



联合国开发计划署、
联合国人口基金和
联合国项目事务厅
执行局

Distr.: General
10 November 2014
Chinese
Original: English

2015 年第一届常会

2015 年 1 月 26 日至 30 日，纽约

临时议程项目 4

评价

审议开发署的评价政策

目录

章次	页次
一. 审查范围	2
二. 方法	2
三. 主要调查结果：开发署和负责执行评价政策相关实体对照变革理论取得的业绩	3
四. 开发署评价政策的主要问题和今后的备选方案	6



一. 审查范围

1. 审查评估 2011 年以来开发署评价职能的绩效，即本组织在多大程度上响应了政策的要求以及政策的执行情况。审查明确了为加强评价职能需要改变政策或做出管理决定的领域。正如职权范围规定和启动审查工作期间所确认的，审查的重点是评价政策的三个要素：开发署的总体评价职能、独立评价和分散评价。¹ 鉴于在分散评价的覆盖面和质量方面反复报告存在不足，而执行局又特别关心这一问题，因此审查特别重视第三个要素，即确认存在的不足之处、探究原因并提出适当的行动方针建议。关于相关基金和方案，即资发基金和志愿人员方案，审查评估它们如何执行评价建议，并对照政策细则和独立评估办公室通过的方法进行比较。

2. 评价政策将主要职能分派给整个开发署系统的不同实体。执行局是评价政策的管理机构。独立评估办公室是评价若干方面的管理机构，特别是负责提供独立评价(主要是评估发展成果及方案评价和专题评价)。署长负责开发署的成果，包括负责确保各局和各国家办事处在高级管理人员的监督下进行适当独立、公正的分散评价。相关基金和方案的评价单位履行其各自机构的评价职能。

二. 方法

3. 审查采用“变革理论”方法，使用混合方法和三角核证法确保数据质量。通过若干调查线索生成与政策及其执行有关的定性和定量数据，这些线索包括：

(a) 对一系列广泛文件和档案进行文件审查；

(b) 对分散评价抽样和独立评估办公室 2012 年以来完成的评价进行元评价(质量评估)，审查开发署评价的质量和评估进程的质量；

(c) 访问开发署总部、非洲、亚洲和太平洋、欧洲、拉丁美洲和加勒比的国家办事处和区域服务中心及位于波恩的联合国志愿人员方案办事处。访问内容包括与开发署、多边组织、双边捐助方和伙伴国家机构的一系列广泛相关利益攸关方进行半结构化访谈或重点小组访谈，访谈后审查了更多文件；

(d) 通过电子调查征求开发署驻地代表、负责评价工作人员和在过去 3 年对开发署进行过分散评价的外部咨询人的意见；

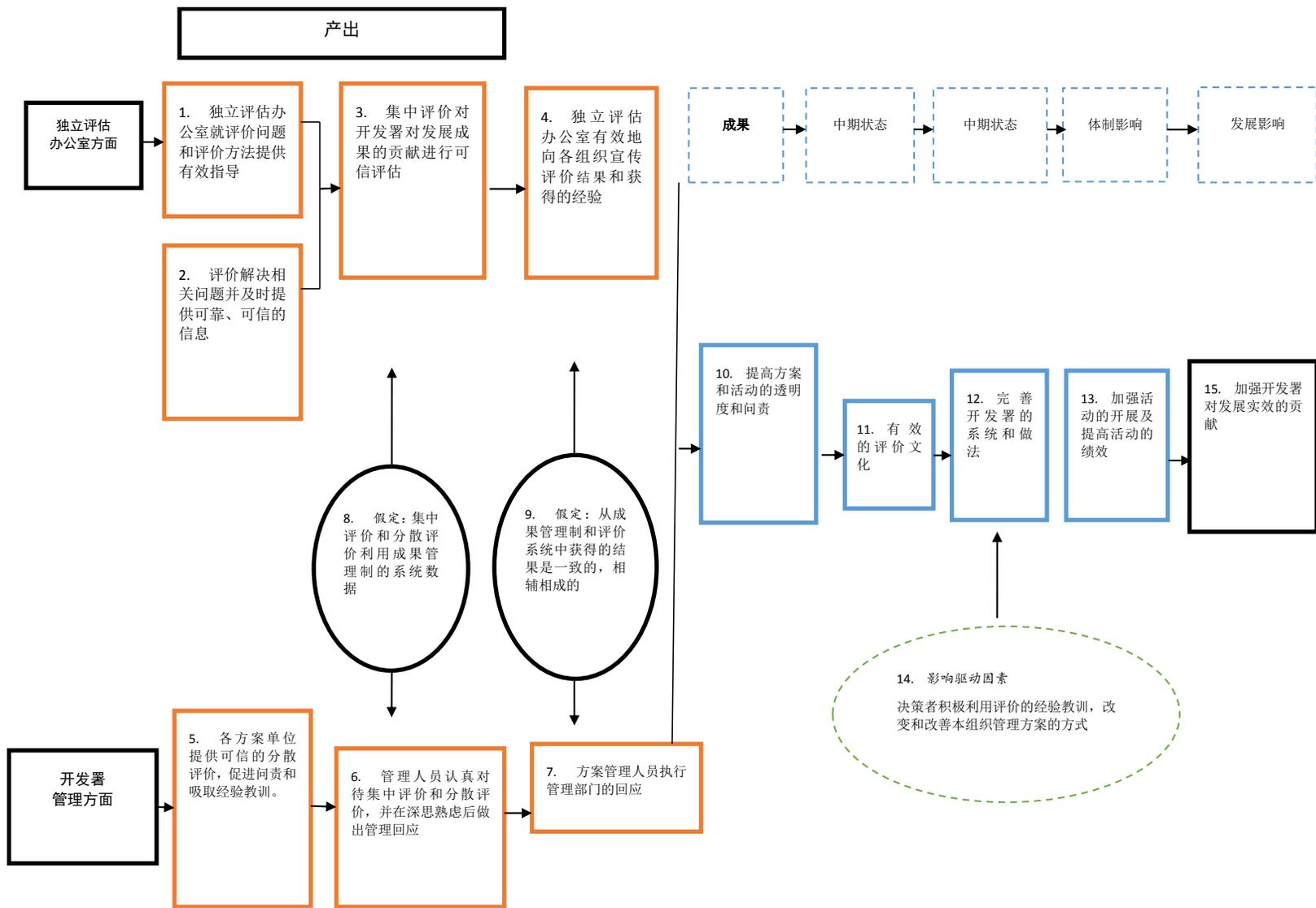
(e) 根据访谈和文件分析对联合国志愿人员方案和资发基金进行小型的评价个案研究。

¹ 独立评价由独立评估办公室进行，而分散评价工作由开发署各方案单位委托独立外部专家进行。

三. 主要调查结果：开发署和负责执行评价政策相关实体对照变革理论取得的业绩

4. 在变革理论框架内对审查的主要结果进行了分析(图 1)。分析表明，政策规定评价产出的两个主要方面将由独立评价办公室(方框 1-4)和开发署管理部门(方框 5-7)落实。根据该理论，这两个方面是相互关联的(方框 8 和 9)。这两个领域的成果应共同形成一个有效的评价文化，推动机构实施和业绩(方框 10-14)。这一过程后面几步的主要动力是决策者积极使用经过核证的评价经验，提高开发署的运作方式。

图 1
开发署评价政策采用的初步变革理论



5. 下表 1 对照图 1 所示成果链对取得的进展进行了详细评级。评级结果是，在开发署明显大幅促进发展实效的预期影响方面进展甚微。这是因为推动该影响的必要主要成果，即“提高方案和活动透明度和问责”没有实现。事实上，考虑到全面评价制度当前面临的各种挑战，即使开发署想要提高其总体发展实效，但也缺乏证明已推动发展实效的可信证据。本审查认为这是一个非常严重的问题，执行局不妨通过紧急修正评价政策和认真监测这一修正带来的结果解决这一问题。

表 1
大致对照评价政策变革理论对进展情况进行评级

理论要素(见上图 1)	评级	简要分析
1. 独立评估办公室就评价问题和评价方法提供有效指导	++	主要通过分发手册和网上培训等方式广泛提供指导，但并未满足整个组织所有工作人员的需求，而且指导需要更新。
2. 通过独立评价解决相关问题并提供及时、可靠、可信的信息。	++	在及时完成方面存在一些问题，但评价质量总体上是好的。
3. 集中评价对开发署对发展成果的贡献进行可信评估。	++	评价质量总体良好，但对开发署贡献分析的可信性存在一些问题，借鉴管理层的成果信息更是如此。
4. 独立评估办公室有效地向本组织宣传评价结果和获得的经验	++	报告质量良好，但宣传产品的范围有限，这限制了获取生成的信息。
5. 各方案单位提供可信的分散评价，促进问责和吸取经验教训	---	超过 38% 的分散评价咨询人报告称，开发署管理人员采取的某种令人无法接受的措施影响了评价的独立性。因此，评价资源中心信息的可信度严重受损，降低了问责或吸取经验教训的价值。还存在价质量参差不齐、低成本评价广泛存在问题。
6. 管理人员认真对待集中评价和分散评价，并在深思熟虑后做出管理回应	-	因为许多分散评价报告是折衷的结果，因此，管理层的回应质量和效用差别很大。对集中(独立)评价的回应通常更为集中。
7. 方案管理人员执行管理部门的回应	-	高级管理一级制定了协调一致的措施，监测管理部门应对的执行情况。但分散评价的质量不可靠降低了这一进程的效力。

理论要素(见上图 1)	评级	简要分析
8. 假定:集中评价和分散评价利用成果管理制的系统数据	++	尽管成果管理系统(包括注重成果的年度报告)中数据的可用性和质量方面存在一些问题,但已将这些数据广泛用于集中评价和分散评价。
9. 假定:从成果管理制和评价系统中获得的结果是一致的,相辅相成的	+	结果大致相符,但许多分散评价报告已经过修订,提交的结果是管理层(特别是在国家办事处一级)期望看到的。
10. 提高方案和活动的透明度和问责	---	分散评价系统的重大缺陷严重削弱了透明度和问责。
11. 有效的评价文化	-	由于分散评价系统存在缺陷以及过度反对一些独立评价,因此在总体评价文化方面取得的进展有限。
12. 完善开发署的系统和做法	-	尽管高级管理层为更好地利用评价结果进行了长足努力,但一直无法有效确保其分析所依据数据的准确性。
13. 加强活动的开展及提高活动的绩效	-	没有足够的公正评价证据核实所报告的任何改进。
14. 开发署为发展做出了更多贡献	-	没有足够的公正评价证据核实开发署做出了更多贡献。
15. 影响驱动因素:决策者积极利用评价的经验教训,改变和改善本组织管理方案的方式	-	尽管有决策者试图利用经验教训改善方案,但分散评价系统缺乏真实性使得无法在整个开发署开展。

四. 开发署评价政策的主要问题和今后的备选方案

6. 本报告提出的调查结果和分析显示,有几个问题执行局不妨通过订正评价政策加以解决。审查建议修订后的政策应提供办法,克服所有已查明的挑战。但这些挑战的困难程度不同,因此需要执行局的投入和关注程度也不同。因此,这一节主要关注有关分散评价问题,这是开发署目前在形成和利用有效评价文化方面面临的最大问题。如果不解决这些问题,执行局将依然缺乏一套有关实地绩效的协调一致的独立、公正证据,也无法据此做出决定。

7. 审查就解决所提出的这些问题提出了初步(有时是备选)办法。审查建议执行局成员间以及在执行局、开发署管理层和独立评估办公室之间通过详细讨论完善这些办法,使它们更加具体。

问题 1. 分散评价的总体可靠性受到严重影响

挑战

8. 政策规定由开发署署长和高级管理层负责确保分散评价的质量，独立评估办公室在“质量评估”方面的作用有限。这一做法行不通。这一做法失败的表现如下：

(a) 管理层(主要是在国家办事处一级)大量² 干预“独立”咨询人的评价报告；

(b) 因此，质量评估系统所评级报告的真实性不详，且真实程度不同；

(c) 评价资源中心汇编的信息来自不可靠报告，因此是不准确的；

(d) 因此，评价资源中心关于分散评价的大部分分析和报告依据的都是不准确或不可靠的数据；

(e) 管理层向执行局提供的有关分散评价的报告依据的是从存在严重缺陷数据库中提取的综合数据，这无助于执行局向开发署提供指导。

9. 要想了解如何最好地改变这一情况，就必须确定哪些因素造成了这一情况，包括有哪些原因，或有哪些导致出现这一情况的体制环境。这些原因可概述如下：

(a) 各方，特别是国家办事处往往认为最后评价没有用；

(b) 极少有进行评价的工作人员接受过这方面的专门培训；

(c) 工作人员高度依赖一本手册，这本手册很有帮助，但是不全面，而且已经过时；

(d) 有关人员常常不理解也不支持分散评价独立性的概念；

(e) 认为评价破坏筹资努力；

(f) 管理人员认为，一些咨询人对开发署有偏见或不了解开发署；

(g) 用于高质量评价的资源往往不足；

(h) 许多国家对应机构不重视评价；

(i) 许多管理人员没有优先考虑发展评价文化(尽管有些管理人员已将此列为优先事项)；

(j) 政策规定由开发署管理机构负责保证分散评价的质量，但这一点并没有按预想得到实施：只有独立评估办公室对报告进行质量评估；

² 约 38% 答复了调查的咨询人(254 名答卷人；回复率为 50.8%；误差 3.6%)报告遭到过管理层某种形式的不可接受的干扰。

(k) 咨询人和客户之间没有解决争端的有效机制;

(l) 由于之前存在争议,再加上资源不足,无法满足职权范围要求,使得许多咨询人未参与开发署分散评价投标工作,导致可选的咨询人有限。

10. 在上述各种促成因素背后是各级工作人员认为,未制定“棍棒或胡萝卜”政策鼓励管理人员的行为有利于独立评价(或任何评价)。每个管理人员都可以自由地对评价采用个人方法,因为业绩评估很少奖励这一领域的积极业绩或惩罚不良业绩。在此方面通常只是评判相对资历较浅的“监测和评价专家”,但他们不够资深,无法促使他们所服务办事处的总体方法实现重大变革。

建议的方法³

11. 评价政策应要求管理层采取和实施有效的分散评价质量保证体系;由独立评估办公室进行核查,对不达标的单位进行惩罚。⁴ 执行局的支持方式应是更新和补充准则,并评估良莠不齐的大量低预算评价给已有的管理信息带来的附加值。

12. 目前,各区域局局长负责“确保评价(计划和)做法的质量和执行”。尽管这些局需要“监测和评价工作人员”,但却在逐步减少使其能够履行这一职能的工作人员。此外,独立审查得到的文件表明,根据开发署的新结构审查,用于评价的时间将比目前用于评价的时间甚至更少。每年约有 300 个分散评价需要五个区域的有效质量保证。鉴于“质量保证”需要首先应对职权范围和可评价性分析的质量等问题,之后按照步骤直至编写最后报告,因此,审查得出结论认为,如果各区域实地等同于全职评价顾问⁵ 少于 10 人,那么开发署管理当局就不能履行这项职能。这些顾问或者由区域工作人员充当,或者由独立评估办公室和各区域联合供资(假设有更多资源)。需要进行矩阵管理,主要责任由独立评估办公室承担,而某种程度的运营管理由各地区提供。

13. 有了必要的评价顾问后,下一步将是独立核查评价报告。在此需要两项措施,其中之一是需要更多资源。第二项措施是开发署-全球环境基金办公室通过一个最终评价进程。这项措施要求独立咨询人将报告草稿同时提交给委托单位(通常是国家办事处)和区域评价顾问。这将使顾问之后可以核实草稿和最后定稿间是否出现了巨大差异。若出现了这种差异,顾问将启动后续调查。第二项措施将需要(或代表)独立评估办公室对一定比例(审查建议 10%)的项目进行后续评价考察。该考察

³ 审查最后报告提出了备选方案,但本执行摘要仅列出了审查建议的做法。

⁴ 一些国际金融机构在该系统中采用的一种惩罚方式是给违反条例或标准的管理单位发“警告红旗”。这些警告可以累积并在今后的执行情况评估中加以考虑。

⁵ 根据“监测和评价专家”的监测和评价身份是全职还是非全职,并根据他们用于评价,而不是监测的时间百分比计算所需的这些专家的人数。

将通过重新评估项目，包括项目评级，⁶ 核实报告的质量。因为只有在提交了最初分散评价后才会通知核查哪些项目，这将给评估进程增加一些“未知情况”，通过确保各区域的核查范围具有可比性，对发现提交的评价报告经过不当影响或修订的区域和国家办事处发出“警告红旗”，加强评估进程。⁷

14. 由于每年约进行 300 个分散评价，因此，每年约需要进行 30 次后续考察。由于这些考察主要是进行核查，因此，时限应比最初评价的时间短，但这些考察加一起仍需要额外资金。

15. 为支持上述控制措施，独立评估办公室和开发署需合作更新和扩大《发展成果规划、监测和评价手册》，特别是在性别平等、建立和维持评价伙伴关系、联合评价，保护独立性以及管理人员评价操守等方面。在更新和扩大《手册》的同时，应提供更多在线和当面评价培训的机会。独立评估办公室还应仔细阅读联合国评价小组开发的各种指导产品，以便将他们的指导信息编制成适合开发署的具体应用并加以推广。

16. 审查表明，如果这些措施都能到位，包括有足够工作人员在岗(而不是仅有“职位”)，那么，开发署经过一段时间将能够制订一个目前没有的可信的分散评价系统。这将使管理部门和执行局能够获得关于国家一级业绩的经过核查的基本数据，以便有效履行各自职责。

问题 2. 关于一些独立评价的不合理争议

挑战

17. 提出的大多数独立评价都没有大问题。但一些评价(过去 3 年中 22 个发展成果评估中的 5 个)已卷入不合理争议，造成最后报告拖延至多一年。审查指出，其中 3 起争议来自一个国家办事处，且最终涉及到相关区域局，而另外两起争议是独立评估办公室程序造成的延误。

18. 这一问题表现为：

- (a) 时间表中断和被拖延；
- (b) 没有为管理目的按计划交付评价；
- (c) 独立评估办公室的资源偏离其他既定工作；
- (d) 一些管理人员不愿意将外部观点纳入工作，降低了独立评价的附加值；
- (e) 另外还有声誉风险，因为争议已经人尽皆知。

⁶ 这项工作将采用的程序与世界银行项目业绩评估体系中使用的一些程序类似。尽管世界银行对之前项目完成情况报告和项目完成情况报告审查(案头研究)评估的 20-25% 的项目开展项目执行情况评价，但目前的审查建议开发署进程仅评估 10% 的项目。理由是，事实上，开发署的分散评价打算由独立专家完成，而世界银行的执行完成情况报告可以由项目实施方完成。

⁷ 审计和调查处已能够向特定办事处提出“审计问题”，之后这些办事处需解决这些问题。

19. 造成挑战的因素是：

- (a) 管理人员认为独立评价的质量不定(可能是如此)
- (b) 管理人员认为，一些咨询人对开发署有偏见或不了解开发署(可能是如此)；
- (c) 评价不如管理人员所期望的正面、积极，管理人员因此认为评价破坏了“沟通战略”；
- (d) 不同立场的利益攸关方对中央评价独立性的概念的解释不同；
- (e) 一些管理人员将提出“消极”调查结果视为不忠。

20. 独立评估办公室同意其也是导致出现这一情况的原因之一，并已采取措施加强办法。其中包括重新平衡独立评估办公室专业工作人员和咨询人的作用，更加强调对这些评价的直接管理。此外，还通过实施时间表制定了加快进度程序。

建议的方法

21. 如管理方面不及时反馈，那么独立评估办公室应获得明确授权，能够根据预先确定的时间表推进独立评价进程的所有步骤。如管理方面持续反对评价结果或建议，那么应将此作为保留列入管理部门的回应，不应允许推迟国家讲习班的举行或推迟印发报告。应通过“警告红旗”系统惩罚阻碍独立评价开展或完成的开发署各单位(如国家办事处)。

22. 这一办法可以在不增加费用的情况下消除在独立评价方面的长期争议和拖延。一个潜在的不利因素是，如果国家合作伙伴(或受专题或全球评价影响的其他方面)收到的是开发署管理层未接受的独立评估办公室报告，那么它们将获得矛盾的信息。相反，这也可以成为一种优势，因为它会迫使管理层迅速作出建设性回应，从而使得按时完成评价。向被认为阻碍该系统的主管人员出示警告红旗也可以加大压力。开发署各级工作人员和管理人员向审查提供了大量证据，他们认为降低评价职能效力的一个关键因素是，该系统对于行为不当者缺乏惩处措施。

问题 3. 在对于独立评估办公室业务独立性至关重要问题方面的政策太过模糊

挑战

23. 政策在对独立评估办公室业务独立性构成严重挑战的若干问题上措辞模糊或保持沉默。问题表现在：

- (a) 独立评估办公室主任的任期“至多四年”，因此，管理人员可以规定任期少于四年；
- (b) 没有具体规定可能延长任期情况；
- (c) 执行局主席不主持也不是主任或副主任任命委员会成员，任命委员会多数人员为管理人员；

(d) 尚不清楚如何评估独立评估办公室主任的业绩，而明确如何评估对于质量保证和对该职位的信心至关重要；

(e) 政策没有具体规定，一旦与管理人员存在严重分歧，独立评估办公室主任在必要时可以直接向执行局报告。

24. 造成这种情况的因素似乎包括：

(a) 假设独立评估办公室和开发署管理部门之间的和谐关系还未实现；

(b) 措辞含糊所产生的影响尚未得到充分认识。

(c) 还未强调通过适当业绩评估确立独立评估办公室主任公信力的重要性。

建议的方法

25. 审查建议执行局修订政策，规定执行局在独立评估办公室主任征聘程序方面的主导作用；该员额任期(依据业绩评估)；延期程序和延期期限；主任必要时直接向执行局报告的权力。这些措施将加强该办公室的结构性独立性，使其成为名副其实的独立评估办公室。

问题 4. 相关基金和方案的独立评价没有保障

挑战

26. 政策规定，独立评估办公室是开发署独立评价“评价职能的管理机构”，并具体规定“相关基金和方案的评价单位是其组织评价职能的管理机构”。这表明，不同评价实体的作用在某些方面是平行的。但尽管该政策指出，独立评估办公室“根据国际最佳评价标准进行独立评价”，但评价单位只是“管理和开展评价”。审查的结论是，虽然评价单位由于各机构的管理系统小，在结构上不能享有和独立评估办公室同等程度的独立性，但也应要求它们根据“国际最佳评价标准”进行运作。

27. 这一挑战的表现包括：

(a) 由于相关基金机构的规模小，因此，评价单位在组织结构上无法与管理部门分开；

(b) 独立评估办公室和相关基金的评价单位都界定为是“评价职能的管理机构”；

(c) 因此，评价单位的地位与独立评价办公室的地位是平行的。他们的独立评价是自我核实，但与独立评估办公室相比职等更低(独立评估办公室是 D-2，它们是 P-4)；

(d) 这些单位的规模小就意味着，他们所有评价工作都是由咨询人完成的，因此它们有效地进行了分散评价；

(e) 政策不能保障相关基金和方案所开展评价的独立性和公正性(尽管目前是可以接受的)。

28. 造成这种情况的因素包括：

(a) 政策中未规定为在小型组织实现独立性而进行结构性调整；

(b) 与独立评估办公室相比，各相关基金和方案的评价专长和主管的权力都被定在了更低水平，而且没有补偿性质量保证；

(c) 独立评估办公室和相关基金和方案评价单位之间的合作程度不够，无法制定协调一致的质量方法。

建议的方法

29. 审查指出，相关基金和方案的评价单位的发展方向极为不同，并建议政策应纳入以下新的灵活办法：应要求评价单位或者向独立评估办公室提交独立评价，以确保质量(以确保遵守具有可比性的“国际最佳评价标准”)；或在管理和报告独立评价时与独立评估办公室直接合作。与这些措施有关的是，还建议政策应要求独立评估办公室更系统地关注相关基金和方案的所有独立评价对开发署成果的贡献。

30. 审查认为，拟议的质量保证措施很合时宜，因为从根本上来说，独立评估办公室主任级员额使该办公室能够在国际最佳评价标准层面运作，而在小型的相关基金和方案的评价单位不太可能任命这一职等职位。这一替代方案被视为是在评价问题上的一种管理协作形式，各基金和独立评估办公室之间可正式商定协作形式，并进行定期审查。例如，这种协作可以是，某一基金或方案资助独立评估办公室(全职或非全职)的一个职位，让该职位主要负责自己机构的评价工作，或只是委托该办公室开展(或联合开展)有关基金或方案的主要评价。

31. 审查的结论是，这些措施将核实这些单位评价的独立性，确保政策涉及的评价实体的标准具有可比性。

问题 5. 该政策概念的相关性需要更新

挑战

32. 政策中有一节是关于审查发现存在问题的定义(第 5 和第 6 段)。这些定义主要摘自经济合作与发展组织 2000 年印发的评价和成果管理关键词汇。根据评价界过去十年讨论和辩论情况，一些定义并不完全符合当前的思路，而另外一些重要专题却没有提及。

建议的方法

33. 审查建议，将政策的定义一节替换为更宽泛的案文，并指出，独立评估办公室将通过业务手册和其他适当手段，依据国际最佳标准，定期更新和分发当前的评价专题和定义。