



# 科技评估

2014年08月25日 2014年第05期

总第70期

## 国际动态专刊之七

- 欧盟“创新型联盟”战略实施的监测评估体系
- UNEG 的评估报告质量检表
- OECD 总结发展评估的 12 条经验

主办：科技部科技评估中心  
责编：田德录 王再进 王越婷

邮编：100081  
电话：010-6216 9515  
传真：010-8823 2615  
网址：<http://www.ncste.org>  
电邮：[news@ncste.org](mailto:news@ncste.org)

## 欧盟“创新型联盟”战略实施的监测评估体系

创新型联盟 (Innovation Union) 是欧洲 2020 战略确定的七项计划的旗舰行动计划，欧盟各成员国据此制定和实施创新政策，旨在摆脱经济衰退，走上可持续、智能化和包容性增长发展的轨道。欧盟委员会通过创新型联盟记分牌 (Innovation Union Scoreboard)、公共部门创新记分牌 (Public Sector Innovation Scoreboard)、区域创新记分牌 (Regional Innovation Scoreboard)、创新型联盟竞争力报告 (Innovation Union Competitive Report)、欧洲研究区观察 (Europe Research Area Watch)、创新联盟信息和情报系统 (Innovation Union Information and Intelligence System I<sup>3</sup>S) 等措施和方法，对创新型联盟的建设进程进行多层次、多角度的监测评估。

### 1. 创新型联盟记分牌

为确保创新政策有效执行，欧盟对原有评估各国创新绩效的工具，即欧洲创新记分牌 (EIS) 进行了改造，并更名为创新型联盟记分

牌 (IUS)。IUS 每年发布一次报告,采用一系列创新指标对欧盟成员国的创新绩效进行广泛的监测和排名,同时部分指标与国际上重要的国家进行比较,分析各国在这些创新绩效方面的相对变化。

欧盟于 2011 年首次发布创新型联盟记分牌报告,使用来自欧洲统计局、OECD、汤姆森·路透等 24 个指标数据,对欧盟成员国及其与美国、日本和金砖四国 (BRICs) 的创新绩效进行了比较分析。创新型联盟记分牌的指标不断修改和完善,2014 年最新创新型联盟记分牌包括 3 类指标 (基础条件、企业活动、产出)、8 个维度 (开放、优异的研究体系、金融支持、企业投资、联系与创业、智力资本、创新、经济影响和人力资源)、25 个与研究创新相关的指标。通过指标综合排名,欧盟创新记分牌将欧盟成员国分为四组:创新领导者,创新水平高于欧盟平均水平的国家;创新追随者,创新高于或接近欧盟平均水平的国家;中等创新国家,创新表现低于欧盟平均水平的国家;适度创新国家,创新表现远低于欧盟平均水平的国家。在 25 个评价指标中有 12 个用来与国际上其他重要国家进行比较分析,但是非欧盟国家不

参与综合排名。

2014 年创新联盟记分牌报告指出,最近几年,欧盟所有成员国的创新能力都在提高,与美国和日本之间的差距也在缩小,但欧盟内部创新能力仍存在很大差异,创新能力下降的区域占比接近 20%。中国的创新绩效快速增长,相对欧盟的绩效能已经从 2006 年的 35% 增加到 2013 年的 44%。

## 2. 创新型联盟竞争力报告

欧盟公布了 2011 年《创新型联盟竞争力报告》,对创新型联盟战略相关任务的实施进展和总体目标的实现程度进行评估,为创新型联盟的政策措施提供证据和分析,就欧盟及成员国研发强度目标和知识密集型企业创新情况等方面进行深度观察。报告还延伸和补充了创新联盟记分牌指标之外的创新评价,包括研究和创新对提高竞争力和应对社会挑战的影响等。报告主要从三个方面对创新型联盟的竞争力进行研究评价:

1) 总体目标:欧洲 2020 战略包括 5 个总体目标,与建设创新型联盟有关的总目标是改善研究和开发的条件,其中包括将公共和私人共同研发投入提高到 GDP 的 3% 的目标。欧盟委员会呼吁成员国采用新的单一综合指标对创新进程进行更好的

监测，还与欧盟统计局和经合组织合作，发展了关于快速增长的创新企业就业比例的指标。

2) 记分牌绩效：对欧盟及各成员国的创新绩效及相对位置进行评价。2011 年年开始发布创新型联盟记分牌(IUS)，此后每年更新一次，将欧盟和成员国的绩效与主要国际伙伴进行比较。

3) 战略分析：分析绩效不足的原因和补救措施。创新联盟竞争力报告(IUC) 每两年对高效的主要特征和社会有效的创新体系进行深度统计和经济分析。

报告研究指出，欧盟正在慢慢地向实现 3% 的研发目标前进，但由于在 R&D 方面投资较少，欧盟和其全球的竞争对手之间的差距在拉大；在研究和创新方面的投资是欧洲未来的增长和创新的主要驱动力。

### 3. 创新型联盟状态报告

欧盟发布了《创新型联盟状态 2012》，总结 2012 年成员国和欧洲层面实现创新联盟进展情况，并总结概括地区仍然需要采取行动以加速创新型联盟建设进程。报告的主要内容包括：一是描述了创新型联盟战略的背景和目的，包括加大在教育、研究和创新的投资，发展创新经济占领新

市场，创造高质量的就业机会等；二是各国家研究和创新体系的状态，包括研发投入、改革提高效率 and 有效性、创新型联盟国家层面的实施进展；三是对进一步推进实施创新型联盟提出措施和政策建议，包括加强知识库建设并减少碎片化、促进科技创新成果的市场转化、建立欧洲创新伙伴关系、支持社会创新和区域合作、加强与欧盟以外的合作等。

还发布了《欧盟成员国及相关国家的创新型国家的状态报告 2013》。该报告根据欧盟统计局(Eurostat)、经济合作(OECD)与发展组织的最新官方统计和其他数据源，对欧盟及各个国家的总体状态进行评价，总结了创新领导者在增加研发投入、改革研究和创新体系、提高效率 and 有效性的经验，同时指出每个国家面临的挑战及其技术优势和创新机会。

### 4. 区域创新记分牌

欧盟认为区域层次的创新绩效监测对区域经济发展和创新政策的设计与实施非常重要，采用《区域创新记分牌(RIS)》的评价指标对区域层次的创新绩效进行监测、比较和评估，帮助理解并推进区域层次的创新。RIS 不定期发布报告，至今已发布 6 次评价报告，分别是 2002、2003、

2006、2009、2012、2014年。

区域创新记分牌的指标主要来源两个部分，一部分来自欧洲创新记分牌(EIS)，另一部分指标来自区域社区创新调查(CIS)，区域创新记分牌的指标不断变化和丰富。由于一些区域没有进行完整的社区创新调查(CIS)不能提供相关数据，无法对各区域进行绝对的排名，但是根据绩效的相似性将各区域进行分组，较早时将各创新区域分成五组，后改为创新型联盟记分牌对各成员国的分类方法，将各区域分成四组：创新领导者、创新跟随者、中等创新者和适度创新者。

2014年区域创新记分牌采用11个指标，其中6个指标与创新型联盟记分牌相同，对欧盟各区域创新的绩效进行分组，其中区域创新领导者34个区域，创新跟随者57个区域、中等创新者68个区域和适度创新者31个区域。

## 5.公共部门创新记分牌

欧盟的公共部门作为监管机构、服务提供者和雇主扮演了一个关键经济角色，相关就业人员占总就业人数的25%以上，在欧盟成员国的经济活动中占很大份额。欧盟委员会于2013年启动《欧洲公共部门创新记分牌(EPSIS)》监测评估试点工作，以更

好地理解和分析公共部门的创新，并改善欧洲公共部门的创新绩效。公共部门创新记分牌的测量框架主要包括三个部分：一是基础条件，包括人力资源，政策服务的质量2个指标；二是活动，包括能力，驱动因素和障碍2个指标；三是产出，包括创新者、对企业绩效的影响、政府采购3个指标，7个评价指标包含了22个二级指标。

欧盟发布了首个《欧洲公共部门创新记分牌2013》，结果表明，欧洲公共部门开展了创新，但仍然面临着许多障碍：公共管理创新对改进用户获取信息、提高用户满意度和更快交付服务有积极影响；改进和创新公共服务管理解决方案有助于提高竞争力和创新公司的绩效。存在的主要问题包括：内部管理和创新文化的障碍不仅对创新产生负面影响也使公共部门的效率一般；公共部门可以通过公共采购做更多的工作来促进创新。

## 6.欧洲研究区观察(ERA WATCH)

ERA WATCH(欧洲研究区观察)是欧洲委员会关于欧洲、国家与区域研究和创新体系以及政策的官方平台，发布欧盟创新型联盟的进展情况，主要包括：

ERA WATCH 报告。该报告主要

包括三个方面的内容：ERA WATCH 分析，涵盖新兴主题和结构的发展趋势，研究和创新的内容和演变系统和政策；ERA WATCH 年度国家报告，提供成员国和区域的研究与创新系统状态和演化的分析评论；ERA WATCH 区域报告，分析区域内成员国在欧洲研究中扮演的角色。

ERA WATCH 欧洲视角。概述了欧洲在研究与创新政策和制度的水平，包括欧盟的总体情况、关键的研究和创新活动、政府间研发合作的总体情况。

ERA WATCH 还提供关键里程碑事件（包括突出的成就和具体的事

件）、出版物和其他活动计划等信息。

## 7. 创新联盟信息和知识产权系统

创新联盟信息和情报系统是欧盟委员会对所有创新型联盟利益相关者开放的信息平台，其主要目的是确保所有利益相关者都能够了解创新联盟战略实施进展情况。通过该信息平台非常容易获取最新的、全面的和可比的信息，以及 2010 年 10 月委员会在创新工作通讯上作出的 34 个创新型联盟承诺的实施进展。该系统还提供每一个承诺详细的背景和相关的情报信息，这有助于进一步探索关键问题的解决，并提供额外的支持数据和分析视角。

（供稿：科技部科技评估中心 陈华雄）

## UNEG 的评估报告质量检表

**译者按：**联合国评估小组 UNEG（United Nations Evaluation Group）是联合国系统内的专业评估组织，其职能是为联合国系统各部门制定评估政策和开展评估活动提供指南。2005 年以来，UNEG 相继制定了评估规范和标准、评估道德准则和行为准则、评估报告质量检表等，形成了联合国系统的评估标准规范体系。这份评估报告质量检表，是联合国系统对评估报告的进行质量控制的技术依据。质量检表中的 8 条要求 48 条质量细则，对于我们撰写评估报告也具有重要的参考价值。

<b>UNEG 的评估报告质量检表</b>	
该检表的目的是帮助评估管理者和评估者确保评估的最终产品——评估报告——符合预期的质量，也可以作为评估实施之前任务大纲（TOR）的一部分，或者用来在评估报告完成后对其质量进行评价。	
评估题目：	
委托机构：	
<b>1. 报告结构</b>	
1.0	报告应具有良好的结构和逻辑，表达清晰且完整。
1.1	报告要有清晰连贯的逻辑结构（例如，在评估发现之前要有背景和目标，在结论和建议之前要有评估发现等）。
1.2	标题页和和开篇的页要提供关键的基本信息： <ul style="list-style-type: none"> <li>1.评估对象的名称</li> <li>2.评估的时间表和评估报告日期</li> <li>3.评估对象所在地点（国家、地区等）</li> <li>4.评估者的姓名和/或所属组织</li> <li>5.评估委托组织的名称</li> <li>6.报告内容目录，并列出图表和附件</li> <li>7.列出缩略词。</li> </ul>
1.3	执行摘要是一个 2-3 页的独立部分，包括1： <ul style="list-style-type: none"> <li>1.评估对象的概况</li> <li>2.评估目标和目标受众</li> <li>3.评估方法论</li> <li>4.最终的发现和结论</li> <li>5.主要建议。</li> </ul>
1.4	附件提高了评估报告的可信性。除了其他事项外，附件可能包括2： <ul style="list-style-type: none"> <li>1.任务大纲（Terms of Reference, TOR）</li> <li>2.人员访谈和实地调研清单</li> <li>3.文献清单</li> <li>4.关于方法论的更多细节，如数据搜集工具，包括关于数据可靠性和有效性的细节</li> <li>5.评估者简历和/或对团队组成的说明</li> </ul>

<sup>1</sup> 执行摘要包含的关键要素列在了《联合国系统评估的 UNEG 标准》(UNEG/FN/Standards[2005])中，第 18 页，标准 4.2 的第 3 条。

<sup>2</sup> 《联合国系统评估的 UNEG 标准》(UNEG/FN/Standards[2005])的第 20 页，标准 4.9 和第 23 页，标准 4.18 描述了附件的内容。

	6.评估矩阵 7.结果框架。
<b>2. 评估对象</b>	
2.0	评估报告要对评估对象进行清晰且充分地描述 <sup>3</sup> 。
2.1	清晰描述评估对象的逻辑模型和/或预期结果链（投入、产出和效果）。
2.2	描述对评估对象产生直接影响的关键社会、政治、经济、人口和制度等因素。 例如，合作伙伴政府的战略和优先事项，国际、地区或国家的发展目标、战略和框架，有关机构的公司目标和优先事项等，都要酌情考虑。
2.3	清晰描述评估对象的规模和复杂性，例如： <ul style="list-style-type: none"> <li>• 包含的行动数量，如果超过一项，要描述每项行动直接或间接服务的目标人口规模。</li> <li>• 地域背景和边界（如相关的区域、国家和/或景观以及挑战）</li> <li>• 目的和目标，对象的组织与管理</li> <li>• 各种来源的资源总量，包括人力资源和预算（例如，有关机构，合作伙伴政府和其他捐助者的捐款。）</li> </ul>
2.4	评估对象实施中涉及的关键利益相关者，包括实施机构和合作伙伴，以及其他关键利益相关者和他们的角色。
2.5	评估报告要确定对象的实施状态，包括实施阶段和任何随着时间推移发生的明显变化（例如，规划、战略、逻辑框架），并解释这些变化对评估的含义。
<b>3. 评估目的、目标和范围</b>	
3.0	充分阐释评估目的、目标和范围。
3.1	明确界定评估目的，包括为什么在该时间点需要开展该评估，谁需要这些信息，需要什么样的信息，如何使用这些信息等。
3.2	评估报告应对评估目标和范围进行清晰的解释，包括评估的主要问题，并描述和解释该评估做什么以及不包括什么。
3.3	评估报告应描述并解释评估者选用评估准则、绩效标准或使用的其他准则 <sup>4</sup> 。
3.4	评估目标和范围也要酌情考虑性别和人权方面的问题。
<b>4. 评估方法论</b>	
4.0	评估报告要坦率地描述评估所运用的方法论，清楚地解释如何进行了评估的具体设计，以反映评估准则和回答评估问题，实现评估目的。

<sup>3</sup> 评估“对象”是干预（效果、计划、项目、项目群、主题、软援助），这些是评估的焦点和报告中提供的评估结果。

<sup>4</sup> 最通用的评估准则如下：OECD/DAC的5个准则，即相关性、效率、效果、影响、可持续性。每项评估都可能有不同的焦点（并非每项评估都必须使用全部评估准则）。每个机构也可能希望在这一工具中增加一个指标，以评价每项准则在某项评估中的使用范围。

4.1	评估报告应描述数据搜集和分析方法，以及选择这些方法理由和局限性。评估报告也要在相关的部分描述参考指标和评估基准。
4.2	评估报告应描述数据来源，以及选择这些来源的理由和局限性。评估报告要讨论如何组合运用多种数据来源，以获得多种观点，保证数据的准确性和克服数据的局限。
4.3	评估报告要描述抽样总表——作为代表的地区和人口、选择理由、选择机制、从潜在受众中选择的数量以及样本的局限性。
4.4	评估报告要完整描述评估中向利益相关者的咨询过程，包括选择特定层面利益相关者的理由和咨询活动的开展情况。
4.5	采用的方法应适合该评估，适合回答评估问题。
4.6	采用的方法应适合分析该评估范围内确认的性别和人权问题。
4.7	评估报告要提供证据，以表明评估采用了充分的措施来确保数据质量，包括关于数据搜集工具可靠性和有效性的支持证据（例如，访谈方案、观察工具等）
<b>5. 发现</b>	
5.0	发现要对评估准则和具体问题做出直接回应，这些准则和问题在评估报告的“范围和目标”部分已经得到描述；同时，发现要以通过数据搜集和分析方法得到的证据为基础，这些方法在评估报告的“方法论”部分已经得到描述。
5.1	报告的发现应反映出对数据进行了系统而适当的分析和解释。
5.2	报告的发现应回答评估范围中确定的评估准则（如效率，有效性，可持续性，影响和相关性）和评估问题。
5.3	发现应基于证据进行客观报告。
5.4	对于数据的差距和局限，以及未预期到的发现，应进行报告和讨论。
5.5	要尽量识别确认成功和失败的原因，特别是持续的制约因素。
5.6	总体发现要清晰、符合逻辑且具有连贯性。
<b>6. 结论</b>	
6.0	结论应以发现为基础并有证据支撑，形成合理判断，围绕评估对象和评估目的提出中肯的见解。
6.1	结论应反映出关于评估关键问题的合理判断。
6.2	结论应有充分的证据支撑，并与评估发现具有逻辑关联。
6.3	结论应为识别和/或解决重要问题或议题提出见解，这些问题或议题关系到评估用户的未来决策和行动。
6.4	结论应以证据为基础，并考虑跨部门的不同利益相关者的观点，提出评估对象（政策，计划，项目或其他干预）的优势和劣势。
<b>7. 建议</b>	



7.0	建议应与评估对象和目的相关，应由证据和结论支持，并在利益相关者的参与下提出。
7.1	评估报告应描述提出建议的过程，包括利益相关者咨询。
7.2	建议应坚定地以证据和结论为基础。
7.3	建议应与评估对象和目的相关。
7.4	应清楚表明每条建议的目标群体。
7.5	建议应明确表达，并明确说明应采取的优先行动。
7.6	建议应具有可操作性，并要慎重考虑委托方的理解和潜在约束。
<b>8. 性别与人权</b>	
8.0	评估报告应表明评估对象设计和实施、对结果的评价和评估过程在多大程度上涉及性别平等和人权问题。
8.1	评估报告应自始至终使用性别平等和尊重人权的表达，包括使用性别，年龄，残疾等数据。
8.2	评估方式和数据搜集及分析方法应体现性别平等和人权，适用于分析评估范围中界定的性别平等和人权问题。
8.3	评估报告应评价评估对象的设计是否以性别分析和人权分析为基础，结果实现是否以性别和人权框架进行监测，以及对性别平等和人权的实际效果如何。
8.4	在评估报告的发现、结论、建议和经验中，应充分提供性别平等和人权方面信息。

（供稿：东北大学秦皇岛分校 杨渝玲 科技部科技评估中心 邢怀滨 译）

## OECD 总结发展评估的 12 条经验

本文介绍了 OECD-DAC 在 2013 年发布的 12 条关于发展评估的经验，这 12 条经验从当前一系列问题出发，总结了 2009-2012 年 DAC 发展评估网络 5 成员的经验 和 DAC 同行评议

的新发现。这 12 条经验旨在到达三个目的：一提高发展合作的数量和质量；二是基于一般原则提供可靠的分析，无论经合组织的国家还是更广阔

<sup>5</sup> DAC 发展评估网络（The DAC Network on Development Evaluation）是 DAC 的一个分支机

构，旨在提高国际发展计划的有效性。该网络汇集了来自 OECD 发展合作机构和多边发展机构的评估管理人和专家，为评估实践提供指导和标准，促进援助机构和合作伙伴的合作。

的国际社会均能够应用的原则；三是使 DAC 的工作人员能够分享经验、分辨好经验和促进协同合作。这 12 条评估经验分为三种类型：战略框架、提供有效的评估和构建学习的文化。每个部分阐述了其对发展合作机构的重要性，具体如下：

### 一、学习和问责的战略框架

#### 经验 1 从证据出发决定发展政策

政策决策者和高级管理部门的决策制定过程需要独立可信的分析和建议。评估可以了解项目进展的效率和效益，辨别其通过各种方式达到何种程度的结果，特别是在决策资金支出、计划设计和战略制定时需要了解哪些干预措施最有可能达到预期的目标，这些都是基于评估发现来展开。

评估提供的必要信息可以让政治家更准确地了解如何实现目标和管控风险。然而目前政治决策的制定更多基于既定的政策制定路线和假设，如在城市行得通的政策在农村也能行得通，或者建更多的学校大楼会提高教育成果等。而不是在在特定环境中针对特定问题提供行之有效的策略。有扎实的证据支持，政策制定者不仅可以更坚定地执行发展合作计划，而且即使在财政紧缩时可以有

效调节支出。基于评估证据作出的决策较少受到政治利益左右，也不会因为人员或政府的变动而改变。

援助管理者需要理解对于组织来说什么是成功和失败，以便作出更好的决策。尽管监控、绩效管理和结果报告系统可以提供一些信息，但是仍然需要进行深入的评估来解释结果是如何实现的。

基于证据来管理项目需要持续的、高质量的证据，同时能及时提供相关课程。建立基于证据的组织需要政治、文化、制度和法规的制定。同时需要高层领导机构的大力支持以保证了解哪些工作是有用的，为何如此并对结果负责。评估单位和管理机构之间的定期互动也可以建立一个有效的对话机制。

#### 经验 2 让学习成为发展合作的文化

拥有强大学习文化的组织会有更大的影响力，更强的适应性和更好的应对失败。这种组织应该鼓励批判性思维、创新、敢于冒险，并且认识到真正的失败是没有从中学习到经验。

学习的文化不仅仅是关于评估本身，而是将评估的思维方式应用到发展合作的各个方面。要注重实效，

基于最佳有效证据来做决策，同时通过工作填补证据缺口。要质疑假设，保持开放的批判性分析在特定环境下是否有效并寻找原因。学习文化的核心是通过不懈的努力得到很好的结果。发展是一个复杂的过程，许多领域仍然充斥着知识空白。很多时候并不知道哪种方式最有效，需要仔细的实验来发掘哪种方式最有效，了解那些环境因素对取得成效是至关重要的。而评估能得到这样的答案。

学习的文化可以鼓励员工和管理层从成功和失败的经验中探索和学习。如果不愿意检验假设或者改变路径，评估发现和建议不会被注意而错误会重复。如果没能理解哪些是成功的就回导致无效的、未达到目标的或者不完整的支持，不仅会浪费资金，而且会让本应从援助中获得帮助的人们丧命或生活环境受损。

持续的学习和提高需要组织和个人都要从评估、研究和结果中学习。鼓励学习的环境不只是改变规定和政策。必须鼓励员工、培养领导和合作者，要有不断提高结果和机构本身的长远眼光。管理者必须能够承担可计算的风险，鼓励员工批判性的将他们的经验应用到将来的评估中。评估的结果并不一直都是正向的，有时

会揭示不是那么有效的项目或者非有意的破坏。政策制定者必须接受并不是所有的风险都可以避免、或者提前降低的。

通过 2010 年比利时的经验，强化学习的环境需要一下关键几步：

(1) 在组织的各个层面都要强调通过评估来制定政策和策略；

(2) 将评估作为一种重要的知识管理工具；

(3) 高级管理者要明确地支持学习，并且能同时接受评估得到的积极和消极结果；

(4) 在组织中创造一个保证每日学习的氛围，包括学习他人的评估。

### **经验 3 为评估做明确定位**

好的发展评估会明确设立目标来保证机构和项目目的完成。评估是通过提供独立可靠的证据来证明发展活动的效益、效率、关联性、影响力和可持续性。这些信息用来支持学习、检测理论，包括如何达成发展目标、强化项目设计和管理、为决策者提供信息。评估应该提供好的证据来满足发展政策决策者的需求，并且帮助员工和管理层设计更好地项目和策略。评估可以使政府和发展伙伴为项目的结果负责。

一个坚实的评估政策应该在一个基础上，这个基础包括学习的氛围，和可以支持这个决定的证据支持。当政策决策者、管理层、员工和参与者理解在发展合作活动中评估的定位，评估的质量和就会提高。

## 二、提供有效的评估

### 经验 4 使目标与有限资源相匹配

评估需要大量的人力和财力的支持，包括能对生产和使用进行评估的资金支持，对管理单位进行评估的资金支持，在委托研究、数据收、知识管理和传播有资金支持。人力支持与财力支持同样重要。评估人员应该是专业的、持续的、有能力应用合适的方法去检测组织的活动和结果，并且能掌控评估过程的有效性。

想要进行高质量的评估必须有根据的进行投资。资源的缺乏不仅会损害评估本身，同时会波及到学习的文化和整个机构的运行。评估可以通过提高有效性和影响力来节省资金，资金既是效益又是成本。DAC 的人员在 2010 年的评估中得到经验认为资源的缺乏，特别是人力资源的缺乏，是影响评估报告质量和合作伙伴的关键。

对于项目的评估可以通过项目预算来提供资金支持，而项目预算中有多少的百分比用来评估是一个讨论的话题。有组织规定了 1%-3% 的大致范围。但这些数字的提出是有些武断的，最重要的标准是能为战略评估提供足够的资源。

常用的保证资金的方式是为核心评估单元提供专项资金。在战略层面，DAC 成果国的核心评估单位总体有 2.4 美元的预算，意味着 0.16% 的 ODA<sup>6</sup>用于支付评估的费用。人数多经费高的评估机构并不意味着每年的评估产出高，因为评估的投入由评估的规模、目标、数据要求和评估方法决定。DAC 成员平均每年每个评估单位的员工参与 1-2 两个重大战略评估。利用研究机构、大学或私人公司也可以帮助提供专业支持，减少直接资金投入。

### 经验 5 加强项目设计和管理系统

评估本身并不是孤立的。它依赖于管理系统提供关键指标数据和详细干预细节。为了有效评估，实施干预措施后必须能够得到可测量的结果。项目设计、监测、性能、知识管

---

<sup>6</sup> ODA (Official Development Assistance)：官方开发援助。是指发达国家政府为发展中国家提供的，用于经济发展和提高人民生活的，赠与水平 25% 以上的赠款或贷款。

理系统能使评估更加完备，是高质高效评估的先决条件。

精心设计和有效监控项目的运行能够增加援助项目的影响力，同时评估也会更简便和省钱。能够有效管理项目结果的组织应该对项目目标有明确的界定，并清楚那些因素会对结果造成影响。很多项目无法评估的原因就是项目的设计和实施过程出现问题。比如说，干预措施考虑不周、目标不明确、没有正确识别和评估风险、干预理论过于笼统、指标没有明确定义、基线数据丢失、没有代表性抽样、不清楚安排的活动是否为目标服务。项目如果出现以上缺陷，则无法有效进行评估。即使进行评估也要耗费大量额外的时间和金钱填补空白。

很多机构已经制定了系统过程，在项目设计阶段就开始审查，保证项目有明确的目标、清晰的项目逻辑、能够被监测和评估。在项目的审批过程中这些因素都会考虑进去。

### **经验 6 提出正确的问题，对预期的结果保持现实**

政策制定者和项目管理者对评估能回答什么问题有很清晰的预期是非常重要的。尽管评估是一个有效的投入，但是评估本身并不能解决机

构自身的学习能力和问责机制。此外并不是每件事随时都需要进行评估。评估的议题应该与机构整体战略规划有密切联系。

发展计划的管理者应与评价单位紧密合作，确定利益的战略问题和着手解决他们的研究议程。决策者应定期参与评估和相关研究，例如对评估工作计划有所了解、与评估负责人讨论战略目标和机构准则。高级管理者应该与评估机构合作，在整体层面上制定明确的政策和战略保证结果的可靠性。

### **经验 7 选择正确的评估工具**

评估应使用恰当的研究方法和统计方法来回答评估问题，对项目活动的效率、效益、关联性、持续性和影响力进行评价。有各种各样的评估方法存在。究竟哪种方法最合适取决于评估目标。评估的类型包括事前评估、事后评估、事中评估、元评估和参与式评估。

没有一种方法是适用所有评估的。经验表明当方法与问题相契合、内容与项目相契合时，这个方法最可能有效。DAC 发展评估质量标准强调在选择评估方法时要注意灵活性。该方法能挖掘有用的、相关的和可靠的

的证据。过于严格的方法会扼杀批判性思维。

OECD DAC 的成员国有评估指南或手册，阐述了一系列评估工具和方法供选择。可能在不同目标上需要权衡。使用者在选择前应先咨询，以确保选择的评估方法能够提供想要的信息。

### 经验 8 合作

评估应涉及到国家利益相关者、政府、其他发展援助提供者、受益人 和非国家行为者。评估可以联合进行，在管理或指导小组进行合作。其他合作方式有共享数据、报告、情景分析、对评估结果的综合调试、元评价、同行评审、制定共同的定义和方法。

合作有助于相互学习，特别是帮助理解高水平的发展结果。此外，在高政治风险和暴力冲突的情况下，合作工作能够使各国政府免收政治压力。评估可以为合作双方提供一个切入点，双方可以共同讨论，在战略目标上达成共识，可以批判性地分析研究方法，共享评估结果。合作计划和实施评估可以创建一个相互问责的基础。从合作数据中得到的评估分析也能准确反映现实状况。

### 经验 9 帮助提高合作国家的能力并加以利用

捐赠人应该支持发展中国家提高自己的成果管理和评价系统，这在巴黎有效援助宣言提出，并在阿克拉行动议程和釜山伙伴关系中重申。要同时提高机构和个人的评估能力，增强在政策制定和支出管理的透明度。

提高这些国家的评估能力不仅确保经费的合理使用，而且还帮助发展中国家减少贫困和促进经济增长，制定更加有效的公共机构和公共政策。公共政策评估对国内和国际间的问责都是一个有价值的工具。

### 三、加强学习文化

#### 经验 10 对评估的发现采取进一步行动

评价结果将政策制定者的决策和行为。因此需要战略规划、方案设计和监督机制来确保适当及时的行动。这通常涉及到一个正式的管理响应系统，项目管理者可以通过这个系统对每一个评价结果进行反馈。非正式的反馈循环也可以帮助管理者了解评估发现。报告和监测管理的实施效果是保证计划实施的关键。

管理响应系统确保相关行动者（项目经理、国家部门、项目人员、高级管理人员、行业主管部门、国家

合作伙伴、议会委员会)对相关评估发现有所了解并有根据的采取正确行动进一步促进项目完成。这样的反馈评价是组织学习环境的一部分,同时鼓励所有人都投身到评估中。在计划设置之处就应考虑建立坚实的反馈系统,这可以避免选择性抑制不好的评估结果,保证透明性。

发展政策的过程是复杂的,除了评估证据以外还被许多因素驱动。因此想要提高评估结果对政策影响绝非易事,需要在以下四个方面加强:优先考虑政策并明确重点、评估者和政策制定者之间建立联系、积极主动的沟通和知识共享。评估单位也需要机会进行宣传,如部长访问特定的国家等。

### 经验 11 有效沟通评估结果

研究发现和结论需要有效的沟通才能实现评价目标。有效的沟通指利益相关者能迅速获得并理解信息和证据。发展机构需要在沟通方面投入足够的关注和资源,否则评估的努力将会白费。良好的沟通会有效支持学习,带来确凿的证据,并使评估发现得到透明的结果。

机构的评估准则应该优先考虑交流评估结果。制定评估计划时,应同时考虑宣传和交流的计划。不同的

评估所针对的对象不同,因此要有针对性的沟通。如,利益相关者对多种类型的信息都有需求,并且沟通时会运用不同方式的媒介;发展机构的人员则更倾向于如何促进工作;政治领导者则对发展的整体成效和是否有声誉风险感兴趣;NGO 则关系投入是否取得了成效。评估管理者应在早期考虑潜在受众的需求。

### 经验 12 对评估者进行评估

高质量的评估应具备可靠性、有效性、准确性、有用性和解释性,这些特征是相辅相成的。发展机构应该检查评估是否达到了目的,是否对政策和实践产生影响从而提高发展成效,包括检查个人评估报告的质量和定期审查评价函数。

严谨的评估更能让人信服。稳健性有助于提高研究的外部效度——这意味着可以吸引更多的个人研究。简而言之,低质量的评价是没有用的,不足的证据、有缺陷的分析或错误的建议会让弊大于利。

2010 年 DAC 发布的《发展评价质量标准》描述了高质量评估过程和评估报告的基本特征。评估报告应该清晰地阐述评估的方法、假设、数据缺失和其他不足,让读者能够判断报告的质量。系统质量检查会根据 DAC

的标准来检查参考文献、报告初稿和最终报告。

关于如何保证评估的高质量：

(1) 检查同行评审的的名单和质量；

(2) 评估专家进行独立评估；

(3) 通过周期性的元评估 (meta-evaluation) 或系统回顾对特定主题的证据进行整体质量把控，发现工作的不足并改进；

(4) 培训员工，包括派遣员工进行外部培训，观察更有经验的评估者如何进行评估，或者借调到其他机构；

(5) 向评估顾问提供培训；

(6) 对评估顾问的表现进行回顾和评价；

(7) 花时间检查参考文献（因为不好的报告经常是设计缺陷的评估或者对参考文献不清楚）。

同行评审或者独立评价可以很好地判断评估是否达到预期目标。同行评审可以用 DAC 的评估网络评估工具进行系统评价。

本文内容参见：OECD 2013. *Evaluating Development Activities: 12 Lessons from the OECD DAC*.

（北京理工大学国防科技创新与教育发展  
战略研究基地 姚思宇、何海燕）