

影响评估国际动议组织



工作文稿 4

关于影响评估的设计： 不同的视角

罗伯特·钱伯斯，迪安·卡兰，马丁·拉瓦雷，派翠西亚·罗杰斯
2009年7月

关于3ie

影响评估国际动议组织（简称3ie）旨在通过支持对客观事实的主体、时间、原因及影响程度的分析和用来改善发展中国家的人民生活水平。**3ie**是一个新的动议组织，可以满足对客观事实的更高要求，并通过制定更具针对性的政策来提升发展的有效性。**3ie**资助高质量的影响评估行动，以向发展中国家建议更好的政策和项目设计。

3ie工作文稿既涉及影响评估的有关概念性问题，又包含了个案研究和情景分析。

本篇工作文稿由**3ie**执行总裁霍华德·怀特博士创作。版权所有**3ie**，2009年。

联系方式

影响评估国际动议组织

转全球发展网络

邮编7510

瓦桑昆

印度新德里110070

电话：+91-11-2613-9494/6885

网址：www.3ieimpact.org

目 录

引言

霍华德·怀特， 影响评估国际动议组织执行总裁

更加重视穷人意见：采用参与式方法进行影响评估

罗伯特·钱伯斯， Sussex 大学发展研究学院

关于发展评估随机实验的思考：开罗评估现场的发言

迪安·卡兰， 耶鲁大学经济学教授， 创新扶贫行动成员； Jameel 贫困行动实验室成员

三类典型干预案例之评估

马丁·拉瓦雷， 世界银行

依据干预性质与评估目的进行影响评估设计

派翠西亚·罗杰斯， 墨尔本皇家技术研究院， 评估研究与咨询， 跨学科研究协作组

引言

近年来，关于影响评估设计方法的争论陷入了僵局。由 3ie、影响评估网络中的网络（NONIE）、非洲评估协会（AfrEA）和联合国儿童基金会（UNICEF）于 3 月 29 日至 4 月 2 日在开罗举办的影响评估面面观国际会议旨在听取不同意见并寻求共识。此会期间的一场重要活动是在一次全体会议上询问代表不同意见方的专家们将如何对三个干预案例进行评估。即附条件的现金转移支付，基础设施投资和反腐败。此场活动意在将争论停止在概念阶段，而一旦进行到某一特定评估设计的具体细节，就能达成更为一致的意见。令我高兴的是四位发言人都同意把他们的意见写出来以进行更广泛的宣传。

感谢休·沃丁顿和瑞兹瓦纳·西迪基协助准备这本文集。

3ie 执行总裁

霍华德·怀特

更加重视穷人意见：采用参与式方法进行影响评估

罗伯特·钱伯斯

Sussex 大学发展研究所¹

摘要

一项评估的起点应是问及为什么开展此项评估，谁将受益以及这项评估本身将产生何种影响及如何产生影响。参与式方式与方法适合多元的、渐进式和可重复的范例。它们包括利益相关者分析、单个讲故事、参与式社会绘图、因果关联、趋势与变化图解、打分以及对项目优缺点的头脑风暴法。精心设计并得到支持的参与式方法是严谨的，除了提供定性见解还能收集一些不可量化的信息，同时生成一些极易被忽视或被认为不能量化的相关方面的统计数据。它们通常考虑了那些受项目影响最大的群体意见，而采用一些更传统的方法方式就不可能做到，因而能使现实与穷人的经验更受重视。

引言

尽管受邀到此深感荣幸，但自感学识浅薄使我不愿参加这一组。参会中的大部分人都设计过评估（事实上，我们真的举手数了一下，其中一半以上的人都有过这种经历）。我只设计过一个评估：它发生在 1970 年，是场灾难。我不太愿意也因为我们生活在一个参与式方法包括监管与评估的探索创新时代，此外，相对于其他项目，如社区发展，农业或自然资源管理，这里选出的三个项目如运用参与式评估并不具有任何优势。作为另一位来自北方的人，我也像其他发言者一样感到很尴尬。一位非洲人在小一些的分会上说：“来自北方的任何事情都是真理”。

让我们认知“真理”中的大部分 – 创造与创新 – 和参与式方法一样产生于南方，并且正在南方、亚洲、拉丁美洲，特别是非洲蓬勃发展。

然而，由于萨利·加利巴在其开幕词中强调“赋权社区”，尔玛·马诺科特呼吁要“开放人民参与赋权之门”，我深受鼓舞。

¹ robertc@ids.ac.uk

会议讨论了三个案例，即附条件的现金转移支付，基础设施开发和反腐败委员会我依旧提倡在这些案例中应用那种多元的、渐进式的和可重复的方法。混合法将被采用。研究将从询问评估的政治经济意义开始：谁将获益？谁将受损？以及这些事如何发生的？特别是它为何预期这些发现会有意义。这也许亟需开一个有出资方职员参与的头脑风暴研讨会。如果他们不愿意或没时间参加研讨或就这些问题发表意见，我希望我会有胆量和资源拒绝这一任务。我将通过协商使备忘录包括其他步骤。其中之一是做利益相关者分析，与这一过程中的相关者协商。

我的讨论主要集中于参与式方法，该方法可以取代其他方法或作为其他方法互为补充。为此，首先要寻找好的创新支持者。为了避免评估计划过早结束，我要争取时间（时间可能长于投资者的预期）开研讨会，不断进行实地考察以完善和检测参与式方法，借用伊琳·吉特所著的一本书的书名，要《追寻惊喜》。总之，通过整个过程，将回答关于影响评估本身的影响路径和成本-效益以及附加值方面结果会怎样的问题。

案例 1：附条件的现金转移支付（CCT）

将尝试采用下列参与式措施和方法：

- 在有代表性的社区进行参与式的人口地图绘制。精心选择社区非常重要。绘图将包括每个社区的全部人口以避免抽样带来的问题。此法有助于识别项目开始前后社区内的学龄女童是否在校。之后将此法生成的统计数据与官方纪录进行核对，官方数据可能还包括女童的学业表现。
- 跟踪调查单个案例，包括表现异常的人和有意见的人。
- 与女孩子、教师、家长、管理者等进行焦点小组访谈和半结构式的访谈。
- 在项目实施前后，分别邀请女孩子和其他人设计和表演自己的生活剧。
- 绘制因果图表。
- 收集各种故事。
- 就项目的优缺点和如何改善进行头脑风暴。

辅助人员和研究人员可以在社区呆几天，并一起对笔记。预计会有新的问题，可将其纳入工作日程。

案例 2：南亚某国交通项目的后评价

我将采用下列参与式措施和方法：

- 在重要的利益相关者和受影响的人群中进行焦点小组访谈和半结构式访谈；这些人包括来自正式或非正式部门的小业主、交通工程承包商、司机、港口工会人员和大老板等。
- 为了评估乡村支线道路的影响，首先要考虑项目实施的不同环境，然后选择有代表性的社区初步非正式地识别受益者和受损者；然后与受影响的人群进行半结构式访谈和焦点小组访谈进一步识别获益者和受损者，并绘制项目实施的时间线、趋势与变化图、前后矩阵评分值、变化的因果关系图，以及各项联系的评分。
- 量化项目的影响，并在合适的时候，仔细研究图表已得到更多启发。
- 量化项目的影响与参与式方法只能生成定性数据的看法相左。然而，最近十几年杂色数字²已经越来越多，这正是参与式方法产生的定量数据（见钱伯斯，2007）。

案例 3：评估对非洲某国反腐委员会的捐赠影响

在例如腐败这样的敏感领域，评估本身将不同寻常地成为一项干预并被当作一次机遇。自始至终，它被认为是能否增加收益并最小化损失。初始步骤是识别腐败的形式和级别（高级，低级），并在捐赠支持和可能发生的变化之间提出因果假说。那些支持者、人权非政府组织、以及记者都将是主要的知情人。

与那些持续开展此类研究的人以及反腐委员会的成员进行下述活动：

- 通过焦点小组进行反思讨论（但参考下述单独会见）
- 对低级别腐败，寻找各种证据来源和见解，发现是否存在可信的关联：
- 市民纪录卡或可得到的类似纪录
- 对愤怒人群进行焦点小组访谈
- 在茶馆等地方进行因果关系方面的讨论。

对高级别腐败，主要部分将是：

- 进行非正式一对一的私下讨论，但不做任何记录。为了做到这点，评估人可以以一种无恶意但不择手段的方式（方便进行访谈但同时可操控）进行，因此通常要给评估人

² 原文为 parti-number,译者注。

提供一些活动费用。

结束语

有人问及如何量化那些不可量化的因素。参与式方法具有很多未被认知的生成数字的能力，这些数字与其它任何统计数字没有差别。通过判断、估算和价值表达，人们可以量化那些定性因子。这些方法的潜力早就该被认识到。就像道德问题总是存在一样。得到良好辅助的参与式方法可以实现双赢 - 赋权于人同时给决策者提供可信且负责任的意见。

在此次会议上，方法的力量和政治经济学意义已经被提升为专门的议题。我听到有人担心，干预措施因影响评估采用的方法在某些方面过于严谨而不得不有所选择。这并不是我能判定的事情。但若这种方法导致人们偏好实施那些简单的标准的能快速投入的项目，而不是那些比较复杂、长期、多元、参与式和赋权的项目，那么穷人的机会成本会很高。

最后，令人吃惊的是在此次会上穷人的能力与优先地位几乎未被提及。影响评估的目的是学习和改变以使穷人生活变好。为达此目的我们需要多种方法和多元化。很多方法和工具可以也应该用于影响评估。

无论他们是什么，他们应时常牢记是那些生活于贫困之中的人、处于弱势之人、被边缘化的人才是他们生活和生计，以及他们如何被影响，的最公正的和最权威的评判者。我们现在知道，如同我们过去 20 多年不知道一样，他们具有比我们想象的大得多的分析能力。我们知道“他们可以做到”。为了有利于他们的自身赋权分析，我们现在有了各种参与式方法。我们需要更多更好地使用它们。值得在此重申：问他们！

参考文献

罗伯特·钱伯斯 (200?) “谁重要？参与和数字静悄悄的革命” IDS 工作文件 296 页。

法尔默：发展研究所

关于发展评估随机实验的思考：开罗评估现场的发言

迪安·卡兰

耶鲁大学/创新扶贫行动/Jameel 贫困行动试验室

摘要

大会要求我们讨论评估三个假设干预的具体方法。本文利用此次论坛谈一下常见的关于随机控制实验的三个误解。首先，没有人认为随机实验适合于所有情景和所有问题。大家都同意应当把问正确的问题摆在第一位。第二，关于度量什么以及如何度量（即是采用定性的或参与式方法，还是采用定量的调查或官方数据）应当与是否采用随机试验无关。第三，除了简单的和静态的干预外，随机实验还可以用于评估复杂的动态过程。评估人应着眼于回答有关未来决策的最重要的问题，并且尽可能真实可信。一旦随机实验可行，研究就会更加可信；并且由于随机实验关注对干预为何有效或无效的潜在理论并对之进行检验，因此，通过随机实验得出的结论可以尽可能地应用到其他环境。

1. 引言

我们为什么做评估？有三个原因：要知道有限的资源都用在哪儿了；了解如何改善项目；以及要激励那些有钱的机构或个人出钱或投资。

我打算从实用哲学家彼得·辛格的一个思想实验开始。如果要弄坏你的衣服或使你错过一次约会而让你付出 100 美元的成本，你会去救一个掉到湖里的落水儿童吗？大部分人回答“会”。但是如果让你现在就拿 100 美元捐给一个贫穷国家的非政府组织去救助一个儿童吗？很多人说“不”，因为没有人知道他们的 100 美元是用于去救助一个孩子了还是被浪费掉了。评估可以反驳这一借口。

相对其他项目评估方法，关于随机控制实验（RCT）的使用有过很多讨论。但是，大多数情

况下，实验方法和其他方法间的“楚河汉界”是人为的，并且这一分界掩盖了实验方法与定量分析方法之间的共同点。在本文中，我首先列举一些普遍存在的对随机控制实验的测量方法、随机控制实验归因和可行性的认识误区。然后，我将列举三个普通的发展项目案例，即附条件的现金转移支付；基础设施和反腐败，并以此为例介绍在何种环境下应当将随机控制实验作为或不作为影响评估策略的一部分随机控制实验。

2. 关于随机控制实验的认识误区

一个普遍的错误认识是，在影响评估时，评估者只能采用定性评估，或者只能采用随机控制实验，两者不可同时使用。这实际上是“归因”严谨的错误表述，将随机控制实验和定性分析方法对立起来。随机控制实验实际上，随机控制实验并非如此，一个好的随机控制实验随机控制实验就包含了对项目如何运作、初始设计、变化理论、受益人参与情况等进行全面的评价。

为明细关于评估方法的讨论，必须将收集数据和结果测量话(测量什么，如何测量，过程中包括谁)与如何建立结果和干预间的因果关系分离开来。随机控制实验通过对反事实的度量建立了因果关系：即如果没有这个项目或政策将会发生什么。正如在医学实验中的标准做法一样，他们通过随机分配试验人到处理组与对照组来取得结果。除了是否提供干预项目，如果样本足够大，处理组和控制组在可观测的和不可观测的特征上都是相似的。

2.1. 建立因果关系

由于存在选择偏误，或者换而言之因为项目的参与者与非参与者具有不同的特征，随机分配是有帮助的。反之，如果我们将其与那些能够参与项目却不参与人相比较，我们将终结于对两组可能完全不同的人进行比较。很容易看到这两组人在某些重要的但却难以度量的方面存在很大的差异。那些参加项目的人也许更有动力追求改变现状，或被赋权或拥有更好的教育背景。或者他们有更多的自由支配时间。研究人员经常试图控制这些差异，但仍然不可避免遗漏变量或变量难以度量等问题的存在，例如如何度量激励。这些差异意味着对干预的影响估计是有偏的，因为处理组和对照组出现的不同结果可能是来自那些未观测到的因素，而非由于干预所致。

2.2. 数据与度量

量化结果的方法对评估是有用的，因为它们允许研究人员评估项目影响的统计显著性。但是随机控制实验不对数据收集作任何方法约束。定量的和定性的数据均可用于随机控制实验框架，并常被综合运用于同一项评估。经济学、社会学和心理学或其它学科的方法均可用于随机控制实验，参与式方法也不例外，它能否倾听当地人意见，(如同 Chattopadhyay 和迪弗洛 2004 所述，若她们村委会的领导是女性，西邦戈尔的妇女极可能参与决策过程。)甚至如钱

伯斯提到的表现异常的人的意见（卡兰和辛曼，2009）。

指导随机控制实验的鼓吹者的一个普遍的错误认识是认为这一方法可以并应该用于任何项目。随机控制实验方法是一种重要的研究工具，通过建立了因果关系能严谨地评估项目影响，因此它有助于了解是否可在其他地方复制这一项目以及如何改善项目。然而，随机控制实验并不总是可行的。其适用之处部分取决于背景情况，也取决于所针对的问题。如同拉瓦雷（2009）所述，人们绝不应该首先考虑方法，然后才想到该问什么。评估人必须首先确定要回答什么问题，然后再选择最适合的工具来回答。一旦可行，随机控制实验便提供了最无偏的项目影响估计量。但可行绝不应该成为采用随机控制实验的理由。仅仅因为可行影响评估而被采用。如果没有其它更给简便的识别策略，随机控制实验无疑是建立可信研究的最实用的方法。

2.3. 随机控制实验方法中的创新

尽管我们强调随机控制实验并非到处可用，但是实际上在很多似乎不可行之时只要稍加创新即变为可行之方法。例如，干预可以经常利用实施的局限之处并在社区或其它地理区域层面上进行随机干预，而不是随机选择个体进入处理组和对照组。有几个小额信贷的评估使用了这样的方法。在其他案例中，在一个随机控制实验中，我们能看到将一个项目推广到不同地区时在推广强度上的差异（鼓励型随机控制实验设计）。随机控制实验的关键标准是样本规模，每组要有足够的样本以能够度量溢出效益和一般均衡效应。如果计划得而且效应未被过度汇总（如国家层次），那么仔细设计的随机控制实验可以评估干预的直接影响，也可以评估对其他人群的正面和负面的溢出效应。这些实际上是所读到的最令人激动的一些随机控制实验，因为它们不仅帮助我们了解了一个想法是否会作用于特定个体，而且它是如何在更大范围内产生直接的和间接的影响。

2.4. 静态与动态的实施方法

随机控制实验另一个常见认识误区是干预必须是同质的和静态的。确实，“突发的，复杂的”或“复合的”干预，例如罗杰斯在此次论坛讨论所提到的，用随机控制实验处理并不比其他方法更困难。认为随机控制实验不能处理复杂的和动态的过程是因为没有认识到随机控制实验真正带给我们的东西。随机控制实验实际上帮助获得一个客观对照组，以此可以比较发生的变化。当然干预本身可以是静态和简单的，也可以是复杂和动态的。若是后者，那么评估当然可以这样描述：某人正在评估一个过程，伴随某些资源的一个机遇，以某种特定路径展开的一个动态的不稳定的干预等。这里的关键在于：评估的是这个过程而不是项目实施中的单个活动。如果项目是去工作，那么需要被复制的是投入资源并帮助他们利用这些资源的过程。这就像很多社区发展干预活动一样，其中为社区提供培训和量身定制的技术支持等资源，并通过一些辅助实践帮助社区发展和繁荣。

我们正在使用随机控制实验的方法做类似的评估，即同时运用定性和定量工具来完成对加纳

反饥饿项目和塞拉利昂社区驱动的发展项目的评估。当然很重要的是要理解这里正在评估的是一个合作过程而不是清晰定义的干预活动。事先不可能知道特定的行动者将选择何种投入，也不期望同一过程在其他地方会产生同样的选择。因此，此类评估得出的教训是关于人们通过这一过程可指望的变化 – 不是来自于行动者的特定选择和投资，而是当他们自己发展项目时，为其提供服务/融资的过程。。也就是说，如果项目官员或项目经理对评估构成干预的各项活动的影响感兴趣，可以设计随机控制实验方法从复杂的干预中得出各项活动的影响。这将需要使干预的各不同因子随机地进入处理组。可能的对照将是利用最基础的服务的一组与那些有一个或多个（累加的）干预元素相互作用的组进行对比。

与其它方法相比，随机控制实验有一个重要的优越之处，因为它解决了很多社会项目中与生俱来的选择偏误，并在有多个处理时评价不同活动的直接影响。例如，如果某人评估业务培训时发现利润有增加，特别是在那些保有良好业务记录的人中，这是否说明业务记录培训应予推进呢？也许业务记录是培训的一个中药组成部分，或者可能有良好素质的企业家天生就爱做业务记录。随机控制实验可以通过指定参与人接受带有或不带有特定业务记录培训模块的培训来理清这些问题。

关于随机控制实验的另一个常见错误认识是它们只度量了对所有人群的平均影响，忽视了干预队总体中的不同人群可能有不同的影响。事实上，如果样本足够大，并在抽样时包括了符合该项目要求的不同人群，随机控制实验能够帮助识别哪些群体受到的影响最大，哪些群体受到的影响不显著或者是负面影响。例如，我们利用随机控制实验评估一个秘鲁小企业主业务培训项目，一个吃惊的发现时那些自称对附加业务培训不感兴趣的人实际上从项目中获益更多（*卡兰和瓦尔迪维亚 2008*）。

3. 三个案例

拉瓦雷在此次论坛上的论文对人们在评估初始为明确评估目的和评估范围必须要问的几类问题和核心研究问题做了很好的概述。如他所述，随机控制实验是否可行取决于随机分配的基本单位。在此恰好可以通过这三个案例说明这一点。我将首先介绍如何对每个案例进行评估，然后具体谈一下即便在采用其他方法评估主要的干预影响时，如何在每个案例中通过随机控制实验回答与具体实施问题有关的辅助问题。这些想法不是要替代整个非实验性评估，但可为获得准确客观数据提供有用的方法，亦有助于解决今后实施中的问题。

首先看第一个案例，附条件的现金转移支付（CCTs），推荐的影响评估方法是随机控制实验，其中涉及定量的和定性的数据收集。这些实验已在若干国家进行。在那些政府只有有限资源来推广附条件的现金转移支付项目的地方，随机分配是一种公平的透明的利益分配方法。近期的研究表明，若设计适当，附条件的现金转移支付可以是实现重要的公共政策目标的有效方法。然而，如何最好地实施这些项目的问题远未解决。例如，实施问题包括支付频率如何

确定，是否需考虑增加储蓄服务，以及是否将教育支出纳入支付范围。³

关于第二个案例，基础设施。有若干设计方案技术上可行，但需要管理这些项目的政府官员做出不同的承诺。我将讨论港口复建、主干道和乡村居民道路建设。很不幸，评估人经常只在太晚的时候被要求作评估。且不论所使用的方法，在事先设定清晰的评估目标和考虑度量什么及如何度量这些结果是最好的。

港口复建与主干道建设评估重点涉及过程评估。第一步是建立一个日志框架，旨在了解等候时间的天数，海运天数，航运时间和陆运时间，航运的商品价值，航运的商品数量，以及进出的船舶、卡车和汽车的数量。是否使用计量经济学工具取决于对不同行业的不同效果，例如关税。这是一个简单的项目监测，但它对实施管理和研究结果都很重要。

在这些案例中，随机控制实验可以被用来回答项目变化理论的一些关键问题，但是不太可能涉及整个的干预。例如，发展中国家的港口复建中的关键问题可能是“较低的运输成本能够带来农村地区较大的产业增长吗？”在这个案例中，可以考虑设计一项随机控制实验，在一些地区随机地补贴运输成本以考量由此导致的经济活动的变化。

通过随机控制实验方法来回答村民道路项目中的主要政策问题技术上是可行的，但是也可能需要决策者方面的大量承诺。这一案例汲取的政策经验对其它国家也有许多益处，但也是那种我们认为政治上很难取得成功的事情。如果有足够的道路，并且地理条件和建设成本允许，则有潜力开展逐步跟进的道路建设的随机控制实验。设想利用 10 年时间来改善或建设乡村道路。

随机排序既公平又容易评估。如果某些道路由于经济和地理或政治原因更重要，可以在十年计划内考虑进道路优先战略。进取的决策者也许会认识到在这种情况下随机控制实验的优点在于它在确定项目实施顺序时能避免政治偏好。（通过一个精心设计的透明的程序决定各轮道路建设；在每轮建设的道路中，道路建设的顺序是随机抽取的，因此是公平的。）在这种情况下，随机控制实验是评估设计的一个方面。也可以考虑使用计量经济学方法，包括倍差方法，之前与之后比较法，以及修路地区与不修路地区的跨区域比较（例如，将距修建的道路 5 英里的城镇与距离未修建的道路 5 英里的城镇进行比较）。

最后一个案例，反腐败的度量通常并不是进行归因分析的评估，尽管一定会有一些可测量的结果，例如，逮捕率或政客被赶下台。但是在公务员如何和为什么诉诸非法计谋之处，有很多可学的东西。进而，评估方法的透明度再次尤为重要。例如，巴西市政审计要在电视上播放。奥肯(2007)在印度尼西亚的工作是另一个好例子。他混合使用定性方法（腐败推断和参与式村民会）与定量方法（实际的道路质量）收集数据，评估了各类反腐方法的相对有效性。

³注 1.关于如何设计选择环境以帮助人们有道德地进行选择方面的探讨，参见理查德·塞勒和凯斯·桑斯坦所著劝说。

4. 结论

随机控制实验最后一个优点是其内在的独立性，因此人们可以在事前建立清晰的统计检验，然后用数据验证某干预是否有效。评估的最终目的应该是帮助决定未来做什么。捐赠人需要知道把钱投向哪里，怀疑论者需要知道项目可以发挥作用，执行者需要知道如何更好地设计他们的项目。一些最令人激动的工作在随机控制实验中结合使用了定性方法，并用随机方法评估动态的和复杂的过程，例如社区发展项目。

在本文中，我重点关注了与影响评估方法有关的争论中的一些核心问题：即用什么来度量与拿什么来比较的问题。审视这些不同的问题，我们可以看到随机控制实验关注的是后者，并且很灵活能包括很多针对前者的参与式的，定性和定量的方法。我也试图消除那些在辩论极端处普遍存在的错误认识。即使随机控制实验的支持者也不提倡在任何地方和任何项目上使用随机控制实验。我冒险做一个猜测，评估预算中用于随机控制实验的不到 1%。我认为应该更多的采用随机控制实验，但也不是 100%（或 99%）。

参考文献

Chattopadhyay, 拉加本德拉和埃斯特·迪弗洛, 2004。妇女的领导和政策决定：在印度进行的全国性随机实验证据。经济矩阵 72 (5), 1409- 1443

迪恩·卡兰 和马丁·瓦尔迪维亚, 2008。教导企业家精神：小额信贷机构和客户业务培训影响。耶鲁大学经济发展研究中心工作文件

迪恩·卡兰 和乔纳森·辛曼, 2009。扩大信贷准入：使用随机提供的决策来评判影响。财政研究回顾

本杰明·奥肯, 2007。控制腐败：印尼田间试验证据。政治经济学期刊 115: 200- 249

马丁·拉瓦雷, 2009。三个典型干预案例之评估。发展有效性期刊, 即将出版
派翠西亚·罗杰斯., 2009。使影响评估设计与干预性质及评估目的相匹配, 发展有效性期刊, 即将出版

理查德·塞勒和凯斯·桑斯坦, 2008。推进：改善关于健康、财富和幸福的决定。新天堂, CT: 耶鲁大学出版社

三个典型干预案例之评估

马丁·拉瓦雷*

世界银行发展研究组

摘要

如同本次会议其他成员一样，我被要求讨论三个典型案例的评估设计：即附条件的现金转移支付、交通部门项目和反腐败委员会案例。本文记录了我的观点，并对一些做了详细阐述，包括参考文献。我将从对有关任何评估在开始时要思考的问题的一般建议开始本文，然后我尝试结合这三个典型案例说明这些观点。

1. 引言

本场会议的参加者被要求讨论如何评估以下三个干预：

1. 在一个中美洲国家的一个附条件的现金转移支付案例，此案例中，如果学龄女性在校上学并符合具体出勤和行为要求，其家庭即可每月获得资金补偿。
2. 在南亚某国的一个交通部门项目案例，此案例建设内容包括港口复建，主干道修建以及投资新建乡村居民道路。
3. 非洲某国一个反腐败委托案例。该项目包括协助设计指南，基础设施更新和考察访问。同样的项目正在六个国家实施。

这三个案例属于在很多方面极为不同的干预案例，但从评估背景来说，最大的差异在于他们参与随机分配的层级，也就是说，一些可观察的单位（居民户，公司，村庄，地区）接受项目而一些不接受。一些相关的差异是溢出效应的程度，即非随机分配的单位受到了项目的影响（正面的或负面的）。事实上，项目被分配的事实当然不意味着非分配单位不受影响。按传统评估方法，溢出效果可能是古典评估方法中严重偏误的起源。⁴

从两个极端看，附条件的现金转移支付项目是一个分配的项目，着眼于特定农户，可能仅有少量的溢出效果；而反腐败委员会项目具有广泛的经济影响，尽管它也可能在某些方面采用派定的方式，例如当其着眼于政府的特定部门和特定公司时。交通项目是介于两个极端

*这些是作者的观点，不代表世行或任何附属组织的观点。地址：mravallion@worldbank.org。作者感谢此次会的参与者所给予的评论意见，也感谢菲尔·基弗，诺伯特·施迪，多米尼克·范·德瓦尔和霍华德·怀特的有益讨论和评论。

⁴例如，陈等（2009）关于地方公共支出对贫困地区发展项目的响应的环境的讨论；奥多等人（2009）关于肯尼亚社会疟疾治疗社会推广项目的溢出效应。

之间，尽管溢出效果可能很大，但大部分干预元素时通过指定的。

尽管评估的基本原则(需要评估干预相对于反事实的影响)是一样的，但是随机分配的项目和非随机分配的项目要采用不同的估计方法。。指定项目便于做观察比较，而非随机分配的单位中选出一些样本常用于在某些特定的识别假定下推测反事实。⁵

在讨论对这些干预的评估问题以前，对每个案例，我将从问一些关于干预的背景情况开始。很自然，由于不了解背景很难做到非常具体，我将尽力争取。首先是关于这三个案例的一般评论。

2. 综合性问题

在任何评估的开始时，我将询问的关键问题如下：

- *为什么做此干预？* 理解具体干预的合理性对设计一个有用的评估非常重要，而且从广义上看它也对政策制定很有意义（从根本上说它是评估的最终目的）。探讨干预的合理性甚至能够引出一个不同的干预。

- *关于这种类型的干预我们目前知道多少？最重要的知识空白是什么？* 几乎总是会有过去的相关经验。在开始阶段，好好地回顾过去的证据会增进了解，也会很好的影响项目设计和实施以及评估所选择的关注点。

- *什么是相关的反事实？* 传统的反事实指没有项目时的情况，但这可能并不是政策制定者最感兴趣的反事实，政策制定者经常花费同样的资源做其他项目。一个特定项目可能比没有任何项目要好，但与其它可行的替代项目相比则很差。在形式上，如果一些替代项目是反事实，评估问题没有本质不同。原则上，我们可以为每个可能的替代项目做相关的“无作为反事实”的重复分析和比较。此法在评估发展项目时不常用，但在卫生和医疗实验中却很普遍（对照组受到现有的干预而处理组受到新的干预）。

- *什么是理想的结果？在哪个时间段出现？什么是潜在的不理想的结果？* 很显然，了解目标和如何将其转化为可量度的成果（定量的或定性的）十分重要。了解在什么时间段出现这些预期成果（证明的或负面的）同等重要。⁶

- *什么是需要估计的有关参数？* 传统评估只关注两个参数，即被给予机会且实际发生了的干预对干预对象的平均影响（意向-处理参数）和对实际接受干预的干预对象的平均影响。然而，政策制定者常常不关心这两个参数。其它有意义的问题包括：此项干预是按其期望的那样进行的吗？什么样的群体获益？什么样的群体受损？有多少参与者获益？把项目推广开来会发生什么？为扩大影响该怎样进行不同的设计？

- *什么是预期的转换机制？* 传统的影响评估在此方面可能是个“黑箱”，在关于项目如何发生或不发生影响方面几乎没告诉我们任何东西。为设计一个能够给黑箱带来光明的评估，我们需要理解项目的理论依据——预期项目能以此方式改善人民的生活。（我将稍后讨

⁵关于对随机分配项目的假定和方法的讨论 及相关参考文献参见拉瓦雷(2008)。

⁶ 参见金和贝曼(2009) 对此观点的有用讨论。

论这样的“基于理论的评估”)。政策制定者常常对于机制比较模糊,由评估者来探索可以帮助于揭示干预的理论依据。有时甚至会发现原来的依据是不可靠的。

● *在给定背景下,有哪些可行的评估方案?* 一些(技术的,经济的,政治的和道德的)因素将限制评估,理想上的,这些均应在一开始加以识别。一些理想评估方法可能在特定的背景下不可行。在可行的方案中,做出的选择应该依赖于对提出的所有问题的回答,而不是按评估人对方法的喜好去选择。虽然这一点显而易见,但事实上并未成为普遍做法。经常发生的是评估人带着他或她所喜好的工具箱到工作现场,然后选择可以用这些工具解决的问题(而不是反过来)。有时候,可接受的工具明显不够,评估者则开始用简单的偏爱的方法(例如随机选择,某些非实验性的计量经济学方法或其喜欢用的定性工具)寻找用这些方法能够解决的问题。原则上,这样做不能产生最有用的评估。

对这些问题的回答通常是设计一个评估的关键,但对每个特定项目和其背景来说要更具体。没有一个简单的正确的方法可通用于每一个有独立背景的给定干预。在这方面,让我来提供一些关于这对三个相当不同的干预案例的评估方面的思考。

3. 附条件的现金转移支付项目

此案例是这三个案例中最容易的一个。有两个原因:首先,附条件的现金转移支付项目是一个指定项目,很有理由推定它的溢出效果很小。第二(部分由于第一个原因),已有很多关于附条件的现金转移支付项目的评估研究,所以我们了解很多有关这类项目的问题以及如何最好地评估它们。然而,知识缺口仍然存在。

附条件的现金转移支付项目的基本立意是接受者家庭必须示范其足够的出勤率和卫生保健;转移支付只在这些条件(有时所谓的“共同责任”被证实)得到满足时发生。⁷ 早先有影响的范例是孟加拉的食品换教育项目,墨西哥的 PROGRESA 项目(现称为机会项目)和巴西的奖学金项目。影响评估的证据显示,这些附条件的现金转移支付项目通过对儿童上学和健康医疗方面的较高投资,在当前收入和未来收入方面,给贫困家庭带来了不可忽视的利益。关于对过去评估的新近回顾见菲什宾和施迪(2009)文章。

附条件的现金转移支付项目的立项依据并不像人们所想象的那样明显。一个长期存在的问题是为什么要强加这些条件而不是实行无条件的转移支付。信用市场失灵导致贫困家庭不能借到孩子上学的资金经常被用来作为附条件的现金转移支付项目的立项依据。但是也许有更好的政策来解决这样的问题,包括无条件转移支付(达斯等,2005)。关于家庭之间不平等和政治经济方面的问题经常比包括信用市场失灵的其它理由更为突出。

关于这些项目重要的知识缺口之一是如何在低收入国家很好地实施项目(大多数项目和评估来自中等收入国家,尤其是拉美国家)。根据近期在评估的附条件的现金转移支付项目,这一知识空缺正在被填补,很可能在今后几年内看到许多新成果。

⁷ 术语“附条件的现金转移支付”是不当称谓,因为几乎所有转移支付和社会保护项目给接受者强加了条件,但此术语只是用于这种特定的项目。

我们需要更多地了解附条件的现金转移支付项目的构成情况，以争取弄明白如何确定转移物的构成（多少，以现金还是以实物）和行为改变的激励（哪些行为受到鼓励）。有些证据表明 PROGRESA 项目将对入学率的补贴预算从小学转移到中学有助于取得更好的教育成就，因为将会有更多的孩子继续上中学（德·杨弗利和赛多利特，2006）。这是我早先提到观点的例证，评估者需要很好地了解项目如何准确地产生其影响，因此我们可以较好地建议如何完善项目设计。

尽管还有一些很好的对行为反应的评估研究，项目设计人关注更多的是有意向的行为人（即与项目共同责任一致的人）。我们需要更多地了解其它反应。包括劳动力供给和家长们储蓄决定。

过去的研究在诱发的行为改变福利的影响方面也留下了知识空缺。这些结果应该包括当前和未来的贫困情况。当前的贫困相对容易测度（尽管仍然不容易测量）。关注“中间结果”例如儿童劳动与学校出勤可能是有用的，但我们几乎不知道最终的结果。可以肯定，当上学的成本通过附条件的现金转移支付降低时，更多贫困家庭的孩子会上小学。但是，他们学到有用的东西了吗？或者影响他们的学习能力是否已经太迟了？对毕业后的收入的影响是什么？

在附条件的现金转移支付项目中与供给方有关的前后关联的因素愈发显得重要。一些过去评估显现的在儿童入学和营养状况方面令人失望的结果可能与供给方因素有关——学校与卫生所的质量差——一个地方与另一个地方不一样。若供给方的条件没有改善，附条件的现金转移支付项目诱发的行为改变也许不足以获得理想的福利结果（菲什宾和施迪，2009）。

重要的知识空缺还与附条件的现金转移支付项目的长期影响有关。这自然需要时间。但是需要现在给予数据方面的投资。可怕的是，其外部性及随之而来的融资与长期影响的持续研究方面面临重重困难。

另一知识差距在于对于需要的改变所能做出明确调整的灵活性。如果需要这些项目发挥安全有效的作用例如像在当前的全球金融危机时期，这是极为重要的。可以预想，给当前受益人发放临时支付款要比临时扩大金融危机中的受益人容易一些。但是需要对此问题以及附条件的现金转移支付项目如何能在实践中更灵活实施做更多研究。

在考虑评估方案方面，深入了解附条件的现金转移支付项目的逐步推进计划很重要。将此类项目一步到位地在全国实施并不总是可行（从预算和技术上考虑），即便项目已不是在试验的基础上开展。一些地理区域可以比其它地方先开展项目。理解如何选择目标地区是关键。在可观察到的、相同的无参与者地区寻找创建一个对照组。或者可以按照一个方案选择哪些地区先行，也可以考虑做不连续的设计，在其毗邻地区的清晰对照点做影响评估。随着项目的扩展，对照地区将加入进来，但在那以前，可以做出识别影响的方案。鉴此，了解对照地区是否知晓其将加入项目和希望何时加入可能很重要，因为这将产生干扰效果，由于期望加入项目对照单位由此受到了影响。

在很多方面，原始的 PROGRESA 评估是其它附条件的现金转移支付项目的典范。其主要设计特点是“随机产生的”组群在逐步推进的项目期间的行为与对照组相像。这个项目的寿命（未受政府更替影响）和其在发展社区中的影响清晰地显现，部分是由于大量的公众的努力，这已经被包含在了它的评估中。有资格加入项目的 1/3 的样本社区被随机选择为初始阶段不参加项目的对照组，在此期间，2/3 的社区接受了项目。公众可接触评估数据为若干宝贵的研究提供了便利。对 PROGRESA 项目的设计概要、执行情况及评估结果可参阅 Skouflas (2005) 文章。

PROGRESA 的设计也有改善的空间。如上所述，更加完整地计量结果变量将更理想。担心对项目的预期（特别是当像 PROGRESA 这样的项目在全国实施并经过评估，很自然会有很高的公共形象）可能污染控制组。评估设计本可以在使人在了解项目设计上做得更多，例如补贴应在小学还是中学阶段发放。供给方因素也在随后的研究中出现，对更好地使用学校和诊所的关联设施调查是有用的。

4. 交通部门项目

评估交通部门项目有若干重要的背景问题需要考虑。逻辑上，第一步是要理解干预的必要性，它与经济发展史、地理和政治经济因素有关。是特定的港口和交通联结选择时的疏忽导致修复重建吗？如果是，为什么？有关公共融资和溢出效应的问题很可能显得严重。该项目是如何供资的？可能存在替代力或捕蝇纸效应吗（尽管只是部分可替代，援助依然堆积在意向部门）？我们在评估一个正确的干预吗？

可以预期该项目有短期的和长期的影响。对劳工收入将发生即时的效应（项目上工作的工人数量以及其他工人由于劳力市场供给偏紧得到好处），在有的背景情况下，这点很重要，例如交通项目是危机反应的一部分。在建设期将产生各种成本，除了建设成本，还有像社会和环境影响的成本（包括移民成本）。

但是对建设完成后的主体影响必须做预期，评估工作计划必须在其时间范围内做好安排。这些影响可能会很宽泛且难以量化。交通成本的节约（包括即时成本、人们旅行的时间价值以及运费）无疑第一重要。对于大型项目，很可能还有更广的（一般均衡）对经济活动的效应，包括地理方面的（我稍后再说）。

评估需要考虑市场功能和含义。我们想知道按地理区域形成的商品交易市场运行得怎么样？通过价格变化提高了识别影响的可能性。（交通项目影响福利的主要方式之一是通过价格变化）。我们还应该了解在给定情形下原有的土地市场运转如何，这将为通过土地价格资产化来识别影响提供空间。（传统的范·杜能模型预测土地租金将随着距市中心的距离而下降，反映了交通成本；对此模型可如何用于评估交通项目的案例参见雅各比, 2000）。

我们还可能想知道干预怎样影响经济活动的地理分布，包括地方市场和制度的发展。乡村道路的改善能够吸引市场、机构和新的经济活动来到落后贫穷地区吗？或者刺激这些活动

在地域上的进一步集中？大型交通项目怎样改变了经济活动的地理布局？

在交通成本对地点选择有明显重要性的情况下，可以预期大的港口项目将改变经济活动景观，使一些地区和活动受益，且很可能导致其他方面衰败。新经济地理学预言，由交通基础设施投资带来的产品（农产品和非农产品）运输成本的降低将增加其他活动（非农活动）在城区的区域集中度以追求聚合经济效应（藤田等，2001 第7章）。世行支持的越南乡村道路项目的一个评估提供了项目在贫困地区刺激当地市场发展的证据——部分地区分散了经济活动；参见穆和范·德·瓦尔（2008）。活动的地域集中对当地农村的影响可能与区域影响完全不同，包括城市地区。

一个理应提出的问题是评估是否只应该关注结果的总体经济效率。有时有这样的意见：使用其他政策工具解决关于平等问题，包括与地域分布有关的问题，例如落后贫困地区，参见世行（2009）相关章节，发展政策的制定不应该关心活动空间布局是否平等，也不应干涉市场导向的布局。如果将此观点用于交通项目，我们将只能考虑它们的效率，用它们对平均收入的影响来衡量。

然而，我认为很多政策制定者未必能够接受这样的观点，并有很好的理由。他们知道他们没有以空间盲法⁸处理平等问题所需要的全套政策工具。确实，在贫穷国家，地理位置取代了更好的关于个体福利水平的信息，成为为再次分配目的识别穷人所使用的最广泛的方法之一。有时在发展中国家，人们可以预计交通项目与政策有关的评估将需要考虑到平等方面，也需要考虑“非收入”方面（包括例如：由于交通改善使母亲得以上诊所而对儿童健康产生影响）。

交通项目的评估设计需要把那些可以被作为指定项目（例如地方乡村道路）的成分与那些可能具有更广的空间影响的成分（例如主干道）加以区分，因此，可以建议采用部门或广义经济的方法。

对指定和非指定成分的评估要求不同的设计。对于指定成分，例如当地乡村道路改善，我们要识别相关的“服务范围”并安全地包容可能的影响（范·德·瓦尔，2009）。我们也需要理解地理定位是如何确定的，并找出可观察的未接受新建或修复道路支出的相同地区。纵向的观察几乎可以肯定是需要的（可以使用“双差分匹配”估计量，参见拉瓦雷，2008，以进一步讨论与示例）。评估人也必须估计选择偏误的可能来源。如果选择很大程度上基于可观测对象做出，应该有可能收集到正确的数据以适当地修正这种偏倚。如果选择是基于不可观测对象做出，就需要考虑可选择的统计方法，例如工具变量估计（显而易见，随机方法几乎不可行）。

如果是大交通项目，那么其服务范围也将是大的。那么，也许难以找到合适的未受到项

⁸ 长期以来经济学已认识到，为了实现总体的社会福利目标，需要很强的假定条件才能严格分离平等和效率工具。但普遍认为，这些条件（特别是非扭曲的一揽子转移支付的可行性）在现实中并不存在。

目影响的对照地区。我们将需要转向很不一样的工具。⁹从一个极端上，跨国增长回归分析已被用于尝试识别基础设施的影响，包括交通项目；参见卡尔德隆 和塞文（2008）实例。关联到居民户和公司的次国别区域数据能够提供清晰的镜面；参见贾兰和拉瓦雷（2002）实例。他们发现，在中国，乡村道路对微观发展进程发生了积极的影响。空间可计算的一般均衡模型，已被新经济地理证实，可以在未来很好地发挥作用（藤田，2001）。

这些模型从概念上非常适合评估大型交通项目影响，尽管它们是复合模型对数据和相应的合作要求很高。欧盟开发了评估交通改善的空间可计算一般均衡模型，其在大型交通项目（荷兰铁路项目）评估上的应用参见纳普和奥斯特汉文（2000）。

必须承认，过去新经济地理模型在如何模拟交通成本方面做出了无可争议的假定，这些假定可以为总结交通改善之影响方面的教训发挥很好的作用。¹⁰可以预期未来将在更加现实的假定基础上建立实用模型。

5. 反腐败委员会案例

这是一个具有非同寻常的权力去调查和检举腐败的独立主体，特别是向国家首脑或议会首脑报告。¹¹在评估反腐败委员会案时需要问两个背景方面的问题。首先，需要估计此项干预是否需要涉及腐败的已知原因。腐败的根本原因在于强加的机制还是政治活动家监督公务员的激励有问题？我们应该谈论受托机构的改革问题特别是公共部门管理问题吗？第二，评估者需要很清楚这次干预的动机。仅仅是捐赠人的意愿要惩罚腐败官员，还是政治领导人热衷于打压反对派，或者此干预源自于对问题的深入分析？

估算腐败影响的评估可以用许多曾用于反腐败案例的工具来设计。奥肯(2007)对印尼道路项目的腐败研究中使用了随机选择的设计，被认为是一个创新的案例。奥肯发现增加政府审计减少了“缺失支出”的程度，其中，“缺失支出”为测量到的正式记录的项目成本与工程师独立估算之间的差异。（相比之下，奥肯（2007）发现参与式方法对监督几乎没有任何影响。）

尽管这样的评估研究可以提供有用的参考，但是它们并没有识别反腐委员会这样的案例的影响。。奥肯的研究告诉我们在印尼通过政府审计的监管可以有助于反腐败；然而，通过至少一个评估看到，印尼自己的成立反腐败委员会这样的努力在促进监管和惩罚方面似乎没有什么作为，除了带不了一些广为人知的逮捕案，这些被捕者都是与政治斗争相关且臭名昭著个人腐败案（夏洛克, 2002）。

典型的反腐败委员会不像前面讨论过的附条件的现金转移支付或交通项目的某些成分那样，它不是一个指定项目。通过给一些地方政府或部门授予反腐败权力，该项目也许可以

⁹ 关于方法学方案的概要可从奥斯特汉文和纳普（2003）找到。

¹⁰ 在此“冰山交通成本功能”意指所支付的货物价格随着航运距离分单元地增长，有证据表明它是不一致的。进一步探讨参见麦卡恩（2005）。

¹¹ 对反腐败委员会的历史与记载的有用概述可在罗斯-阿克曼（1999）第9章找到。

逐步扩展到其他地区或部门。如果逐步扩展是可行的，且预计影响不是很严重，将接受了反腐败审查的政府机构与一些未开始反腐败活动的地区的观察特点相似的政府机构之间的客观的腐败测试结果（例如奥肯 2007 所使用的）进行对比可能有助于评估反腐败的影响。然而，即使不考虑可行性（政府一开始也许非常愿意给反腐败委员会以自由空间），评估者也要关注期望的效果和选择偏误。（对背景情况不够了解使选择偏误的迹象含糊不清；总统可能会从他认为问题最大的地区或部门开始反腐行动；总统使用反腐败委员会作为政治工具时将会做出不同选择，甚至可能避开了腐败中心）。

比较有前景的（可能是辅助性的）一个方法是考虑反腐败委员会产生影响的可能渠道，建立一个或多或少清晰的理论模型说明为什么存在腐败。然后评估工作集中于预计因果链中的关键环节，类似于魏斯(2001)，罗杰斯(2009)和怀特 (2009)所探讨的“基于理论的评估”类型。例如，克里特戈德(1998)提出腐败是三个因素的结果：对某些资源的垄断；政府官员允许接触此类资源的自行决策权；以及追究那些官员责任的机制缺失或失灵。然后，通过改变官员在特定体制环境中面临的激励，可以开始评估反腐败委员会怎样处理每一个因素。这将需要评估各个官员面临何种激励和随着反腐败委员会工作的开展它如何发生变化。这绝不是一个容易干的事情，因为这些官员也许根本没有向评估人揭示真相的动力。但评估至少能帮助理解为什么我们重视腐败（远超出对官员不朽的呼吁），并有机会确定反腐败委员会是否正在对腐败问题产生真正的影响。

反腐败委员会喜欢引用被起诉官员数量作为其成功的度量。给定反腐败委员会可能有选择地以政治对手为目标，因此，这一度量结果可能具有欺骗性。对此，评估可以直接关注对反腐败委员会的政治支持或反对。评估者可以收集反腐败委员会瞄准的具有政治背景的机构的数据，包括雇员与政客联系的调查，离去雇员的离开原因。然后，要通过追踪调查（基线调查之后）询问居民，公司和政府官员，与其它机构例如中央银行，财政部和公诉机关相比较，反腐败委员会的政治化的程度如何。

6. 结论

好的评估艺术在于在一开始就以弥补知识缺陷为目的询问正确的问题，并以量身定做方式使用数据和分析来回答特定情境中的这些问题。不能指望一篇类似于此的文章能概括某项特定评估的所有最重要的问题，或者给出评估方法的全部方案。两者很自然地取决于特定情境。然而，本文希望为此提供一些有用的思考启发。

参考文献

卡尔德隆，塞萨尔和路易斯·塞文, 2008, 在非洲次撒哈拉的基础设施建设与经济发展, 政策研究文件 4712, 世界银行, 华盛顿特区

陈少华, 穆仁和马丁·拉瓦雷, 2009, 对贫困地区援助的持续影响存在吗? 来自中国农村的证据, 公共经济学期刊 93: 512-528

达斯, 伊松努, 葵-道安 杜和伯克·厄兹莱尔, 2005, 附条件资金转移计划的福利分析, 世行研究观察员, 20 (1): 57-80

德·杨弗利, 阿兰和伊丽莎白·赛多利特, 2006, 使附条件资金转移项目更有效: 设计效应最大化的条件, 世行经济评论 20 (1): 1-29

费斯贝恩, 阿里尔和诺伯特·施迪, 2009, 抗击目前与未来贫困的附条件资金转移, 世行政策研究报告, 世行, 2009

藤田, 昌久, 保罗·克鲁格曼 和安东尼·维纳布尔斯, 2001, 空间经济, 剑桥, Mass.: MIT 出版社

雅各比, 哈南, 2000, 市场进入与乡村道路的收益, 经济期刊 110: 713-737

贾兰, Jyotsna 和马丁·拉瓦雷, 2002, 地理贫困陷阱? 一个中国农村消费增长的微观模型, 实用数量经济学期刊 17 (4): 329-346

金, Elizabeth M. 和杰雷·河·贝曼, 2009, 社会项目评估的披露时机与期限, 世行研究观察员 24 (1): 55-82

Klitgaard, 罗伯特, 1988, 控制腐败 加利福尼亚大学出版社, 伯克利 加利福尼亚

纳普, 蒂吉斯和简·奥斯特汉文, 2000, 新基础设施的福利效应: 评估荷兰新铁路连接的经济地理方法, 伊拉斯姆大学诱因高稿, 鹿特丹

麦卡恩, 菲利普, 2005, 交通成本与新经济地理, 经济地理期刊 5: 305-318

穆, 仁 和多米尼克范德瓦尔, 2008, 乡村道路与当地市场发展, 越南政策研究工作文件, 世界银行

奥多, 雅各布, 安妮卡·马乌和埃文·马森格, 2009, 评估小额捐赠抗疟药物对肯尼亚疟疾死亡率与发病率分布的影响, 发展有效性期刊, 本期

本杰明·奥肯, 2007, 监控腐败: 印度尼西亚实地实验证据。政治经济期刊 115 (2): 200-249

奥斯特汉文, 简和蒂吉斯·纳普, 2003, 交通基础设施投资的空间经济影响, 交通项目, 规划与政策 (A.D. 皮尔曼, Peter J. Mackie, John Nellthorp 编辑) 社会科学跨学科比较研究中心, 阿什盖特出版公司

马丁·拉瓦雷, 2008, 评估反腐败项目。在鲍尔·舒尔茨 和约翰·斯特劳斯编辑的发展经济学手册第 4 卷, 阿姆斯特丹, 荷兰北部

_____, 2009, 在发展实践中评估, 世界银行研究观察员 24 (1): 29-54

罗杰斯, 帕特里夏, 2009, 依据干预性质与评估目的进行影响评估设计, 发展有效性期

刊，即将出版

玫瑰阿克曼, 苏珊, 1999, 腐败与政府: 原因, 结果和改革, 剑桥大学出版社, 剑桥

夏洛克, 斯蒂芬, 2002, 与印尼的腐败斗争? 检察官和资产审查委员会, 印尼经济研究公告, 38 (3): 367-83

Skoufias, 埃马纽埃尔, 2005, PROGRESA 和它对墨西哥农村居民户的影响, 研究报告 139, 国际粮食研究所, 华盛顿特区

范德瓦尔, 多米尼克, 2009, 乡村道路项目的影响评估, 发展有效性期刊, 1 (1): 15-36

魏斯, 卡罗尔, 2001, 基于原理的评估: 减贫项目的变化原理。在 Osvaldo Feinstein 和 Robert Piccioto 编辑的评估与减贫, New Brunswick, NJ: 翻译出版物

霍华德, 2009, 基于原理的影响评估: 原则与实践, 发展有效性期刊, 本期

世界银行, 2009, 重塑经济地理, 世界银行, 华盛顿特区

依据干预性质与评估目的进行影响评估设计

派翠西亚·罗格斯

澳大利亚墨尔本皇家技术学院公共部门评估教授¹²

摘要

合理的影响评估设计需要结合实际——设计应该依据特定案例的需求、约束与机会而定。这种设计需要反映干预的性质和影响评估的目的。特别是，影响评估需要包含简单的、复杂的和复合的不同层面。简单层面可以进行简洁地描述并采用标准化处理，复杂层面需要考虑前因后果；复合层面需要结合动态和适应性方法。每个案例都需要不同的评估设计，包括随机控制实验，非连续回归，非结构化社区访谈，参与式故事报告以及开发式评估。

影响评估的情景回应方法

本场会议的目的是示范影响评估设计中各种方法的应用。讲座侧重于度量和因果分析，但需要强调的是进行影响评估还必须完成其它方面的任务。（罗杰斯, 2008a）。完整的影响评估设计需罗列出开展评估中涉及到的全部工作，包含重要影响因素的识别界定，系统而严谨的数据收集和对这些影响的证据分析，合理地推断干预对影响的贡献，以及对评估的有效管理，这其中包括公开所使用的分析方法以及在某些适用场合采用正式的荟萃评估”（NONIE 2 分组，2008）。事实上每一项任务均有一系列可选择的方案，影响评估设计需要选择最适合的方法或者综合几种方法来完成某项任务。

影响评估中一些方法强调运用特殊设计的数据收集和分析方法，或是监管方式。对我采用的影响评估方法设计最为恰当的描述应为“情景回应”（巴顿, 2008a）。

对发展和评估的其它方面的认知在与日俱增。不同的评估情景将通过适当地使用一系列的方法和技术来处理。例如，NONIE1 在其 2008 年关于影响评估的陈述中这样说：“NONIE 赞成以一个兼收并蓄并且开放的思路来选择进行影响评估最好方法——从不同学科众多的技术中来选取。（NONIE, 2008）。

在世行独立评估组会议“衡量发展有效性：进步与约束”上，迈克尔·奎·巴顿应邀作了报告“衡量发展有效性的学术进展”，他让我们重新认识了影响评估“黄金标准”。“任何特

¹²联系地址：Patricia.Rogers@rmit.edu.au，拉澳大利亚墨尔本特罗布大街 124 号维也纳国际中心 3000。

定的方法都不能成为方法学的黄金标准，判断黄金标准的关键是方法是否恰当。”（巴顿，2008b）。

情景回应是指在设计时应该同时考虑到特定环境下的需要、约束条件和机会。在完成影响评估设计之前需要回答两个关键问题：“干预的本质是什么？”和“为什么要进行影响评估？”。

干预的本质是什么？

干预的类型和规模是决定影响评估设计是否合理的关键。它是一个可能被复制和推广的小示范项目吗？它是一个可能会以某种形式继续进行的在建项目吗？它是单一的干预还是集合了多种不同动议的干预？这项干预是严格遵照统一标准执行呢，还是因地制宜、根据不同地点、条件、需求和机会进行调整呢？这些问题对于影响评估的类型选择，对于资源的可得性和研究设计方案均具有意义。

我们所探求的影响的本质是什么？影响是干预的直接结果（发散型），还是间接造成的（波浪形）？是在项目执行和评估期间内显现的短期影响（比如儿童在校表现）呢，还是若干年后才能显现的长期影响（例如教职员工的培训，或土地盐碱化的恢复）？影响是暂时性、过渡性的吗？是不可逆转的（例如学会如何阅读或骑自行车）还是脆弱的、容易反弹的（例如提供充足的营养或保障女童上学）？影响是不顾条件采用“银弹”干预的结果呢？还是“填鸭式”的干预结果，仅在有利条件下才能取得效果？或许还有其它的干预方法？

这些不同的特点可以归纳为三种类型：简单的，复杂的或复合的（斯塔塞 1992；格洛布曼，2001；格洛布曼 和齐默曼，2002，库尔茨和斯诺登，2003）。实际在规划和分析评估中已经显现了分类是很有用的（盖特，2008，巴顿，2008a；罗杰斯，2008b），尤其在划分干预的不同方面而非干预整体时，分类更显得更有用。

在分类学中，“复合”这个概念具有特定、重要的意义，在实际使用中往往有不同的含义。有时它可作为“复杂”的同义词，有时用来特指那些困难的事，有时还被作为不适当计划的借口。在这种分类中，“复合”还意味着在遇到突发紧急的需求和机会的时候，能恰当地做出反应并采取紧急干预措施。简单的干预层面可以明确地界定并标准化，例如，某个特定的产品、技术或过程。干预的复杂层面具有多元成分，其本身又是更大的多元成分干预的组成部分，或是作为更大的因果包的组成部分开展不同工作，例如处于特殊的实施环境，为特殊类型的参与者而作，或者结合其它干预进行。

这些不同的干预层面对于如何实施干预、如何理解它们以及我们如何使用都具有重要含义。参见表 1 述。

表 1 干预的简单、复杂和复合层面

| 层面 | 含义 | | |
|-----------------|------------------|--------------------|---------------------|
| | 干预的实施 | 因果程序 | 报告和影响评价结果的使用 |
| 简单 | 单一组织 | 能产生影响的必要的因果链 | 简单信息— 做什么 |
| 复杂（多元成分） | 通过契约进行明确分工的多个组织 | 能产生影响的多重因果链： | 偶然信息— |
| | | 系列干预 | 实施条件 |
| | | 同期多重干预 | 做什么 |
| | | 不同层次的干预 | 目标人群 |
| 复合（动态的、突 然的） | 多个发展中的 合作伙伴关系 | 或者不同背景下不同机制的干 预 | |
| | | 回归关系，紧急反馈 | 动态的、突发的信息—正在 做什么 |
| | | 突发事件—总体效应 | |
| | | 超过分项加总 | |

为什么要做影响评估？

在设计影响评估方案时需要考虑影响评估的目的。谁是此实施评估的主体？在所纳入的影响方面，什么是被认为可信的证据，采用什么样的因果分析方法与方式？由哪些人来决定评估参数？

评估中使用谁的价值观？无论是正面还是负面，什么被认为是重要影响？理想的成本效益分配是什么样的？重点应放在平均效果还是最弱势群体上？评估的目标效用是什么？是为了净效益的可信估算？还是为了因推广试点而要了解其工作机制的足够信息？还是为调整项目或政策提供依据？

最后，在设计影响评估之前，后勤方面应做好安排。何时需要提交报告？已经掌握的此项干预和同类干预的证据有哪些？有什么额外的资源可利用于影响评估？

如果要采用情景回应式方法，只有在我们回答了以上这些问题才能够着手解决“应该如何做”的问题。显然，在一个真实评估中，所有这些问题将被事先解决，或者在设计的过程

中解决。在本次设计现场研讨中，我们试图提供更多特定案例以填补案例描述方面的空白。

1. 附加条件的现金转移支付案例

相关案例描述如下：

在中美洲一个国家的附条件的现金转移支付案例中，如果居民户中学龄女童坚持上学并满足规定的出勤和行为要求，该户即可得到每月的转移支付。

附加条件的现金转移支付项目具备简单干预类型的基本特征：离散的，标准化的，在不同地点均采用统一标准实施干预。然而，对附加条件的现金转移支付项目的评估发现：其结果有很多异质性（例如，苏亚雷斯等，2009）。附加条件的现金转移支付项目或许需要辅助其它的干预手段才能取得预期的效果。例如，附加条件的现金转移支付最初运用到提高就学率上很成功，但在学生的学习能力，毕业和就业等方面的长远影响则需要依赖于到位的有效教育。所以，考虑干预的复杂方面很重要，包括衡量和分析其它重要的影响因素（例如实施环境的特性和参加者的特点）。

在这个案例中，假定这个影响评估的目的是决定是否推广试点项目，那么随机控制试验就比较适用于做因果分析设计，尤其是在需要结合其它影响因素的数据系统收集时（例如为取得既定影响目标所需要的学校教育质量和为识别贡献而做的拆分分析）。开发和检验包括额外因子、差别效果和中间结果的程序理论将改善分析的质量。如果说附加条件的现金转移支付有助于提高学校成绩，但必须结合有效的学校教育，他们不是单独快速适用的。换言之，因为合格参加附条件的现金转移支付项目的条件设在一定收入水平，断点回归方法可以比随机控制试验以较少花费提供因果归因的有力证据，尽管它也需要实施环境拆分。

除了衡量既定的影响目标外，考虑其它影响在评估中也很重要，可以参照发展援助委员会对影响的定义，包涵了“正面的和负面的，由发展干预产生的主要和次要的长期效应，直接的和间接的，既定的和非目标的”等各类影响。

一些潜在的非目标的影响可以被预期并包括在目标数据收集中，使用“负面程序理论”（魏斯, 1977）。无法预期的影响，特别是负面影响，最可能通过参与式方法观察到（钱伯斯, 2009）。之后，案例研究需要通过重复访谈，观察和文献查找等方式来揭示非目标性的影响（正面的或负面的），并且需要通过大规模的调查或官方统计数据来估算这些非目标影响的频率。

2. 交通基础设施项目

此案例描述如下：

对南亚某国一个交通项目的事后评价，包括港口复建，主干道复建和乡村居民

道路的新投资。

对此案例，我们设想其影响评估的目的是了解该项投资的总体影响，主要用于向供资者也包括向公众报告，但在项目开始时并未作评估设计。这一缺陷加上干预的性质（多层面，多元化并影响整个地区）给因果分析带来相当的挑战，因为传统的实验性和类似实证式方法似乎不可能有效地用于此案例（拉瓦雷，2009）。

此案例影响评估的方法需要收集大量多元成分的证据并在没有对照组或控制组的情况下处理归因问题。对此案例，比较合适的方法是参与式故事报告会。本会议上的演示（达特，2009）表明该方法能够发挥这些作用。由于一些人认为，“参与式”在数据收集和分析方面不够严谨，所以在运用该方法时强调实证基础是很重要的。

参与式故事报告会是多线性和多等级证据的开发和系统化，在因果分析中这一方法用于高风险的情况，此时不可能再构建对照组或控制组，但有相当多的可用的高质量证据并且需要做系统的因果分析。

“线性证据是证据的一种类型，例如一种生态系统属性（鱼类的丰度，巨型无脊椎物种的丰度，巨型生物），它在受到应激或者干预时被调查。证据等级是证据的强度值，用于案例分析中推断某个特定的人类活动将会给生态或地理生态带来的变化”（考汀汉姆等，2005）。

虽然在大部分影响评估指南中没有提及这一方法，但它的使用却可追溯到1960年关于吸烟与肺癌的关系证据分析战略。面对大量的科学数据，随机控制试验却没有证据，尽管如此，能够做出吸烟对健康有影响的结论十分重要。近些年来，多线性等级数据已被用于人类和生态风险评估以及自然资源管理（例见凯等，2002，博伊斯，2006）。

通过分析观测到的关联数据，并识别和排除其它可能的解释，多线性等级证据可系统地判断干预的强度和结果。关于吸烟与肺癌之间的关联，希尔（1965）考虑了其强度、连续性、特异性、临时性、与其它所接受证据的一致性、合理性以及与类似干预的相似性。为了调查管理干预引起的生态反应，诺里斯等人（2005）考虑了生物合理性、生物反应、关联的剂量与连续性。考虑到专家和科学证据经常跨学科的特性，此项调查由跨领域的有信誉的专家组来执行，要求他们评判证据的可信度并作因果分析。参与式故事报告会多线性等级证据方法扩展到包括主要数据收集和利益相关者会议，通过对现成的数据、另外收集的原始数据，专家组分析结果等进行汇总，最终完成意见一致的影响评估报告，它包含经得起审查的一批证据（例见克里尔·霍瑞珍等，2008）。

对交通基础设施案例使用参与式故事报告会方法可从开发结果链或程序理论开始，识别目标影响以及带来这些影响的一系列中间产物。第二，定位项目的可利用证据，中间结果和最终影响以及项目对这些方面的贡献，并评估证据的质量。证据评估要依据与信息类型相关的质量，而不是依据研究设计证据等级的评估。例如，与交通使用人的访谈将从所用抽样方

法、所提问题、所用减少偏差的过程以及数据的审计方面进行评估。第三，额外的证据将被收集以弥补部分数据缺失造成的影响。例如，如果可利用数据大部分为官方统计数据，例如到港货船的数量和类型，使用道路的交通工具数量和总的经济活动数量等等，那么收集交通项目对个人、居民户和村庄带来影响的小案例作为补充数据是很有用的。通过对这些附加信息的收集，也能了解很多重要但非设定目标的效应信息，无论这种影响是正面还是负面的。例如，当地道路的改善可以增加去中心医疗机构的便利，从而让孩子出生时的并发症降到最低。从这些少量的实证分析中得出的结论可以作为检验官方健康统计数据的参考。

接下来，来自多领域的专家小组将考察基于这些数据描述的影响及其归因分析（可以是单一的，也可结合其它因素）从而对数据进行评估。尽管参与式故事报告会方法没有规定专家检验数据必须与额外的数据收集相结合，但最好专家检验数据过程中能包含这些另行收集的数据，并发现数据的不足，从而为进一步的数据收集与分析工作提出建议。最后，将召开所有主要利益相关群体的代表会议来审议影响评估报告。最终报告将以不同版本出版 – 总结报告记载主要发现和部分实证案例；详细报告则包括所有支持结论的证据。比起其他方法，参与式故事报告会这种方式会给不同的利益相关者包括社区提供了更多的论据，证实交通项目带来的不同影响。

3. 反腐败项目

此案例描述如下：

捐赠人支持非洲某国的一个反腐败委员会。该项目包括协助制定准则、基础设施更新和考察出访。类似项目正在六个国家实施。

这不是一个标准化的干预，事先也没有很严格的描述。反之，项目的具体目标和实现目标的手段随着项目的进展才逐渐显现，对于优先序和各种存在的可能性也有了更好的了解。鉴于此原因，我建议采用“开发式”的评估方式（巴顿，1994，2008），此法目的不是在完成标准、固定的干预后提供报告，而是希望在执行连续变化的、综合干预的过程中提供信息。

“开发式评估是长期的，且评估者与从事创新和开发的行动者之间是伙伴关系。开发式评估过程包括询问评估问题、收集信息为发展式决策制定提供反馈和支持、并且在突发情况下调整方向。评估者是团队的一部分，成员们在长期的、不断完善、不断适应、目标不断变化的过程中共同提出概念、设计并尝试各种新方法。评估者在团队里的主要职能是解释研讨评估问题，数据和逻辑，还有协助基于数据的有关事实评估，事实如何被揭露，哪些方向应该坚持，哪些方向应舍弃，应作哪些新实验 – 换句话说，在揭示和创新的开发过程中做基于数据的决策。（韦斯特利等，2006）。

在此案例中，影响评估首先必须清楚项目执行者和参与者对腐败的关键问题怎样认识，并且了解他们想要如何解决问题。这种认识作为项目的支持理论，将针对不同阶段项目的具体动议，与检验因果归因的方法一起得到发展完善。例如，如果某动议关注公众匿名报告腐

败案的热线，那么影响评估应该跟踪这些案例（有可能全部案例，或可能分层次随机取样，取决于案例数量和可用资源）以发现所指案例发生了什么，做了什么调查，是否是持续指控以及如果是的话，作恶者的下场是什么。

通过增加动议完善的理论，它的变化是很容易识别的。例如干预是想通过识别、惩罚和开除腐败官员来提升官员们的廉洁意识而发挥作用呢？还是想通过增加正式审计侦查风险阻止腐败或使公众检举腐败更容易？理论在惩治不同类型的腐败或不同类型项目中的腐败时是否及时做了修正？

如果评估的主要目的是为了识别良好的实践经验并借鉴到其他环境中，则一种叫“正向偏差”的方法可能是有效的。这需要我们识别出即将获得超常结果的地点或案例并加以证实，分析是什么导致结果的产生。“正向偏差”方法的显著特点是通常进行此类调查的群体都是为了学习良好的实践经验。它不是由评估团队做出的，这些经验将在随后传递给使用者。正向偏差已用于公共卫生、营养、妇女绝育术、教育和农业发展，在腐败案件中特别是公职人员索贿也有相关的案例报告，(霍洛维茨, 2006)。

评估反腐败项目的一个挑战是腐败本身对评估人员收集准确的证据的能力的影响，并在发现后采取适当的行动。在这种案例中，需要特别注意评估管理以确保评估的独立性和调查人的安全。这将包括直接向一定级别的政府部门（保证可信的）汇报以防发现的事实被掩盖或篡改，支持民间舆论组织收集或调查相关证据，检举腐败。

4. 结论

有意思的是，这些案例中都包含基于理论的评估成分，尽管项目理论的使用方式在实践中大相径庭。在第一个案例中，它需要识别哪些中间结果将会带来长远影响，并且判断在分析中还有哪些情景因素需要被调查。在第二个案例中，它用作概念框架将多种成分的不同证据拢在一起。在第三个案例中，它用作概念框架来指导渐进的设计，通过收集和分析数据，随时通报最新变化。

表 2 总结了对这些案例的分析及其对于影响评估的含义。虽然这个表重点总结了干预的几个重要层面，但不同程度上都会包含其它的层面（例如，所以案例都可能涉及到的一些简单层面）。

这三个案例表明“情景回应”设计需要掌握大量的方法和技术。即使影响评估的实际执行需要外包给有相关经验和技能的外部评估人，任务执行机构也必须对方法充分地了解，能够拟写适合的任务大纲、选择适合的咨询顾问、有效地管理合同，包括评估工作的质量。这些对于评估队伍的建设有很大的意义，我们对评估人员的要求既需要深度（熟练掌握某种特殊技能），也需要广度（广泛了解各种方法，并知道恰当地运用）。

表 2 案例特征与目标分析

| | 干预的主要层面 | 评估目的 |
|---------------------|------------------------------------|-----------------------------|
| 案例 1 附加条件的现金转移支付 | 简单层面-标准化干预; 复杂方面-与其它因素和项目结合发挥作用 | 了解是否生效——在什么条件下起作用,为当前政策提供信息 |
| 案例 2 交通基础设施 | 复杂的-要求不同的多元成分有效结合 | 评估已完成项目的总体影响 |
| 案例 3 反腐败行动 | 复合的- 适应与紧急干预, 依照需求、问题和机会快速做出调整反应 | 理解和改善在建的和变化的项目 |

致谢

感谢迈克尔·巴顿, 凯·史蒂文斯, , 霍华德·怀特和鲍勃·威廉姆斯对本文早期版本提供了宝贵的意见。

参考文献

博伊斯, B. (2006)。决定和管理肖尔黑文河环境流域, 环境流域知识回顾-报告 1。
NSW 自然资源部, 2006 年 5 月。2009 年 5 月 14 日从
www.dwe.nsw.gov.au/water/pdf/monitor_sholahaven_shoo3.pdf 检索

钱伯斯, 罗斯.(2009) 参与式方法。发展有效性期刊, 这一期。

Clear Horizon 公司和奥康纳自然资源管理局 (2008) 表演故事报告: 南 Emu-Wren 高山牧场和 Fleurieu 半岛沼泽地恢复项目以及它如何对阿德莱德与高山自然资源管理区的生物多样性做出贡献。堪培拉: 澳大利亚共同体。2009 年 5 月 16 日从
<http://www.nrm.gov.au/publications/books/pubs/psrmout-lofty.pdf>

达特, 杰西卡, J.(2008) “让人人参与的结果报告: 参与式故事保护会技巧”。在 2008 年澳大利亚评估协会会议上的论文演讲, 珀斯。2009 年 5 月 16 日从
<http://www.aes.asn.au/conferences/2008/papers/p100.pdf> 检索

达特, 杰西卡(2009) “参与式故事保护会技巧” 2009 开罗影响评价会议参会论文

唐斯, B.J.L.A. 巴尔穆塔, P.G. 费尔韦瑟, D.P. 费斯, M.J., 凯, P.S.雷克, B.D.麦普斯通 和 G.P. 奎恩(2002) 监控生态影响: 在流动水域的概念与实践 剑桥: 剑桥大学出版社

克鲁博曼, S.(2001) “健康政策的新展望”, CPRN 研究 H/03 卷, 加拿大政策研究网络公司, 渥太华

克鲁博曼, S.和 B. 齐默曼(2002) 复杂和复合系统: 成功的医疗改革是什么样的? 加拿大未来医疗委员会, 讨论稿 8。2009 年 5 月 14 日从
[http://www.hcsc.gc.ca/english/pdf/romanow/pdfs/8_Glouberman E.pdf](http://www.hcsc.gc.ca/english/pdf/romanow/pdfs/8_Glouberman_E.pdf)

吉特, I (2008) “寻找惊喜: 农村资源管理集体学习监控的再思考”。博士论文。瓦格宁根, 荷兰

希尔, A.B. (1965) “环境与疾病: 联系或原因”, 医学皇家学会出版物 1965 年 5 月, 58:295-300

好罗魏斯, B. (2006) 布里奇 Ogres, 小家伙和正向变异: 一对一的防止公职人员勒索要求。2009 年 5 月 14 日从 [http://www.positivedeviance.org/projects/law/Bridges final.doc](http://www.positivedeviance.org/projects/law/Bridges_final.doc) 检索

库尔茨 C, 斯诺登 D. (2003) 战略新动态: 在复合和复杂世界中的意义。IBM 系统期刊 2003; 42 (3): 462-483。

澳大利亚土地与水 改善区域自然资源管理系统。堪培拉: 澳大利亚政府出版服务。5 月 14 日从 <http://www.rkrk.net.au/imag3/3/34/PR061220.pdf> 检索

NONIE 2 分组 (影响评估网络的网络) NONIE 影响评估指南。2009 年 5 月 14 日从 [http://www.worldbank.org/ieg/nonie/docs/NONIE SG2.pdf](http://www.worldbank.org/ieg/nonie/docs/NONIE_SG2.pdf) 检索

NONIE (影响评估网络的网络) 影响评估陈述。2009 年 5 月 14 日从 [http://www.worldbank.org/ieg/nonie/doce/IE statement.doc](http://www.worldbank.org/ieg/nonie/doce/IE_statement.doc) 检索

莫里斯, R.: 利斯顿, P.; Mugodo, J.; 尼科尔斯, S. (2005) 探查生态对管理干预反应的多线性等级证据。在美国地理协会 2005 年春季会议上的论文。

巴顿, 迈克尔·奎恩. (1994): “开发性评估”, 评估实践, 15 卷, 3 期: 311-319

巴顿, 迈克尔·奎恩 (2008a) 重点评估的使用 第四版. 课本。千年橡树, CA: Sage 出版物

巴顿迈克尔·奎恩(2008b)。测量发展援助的艺术 给世行独立评估组的意见 2008 年 4 月 10 日会议, 华盛顿特区。2009 年 5 月 14 日从
<http://www.worldbank.org/ieg/conference/results/patton.pdf> 检索

拉瓦伦, M. (2009) 三个典型干预案例之评估 发展有效性期刊, 此问题。

罗杰斯, P.J. (2008a) 复合干预影响评估的四个主要任务, 主旨发言。影响再思考研讨会。理解贫困与变化的复杂性, 国际农业研究磋商组织, Cali-Colombia。2009 年 5 月 14 日从
<http://www.prgaprogram.org/riw/files/papers/Roger%20material%20for%20workshop.ppt> 检索

罗杰斯, P.J. (2008b) “用于复杂和复合项目的程序理论” 评估: 理论与实践国际期刊 14 (1): 29-48

苏亚雷斯, F.V, 佩雷斯, G.I. 平田 (2009) “附条件资金转移的成果与失败: 巴拉圭 Tekopla 项目的影响评估。2009 影响评估会议的提交论文, 开罗

斯塔塞, R. (1992)。管理未知项。旧金山: Jossey-Bass

魏斯, C. (1997)。评估: 研究项目和政策的方法。(第 2 版) Upper Saddle 河, NJ: Prentice Hall。

韦斯特利, F., B. 齐默曼, 巴顿, 迈克尔·奎恩 (2006) 实现也许: 世界如何变化。加拿大 Random House。2009 年 5 月 14 日从 <http://innovationlabs.com/r3public/rtr3/pre/pre-read/Patton.Developmental%20Eveluation.pdf> 检索。

注 1 NONIE 是影响评估网络的网络, 由经济合作与发展组织的发展援助委员会评估网络、联合国评估组、评估合作组和国际评估合作组织-区域评估协会之网络组成。