Australasia (Australasian Evaluation Society)
- Evaluation and Health Professions (Sage Publications)
- Educational Evaluation and Policy Analysis (American Educational Research Association)
- Assessment and Evaluation in Higher Education (Carfax Publishing Ltd.)

项目和政策评估的专业组织

- American Evaluation Association (Web page: <http://www.eval.org/>)
- Association for Public Policy Analysis and Management
- American Educational Research Association (Evaluation Division)
(Web page: <http://aera.net>)
- Canadian Evaluation Association
(Web page: <http://www.unites.uqam.ca/ces/ces-sce.html>)

- Australasian Evaluation Society (Web page: <http://www.parklane.com.au/aes/>)
- European Evaluation Society (Web page: <http://www.europeanevaluation.org>)
- UK Evaluation Society (Web page: <http://www.evaluation.org.uk>)
- German Evaluation Society (Web page: <http://www.fal.de/tissen/geproval.htm>)
- Italian Evaluation Society (Web page: <http://www.valutazione.it/>)

消费者观点的介入,使得评估研究不再只限于让应用社会科学研究者对社会项目进行研究。同时,也使评估变成了政治和管理等复杂的活动。评估的目的就是,使政策决策、资源利用、项目设计、项目实施和延续变得更加有利于人的发展。在这样的意义上,评估研究应该是社会政策和公共行政运动的有机组成部分。

社会政策和公共行政运动

社会项目和评估活动出现的一个重要背景是晚近一段时期以来社会和环境状况以及国民生活素质等方面的责任向政府的转移。正如布伦纳(Bremner, 1956)所描述的那样,第一次世界大战以前,除了战争老兵以外,公共服务基本上是个人的和志愿组织的义务和责任。地方慈善组织善事活动的对象主要是穷人、残疾人 and 有麻烦的家庭。在我们头脑中,从事慈善活动的志愿者往往是这样一种形象——用篮子装上食品和衣物,施舍给那些不幸的人们。和民间团体、地方性慈善医院、县和州政府收容所、地方性公立学校、州师范学院以及养老院一起,志愿者们筑起了我们的公共服务“系统”。的确,20世纪30年代以前,政府规模相对较小,尤其是联邦政府。譬如,20世纪20年代,如果一年要花几十亿美元为老年人或穷人购买医疗设备和药品,就会使政府官员感到荒唐。联邦财政预算中,用于公共教育的支出也少得可怜——现在联邦政府用于公共教育的几个月支出就远远多于20世纪前十年的总和。

相应地,社会和经济信息的需求也很小。即使到了20世纪30年代末期,联邦支出中,每年用于社会科学研究和统计研究的份额,也只有4 000~5 000万美元,而今天的份额已经是那时的很多倍。同时,那时的公共服务和政府运作规则也与今天的不相同。重要政府官员的选择往往不考虑客观的竞争性标准;事实上,那时候用来评定个人能力的客观方法几乎没有。那时,专业的公共服务仅仅相当于现有总量的很小一部分,大多数的公共服务工作都不需要专门的技能;同时,服务人员也没有受过什么正规的训练。

所有这些,在20世纪30年代以来都得到了改变。随着大萧条的到来,公共服务获得了飞速发展,当然,政府部门也是如此。一部分原因是,公共项目的快速发展,形成了应用科学管理理念和技术的强大压力;特别是科学管理在工业领域已经有了很好的应用,给政府项目和活动的压力自然就很大。首先采用科学管理理念的是国防部,接下来就在政府其他组织和部门扩散开来,包括社会服务机构。类似于计划、预算、质量控制、责任以及后来更加复杂的如成本—收益分