

第 1 章

财政支出政策绩效评价的 概念与性质

绩效评价需要结合技术、管理与政治智慧，因为它使用的是高度理性、科学的观察与分析，所以它需要具备专业技术知识。

——戴维·奥斯本，2000

正确地认识和理解财政支出政策绩效评价是科学开展财政支出政策绩效评价的基础。为此，我们首先需要对财政支出政策绩效评价的概念、性质与分类等问题做一分析和讨论。

1.1 财政支出政策绩效评价的概念

财政支出政策绩效评价是一种典型的公共政策评价（public policy evaluation）。作为一种公共政策评价，其绩效评价的对象是一种特定的公共政策，即涉及财政资金使用或支持的公共政策。这种特点的公共政策大多是分配性或再分配性的政策^①。政策学家詹姆斯·安德森认为，“公共政策实施的有效性很大程度上依赖于政府对其的资金支持。”^②因此，可以说，大部分公共政策评价都是财政支出政策绩效评价。为了全面地认识和理解财政支出政策绩效评价，我们从三个方面进行分析和讨论：首先，分析和研究财政支出政策的概念、特点及其载体；其次，分析和讨论政策评价的内涵及其特点；最后，给出财政支出政策绩效评价的定义。

1.1.1 财政支出政策及其载体

政策一般是指，国家、政党为了实现一定历史时期的路线和任务而规定的行动准则。^③据此，我们认为，财政支出政策是国家制定的指导财政支出分配活动和处理有关财政支出分配关系的基本方针和准则。它主要通过运用预算、投资、补贴等分配手段，贯彻国家的路线、方针及其政策意图，服务于国家社会、经济发展的目标和任务，为国家履行各项政治经济职能提供财力保障。^④所以，财政支出政策不仅是公共政策的一个主要组成部分，也是国家的一个重要经济调控手段。

按照经济性质的不同，财政支出政策可以划分为两大类：购买性支出政策和转移性支出政策。前者表现为政府购买商品和服务的活动；后者则直接表现为资金无偿的、单方面的转移（如政府用于养老金、补贴、失业救济金等方面的支出）。从政策结果的影响来看，购买性支出政策对社会的生产和

① 分配性政策涉及将服务和利益分配给人口中的特定部分——个人、团体和社区。它一般只产生得利者，而没有明确的失利者。与之对应的是管制性政策，即对个人或团体的行动加以限制和约束的政策。

② 詹姆斯·安德森. 公共政策制定[M]. 5版. 谢明, 等译. 北京: 中国人民大学出版社, 2009: 第194页.

③ 在学术界，政策有多种定义。该定义是《辞海》关于“政策”的定义。引自陈庆云. 公共政策分析[M]. 2版. 北京: 北京大学出版社, 2011: 第2页.

④ 苏明. 财政支出政策研究[M]. 北京: 中国财政经济出版社, 1999: 第4页.

就业有直接影响。此类的政策也影响分配，但是这种影响是间接的。转移性支出政策，直接影响收入分配，而对生产和就业的影响是间接的。联系财政的职能来看，购买性支出政策主要执行资源配置的职能；转移性支出政策，则主要执行收入分配的职能。资源配置职能旨在促进效率，而收入分配职能则旨在促进公平。所以，各种财政支出政策虽然无一例外地均表现为资金从政府手中的流出，但是不同的财政支出政策对经济和社会的影响却存在着差异（陈共，2009）。

根据上述的定义，财政支出政策是指导财政支出分配活动的基本方针和准则。在各国的政策实践中，其主要载体一般为各类财政支出项目（program）或者项目组合（a collection of various projects and programs）。^①这里，所谓项目，是为解决某些问题的一系列行为的集合。由于这些项目或项目组合有着共同的政策目标（即可被用于解决某个社会问题或者满足某种社会的需要），因而，构成具体的财政支出政策。换言之，财政支出的项目或项目组合^②是财政支出政策的具体化（参见专栏1-1）。

专栏1-1 政策（policy）与项目（program）的关系

政策（policy）说明的是相对来说比较宽泛的社会导向目标，这些目标以更高层次的理想为基础，而项目（program）则与定义更清楚、更具体的目的有关，这些目的的基础乃是总的政策目标。例如，自由教育是一种理想，学校教育是一个目标，而上学则构成特定的目的。

要想纳入政府部门的议事议程，必须把政策转化成项目。项目将确定并列出的总的目标所衍生出的特定目的。例如，州际公路项目是全国交通（也包括公共交通、航空等）政策的规划性目标之一。因此，政策回答的是社会系统里广泛存在并且常常互相交织的问题，而项目则针对的是这一系列问题的特定情况。

资料来源：弗兰克·费希尔《公共政策评估》，2003

正是由于政策（policy）与项目或项目群（program/projects）之间存在着

① Ministry of Foreign Affairs of Japan, *ODA Evaluation Guideline (7th Edition)*, April 2012, p. 20; 彼得·罗希. 评估：方法与技术[M]. 7版. 邱泽奇, 等译. 重庆: 重庆大学出版社, 2007: 中文版前言.

② program可译为计划、方案、项目或者项目群等。program通常代表的是多个干预（interventions），正如戴维·罗伊斯所说，program是“为解决某些问题的一系列有计划的行为的集合”。而project通常表示的是单个干预（intervention）。

密切的联系，在使用中人们常常将二者进行连用或互换。比如，在《中共中央 国务院关于全面实施预算绩效管理的意见》（2018）中，即多次出现“政策、项目”或“重大政策、项目”的表述（专栏1-2）。在国外许多的《公共政策学》和《公共项目评价》教材中，政策和项目也常常进行连用或互换。例如，弗兰克·费希尔在《公共政策评估》一书中写道，“为了提供复杂的社会问题和经济问题的信息，为了评价解决问题的过程，评估将焦点集中在政策或项目的结果（结果评估）或政策项目的形成与实施的过程（即过程评估）上。”^①彼得·罗希在《评估：技术与方法》中认为，“评估的任务就是了解项目的形成，即形成政策的特定形式。”^②詹姆斯·安德森在《公共政策制定》中也如此写道，“一项政策或一个项目的支持者，他们不可能仅仅满足于这些政策或项目通过立法批准。更重要的是，他们必须努力争取以确保该政策或项目有足够的、持续的资金支持，一直到该政策或项目的目标能够圆满地完成。”^③事实上，许多国家并未将项目评估和政策评估（评价）加以区分。例如，法国的“公共政策评估”指的是项目评估和政策评估（评估技术委员会，1996）。^④由此可见，在西方国家中，政策（policy）和项目（program）两个术语在多数情况下可以互换。但有些情况下，“政策评价”的范围更广，因为它可以涵盖多个项目、管理框架以及项目和规章制度之间的相互关系等方面。^⑤

专栏1-2 中共中央（2018）34号文中有关“政策、项目”的表述

- 坚持问题导向，聚焦提升覆盖面广、社会关注度高、持续时间长的重大政策、项目的实施效果。
- 既关注新出台政策、项目的科学性和精确度，又兼顾延续政策、项目的必要性和有效性。
- 对新出台重大政策、项目开展事前绩效评估。

① 弗兰克·费希尔. 公共政策评估[M]. 北京：中国人民大学出版社，2003：第2页.

② 彼得·罗希等. 评估：方法与技术[M]. 7版. 邱泽奇，等译. 重庆：重庆大学出版社，2007：中文版前言.

③ 詹姆斯·安德森. 公共政策制定[M]. 5版. 谢明，等译. 北京：中国人民大学出版社：第194页.

④ Richard Allen & Daniel Tommasi. 公共开支管理——供转型经济国家参考的资料[M]. 章彤，译. 北京：中国财政经济出版社，2009：第334页.

⑤ 同上书，第333页。

- 支出方面，要重点关注预算资金配置效率、使用效益，特别是重大政策和项目的实施效果。
- 对实施期超过一年的重大政策和项目实行全周期跟踪问效，建立动态评价调整机制。
- 建立重大政策、项目的绩效跟踪机制，对存在严重问题的政策、项目要暂缓或停止预算拨款，督促及时整改落实。

资料来源：《中共中央 国务院关于全面实施预算绩效管理的意见》，2018.9

1.1.2 政策评价的内涵及其特点

如上所述，财政支出政策绩效评价是一种典型的公共政策评价，因此认识财政支出政策绩效评价首先要了解政策评价。关于政策评价的定义，学术界主要有以下四种观点：

第一种观点认为，政策评价是对政策方案的评价，即通常所说的前评价。这种观点强调的是对政策方案的评价，目的在于分析、比较各种政策方案的可行性以及相对的优缺点。该评价也可以称为事前评价，是政策执行之前进行的带有预测性质的评价。事前评价的重要作用是将评价从单纯的事后检测变为事前控制。它主要对两种情况做出预测：一是实施政策会有什么结果；二是不实施这项政策会有什么结果。

第二种观点认为，政策评价是对政策全过程的评价，既包括对政策方案的评价，还包括对政策执行以及政策结果的评价。我国台湾的两位学者林永波、张世贤认为，政策评价是“有系统地应用各种社会研究程序，搜集有关的资讯，用以论断政策概念与设计是否周全完整，知悉政策实际执行的情形、遭遇的困难，有无偏离既定的政策方向，指明社会干预政策的效用”。^①

第三种观点认为，政策评价的实质是发现政策问题和解决政策问题。持这一观点的主要是我国台湾学者朱志宏，他认为，“就一项公共政策而言，发现政策误差、修正政策误差就是政策评估，换言之，政策评估的工作就是发现并修正政策的误差。”^②

第四种观点认为，政策评价就是对政策效果的研究，即通常所说的后评

① 林永波、张世贤. 公共政策[M]. 台北：五南图书出版公司，1982：第499页.

② 朱志宏. 公共政策[M]. 台北：三民书局，1995：第299页.

价。美国学者托马斯·戴伊在《自上而下的政策制定》一书中提出，“政策评价就是了解各个政策所产生的效果的过程，就是试图判断这些效果是否是所预期的过程，就是判断这些效果与政策的成本是否符合的过程。”^①詹姆斯·安德森也认为，“系统（政策）评价主要寻求政策和项目的结果信息，直接关注某一政策或项目（计划）对公众需要或其解决问题的作用。”^②

在上述四种观点中，第四种观点是学术界的主流。^③美国学者苏珊·韦尔奇指出，“当社会科学家谈到项目或政策评估（我们将交替使用这两个术语）时，通常是指对项目结果的评估。问题的答案既不是项目拥有多少资源，也不是项目是如何运行的，而是项目是否已经达到了既定的目标。”^④因此，财政支出政策绩效评价本质上是一种结果为导向的评价，它主要关注财政支出政策的实施结果，关注政策产出（output）、成效（outcome）和影响（impact）的实现。这一评价旨在回答以下基本问题：

- 政策实施后，是否实现了（或者可能实现）预期的政策目标？其实现的程度和范围如何？
- 政策对社会经济和生活带来了（或可能带来）哪些影响？正面和负面的影响各是什么？
- 财政支出与政策效果是否相称？政策是否为满足一系列特定需求所能采取的最具成本效益性和可持续性的方案？
- 如果政策的实施效果不理想，那么阻碍政策效率和效果实现的主要因素是什么？
- 政策的未来走向如何？是继续、调整或是终止？如需调整，如何调整？

那么，财政支出政策评价为什么需要主要关注结果？大家知道，一项支出政策是否合理，主要是看政策实施效果是否达到了其预定的目标。从实践来看，由于政策活动本身所具有的复杂性，政策规划的目标和手段在政策的实施中可能遇到这样或那样的问题，政策预期的目标并不是都能够顺利地实

① 托马斯·戴伊. 自上而下的政策制定[M]. 北京：中国人民大学出版社，2002：第203页.

② 詹姆斯·安德森. 公共政策制定[M]. 5版. 北京：中国人民大学出版社，2009：第297页.

③ 这里需要注意的是，这并不意味着其他观点没有意义。在第5章中我们会看到第一种观点（即前评价）的作用和意义，在第6章的事中评价中会看到第三种观点的影响，而在第4章的全过程指标体系框架中可以看到第二种观点的影响。

④ 苏珊·韦尔奇、约翰·科默. 公共管理中的量化方法：技术与应用[M]. 3版. 北京：中国人民大学出版社，2003：第5页.

现。有些时候，还可能导致政策的失败。造成政策失败的原因，从管理学的角度来分析，既可能是因项目不能按事先设计的那样被执行而导致失败，也可能是按原来设计的要求来执行，但是由于基本理论的错误而达不到预期的结果。从经济学角度来分析，主要是因为财政支出可能会产生不同性质的政策效应，例如挤出效应、替代效应、收入效应和分配效应等。^①这些政策效应可能有助于达到预定的政策目标，也可能妨碍达到预定的政策目标。为此，就需要通过专门的政策评价来探究和了解政策的实施效果。可以说，探究政策效果正是政策评价的主要使命所在。进一步来说，一个完整的政策过程，除了科学的制定和有效执行外，还需要对政策执行以后的结果进行政策评价，以确定政策的成效或是价值，发现政策存在的主要问题，从而决定政策的延续、革新或终结。换言之，对于结果的政策评价是政策运行过程中不可缺少的一环。

政策评价主要是对政策的实际效果进行评价，它要求对政策给社会所带来的实际变化加以测评。开展这一测评，比较理想的做法是，在引进政策前后进行比较衡量，并用没有受到政策影响的控制小组进行衡量。也就是说，不仅要测量“政策给社会所带来的实际变化”（“what have happened”），而且要测量政策如果没有实施社会的发展状况（“what would have happened if the policy had never been implemented”）。为了对政策给社会所带来的实际变化加以测评，评价者在评价中往往需要采用实验性设计。美国著名的政策学家詹姆斯·安德森在《公共政策制定》中提出了三种基本的研究方法：实验研究方法、准实验研究方法和前后对比方法（专栏1-3）。由于政策评价的中心问题是回答政策是否有效（这是一个因果性问题），因此，国外的许多学者，特别是提倡实验设计的学者认为，随机实验是回答因果关系问题的最佳途径。^②尽管现实的条件经常会限制实验研究方法的使用，但是，随机实验的逻辑是所有政策评价设计和数据分析的基础。所以，评价人员仍然有必要学习并熟悉它们。关于实验研究的具体方法，我们在第7章（“基于因果分析的影响评价”）中还有更深入的讨论。除了实验研究的方法，还有其他的

① 通常所说的财政挤出效应是指政府的财政支出会通过税收机制减少私人部门可支配收入，而公共支出的财政挤出效应则是指政府的旨在增加社会福利的某些支出项目可以减少公众有关的支出，从而使得公众实际可支配收入趋于增加。参见华民. 公共经济学教程[M]. 上海：复旦大学出版社，2000：第151页。

② 参见詹姆斯·麦克戴维，劳拉·霍索恩. 项目评价与绩效测量：实践入门[M]. 李凌艳，等译. 北京：教育科学出版社，2011：第336-337页。

方法可以帮助了解政策的效果，我们在第3章、第4章和第6章中对此都有详细的论述。

专栏1-3 政策评价的三种基本方法

- 实验设计（experimental design）：引入控制组，并且随机分配实验对象。
- 准实验设计（quasi-experimental design）：采用一定的操控措施，但是没有随机分配实验对象到实验组和控制组。
- 前后对照设计（before and after studies design）：将计划被执行一段时间之后的结果与计划初存在的条件进行对比。

资料来源：詹姆斯·安德森，2009

1.1.3 财政支出政策绩效评价的概念界定

通过上面的分析可以知道，政策评价“就是对政策效果的研究”。因此，财政支出政策绩效评价首先要关注政策的实施效果。除此之外，由于该项政策涉及财政资金的使用或支持，因而它还需要关注资金的成本和使用效率。

2020年2月，财政部预算司修订和颁布了《项目支出绩效评价管理办法》（财预〔2020〕10号）。在该办法中，财政部预算司给出了如下的项目支出绩效评价的定义。我们认为，财政部预算司的这一定义也适用于财政支出政策绩效评价。

- 项目支出绩效评价（简称绩效评价）是指财政部门 and 预算部门（单位）根据设定的绩效目标，对项目支出的经济性、效率性、效益性和公平性进行客观、公正的测量、分析和评判。

根据该定义，项目支出（财政支出政策）绩效评价的基本内容为：经济性（economy）、效率性（efficiency）、效果性（effectiveness）^①和公平性（equity），亦即“4E”。

- 经济性（economy）

“经济性”关注的主要是投入和成本，它要求以尽可能低的投入和成

^① “effectiveness”有时也译为有效性、效益性。它有狭义和广义两种用法，狭义用法是指项目或政策目标的实现程度，与“efficacy”同义；广义用法还包括了相关性和效率性，此时译为“有效性”更为准确。

本，提供与维持既定数量和质量的产品或服务。换句话说，经济性意味着“节约和没有浪费”。经济性评价关心的主要问题是：“政府在既定的时间内，究竟花费了多少钱？对一项产出来看，该花费是不是物有所值？”（见专栏1-4）

专栏1-4 经济性

经济性是指，对一特定产出而言，没有浪费。

注：当一项活动所使用的稀缺资源成本接近于为实现计划的目标所需的最低成本时，该活动就是经济的。

资料来源：OECD，2002

● 效率性（efficiency）

经济性关心的主要是投入和成本问题，而效率性则是一个比较的概念，它主要对比产出或服务与资源投入的关系，关心的主要是投入和产出的关系问题（专栏1-5）。“从给定的投入得到最大的产出”或“用最小的成本来达到目标”就是效率的最基本的定义。简言之，效率性主要关注投入是否得到充分有效的利用（技术效率）。

专栏1-5 效率性

效率性是对资源或投入（包括资金、专业知识和时间等）转化成产出或结果的经济性的衡量。

资料来源：OECD，2002

● 效果性（effectiveness）

效果性表达的则是产出与效果之间的关系，它通过对比资金支出后所实现的实际效果与预期效果之间的关系，保证资金达到理想的效果（专栏1-6）。效率告诉你用最小的努力来达到目的，并不告诉你方向；而效果性则衡量项目或政策目标的实现程度，告诉你工作努力的方向。因此，相比于

效率性而言，项目或政策的效果性是绩效评价更为重要的内容。如果没有效果，无论是什么政策，即使再具有效率，也是没有意义的。

专栏1-6 效果性

效果性是指，发展干预活动的目标在多大程度上实现了，或者根据其重要性预计这些目标会在多大程度上可以实现。

注：也可用来综合衡量（或判断）一项活动的优点或价值，例如某一活动在多大程度上实现了或预计会实现其目标，其主要相关目标是否具有可持续性，是否会给机构的发展带来积极影响。

资料来源：OECD，2002

● 公平性（equity）

公平性与法律和社会理性密切相关，是指效果和努力在社会不同群体中的分配。一项公平的政策是指效果（如服务的数量或货币化的收益）或努力（如货币成本）能被公平或公正地分配。一项政策方案可能既有效益（效果），又有效率，但是由于成本和收益的不公平的分配而缺乏公平性。比如，那些最需要的人并没有得到与他们的人数成比例的服务，最没有支付能力的人却要超出比例地分摊成本，或者最得益者并未支付成本。

上述“4E”中的经济性（economy）、效率性（efficiency）、效果性（effectiveness）（简称为“3E”）代表了项目结果链中的三种联结关系。如图1-1所示，经济性反映的是资源与投入之间的关系，效率性反映的是投入和产出的关系，而效果性则反映的是产出与效果之间的关系。

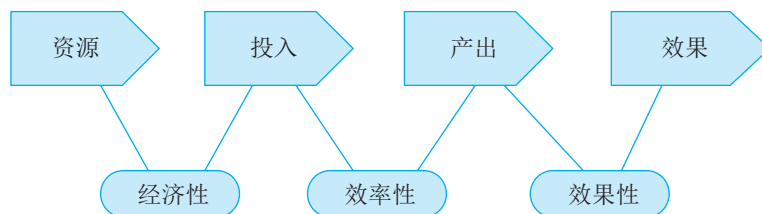


图1-1 “3E”关系图

定义是对于一个事物本质特征或一个概念的内涵和外延所做的一种简要

说明。根据上述定义，我们知道，财政支出政策绩效评价的主要任务就是要研究和回答财政支出政策的经济性、效率性、效果性和公平性问题。这告诉我们，无论是开发评价指标体系，还是实施绩效评价，都应当主要围绕经济性、效率性、效果性和公平性四个问题来进行。2018年9月中共中央发布的《中共中央 国务院关于全面实施预算绩效管理的意见》就明确指出，全面实施预算绩效管理要“更加注重结果导向、强调成本效益、硬化责任约束”。这就是说，预算绩效管理与评价的核心是效果（结果导向）与效率（成本-效益）。国务院秘书长肖捷（2017）也发文强调，要“注重成本-效益，关注支出结果和政策目标实现程度”^①。然而，在我国财政支出绩效评价的实践中，我们对这两个方面的“注重”和“强调”都还远远不够。从财政部预算司发布的《项目支出绩效评价共性指标体系框架》^②（见表2-1）来看，其评价的基本内容包括四项：投入、过程、产出和效果。但是，很显然，其评价重点不是产出和效果，而是投入和过程，因为在共性指标体系框架中，投入和过程指标占了近70%，产出和效果的指标不到30%，成本-效益分析则更是欠缺。显然，该指标体系框架与《中共中央 国务院关于全面实施预算绩效管理的意见》的基本要求不相符合。为此，上述共性指标体系框架应大幅减低投入和过程指标的权重，转而加大产出和效果指标的权重，并增加成本效益分析（效率）的指标。简而言之，财政支出政策绩效评价应当重点关注“效果”和“效率”两个方面（见专栏1-7）。

专栏1-7 效果与效率要并重

歼敌三千，自损八百太蠢！一个指挥员不仅要负歼敌三千之责，也要负自损八百之责。

资料来源：刘伯承，1947

① 肖捷. 加快建立现代财政制度[N]. 人民日报, 2017年12月20日.

② 参见财政部关于印发《预算绩效评价共性指标体系框架》的通知, 财预〔2013〕53号.

1.2 财政支出政策绩效评价的性质

1.2.1 财政支出政策绩效评价的性质是一种评价研究

财政支出政策绩效评价在性质上属于一种特殊的应用研究——评价研究（参见专栏1-8）。作为评价研究，它主要对财政支出政策或项目产生的结果（效果和效率）进行系统性分析。其主要的特点是，重视实证主义的研究设计，重视对过程与结果的监测，重视应用统计抽样技术收集证据，重视对收集的资料进行分析和解释，以及重视建立可预测的因果关系模型等。在评价的方法上，它表现为一系列实证分析技术的结合：观察法、问卷调查、实地访谈、焦点小组、专家评判法、德尔菲法、统计抽样、成本-效益分析、实验研究、准实验研究、多元回归分析以及系统分析等。著名社会研究方法专家劳伦斯·纽曼认为：“评价研究是一种广泛使用的应用研究。这种研究被广泛地用于大型科层组织，去发现一个项目、一种新的做事方法、一种营销活动、一个政策等是否有效。评价研究测量一个方案、政策、做事方式的有效性。”^①

同基础研究一样，应用研究也是一个有逻辑、有顺序的调查过程。二者的区别在于对数据的期望和数据的使用目的。基础研究的目的是产生一般化的知识（理论），而应用研究所带来的信息一般只适用于一个特定的政策或项目。詹姆斯·安德森认为，政策评价“运用社会科学的方法论来测量政策或项目的社会影响以及它们达到预定目标的程度”。^②侧重结果的评价研究在很大程度上需要依赖于社会科学方法，是因为这些方法不仅能回答社会科学家提出的一般的理论性问题，同时，也能够回答政策分析人员提出的具体的应用性问题。具体来说，社会科学方法能应用于政策评价是因为：

- 在某种程度上为可解答问题的形成提高了可能性（形成假设）。
- 能够提出收集答题信息的策略（研究设计）。
- 为确定答题信息类型以及如何收集图提供指导方针（策略和设计数据）。
- 提供适当的数据分析技术（统计分析）。^③

① 劳伦斯·纽曼. 社会研究方法——定性和定量的取向[M]. 5版. 郝大海, 译. 北京: 中国人民大学出版社, 2007: 第32页.

② 詹姆斯·安德森. 公共政策制定[M]. 5版. 北京: 中国人民大学出版社, 2009: 第297页.

③ 苏珊·韦尔奇、约翰·科默. 公共管理中的量化方法: 技术与应用[M]. 3版. 北京: 中国人民大学出版社, 2003: 第5-6页.

从评价研究的角度来说，财政支出政策绩效评价需要研究和分析财政支出政策的所有重要方面，具体包括^①：

- 项目需求评价：分析政策或项目运作所需的社会条件以及需求程度等问题。
- 变革理论评价：分析政策或项目的概念化和设计等问题。
- 项目过程评价：分析政策或项目的实施及服务送达等问题。
- 项目效果评价：分析政策或项目的成效和影响等问题。
- 项目效率评价：分析项目的成本-收益和成本-效果等问题。

(1) 项目需求评价。项目需求评价通常被作为设计和规划新项目或重组既有项目的第一步，用来提供需要什么服务和怎样最有效向服务的需求者送达服务的信息。对已经建立、稳定的项目来说，需求评价也很重要，能够检验服务是否满足了目标群体的实际需要，从而为项目改进提供指导。需求评价的重要方式是，评估某个社会问题的性质、重要性和分布状态，评估针对这一问题实施干预的必要性，分析现有环境对于干预的概念化和设计的意义。进行需求评价，评估者经常要分析和利用现有的统计资料或者进行深入的社会调查。

(2) 项目理论评价。项目理论指的就是项目逻辑框架（或结果链）。项目理论评估（assessment of program theory）关注的是与项目的概念化和设计方法相关的问题。第一，项目理论评估包括以外化的和具体的书面或图表形式呈现项目理论。第二，可以用各种方法检验理论的合理性、可行性、道德性，甚至其他方面。项目理论评估是项目初期（在项目刚开始或规划阶段）基本的工作，当然也是在事前评估阶段需要做的基本工作。对于已经建立的项目而言，项目理论评估也是适用的，尤其是需要说明开展的项目服务在何种程度上达到了他们尽量去满足的社会需求的时候。也就是说，开展项目的事中评价和事后评价，也都需要进行项目理论评价。

(3) 项目过程评价。项目过程评价就是评估项目的活动和操作情况。过程评价致力于解决那些和项目如何运作相关的问题，包括服务和项目目标的一致程度、服务送达干预对象的好坏、项目管理的绩效、项目资源的使用情况，以及其他类似方面的问题。过程评价可以用作独立的评价方法（即实施独立的事中评价），也与结果评价一起用作比较复杂的综合评价的一部分。

^① 彼得·罗希，等. 评估：方法与技术[M]. 7版. 重庆：重庆大学出版社，2007：第39-45页.

作为独立评价，过程评价能够提供高质量的信息，即根据既有的标准，评估项目按照预先设计实施的程度。对于结果评价（事后评价）来说，过程评价也是不可缺少的。事后评价提供的是项目效果方面的信息。但是，如果没有过程方面的信息，那么对于项目效果的说明就是不完整的和不明确的。如果项目产生了效果，过程评价有助于确认项目效果是否是项目活动的结果，而非虚假的关系。

（4）项目效果评价。效果评价是项目评价的核心。除非项目拥有经过证明的结果（或效果），否则很难给出将项目执行或维持下去的理由。项目效果评价主要用来评估在一定社会环境中的项目产生了哪些预想效果，项目对社会环境的干预是否发生了作用，项目影响中是否包含意想不到的效果。我们评价某个具体项目效果的最大困难，就是其成效往往是由与项目不相关的因素所造成的，为此，效果评价的基本目标就是对干预活动的净效果进行估计，即是说，估计在没有其他过程和事件的影响下干预的纯粹效果。其通常使用的方法是：①运用实验设计或者有控制组的前测和后测设计方法等进行研究设计；②运用测量工具（如指标、指数、量表）进行数据收集；③利用前述收集的数据进行变量关系的分析（如回归分析），以确定政策或项目的净效果。如果上述方法缺乏实施的条件，评价者也可以现场的观察、相关资料的分析、个案研究等方法，来探寻和了解项目或政策实施效果。但是，这一方法在揭示造成效果的原因方面，没有前一方法（即实验方法）更为科学、严谨。

（5）项目效率评价。只有效果的信息通常是不够的，项目的效果还必须经过和成本对比再做出判断。美国评价专家彼得·罗希说，“项目成功执行的程度如何，获得的成效如何，对项目经理、项目各方和决策者而言，都是不可缺少的信息。然而，几乎在所有的情况下，最重要的是要知道与项目成本相比较，项目的成效如何。在事实上，项目是否给人留下深刻的影响，正如大多数日常生活判断或正式决策过程一样，在决定是否应该扩展、继续或结束项目时，比较项目的成本与收益（即效率）是最重要的考虑。”^①效率评价主要分析项目的成本与效果之间的对比关系，解决这类评价问题的两类密切关联的方法是：成本-效益分析方法（cost-benefit analysis）和成本-效果

① 彼得·罗希，等. 评估：方法与技术[M]. 7版. 邱泽奇，等译. 重庆：重庆大学出版社，2007：第231页.

分析方法（cost-effectiveness analysis）^①。其中，成本-效益分析需要将成本与收益均量化为货币，计算是否存在社会净收益。而成本-效果分析则是将成本与单一的非货币化产出（即产出单位）进行比较。这类评价的典型问题主要包括：（1）“相对于付出的成本而言，项目是否产生了足够的收益？”（2）“项目创造的收益是否比其他致力于相同目标的干预所消耗的单位成本要低？”

专栏1-8 评价研究的早期历史

项目评价（program evaluation）和评价研究（evaluation research）是新近创造的词汇，但却不是刚刚才有的活动。评价活动可以追溯到科学的初创期，科隆巴赫（Cronbach, 1980）及其同事指出，三个世纪以前，霍布斯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，作为当代评价研究面临的最大的技术挑战，也不是最佳的发明。最早的“社会实验”发生在18世纪初期。那时，一个英国船长观察到，在地中海地区的诸多国家的船只上工作的海员很少患坏血症；他同时注意到，柑橘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。按照现在的医学知识可知，患坏血病是因为缺乏维生素C，而柠檬富含维生素C。总之，干预发生了作用。由此，英国船员都被迫食用柑橘类果实，这就是柠檬茶流行的由来。巧合的是，人们花了大约50年的时间才使创造的“社会项目”得到广泛利用。

资料来源：彼得·罗希等，《评估：方法与技术》，2007

1.2.2 政策评价的主要目的是辅助政府决策

在公共管理中，决策者往往需要关注和回答以下问题：一项政策（项目）是否达到了目标，是否已经解决了问题，投资是否物有所值，或者是否

^① 成本-效果分析方法（cost-effectiveness analysis）在国内常常称为最低费用（成本）法。其基本含义是，不对效果进行量化，仅对实现同一效果或目标的不同方案成本进行比较，选择其成本最低的方案。

有达到同样效果的、花费更少的项目，等等（专栏1-9）。为此，决策者经常需要获取政策结果信息。评价研究的一个主要任务即是帮助决策者收集、整理和分析这些政策结果的信息，以让人们在某些政策行动和结果了解得更清楚。由于通过政策或项目评价提供的客观的或者“硬数据”指导下的决策比没有任何评价数据下的决策要胜出一筹，^①因而管理者进行公共决策时往往需要依赖于政策评价。用世界银行专家库赛克的话来说，就是“一个好的决策往往要基于监测、衡量和评价”。^②对于政策评价在决策中扮演的这一角色，英国财政部在其《红皮书》（2011）中这样写道，“政策评价检查一项政策的实施情况和效果，评估政策预期的效果及成本效益是否实现。评价发现能告诉决策者哪些政策是成功的，哪些是不成功的（发现问题），它总结好的政策经验，揭示政策的非预期结果，展示资金投资的价值，并能够反馈给事前评估过程，改进未来的政策决策。”^③美国评估学家彼得·罗希说得则更为明确：“项目评估的角色就是向人们提供答案。它所要回答的问题就是项目怎样做才是有价值的，项目实际上如何被实施。对评估研究而言，这是最基本的，即评估研究的基本目标就是知会社会行动。因此，对评估研究特别关注的，就是那些根据评估结果来决策和采取行动的各类对象。评估的发现也要有利于决策者的决策，譬如对项目的某个方面进行调整，或者启动新的项目或继续已有的项目。”^④为此，政策评价者必须要收集、管理和分析数据，并且所有这些数据都必须是“基于实证的事实”，而不能仅是一部分人的“主观看法”。

进入大数据时代，决策最重要的依据将是系统的、成片的、动态的数据流，而不是个人经验或长官意志。过去深入群众、实地考察的工作方法虽仍然有效，但对政府和国家的治理而言，系统采集的数据、科学分析的结果更重要。我国著名大数据专家涂子沛提出，数据治国就是要凭借对数据的有效收集、处理和分析来治理国家，决定国家的大政方针和具体政策。他还说，数据治国可以理解为“用定量分析、实证研究的方法来治理国家”。^⑤为此，

① 戴维·罗伊斯，等. 公共项目评估导论[M]. 3版. 北京：中国人民大学出版社，2007：第19页。

② 乔迪·库赛克，等. 十步法：以结果为导向的监测与评价体系[M]. 北京：中国财政经济出版社，2011：第151页。

③ 英国财政部红皮书：评价指南（HM Treasury: *The Magenta Book: Guidance for evaluation, 2011*），第7页。

④ 彼得·罗希，等. 评估：方法与技术[M]. 7版. 邱泽奇，等译. 重庆：重庆大学出版社，2007：第14-15页。

⑤ 涂子沛. 大数据[M]. 3.0升级版. 南宁：广西师范大学出版社，2015：新版自序。

政策决策者需要的是全面、系统的，并且是能够证实的证据或数据，而不是一些片面的、零碎的数据或主观观点。对于政策评价来说，人们希望它能基于事实证据而不是个人观点客观地得出，某个政策或项目是好的，而另外一个不是好的结论。并且该结论必须是能够重现的（可以再生的），换成其他的人也能够得到同样的不受约束的结论。这就要求，政策评价人员要能够掌握一系列的实证分析技术方法，如观察法、调查法、实地访谈、专题小组、案例研究、统计抽样、成本-效益分析、实验研究、多元回归分析以及系统分析等。尽管这些方法有时是麻烦的和不完美的，但是以主流科学的方法和哲学为指导，同时主张定性和定量研究方法的、系统化的、实证类型的评估工作是无可替代的。^①

专栏1-9 政策制定的老手要看实际的评价结果

尽管不断有行政官员说服我们，要增加这个或那个项目的经费，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儿，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6 000万美元。在过去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。对比，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”的文章，文章似乎是华盛顿社会

^① 戴维·罗伊斯，等. 公共项目评估导论[M]. 3版. 北京：中国人民大学出版社，2007：第22页.

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《伊利诺伊家庭优先防护项目评估总报告》。这是自1987年伊利诺伊家庭维持法案后，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4 500份案例资料和服务特点资料：大约1 600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，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·莫怡尼安参议员 敬上

资料来源：彼得·罗希等，《评估：方法与技术》，2007

遗憾的是，在我国政策绩效评价实践中，有一些评价只是简单地根据个人感觉与判断进行打分或评级。科学性和客观性是政策评价的基本要求，简单打分式的评价显然达不到这一基本要求。为此，我们有必要改变现行简单化、主观性的“打分”方法，采用科学性、实证性的研究方法。正如著名科学家爱德华·威尔逊所说，“感觉良好并不是科学。正确实验，然后根据经过测试和仔细权衡的客观知识做出社会决策，这才是科学的意义所在。”^①

▶ 1.3 财政支出政策绩效评价的分类

按照不同的标准，财政支出政策绩效评价可以分成不同的类型。本节主要介绍其两种重要分类：一是内部评价与外部评价；二是事前、事中和事后评价。前者是按评价主体进行的分类，后者则是按评价时点进行的分类。

1.3.1 内部评价与外部评价

财政支出政策绩效评价的主体既可以是评价对象内部的组织或个人，也可以是评价对象外部的组织或个人。由前者进行的绩效评价，称为内部评价，包括由业务部门内执行人员实施的评价和由专职评价人员实施的评价两种形式。由后者进行的绩效评价，称为外部评价，包括由财政部门、专业学术团体和研究机构实施的评价以及目标群体进行的评价，向外部门的委托评价也属之，被委托的对象可以是研究机构、学术团体、高等院校和专家学者等。

^① 原文为：“...feeling good is not what science is all about. Getting it right, and then basing social decisions on tested and carefully weighed objective knowledge, is what science is all about.”引自詹姆斯·莱斯特，等。公共政策导论[M]. 2版. 北京：中国人民大学出版社，2004：第125页。

内部评价和外部评价各有利弊。一般而言，内部评价人员通常较外部人员更为了解评价对象的情况，因而能够提出更为相关与恰当的问题，并能获得更重要的发现。然而，上述优势也可能成为其劣势。由于置身评价对象之中，内部评价人员往往无法看清真相，或是无法像他人那样发现解决方案。由于缺乏独立性，内部评价更可能受到来自评价对象内部的各种压力或影响，因此，其评价结果可信度往往较低。相反，外部评价则具有较强的独立性，并且大多数人员均具备评价所必需的专业技能，所以，外部评价结果通常具有更高的可信度。但是，如果外部评价人员与评价对象存在某种关系，其评价可信度也会大打折扣。基于此，戴维·奥斯本感慨地说：“也许评估界最大的争论就是评估应该由<局内人>（如管理者）还是<局外人>（如来自中央财政部门的审计人员或评估员）来控制。”^①

为了提高和保证评价工作的质量，OECD成员国及国际组织大多选择了外部评价。他们多数的成员国及国际组织都成立了独立评价机构（专栏1-10）。独立评价机构不受项目管理部门的管辖，而是直接对其最高管理当局负责。比如，世界银行成立了独立评价局（IEG），直接对其执董会负责。有些国家（如智利、哥伦比亚）则是通过签订合同将绩效评价外包给学术界或咨询公司。这些国家将评价外包而非让政府官员来从事评价的一个理由就是，要实现更高层次的独立性、客观性和可信性，以及避免可能会因自我评价而产生的潜在利益冲突，或显在的利益冲突。

专栏1-10 绩效评价的组织形式

- **世界银行。**在世界银行内部有两个负责对项目质量进行评价的部门：质量保证小组和独立评价局（IEG）。前者主要对世行管理当局负责，其职责是对处于准备和实施阶段的世行贷款项目进行质量审查；后者是直接对执董会负责的独立评价机构，其主要职责是对世行贷款项目进行评价，即对已经完工的项目进行评价。
- **亚洲开发银行。**1978年，亚洲开发银行成立了业务评价处（Operations Evaluation Department, OED）。2004年，OED成为一个独立的部门，并通过发展有效性委员会向董事会报告；

^① 戴维·奥斯本，等. 政府改革手册：战略与工具[M]. 北京：中国人民大学出版社，2004：第53-54页.

2008年，业务评价处更名为独立评价局（Independent Evaluation Department, IED）以彰显其独立、公正的特征。IED要对1/4的各类亚行项目进行后评价，以期总结经验，改善后续投资项目的管理。

- **联合国开发计划署**。1966年，联合国开发计划署（UNDP）正式成立。为了推进其职责的履行，UNDP建立了一个由执行委员会、评价办公室、项目经理、项目官员组成的绩效评价体系。在该评价体系中，执行委员会、评价办公室、项目经理、项目官员等各司其职，定期开展评价工作。UNDP的评价有两种形式：一是评价办公室组织的独立评价；二是项目单位委托独立外部专家开展的评价。
- **法国**。法国于1999年设立国家评价委员会，负责根据各级部门和地方政府的建议制订年度的评价计划。评价委员会由科学家、其他专家和地方的代表组成。评价研究由国家评价基金出资进行，并公布评价结果。除了国家评价委员会组织进行的评价外，各级部门和地方评价委员会也进行评价研究。
- **荷兰**。在荷兰，各级部门的预算制定者负责协调评价研究的进行，并负责提供必要的建议、指导和研究专家。**各部门的预算制定者为项目单独起草评价计划，并鼓励对政策的定期评价，确保研究分析的质量和实际应用。**

资料来源：世界银行（2011），OECD（1999）

近年来，在我国的政府绩效评价实践中，非政府组织、高等院校、独立研究和调查机构以及社会公众等在政策绩效评价的发展中起到了越来越重要的作用。比如，我国新一届政府组织的第三方评估就体现了这一特点。第三方评估是新一届政府管理方式的一个重大创新^①。李克强总理多次强调，评价政府工作不能“自拉自唱”，要让第三方独立公正评估，让社会和群众来评判，更好地推动国务院各项政策的落实。2015年7月，为全面了解党中央、

^① 为推动已出台政策措施落实，2014年6月国务院启动全面大督查，并在自查和实地督查基础上引入第三方评估。根据部署，第三方评估主要邀请全国工商联和部分研究咨询机构，围绕简政放权、棚户区改造、精准扶贫、重大水利工程等部分重点政策措施落实情况展开评估，以便与自查和督查情况进行对比分析。在督查中引入第三方评估，是政府管理方式的重大创新，体现了政府更加开放、乐于接受监督的胸怀。让专业部门评估政府工作，用群众的眼睛监督政府，有助于避免政府在自我评价体系中“既当运动员又当裁判员”；同时体现了政府抓铁有痕、勤政务实的新风。有了第三方评估情况与自查、督查情况进行“对比分析”，能够反映最真实的底层声音，借助外部力量帮助政府找准症结，对症下药，增强督查实效。

国务院“稳增长促改革调结构惠民生”一系列重大政策措施贯彻落实情况，国务院办公厅组织了对部分政策措施开展的第三方评估。为确保第三方评估取得更好效果，国务院办公厅在总结上年评估工作的基础上，进一步吸纳更加广泛的专业力量参与，通过竞争性遴选，确定委托中国科学院、国务院发展研究中心、国家行政学院、中国科协、全国工商联、中国国际经济交流中心、北京大学、中国（海南）改革发展研究院等独立第三方开展评估。政府用第三方来独立评价政府绩效表明，民间第三方机构已成为我国一种新的、重要的评价力量。从国际经验来看，政府将绩效评价外包给学术界或咨询公司，而非让政府部门或官员来从事政府绩效评价，能够实现更高层次的独立性、客观性和可信性。我们认为，这种第三方评估模式可以作为我国财政支出政策绩效评价主体发展的一个方向加以推广。

第三方评估事实上是一种政府购买公共服务制度。2018年7月，中共中央、国务院印发了《法治政府建设实施纲要（2015—2020年）》。该纲要明确提出，“建立健全政府购买公共服务制度，公开政府购买公共服务目录，加强政府购买公共服务质量监管。推进公共服务提供主体和提供方式多元化，凡属事务性管理服务，原则上都要引入竞争机制向社会购买；确需政府参与的，实行政府和社会资本合作模式。”可以预计，第三方评估模式将会在我国得到更为快速、健康的发展。

1.3.2 事前、事中与事后评价

弗兰克·费希尔在《公共政策评估》一书中说，“为了提供复杂的社会问题和经济问题的信息，为了评价解决问题的过程，评估将焦点集中在政策或项目的结果（结果评估或影响评估）或政策项目的形成与实施的过程（过程评估）上。而且，这种评估可以侧重于对政策预期效果的评估（事前评估），或者侧重于对其运用之后的实际效果的评估（事后评估）。”^①因此，公共政策评价不仅仅是事后评价。按照评价时点的不同，我们可以将财政支出政策绩效评价分为事前、事中和事后三类评价。

事前评价一般在政策实施之前进行，主要评价拟实施政策的可能后果（包括负面影响），并分析潜在的风险因素可能会对政策预期成果产生影响的程度。一般情况下，可以根据不同的假设设计多个备选方案，并进行情景分析（scenario analysis）来辅助确定具体的政策实施方案。必要时，应对拟

^① 弗兰克·费希尔. 公共政策评估[M]. 北京：中国人民大学出版社，2003：第2页.

实施的政策进行多次事前评价，根据新的信息不断更新、调整相关的政策或环境参数（如经济增长率、汇率、折现率等），修订相应的政策目标，以提高预期目标的可实现程度。

事中评价也称为过程评价，在政策的实施阶段进行，侧重于评价政策的实施、执行情况以及存在的问题。按照谢尔（Scheirer, 1994）的解释，过程评价解释“证明项目是什么和是否按照预期被送达给既定的接受者”。事中评价关注的范围一般包括：结果链中从投入到产出（也可包含短期成效）等阶段，重在研究政策实施、执行的方式和效果，主要检测事先假设的要素（投入、活动、产出和成效等）之间的“运行逻辑”与实际情况是否一致。如不一致，应分析成因，并提出问题解决方案和目标调整计划。

事后评价也称为结果评价，在政策的实施完成之后进行，侧重于评价政策的预期结果和影响是否已实现，并对它们的实现情况和程度进行归因分析，目的主要在于总结经验或教训，为未来相关政策的实施建设提供经验和教训，辅助相关部门进行科学决策。事后评价关注的范围包括结果链中的投入到产出再到短期成效（有时甚至包括影响）的全过程，主要评价政策预期结果的实现程度。如预期结果与实际不一致，应分析成因，为相关政策的实施建设提供参考。

在国际上，事前评价、事中评价和事后评价通常称为前瞻性评价、形成性评价和总结性评价（专栏1-11）。其中，前瞻性评价（prospective evaluation）一般在项目或政策实施之前进行，主要评价的是所建议的项目或政策的可能后果。形成性评价（formative evaluation）绝大多数在项目或政策的实施阶段（过程）进行，注重于项目和政策的执行和改进，所以其主要作用在于促进“学习”。总结性评价（summative evaluation）主要在项目或政策结束之后进行，注重评价项目或政策的结果和影响，目的主要在于总结经验教训，以为未来项目或政策的建设提供经验和决策参考。

专栏1-11 三种评价的不同特点

- ◆ **总结性评价（事后评价）**
注重（项目或政策）结果。
- ◆ **形成性评价（事中评价）**

注重（项目或政策）执行前和执行中绩效的提高。

◆ 前瞻性评价（事前评价）

- 对提出的干预活动所产生的可能结果进行评估。
- 该计划/项目/政策值得进行评价吗？
- 评价获得的收益值得付出这些努力和资源吗？

资料来源：伊玛斯·琳达，2009

如图1-2所示，从投入到产出（也可包含短期成效）主要是形成性评价关注的范围；而从短期成效到影响（长期成效）主要是总结性评价关注的领域。形成性评价重在研究政策或项目执行的方式，主要检测和分析项目或政策事先假设的“运行逻辑”与实际情况是否一致，其关心的主要问题包括相关性、效率、效果以及经验教训等。该评价如果是在中间阶段进行，也可称为中期评价（midterm evaluation）。中期评价是形成性评价的一种特殊类型。总结性评价经常被称为结果或影响评价，一般在项目或政策完成后进行，主要评价项目预期结果的实现程度。其关心的主要问题包括：相关性、绩效、影响、可持续性、外部效用以及经验教训等。

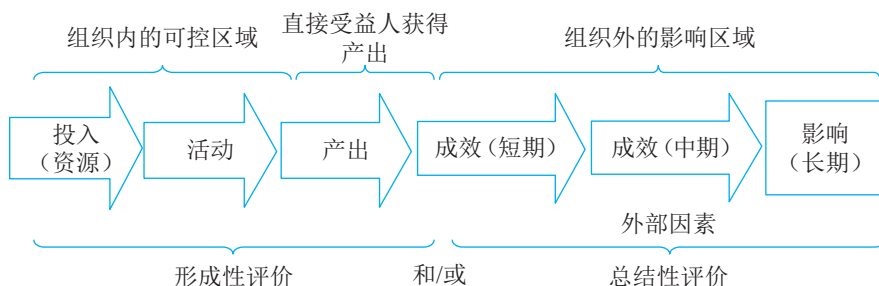


图1-2 形成性评价和总结性评价

2018年9月，《中共中央 国务院关于全面实施预算绩效管理的意见》发布。该意见提出，“对新出台重大政策、项目开展事前绩效评估。……投资主管部门要加强基建投资绩效评估，评估结果作为申请预算的必备要件。各级财政部门要加强新增重大政策和项目预算审核，必要时可以组织第三方机构独立开展绩效评估，审核和评估结果作为预算安排的重要参考依据。”还提出，“通过自评和外部评价相结合的方式，对预算执行情况开展绩效评价。”可见，我国要建立对重大政策和项目的事前、事中与事后续效评价机制。事前的绩效评估不同于现有的项目预算评审。预算评审的重点是审核项

目或政策支出预算的真实性、合理性和准确性，影响的一般只是新增项目或政策预算的多少问题，通常不会影响公共支出项目或政策的设立与存废。而事前绩效评价的主要任务是，评估项目或政策该做不该做，主要解决的是财政支持与否的问题。毫无疑问，事前绩效评价机制的建立将极大提高绩效评价在政策和预算管理中的地位与作用。

本章小结

1. 财政支出政策绩效评价是一种结果导向的评价，它主要关注财政支出政策的实施结果，关注政策产出（output）、成效（outcome）和影响（impact）的实现。这一评价旨在回答以下基本问题：政策实施后，是否实现了（或者可能实现）预期的政策目标？其实现的程度和范围如何？政策对社会经济和生活带来（或可能带来）哪些影响？正面和负面的影响各是什么？财政支出与政策效果是否相称？政策是否为满足一系列特定需求所能采取的最具成本效益性和可持续性的方案？如果政策的实施效果不理想，那么阻碍政策效率和效果实现的主要因素是什么？政策的未来走向如何？是继续、调整或是终止？如需调整，如何调整？

2. 财政支出政策绩效评价在性质上属于一种特殊的应用研究——评价研究。作为评价研究，它主要对财政支出政策或项目产生的结果（效果和效率）进行系统性分析。其主要的特点是，重视实证主义的研究设计，重视对过程与结果的监测，重视应用统计抽样技术收集证据，重视对收集资料进行分析和解释，以及重视建立可预测的因果关系模型等。在评价的方法上，它表现为一系列实证分析技术的结合：观察法、问卷调查、实地访谈、专题小组、专家评判法、德尔菲法、统计抽样、成本-效益分析、实验研究、准实验研究、多元回归分析以及系统分析等。

3. 在我国的财政支出政策评价实践中，许多的绩效评价只是简单地根据个人感觉与判断进行打分。科学性和客观性是政策评价基本的要求，这种简单打分式的评价显然达不到这一基本要求。为此，我们有必要改变这种带有简单化的“打分”方法，采用科学性、实证性的研究方法，因为只有建立在实证研究基础上的绩效评价，才可能对财政支出政策及其结果做出科学的评判。

4. 按照评价时点的不同（或者政策生命周期的不同阶段），财政支出政策绩效评价可以分为三类：事前评价、事中评价和事后评价。事前评价一般在政策实施之前进行，主要评价拟实施政策的可能后果（包括负面影响），

并分析潜在的风险因素可能会对政策预期成果产生的影响。事中评价也称为过程评价，在政策的实施阶段进行，侧重于评价政策的实施、执行情况以及存在问题，其目的是支持并改进项目的管理和实施。事后评价也称为结果评价，在政策的实施完成之后进行，侧重于评价政策的预期结果是否已实现，并对它们的实现情况和程度进行归因分析，目的主要在于总结经验或教训，为未来相关政策的实施建设提供经验和教训。或者说，事后评价是在政策或项目制定后一段时间内对其有效性和总体价值做出的评估。它一般用于分配资源和加强责任制，同时还要涉及项目效果和总体相关性等问题。

5. 内部评价和外部评价各有利弊。一般而言，内部评价人员通常较外部人员更为了解评价对象的情况，因而能提出更为相关与恰当的问题，并能获得更重要的发现。然而，上述优势也可能成为其劣势。由于置身评价对象之中，内部评价人员往往无法像他人那样发现解决方案。由于缺乏独立性，内部评价更可能受到来自评价对象内部的各种压力或影响，因此，其评价结果可信度往往较低。

复习与思考

1. 财政支出政策评价（policy evaluation）和财政支出项目评价（program evaluation）是否是一回事？如果说不是一回事，那么二者有何区别和联系？

2. 财政支出政策绩效评价在性质上是一种特殊的应用研究——评价研究。你如何来理解评价研究？请联系我国当前的政策绩效评价实践来进行分析。

3. 美国评价专家斯克莱文（Scriven）经常用一个比喻来说明项目评价工作，他说：当我们做项目评价工作时，把自己当作是专家证人。实际上，项目评价工作就是让我们来为一个项目做证，我们提供自己的专业意见，“法庭”（客户）根据我们提供的内容来决策。你是否认同这一看法？为什么？

4. 在财政支出政策绩效评价中，政府为何要提倡第三方评价？第三方评价是否一定是独立评价？

5. “批评和自我批评”是中国共产党的一个优良传统。如果我们把内部评价理解为“自我批评”，而把外部评价理解为“批评”，你认为这种观点是否正确？

6. 财政支出政策事前评价、事中评价和事后评价是否能用一套绩效评价指标体系？如果说不能，为什么？