



联 合 国

对联合国系统各组织分散评价的审查

联合检查组的报告

艾琳·克罗宁编写



对联合国系统各组织分散评价的审查

联合检查组的报告

艾琳·克罗宁编写



联合国 • 2026 年，纽约

项目组

艾琳·克罗宁，检查专员

Kamolmas Jaiyen，评价和检查干事

Irina Ciocirlan，评价和检查干事

Eleyeba Bricks，研究助理

Asyanadiva Najwa Fazrany，实习生

目录

	页次
简称和缩略语	iv
一. 导言	1
A. 目标、范围和预期影响	1
B. 方针和方法	1
二. 集中评价和分散评价职能	3
A. 定义	3
B. 评价活动(2021-2025 年)	6
三. 联合国系统各组织的集中评价状况	8
A. 中央评价实体的结构布局	8
B. 职能独立性和行政独立性	8
C. 资源	9
四. 分散评价的体制框架	11
A. 分散评价的体制定位	11
B. 分散评价的规范性监管机构	11
C. 进行分散评价的体制安排	12
五. 分散评价的实施	16
A. 分散评价的需求和驱动因素	16
B. 分散评价的可信度	17
C. 分散评价的使用和效用	20
D. 进行分散评价的效率	23
六. 分散评价的未来展望	25
附件	
各参加组织根据联合检查组的建议应采取的行动概览	27

简称和缩略语

AI	人工智能
粮农组织	联合国粮食及农业组织
全球环境基金	全球环境基金
原子能机构	国际原子能机构
国际民航组织	国际民用航空组织
劳工组织	国际劳工组织
海事组织	国际海事组织
国贸中心	国际贸易中心
国际电联	国际电信联盟
联检组	联合检查组
监督厅	联合国秘书处内部监督事务厅
艾滋病署	联合国艾滋病毒/艾滋病联合规划署
贸发会议	联合国贸易和发展会议
开发署	联合国开发计划署
教科文组织	联合国教育、科学及文化组织
环境署	联合国环境规划署
人口基金	联合国人口基金
人居署	联合国人类住区规划署
难民署	联合国难民事务高级专员公署
儿基会	联合国儿童基金会
工发组织	联合国工业发展组织
毒罪办	联合国毒品和犯罪问题办公室
项目署	联合国项目事务署
近东救济工程处	联合国近东巴勒斯坦难民救济和工程处
可持续发展合作框架	联合国可持续发展合作框架
联合国世旅组织	世界旅游组织
妇女署	联合国促进性别平等和增强妇女权能署
万国邮联	万国邮政联盟
粮食署	世界粮食计划署

世卫组织	世界卫生组织
知识产权组织	世界知识产权组织
气象组织	世界气象组织

一. 导言

1. 应各参加组织的请求，本次审查纳入了联合检查组(联检组)2025 年工作方案。审查主题主要属于 2020-2029 年联检组战略框架中问责和监督职能的专题领域。

2. 本次审查扩展了联检组以往的工作，特别是联检组 2014 年针对该专题进行的题为“联合国系统评价职能的分析”的最新审查，联检组在该审查中提出了一项专门针对分散评价职能的正式建议。¹

A. 目标、范围和预期影响

3. 由于过去十年各参加组织都增加了分散评价，因此必须审查投入其中的资源和努力是否得到高效率、有成效的利用。在此背景下，本次审查评估了如何加强分散评价才能产生可信、有用和及时的证据，促进建设更强有力的成果和循证文化，并加强组织学习和问责。本次审查审视了分散评价作为一项职能的体制安排、资源、能力和效率，以及评价结果和评价证据在问责和学习方面的有效性、可信度和使用情况。本次审查仅涵盖符合联合国系统评价规范和标准所述评价定义的评价活动。²

4. 本次审查审视了 28 个联检组参加组织 2021 年至 2025 年的政策和实践。对于联合国秘书处而言，这项全系统审查不能涵盖秘书处所有实体(超过 70 个)的评价实践。对六个选定实体³进行了审查以提供实用见解，而联合国贸易和发展会议(贸发会议)、联合国环境规划署(环境署)和联合国人类住区规划署(人居署)则作为联检组参加组织分别接受了审查。本次审查还涵盖了集中评价的状况，但没有那么深入，因为不能孤立地评估分散评价。虽然本次审查审视了消除 2014 年联检组报告确定的关键差距方面取得的进展，但其主要重点是 2021 年以来的事态发展。鉴于目前影响各参加组织的不稳定财务环境，本次审查还力求总结集中评价和分散评价职能的正在发生的变化和预期的变化，确保其审查结果和建议在 2026 年底发布报告时仍然具有相关性。

B. 方针和方法

5. 本次审查是根据联检组章程及内部程序进行的。根据联检组的协作方法，各参加组织在审查的各个阶段提供了意见。为了确保审查结果的有效性和可靠性，检查专员使用了不同来源的一系列定性和定量方法。支持本次审查的数据和证据是在 2025 年 5 月 1 日至 2026 年 3 月 31 日间使用以下工具收集的：

¹ JIU/REP/2014/6，建议 9(81%的参与组织接受了该建议，其中 95%的组织报告称已执行该建议)。

² 联合国评价小组，“评价规范和标准”，2016 年。可查阅 https://www.unevaluation.org/uneg_publications/uneg-norms-and-standards-evaluation-un-system。

³ 发展协调办公室、非洲经济委员会、欧洲经济委员会、亚洲及太平洋经济社会委员会、人道主义事务协调厅和联合国人权事务高级专员办事处。

- 对相关文件进行广泛的案头审查和深入的比较分析，包括评价政策和其他相关政策、年度报告、手册和准则、评价职能的同行审议报告以及抽样分散评价报告
- 通过一份机构调查问卷，收集所有参加组织的集中评价和分散评价数据
- 对 215 名个人进行 110 次半结构化访谈，其中包括各参加组织在总部、区域和国家各级的评价、方案、伙伴关系和高级管理职能干事以及捐助方代表
- 对 20 个国家 13 个参加组织的国家办事处进行一次在线调查，借鉴其最近在联合国可持续发展合作框架(可持续发展合作框架)、国家方案和项目评价方面的经验。调查收到 146 份回复，回复率约为 60%，令人满意。

6. 主要局限是大多数组织没有系统地跟踪分散评价，包括其数量、成本或用途。这限制了跨组织的比较且导致难以衡量其整体范围。访谈和调查被用来帮助消除这些差距。

7. 检查专员谨向通过访谈、机构调查问卷和调查做出贡献的联检组参加组织和其他利益攸关方的所有官员表示感谢。受访者的私密性得到适当保护。

8. 根据《联检组章程》第十一条第 2 款，对报告草稿进行了联检组检查专员内部同行审议以保证质量，并利用联检组的集体智慧检验各项建议。本报告是在考虑了已收到的所有评论后定稿的，但对其内容的责任完全由作者承担。

9. 在履行专业和道德义务方面，联检组团队在本报告的规划、执行和起草阶段未受到任何可能干扰其独立性、公正性、中立性或职业操守的外部影响。

10. 本次审查产生了两项产出：(a) 本报告，以联合国六种正式语文提供；(b) 一份补充分析，以 [JIU/REP/2026/5\[Expanded report\]](#) 文号发布，载有本报告的基本证据。

11. 本报告载有六项正式建议，其中一项提给立法机构，五项提给行政首长。为了促进建议的执行，附件包含一个表格，具体说明每项建议是否需要立法机构或行政首长采取行动。除了正式建议外，还有 11 项非正式建议(粗体)提供了额外建议和(或)补充正式建议。

二. 集中评价和分散评价职能

12. 本节列出了本次审查期间评价活动的关键定义和概述，为后续章节的分析提供了框架。

A. 定义

13. 28 个联检组参加组织采取不同的评价方法，使用不同的术语和定义。由于相同的术语在不同的组织中可能具有不同的含义，在本次审查中使用以下定义。

14. “集中评价”系指由中央评价机构或部门规划并开展的评价。然而，对于什么是中央评价机构并没有商定的定义，而且各参加组织的此类机构在规模、形式和任务方面差异很大。根据主流做法，本报告采用以下工作定义：

中央评价机构或部门是一个正式成立的组织实体，由评价政策或同等机构框架管理，有明确的任务授权以及进行涵盖所涉组织任何领域的评价所需的资金和人力资源。它与其他组织实体的区别在于，评价是其首要目的，其规模可大可小，可以只有一名评价干事，也可以是一个专门的评价部门，根据组织的需要，可以作为独立实体设立，⁴ 也可以嵌入其他组织职能。并非所有中央评价实体在结构和(或)职能上都独立于管理层。

15. 根据其任务规定，中央评价实体可以履行以下一项或多项职责：

- 为理事机构或立法机构提供独立保证和监督(内部审计师协会三线模型中的第三线职能)⁵
- 生成评价证据，为高级管理层决策提供参考
- 支持跨方案或跨职能的组织学习和问责制
- 建设评价能力并促进组织内部的循证决策和成果管理制文化。

16. 分散评价没有商定的定义。常用的广义定义系指由中央评价机构或部门以外的组织实体进行的评价。⁶ 然而，这样的定义没有明确组织内部分散评价的独特职能和目的，也没有将分散评价与集中评价进行明确区分。

⁴ 独立是指独立运营的机构或部门，而不是嵌入另一个司局或处室。

⁵ 内部审计师协会的三线模型区分业务管理职能(第一线)、支持、监测和指导管理的职能(第二线)和独立保证职能(第三线)。虽然该模型主要为治理、风险管理和内部控制而开发，但它为考虑分散和集中的评价职能在组织问责安排中的定位提供了一个有用的分析框架。联合国系统行政首长协调理事会在 CEB/2014/HLCM/14/Rev.1 号文件中建议采用内部审计师协会以前的框架，即三道防线模型，并根据联合国系统的风险管理、监督和问责安排进行调整。

⁶ 包括联合国评价小组最近关于分散评价的出版物“联合国系统分散评价职能的评估框架”(2025 年)。

17. 通过对各参加组织进行分散评价的政策和实践进行分析，提出以下更全面的分散评价定义：⁷

分散评价是由中央评价实体以外的组织单位规划、资助、委托和管理的评价。这些单位可在所涉组织的任何一级运作，包括总部各部门、区域办事处、国家办事处以及方案或项目管理结构。与中央评价实体不同，这些单位不承担全组织评价任务；其范围仅限于各自的责任领域，包括但不限于项目、方案、群组、主题和机构间活动。

分散评价在部门管理层的授权下得到管理和批准，因此由那些在结构上和(或)行政上不独立于所评价的活动、方案或项目的职能部门进行委托和管理。分散评价通常由外部咨询顾问进行，外部咨询顾问应行使专业独立性和公正性，并遵守适用的组织评价标准。

分散评价是一个组织监测和评估职能的组成部分，理想情况下应位于内部审计师协会三道防线模型的第二线。其主要目的是支持业务学习、管理和外部问责以及为方案和项目改进生成证据，而不是提供通常由中央评价实体在机构层面履行的独立保证和监督。

18. “分散评价”一词已纳入 14 个组织(国际民用航空组织(国际民航组织)、国际劳工组织(劳工组织)、国际电信联盟(国际电联)、联合国开发计划署(开发署)、联合国环境规划署(环境署)、联合国教育、科学及文化组织(教科文组织)、联合国人口基金(人口基金)、人居署、联合国难民事务高级专员公署(难民署)、联合国儿童基金会(儿基会)、联合国近东巴勒斯坦难民救济和工程处(近东救济工程处)、联合国促进性别平等和增强妇女权能署(妇女署)、世界粮食计划署(粮食署)、世界卫生组织(世卫组织)和世界知识产权组织(知识产权组织))的评价政策。联合国项目事务署(项目署)和万国邮政联盟(万国邮联)开展分散评价，但没有评价政策。一些组织使用不同的术语来指代与上述分散评价定义相对应的评价活动，包括在预算负责人的责任范围中进行的评价、自我评价、项目管理评价和管理层主导的最终审查。

19. 根据上述定义，20 个组织既开展集中评价也开展分散评价，5 个组织只开展集中评价，2 个组织只开展分散评价，1 个组织不开展任何评价(见表 1)。在联合国秘书处各实体中，可持续发展合作框架评价被归类为分散评价，并由联合国秘书处负责审议，因为其实体之一——发展协调办公室⁸负责监督这些评价。可持续发展合作框架评价是联合国秘书处内以最正式方式确立的分散评价类型。

⁷ 关于分散评价职能在组织内的定位，联检组的立场与联合国评价小组 2025 年出版物“联合国系统分散评价职能的评估框架”中的立场不同。

⁸ 联合国评价小组针对联合国可持续发展合作框架的评价准则(2021)表明，可持续发展合作框架评价是分散的，由发展协调办公室提供监督和技术支持。可查阅 www.unevaluation.org/uneg_publications/guidelines-evaluation-united-nations-sustainable-development-cooperation。

表 1
联检组参加组织评价职能概览

组织	设有集中评价部门	存在分散评价
粮农组织	是	是
原子能机构	是	否
国际民航组织	是	是 ^a
劳工组织	是	是
海事组织	是	否
国贸中心	是	是
国际电联	是 ^b	是
联合国秘书处	是 ^c	是 ^d
艾滋病署	是	否
贸发会议	是	否
开发署	是	是
环境署	是	是
教科文组织	是	是
人口基金	是	是
人居署	是	是
难民署	是	是
儿基会	是	是
工发组织	是	是
毒罪办	是	否
项目署	否	是
近东救济工程处	是	是
联合国世旅组织	否	否
妇女署	是	是
万国邮联	否	是
粮食署	是	是
世卫组织	是	是
知识产权组织	是	是

组织	设有集中评价部门	存在分散评价
气象组织	是	是
共计	25	22

资料来源：联检组机构问卷。

^a 分散评价已在政策中阐明，但尚未付诸实施。

^b 从 2026 年开始。

^c 监督厅是联合国秘书处内部(包括评价工作)的内部监督机构，也是秘书处唯一的全组织第三线评价职能部门。然而，许多联合国秘书处实体，包括区域委员会和总部以外办事处，都有自己的评价部门，这些部门独立于监督厅，但不履行第三线职能。

^d 仅指由发展协调办公室协调的可持续发展合作框架评价。

B. 评价活动(2021-2025 年)

20. 在本次审查期内，25 个组织的中央评价实体⁹ 共进行了 1 381 项集中评价。这些机构的支出总额(包括人员)为 423 725 345 美元，约占组织支出的 0.19%。¹⁰ 22 个组织进行的分散评价的数量没有得到很好的跟踪：在有可用数据的 18 个组织中，在本次审查期间共完成了 3 703 项分散评价，只有 12 个组织¹¹ 可以提供支出数据，金额为 382 902 998 美元，约占其组织支出总额的 0.24%。

21. 根据已知数字，平均每年进行约 1 000 项评价，其中分散评价的数量始终比集中评价至少多出两倍。已报告的分散评价支出似乎与集中支出相当，但可能被严重低估，因为许多组织不跟踪其支出，而且可用数据通常不包括工作人员管理该流程的时间。有关具体组织的数据，见下文图一和二以及补充分析文件(JIU/REP/2026/5[Expanded report])。

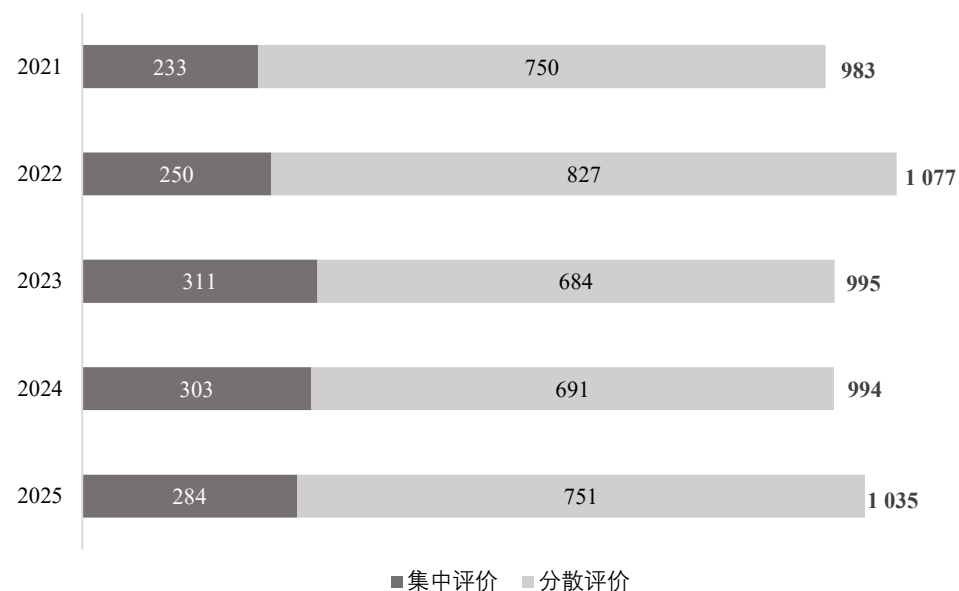
⁹ 就联合国秘书处而言，仅包括监督厅的评价报告。

¹⁰ 基于 2021 年至 2024 年间 22 个中央评价实体的总支出。联合国秘书处被排除在计算之外，因为本次审查没有列出除监督厅外的评价实体的支出。贸发会议也被排除在外，因为在联合国系统行政首长协调理事会记录中其组织费用并未与联合国秘书处的组织费用分开。2025 年的组织费用尚未提供。

¹¹ 劳工组织、国际贸易中心(国贸中心)、联合国秘书处(可持续发展合作框架评估)、开发署、教科文组织、人口基金、儿基会、妇女署、粮食署、世卫组织、知识产权组织和世界气象组织(气象组织)。国际民航组织和国际电联在本次审查期内尚未进行分散评价。

图一

联检组参加组织的评价数量(2021-2025 年)

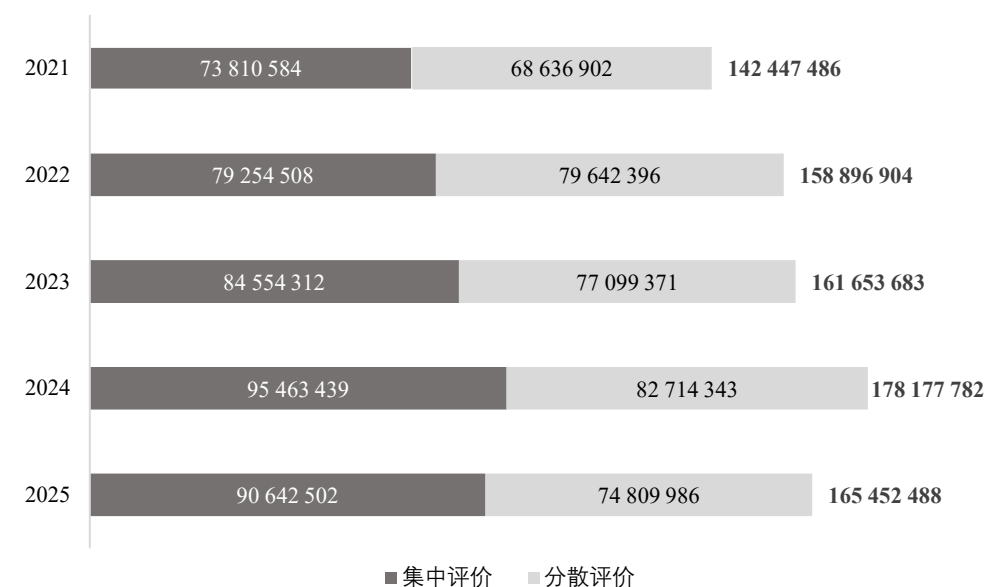


资料来源：联检组机构问卷。

图二

联检组参加组织的评价支出¹² (2021-2025 年)

(美元)



资料来源：联检组机构问卷。

¹² 集中评价的支出金额包括评价活动和人事费。

三. 联合国系统各组织的集中评价状况

22. 自 2014 年以来，联检组参加组织的中央评价实体发生了很大变化。尽管本次审查的重点是分散评价，但中央评价实体仍然是塑造和支持该职能的关键。鉴于各项安排不断演变，本报告的审查结果反映截至 2026 年初的情况。本节将当前安排与 2014 年联检组审查的结果进行比较，并为分散评价分析提供背景。¹³ 按组织列出的数据见补充分析文件(JIU/REP/2026/5[Expanded report])。

A. 中央评价实体的结构布局

23. 随着各组织继续调整其职能以适应不断变化的需求，中央评价实体的结构布局仍然不稳定。在联合国艾滋病病毒/艾滋病联合规划署(艾滋病署)、贸发会议和难民署，曾经安排在行政管理或方案结构内的中央评价实体如今已成为独立的机构。更多组织将中央评价实体作为内部监督办公室内部的独立单位运营，其目的要么是为了加强一体化结构内的自主性(国际原子能机构(原子能机构)、知识产权组织)，要么是为了节省成本而合并以前的独立机构(联合国工业发展组织(工发组织)、妇女署)。2014 年，国际电联、项目署、世界旅游组织(联合国世旅组织)和万国邮联都没有中央评价实体，目前只有国际电联正在建立一个中央评价实体。如前所述，¹⁴ 检查专员鼓励项目署和万国邮联考虑设立一个负责监督评价并提供政策指导的职能部门。

24. 联检组 2014 年审查发现，结构布局主要以独立性、效率、相关性和效用为指导；最近的变化表明财务考虑现在发挥更大的作用。在某些情况下，重组先于评价政策的相应更新。例如，2025 年世卫组织将其独立的评价办公室重新纳入监督职能，而人居署以前的独立办公室如今在行政意义上向机构管理司报告工作。

B. 职能独立性和行政独立性

25. 本次审查采用与 2014 年审查相同的关键参数，评估了中央评价实体的行政独立性和职能独立性。由于独立性存在于一个范围内，这些参数仅表示相对定位，而非最终结论。

26. 职能独立性已得到加强，现在有 15 个中央评价实体能够独立于管理层确定其评价议程，而 2014 年只有 11 个。在 24 个中央评价实体中，除国际海事组织(海事组织)和国际贸易中心(国贸中心)外的中央评价实体外，其他都可以在没有管理层参与的情况下发布评价报告和其他产出。¹⁵

¹³ 对于联合国秘书处而言，本节仅提及监督厅，它是秘书处的中央评价部门。

¹⁴ 在 JIU/REP/2017/4 号文件中，检查专员鼓励万国邮联制定评价政策和工作计划；JIU/REP/2018/3 号文件承认，项目署没有评价政策或职能部门，因为其评价由合作伙伴管理。然而，情况已不再如此。

¹⁵ 然而，人居署评价部门报告称，这一点在实践中尚未得到充分执行。

27. 理事机构在挑选和任命评价负责人方面也发挥了更大的作用。12 个独立的中央评价机构中有 8 个让其理事机构参与任命负责人。其中，粮食署是唯一有政策规定由其执行局批准评价主任任命的组织。在其他七个组织——联合国粮食及农业组织(粮农组织)、联合国秘书处(内部监督事务厅(监督厅))、艾滋病署、开发署、人口基金、难民署和儿基会，此类行政首长的任命需要与理事机构协商。但总体而言，与 2014 年相比，如今有更多的组织在没有理事机构提供意见的情况下任命评价负责人。这种趋势主要出现在妇女署和世卫组织等组织中，这些组织已将以前独立的评价机构纳入更广泛的监督部门或其他单位。

28. 在设有中央评价实体的 25 个组织中，¹⁶ 只有 10 个对评价负责人实行任期限制和重新进入该组织的限制。**检查专员提醒各组织，此类限制减少了可能影响评价负责人公正性的实际或感知到的职业激励，从而加强了评价负责人的独立性。**

29. 关于行政独立性，17 个中央评价实体报告称，其对分配给它们的财务和人力资源的管理拥有控制权，包括工作人员和咨询顾问的选择以及经常性行政职能的批准。其余七个机构——粮农组织、国贸中心、艾滋病署、贸发会议、环境署、人居署和难民署都报告了这方面的局限性。

C. 资源

30. 自 2014 年联检组审查以来，整个联合国系统的评价政策在资源配置处理方面明显成熟，从只获得不稳定临时供资的资源不足的职能发展成为正规化预算规范。9 个组织在其政策中明确提及 2014 年联检组建议的组织支出 0.5%-3% 的基准，许多其他组织也制定了自己的百分比目标，主要用于集中评价和分散评价，从 2%-3% 的高端承诺(教科文组织、妇女署以及联合国毒品和犯罪问题办公室(毒罪办))到 1% 左右的中等目标(开发署、儿基会、人口基金、艾滋病署和世卫组织)和较低水平的调整(粮农组织、近东救济工程处和粮食署)，不一而足。政策也日益区分中央评价机构的核心供资与预算外项目中的强制性评价预算额度，而一些组织(粮农组织、劳工组织、儿基会和粮食署)已设立了用于不同目的的集合基金或信托基金，如为创造规模经济、避免交叉补贴以及促成专题评价、联合评价和影响力评价而设立的基金。

31. 然而，政策承诺和执行之间仍然存在重大差距。2021-2023 年，大多数组织都收到了所要求的拨款，通常低于正常水平，但从 2024 年开始，越来越多的组织报告了短缺问题，导致员额冻结或降级以及非员额资源大幅削减。¹⁷ 全系统的趋势是比例下降：集中评价的平均支出占比从 2012-2013 年的 0.30% 小幅下降至 2023-2024 年的 0.28%。虽然有 12 个组织出现增长，但许多实体的占比都有所下降，其中包括一些以前投入水平较高的机构。总体而言，情况令人担忧：更高的正常水平尚未转化为可预测的供资，使该职能的可持续性和独立性面临风险。

¹⁶ 就联合国秘书处而言，这里指的是监督厅。

¹⁷ 供资与员工职位无关。

32. 在人力资源方面，中央评价实体的规模和资历都有提高。2012 年至 2025 年间，25 个参加组织的人员配置增加了约 73%，全职人员从 197.95 名增加到 343.10 名。多个组织实现了大幅增长，其中包括开发署(从 24 名增加至 47 名)、儿基会(从 17 名增加至 26 名)、毒罪办(从 5 名增加至 11 名)和粮食署(从 19 名增加至 58 名)，其中一些组织还聘请了长期编外人员。¹⁸ 所有专业职等的员工人数都有所增长，但一般事务人员的配置水平在绝对数量上基本持平(虽然工作人员总数有所扩大)，这意味着一般事务人员在员工队伍中的占比下降。按比例计算，早期职业工作人员(P-1 至 P-3)增长最快，而中高级(P-4 及以上)和主任级工作人员增长较为温和。然而，这些趋势值得谨慎对待，因为尽管人员配置增加，但集中评价的数量仍然大致相似，而且数字工具、远程数据收集和不断变化的任务授权等因素使得人员数量不再是衡量充足性的完美指标。差异也持续存在，因为难民署报告称，2025 年评价专业人员的数量(12 人)远低于儿基会(21 人)和粮食署(34 人)等可比人道主义机构。

¹⁸ 集中评价实体的范围和任务差异很大，因此人员和预算的比较应考虑到这些差异。

四. 分散评价的体制框架

33. 2014 年联检组审查得出的结论是，大多数组织缺乏明确界定的分散评价体制框架，导致该职能不稳定。本节审视了自 2014 年以来在将分散评价制度化为一项关键组织职能方面取得的进展。

A. 分散评价的体制定位

34. 各组织对分散评价与集中评价不同的认识仍然不均衡，许多政策框架没有明确定义其目的或体制地位。这两项职能最终均服务于学习和问责，但方式不同、层次不同、用户也不同。这种区别是如何负责评价、如何为评价提供资源以及如何开展和使用评价的基础。若能明确阐述这种区别，就能向部门管理人发出信号，表明分散评价是其自身问责和学习责任的一部分，从而推动对评价结果的自主权和使用。

35. 在既进行集中评价又进行分散评价的 20 个组织中，13 个认识到分散评价的独特之处不仅在于由谁进行评价，通常还在于评价的范围、层级和预期用途：集中评价被定位为战略性的全组织活动，而分散评价则侧重部门管理人的业务和方案学习、问责和改进。虽然粮农组织、劳工组织、国际电联、环境署、儿基会和世卫组织的政策在结构或行政上对两者进行了区分，但它们并没有阐明各自目的、预期用户或对组织学习和问责的不同贡献的差异化理由，导致这些组织缺乏足够的政策基础，无法在两个职能范围内开展一致又独特、互补又充分自主的评价实践。

36. 除了认识到分散评价的独特价值之外，只有开发署、难民署、粮食署和世界气象组织(气象组织)在其问责或企业风险管理框架中明确指定，由部门管理层委托进行分散评价是第二线职能。工发组织将其分散评价(称为自我评价)视为第一线和第二线职能，因为它们可能由负责干事¹⁹或外部咨询顾问进行。检查专员建议尚未这样做的参加组织考虑将分散评价明确列为第二线职能，以明确是部门管理层而非中央评价机构主要负责委托开展这些评价，并协助保护中央评价实体独特的第三线作用。如果没有这种区分，两者之间的界限就会模糊不清，而中央评价机构的所谓独立性就可能受到损害。

37. 在劳工组织和环境署，分散评价更多是被视为一种委托开展的集中评价职能。其中央评价机构保留最终签署权、可以修改报告或将报告退回给外部评价人员，从而将自主权从部门管理层转移出去。环境署的安排类似于毒罪办的项目评估，但毒罪办并未将其归类为分散评价，考虑到中央评价机构实际上拥有最终成果，这是一种更为坦诚的立场。

B. 分散评价的规范性监管机构

38. 无论评价由哪个部门委托进行，明确指定分散评价的规范性监管机构对于确保所有评价遵循连贯一致的标准、质量、监督和问责都至关重要。监管机构

¹⁹ 其自身干事进行的评价可能不符合联合国评价小组对评价的定义。

应承担组织责任，拟定该职能的体制安排和标准、制定规范、通过能力发展和质量保证支持质量并监督覆盖范围和使用，但不管理各项评价工作。理想情况下，规范性监管机构促成并保障评价职能，在组织标准与所有权和自主权之间取得必要的平衡，从而使分散评价成为真正的直线管理工具，而不是合规性活动。

39. 在接受审查的组织中，有 12 个在其评价政策中指定其中央评价实体为总体评价职能的监管机构，这意味着中央评价实体对集中评价和分散评价都进行监管。只有三个组织(人居署、近东救济工程处和妇女署)明确对分散评价职能进行规范性监管。在劳工组织，评价机构的监管范围更窄，仅涵盖强制性的分散评价。

40. 开发署和儿基会采用反映全组织方法的共同问责模式。在开发署，责任由中央评价办公室和政策 and 方案支助局分担。中央评价办公室保留评价标准和质量评估的规范性权力，政策和方案支持局则负责促进分散评价质量、加强权力下放机构的评价能力并监测相关机构指标。在儿基会，评价办公室保留对评价政策、标准和质量评估的规范性责任，而各司司长、区域主任和国家办事处代表则对分散评价的资源配置、人员配置、实施和使用承担明确的管理责任。虽然共享模式可以扩大所有权，但它们需要特别明确地划分角色和责任，以避免问责差距。

41. 六个组织(粮农组织、国际电联、工发组织、项目署、万国邮联、知识产权组织)没有明确指定分散评价的规范性监管机构。检查专员指出了一项风险，即，如果没有指定监管机构来制定标准、确保质量并监测覆盖范围和使用，那么分散评价职能的一致性和可信度就难以维持。

42. 执行以下建议有望提高透明度和问责制。

建议 1

到 2027 年底，联合国系统各组织行政首长应审查其政策，以：(a) 在其问责框架内纳入分散评价的明确定义；(b) 指定分散评价的规范性监管机构；(c) 界定分散评价所有关键阶段的作用和责任。

C. 进行分散评价的体制安排

43. 各参加组织的政策框架通常将分散评价阐述为操作上的分散，由委托单位负责评价的规划、预算编制、管理和使用。但在实践中，中央评价实体参与分散评价过程的程度各不相同。

44. 主要出现了三种形式。在国际民航组织、劳工组织、国贸中心、国际电联和环境署，中央评价实体必须参与整个过程的各个环节，特别是工作计划、职权范围的批准和评价报告的最终定稿。鉴于这种深度参与，劳工组织将其方法描述为混合式分散评价。对于三个组织而言，中央评价实体的作用仅限于特定阶段。在粮农组织，捐助方合规要求推动评价办公室对工作计划和职权范围进行前端审查，对于可持续发展合作框架评价，联合国秘书处发展协调办公室

根据其指南审查并批准职权范围。²⁰ 在难民署，中央评价实体必须在最终报告发布前予以批准。在其他组织，责任完全由委托单位承担，中央评价机构则发挥咨询作用，重点是指导、能力发展和质量保证。表2汇总了中央评价实体参与分散评价过程的程度。

表 2

联检组参加组织分散评价过程的体制安排

模式	说明	组织
要求中央评价机构或部门予以批准的分散评价	委托单位管理分散评价，但在周期的大多数或所有关键阶段不拥有独立的决策权	民航组织、劳工组织、国贸中心、国际电联、环境署(5)
前端中央控制	分散评价的规划和(或)设计需要中央评价机构或部门的批准；后续阶段是分散的	粮农组织，联合国秘书处(发展协调办公室)(2)
最后产出阶段的中央控制	最终的分散报告在发布前需要中央评价机构或部门予以审批或批准；之前各阶段都是分散的	难民署(1)
操作上完全分散	委托单位对分散评价周期的所有阶段拥有全部权力	开发署、教科文组织、人口基金、人居署、儿基会、工发组织、项目署、近东救济工程处、妇女署、万国邮联、粮食署、世卫组织、知识产权组织、气象组织(14)

资料来源：联检组机构问卷。

45. 各组织对于评价能力的归属也存在分歧。在拥有大量外地存在的 10 个组织中，评价能力已在区域一级建立，更接近于大多数分散评价的开展地点，但有两种不同的所有权模式(见表 3)。在第一组(劳工组织、开发署、人口基金、儿基会、世卫组织)中，分散评价的区域技术能力完全或主要由区域办事处所有，儿基会也拥有国家一级的评价能力：它们资助和指导自己的评价人员，中央评价机构仅限负责设定规范、制定指导意见和建设整体组织评价能力。这就支持部门管理层真正拥有该职能。²¹ 在第二组(粮农组织、教科文组织、难民署、妇女署、粮食署)中，区域评价能力继续由中央评价实体提供，该实体向各区域部署自己的人员。

²⁰ 见联合国评价小组关于评价联合国可持续发展合作框架的准则(2021 年)。

²¹ 劳工组织是一个部分例外，因为其中央评价办公室仍然控制着分散评价周期以及区域评价干事的任命和管理，因此评价人员的区域归属尚未转化为对该职能的全面的分散所有权。

表 3
2025 年区域和国家层面具备评价专业知识的人力资源

组织	职位	专门用于分散评价相关活动的职责和时间	上下级关系		
			中央评价机构	区域办事处	国家办事处
粮农组织	评价干事(外派到各区域)	根据需要，15-20%的时间专门用于技术支持	是	否	否
劳工组织	区域评价专家	全职提供技术支持	次要	是	否
开发署	区域评价顾问	20-40%的时间专门用于技术支持，主要是为区域监测和评价协调中心提供技术支持	是	否	否
	区域监测和评价/成果管理制/评价干事	20-40%的时间专门用于国家办事处的技术支持和能力发展	否	是	否
教科文组织	区域评价干事	全职提供分散评价的技术支持或管理	是	次要	否
人口基金	区域规划、监测和评价顾问	20%的时间专门用于技术支持	次要	是	否
难民署	高级区域评价干事	50%的时间专门用于技术支持	是 (2024 年前为次要)	次要 (是，直至 2024 年)	否
儿基会	区域评价顾问	全职提供技术支持，主要是为国家一级评价干事以及监测和评价协调人提供技术支持	次要	是	否
	多国评价专家、(国家一级)评价干事	全职提供分散评价的技术支持和管	否	次要	是
妇女署	区域评价专家	20%的时间专门用于技术支持	是	否	否
粮食署	区域评价干事	全职	是 ^a	否	否
世卫组织	评价/规划/技术干事	25-100%的时间专门用于分散评价的技术支持或管理	否	是	否

资料来源：联检组机构问卷。

^a 截至 2025 年，大多数区域评价干事都向各自的区域办事处报告工作。

46. 若评价能力属于中央评价实体以外的机构，则反映了更加成熟的分散评价职能。实现和维持这一目标取决于中央规范和指导、实地技术支持和政策支持的能力建设投资，明确职责、质量标准、支持安排、人员配置和资金承诺、培训和促进使用评价结果的机制的政策框架则是其基础支撑。这种成熟度反映了持续的机构关注，包括对加强成果管理制、评价文化以及监测和评价能力的投入，从而使中央评价机构得以向赋能作用转型。这一点在开发署、儿基会和人

口基金中体现得最为明显，大多数区域和国家办事处都投资并维持自己的评价能力，表明该职能已真正实现分散。

47. 在中央评价实体仍然参与各项分散评价工作的组织中，来自中央评价实体的受访者将此视为风险管理，以应对委托进行分散评价的单位的能力不均、质量参差不齐、声誉风险和捐助方要求，而不是反映对分散的不同解读。只有在存在足够的管理能力、可靠的监测系统以及既定的评价和成果管理制文化时，委托单位完全拥有分散评价周期的所有权才是可行的。与此同时，受访者承认，即使中央评价实体发挥的作用有限，也会削弱委托单位对分散评价的主人翁感以及目的和效用的归属感。

48. 这些发现对分散评价应如何发展产生明确影响。检查专员强调，在评价文化和能力刚刚起步的情况下，更密切的中央支持是适当的，但这种支持应用于发展能力而不是施加控制。随着成熟度的提高，目标应该是逐步过渡到真正的自主，使分散评价成为内部化的直线管理工具，而不是集中管理的要求。中央评价实体继续参与关键分散评价步骤的组织应明确阐述退出此类操作角色并将其重新定位为标准和质量的规范性监管者的路径。这将保留中央评价机构的第三线保证作用，同时让分散评价发展为一项完全成熟的第二线职能。

49. 以下建议旨在改善协调、加强第二线管理分散评价的能力并更好地定义中央评价实体的作用和职责。

建议 2

到 2028 年底，联合国系统各组织行政首长应制定一种加强分散评价作为一项成熟的第二线职能的方法，涵盖治理安排、质量标准、专用资源配置、指导材料和能力建设措施。

五. 分散评价的实施

A. 分散评价的需求和驱动因素

50. 委托进行分散评价通常出于三个不同(有时是重叠的)原因：(a) 政策覆盖规范；(b) 捐助方的要求；(c) 委托单位本身的管理需要，如问责、学习和方向修正。在实践中，若干组织的政策覆盖规范首先触发评价，而且，如果捐助方的要求也得到满足，则不单独予以报告，这意味着捐助方驱动的数字可能低估了其真正的份额。在追踪委托进行分散评价的原因的 12 个组织中，约 82% 的分散评价由政策覆盖规范驱动，24% 由捐助方的要求驱动，有些则由两者共同驱动。劳工组织、教科文组织、人口基金和难民署的所有或几乎所有分散评价均由政策授权开展，所有可持续发展合作框架评价也是如此。尽管数据不完整，但采访和调查回复证实，分散评价主要由强制要求驱动。

51. 对分散评价的强制性最低要求因组织而异，并且有别于捐助方授权开展的评价，后者在整个过程中被认为具有强制性。对于项目评价的要求通常分为六类，这些要求并不总是相互排斥，而是可以组合出现：(a) 预算门槛，这是最常见的类别，有时会分级，以便更高的预算触发不同形式或数量的评价(如劳工组织、国际民航组织、近东救济工程处、开发署)；(b) 与方案周期、机构规模或紧急情况类型相关的基于频率的规范(如儿基会、难民署、粮食署、妇女署)；(c) 对于未以其他方式选择进行集中评价的项目，默认“评价一切”方法(如工发组织)；(d) 基于标准的选择，根据战略重要性、捐助方要求或地理覆盖范围等因素选择项目(如环境署、世卫组织)；(e) 源自理事机构决定的基于治理或任务授权的触发因素(如知识产权组织)；(f) 咨询要求或无正式要求，鼓励分散评价但不强制要求(如气象组织)或尚不存在评价政策(如项目署、万国邮联、国际电联)。对于国别方案评价，其要求与方案周期一致，可以是集中评价(如粮农组织、劳工组织、人居署、难民署、妇女署、粮食署)，也可以是分散评价。有些组织允许根据具体情况灵活采用任何一种方式(如开发署、世卫组织)，而其他一些组织则规定了固定的频率，如每个方案周期一次或每两个周期一次(如人口基金、儿基会)，或在所有国家级框架中有系统地强制要求进行评价(如可持续发展合作框架评价)。

52. 案头审查和捐助方访谈揭示了捐助方评价要求方面的显著差异。并非所有提供严格指定用途资金的被审查捐助方都要求进行接收方评价。此类要求在某些政府捐助方(如德国、大韩民国、瑞典，在某些情况下还有大不列颠及北爱尔兰联合王国)以及全球环境基金和绿色气候基金等垂直基金中最为常见。相比之下，慈善基金会和欧盟委员会通常不要求进行接收方评价，而是依赖其自己的评价职能。总体而言，对分散评价的需求似乎在很大程度上独立于捐助方的要求。垂直基金提出了最严格的要求，而基金会往往依赖监测数据和自己的评价。

53. 这些发现引出了一个关键问题：组织评价政策是否受到捐助方范围有限或未明确定义的要求的影响？证据表明，联合国系统内的一些组织可能存在评价要求和相关范围超出捐助方要求或计划用途的情况。虽然此类要求可能反映了与问责、治理和机构学习相关的合法组织目标，但相关发现表明，必须审查其

范围和资源影响是否与利益攸关方的需求和预期增值效应相称。因此，检查专员请各参加组织审查与捐助方供资协议相关的分散评价政策触发因素，确保其契合目的并基于捐助方的实际要求而不是假设的期望。

54. 自发开展的分散评价仍然不常见。即使有这方面的需求，也常常受到相互竞争的优先事项、评价能力有限和预算不足的制约。自发开展的分散评价通常是出于其对决策和学习的价值、对良好做法的记录、战略一致性以及弥补证据缺口的考虑。由于缺乏明确的激励或要求，部门管理人员往往优先考虑执行而非评价，尤其是在资金限制和运营压力最大的人道主义环境中，因此自愿分散评价经常被降低优先等级。这种模式并不普遍：更有力的领导层参与和感知到的发现效用往往会增加主人翁感并减少阻力。然而，总体而言，分散评价在很大程度上仍然由要求驱动，这反映了薄弱的评价文化，在这种文化中，分散评价之所以进行，是因为人们期望或要求进行评价，而不是将其作为管理工具优先予以考虑。

B. 分散评价的可信度

55. 评价的可信度是决定相关发现是否得到使用和建议是否得到执行的关键因素，它取决于两个基本要素：独立性和质量。2014 年联检组审查发现，分散评价的可信度有所提高；尽管从设计上看，该职能在结构上仍然没有独立于直线管理，但由于采取了各种措施，公正性有所提高；报告质量有所改善，但仍低于预期标准。

1. 分散评价的独立性方面

56. 虽然对于分散评价而言，结构上和行政上的独立既不可期望也不切实际，但多个组织已经建立了促进行为独立性和公正性的机制。但它们的有效性却大相径庭：在一些组织中，利益冲突、资历差距和非正式压力仍然令人担忧，而在另一些组织中，保障措施仍然有限。

57. 分散评价管理人员在促进评价独立性方面发挥至关重要的作用。在大多数组织中，该职责被分配给负责干预和(或)监测的管理人员以及评价人员。国际民航组织、劳工组织、开发署、难民署和粮食署要求评价管理人员不得参与接受审查的干预措施，而难民署还要求具备最低程度的资历，以帮助抵制管理压力。儿基会则更进一步，保留了专门的国家和区域一级评价干事，以管理大多数区域的分散评价(见表 4)。

表 4
分散评价的典型管理人员

分散评价管理人员	组织
被评价项目或方案的管理人员	粮农组织、国贸中心、国际电联、环境署、教科文组织、人居署、工发组织、项目署、妇女署、万国邮联、知识产权组织、气象组织(12)

分散评价管理人员	组织
被评价项目或方案的监测和评价人员或协调人	人口基金、近东救济工程处(2)
与被评价项目或方案没有联系的人员	国际民航组织、劳工组织、开发署、难民署、粮食署(5)
区域或国家一级专门进行分散评价的人员	儿基会 ²² (1)
其他安排或混合	联合国秘书处(用于可持续发展合作框架评价)、世卫组织(2)

资料来源：联检组机构问卷。

58. 过去十年来，大多数组织都发布了评价管理人员指南，而包括劳工组织、开发署、环境署、人口基金、难民署、儿基会、妇女署、粮食署和世卫组织在内的一些组织已通过培训和同业交流群加强了能力。尽管做出了这些努力，但调查和采访表明能力限制持续存在。评价管理通常被列为次要职责，因此限制了评价过程可用时间。调查受访者报告了技术专业知识方面的差距，特别是在评价设计和质量保证方面的差距，并表示强烈要求提供进一步的培训和指导。这些弱点可能会降低对与独立性相关的评价规范和标准的理解，并限制对外部评价人员的支持。

59. 使用没有利益冲突的外部咨询顾问是促进分散评价独立性的基石。大多数组织禁止评价人员事先参与被评价干预措施，并采取利益冲突筛查、行为守则和投诉机制等措施。与中央评价实体的访谈表明，对评价人员施加明显压力的情况很少见。更常见的是不易察觉的风险，在合格咨询顾问库有限的小国，这点尤其突出，导致很难避免事先建立的关系。咨询顾问也可能为了维护职业关系和未来机会而对相关发现进行调整，这种风险很难通过正式政策来发现或解决。

60. 受访者强调了诸如上报区域或中央评价实体、由区域或中央评价机构予以监督以及对评价管理人员进行能力建设等保障措施。然而，这些措施的有效性参差不齐，而且往往依赖非正式做法而非一致的保障措施。相比之下，接受调查的国家办事处有 96%认为自己有效保护了评价人员免受不当影响，并列举了参考小组和将评价管理与方案执行分开等措施。这表明，认知层面的独立性和实际体验到的独立性之间存在差距，而区域和中央评价机构则确定了与上下级关系、资历和组织文化相关的更微妙的影响形式。

61. 对于供资协议要求的分散评价，大多数捐助方将独立性安排留给各组织，并且通常不要求结构上的独立性，但某些垂直基金的情况除外。例如，全球环境基金要求评价独立于政策制定、实施支持、中期审查和方案管理。对于全球环境基金执行机构(粮农组织、开发署、环境署和工发组织)而言，这意味着组织、操作和行为上的独立性，并辅之以利益冲突保障措施、运营职能和评价职

²² 但并非在所有区域。

能的分离以及对评价人员的合同保护。尽管方案或项目工作人员可以管理评价工作，但捐助方可认为，所涉组织的中央评价部门参与质量控制加强了评价过程的独立性。

62. 综合来看，相关发现表明，分散评价应始终受到某种形式的监督以保障独立性，无论是通过评价参考小组、区域或中央评价机构的支持还是类似安排，莫不如此。本次审查发现了五个缺乏此类机制的组织。**检查专员鼓励工发组织、项目署、万国邮联、知识产权组织和气象组织制定最低保障措施，确保分散评价在行为上的独立性。**额外措施应与风险相称并适合评价目的。不必要的保障措施几乎没有任何价值，而且可能将有限的资源从更优先的需求中转移出去。

2. 分散评价的质量

63. 本次审查的证据证实，外部评价人员的可用性和能力对于分散评价的质量仍然至关重要。尽管自 2014 年以来评价数量大幅增加，但各组织仍然面临合格评价人员短缺特别是掌握当地语言技能和背景知识的评价人员短缺的问题。挑战已从评价职能不完善变为评价职能因合格外部评价人员数量有限而受到制约。与 2014 年一样，相关发现表明需要系统性的解决方案而不是针对特定组织的解决方案。**检查专员建议各参加组织制定外部评价人员长期能力（特别是国家评价能力）发展计划，最好通过联合国评价小组等机构间论坛做到这一点。**

64. 评价质量仍然受到变化理论不充分和高质量监测数据不足的制约，这与 2014 年审查结果相呼应。在 71 项抽样评价中，75%指出监测和评价数据不足和(或)项目设计薄弱是评价受限和(或)需要提出建议的主要问题。在许多情况下，评价被用来弥补监测数据方面的差距。因此，提高评价质量需要更强有力的项目和方案设计以及资源配置更充足的监测系统；如果不具备这些条件，相关组织应评估是否有理由开展评价。**检查专员敦促各组织确保项目和方案以成果管理制为基础、得到资源充足的监测框架的支持并且只有在充分可评价的情况下才进行评价。**

65. 分散评价质量还受到管理能力有限、角色不明确和工作人员时间不足的限制。评价管理的作用自 2014 年以来变得更加正式，但评价管理仍然经常被分配给没有专门时间、专业知识或两者都不具备的项目和方案人员。调查结果就反映了这一点：42%的机构表示时间有限，19%确定评价专业知识不足是主要限制因素。相比之下，教科文组织报告称，引入区域评价干事来支持评价管理人员，显著提高了分散评价的质量，体现了专门能力支持的价值。

66. 自 2014 年以来，质量保证体系有了大幅改善。目前，拥有分散评价职能的 22 个组织中有 11 个正在进行正式的形成性质量控制和(或)事后质量评估并跟踪质量评级，覆盖范围广泛，对有效性的看法积极。通过这些质量保证活动发现的评价质量常见弱点包括性别和人权的融入有限、方法上有缺陷、数据质量差以及相关发现或建议缺乏足够的证据或针对性。2025 年，获得满意或以上评级

的分散评价占比从开发署的 57%到妇女署的 95%不等。²³ 质量保证制度仍然是提高分散评价质量的最有效手段，尽管其影响因各组织采用程度参差不齐和监测不一致而受到限制。

67. 人工智能辅助的质量保证提供了潜在的增效和更一致的质量评级，但不同的标准和评级系统可能会产生碎片化的解决方案。一个前景看好的例子是开发署独立评价办公室 2025 年开发的分散评价人工智能质量评估应用程序，该应用程序也得到了粮农组织、劳工组织、教科文组织、人居署、难民署、儿基会、粮食署和世卫组织的测试。要使此类工具能够支持全系统的协调统一，各组织需要就其评价工作的通用质量保证框架的核心要素达成一致。因此，各参加组织应通过联合国评价小组确定可以支持人工智能辅助的质量保证的共同质量保证框架的基线要求。

68. 总体而言，一些组织仍然缺乏正式的质量标准和保证机制，但仍然有机会协调和精简各组织的现有方法，以减少重复并产生规模经济，特别是通过使用数字化工具做到这一点。以下建议旨在加强评价质量和全系统协调。

建议 3

到 2028 年底，联合国系统各组织行政首长应实施分散评价的质量标准和保证机制。

C. 分散评价的使用和效用

69. 评价的价值不仅在于其产生的证据的质量和相关性，而且最终在于其利用。然而，关于分散评价对业务或方案有效性和效率的实际影响的数据仍然有限。因此，对使用情况的评估主要依赖代理数据和基于感知的数据。其中一个代理数据反映促进和了解评价使用情况所需的最低基础设施的到位程度，包括从公开披露到管理层回应再到跟踪实施的一系列行动。本次审查发现，大多数组织都确立了这一链条的一些要素，但很少有组织全面一致地予以落实。

70. 22 个组织中有 15 个要求公开披露所有分散评价工作，以认识到它们对透明度、问责制和学习的价值，而 7 个组织仅发布了部分报告或未发布任何报告。在某些情况下，这反映了合法的限制，如敏感内容或特定任务的考虑，即使在其他情况下需要披露时，这些限制也可能适用，但评价结果应当披露仍然是常态。16 个组织要求管理层对所有分散评价做出回应，以帮助将建议转化为组织承诺，而 6 个组织则没有这样做。只有 10 个组织跟踪了建议的执行情况。共有 9 个组织建立了所有三个机制，为跟进和使用评价提供了更坚实的机构基础(见表 5)。

²³ 由于各组织采用不同的质量保证评级机制，其评级不具有直接可比性。

表 5

公开披露、要求管理层做出回应和跟踪建议执行情况的政策和做法²⁴

组织	公开披露所有分散评价报告	要求管理层做出回应	集中跟踪建议的执行情况
粮农组织	无正式政策；发布了一些报告	是	否
国际民航组织	是	是	是
劳工组织	是	是	是
国贸中心 ^a	是	是	是
国际电联	否，所有报告均为内部报告	否	否
联合国秘书处(可持续发展合作框架评价)	是	是	否
开发署	是	是	是
环境署	是	是	否
教科文组织	是(政策没有得到一致采用)	是	否
人口基金	是	是	是(由方案司执行)
人居署	是(仅限于符合规定评价标准的报告)	是	否
难民署	是	是	否
儿基会	是 ^b	是	是
工发组织	否，所有报告均为内部报告	否	否
项目署	一些，前提是与捐助方和(或)政府达成了协议	否	否
近东救济工程处	一些，须经管理层批准	是	是
妇女署	是	是	是
万国邮联	否，所有报告均为内部报告	否	否
粮食署	是	是	是
世卫组织	是	是	是

²⁴ 可能不适用于联合评价。

组织	公开披露所有分散评价报告	要求管理层	集中跟踪建议的执行情况
		做出回应	
知识产权组织	是	否	否
气象组织	否，所有报告均为内部报告	否	否

资料来源：联检组机构问卷。

^a 从 2026 年开始，继修订评价政策后。

^b 敏感报告除外。

71. 执行以下建议有望提高分散评价的效用和效力。

建议 4

到 2028 年底，联合国系统各组织行政首长应确保尽可能公开披露所有分散评价报告，应要求管理层做出回应，并应系统跟踪建议的执行情况，从而使评价充分发挥作为改进问责、学习和组织工具的作用。

72. 由于分散评价在很大程度上由强制性要求而不是内在需求驱动，因此对评价发现的参与往往比较薄弱，其使用也比较机械，仅限于根据建议采取行动，而不是吸取更广泛的经验教训。这种要求驱动的逻辑还可能扭曲哪些干预措施得到评价：如果评价主要由财务门槛触发，那么一些可能产生关键学习经验的小型干预措施可能会得不到审查。即便如此，访谈和调查数据表明，这些评价在概念和战略上有其用途，如跨方案学习、为战略和供资优先事项提供参考以及加强组织系统，但这些惠益没有系统性，在很大程度上取决于领导力、文化、制度和激励措施等有利条件。

73. 这种喜忧参半的现象在各组织都明显存在。调查和访谈的受访者自我报告称，他们对分散评价的使用程度为中等到高度，尤其是在方案设计、国家战略、适应性管理和中期调整方面。在国家办事处受访者中，77%报告了高度使用或中等使用，而 42%则表示评价结果通常会为决策、规划和报告提供参考。然而，实际使用的证据有限，综合研究几乎没有，总部和区域办事处的受访者持更为批评的态度，指出评价结果经常会出具，但并没有得到实质性使用，有时使用仅限于合规和报告目的。

74. 使用情况也因背景不同而存在重大差异。儿基会 2025 年有影响力的评价研究²⁵发现，正式质量评级本身与影响力没有统计学关联；影响力由领导力、文化、利益攸关方参与、运营能力、时间安排和建议的可操作性等因素的更广泛组合塑造。

²⁵ 儿基会评价办公室“2025 年有影响力的评价研究”（纽约，儿基会，2025 年）。

75. 访谈和调查证据发现，质量不均衡、知识共享薄弱、后续行动有限、与项目周期不一致都是使用障碍。由于终期评价通常是在项目结束后完成，因此往往被视为不如期中评价或策略性择时评价有用。这一问题因终期评价通常是强制性的而中期评价通常是可选项而变得更加复杂。总体而言，在存在领导层要求、利益攸关方参与、机构后续行动以及及时可操作评价结果的情况下，更有可能使用评价。

76. 这种评价使用模式也适用于捐助方。除了汇总评价结果的垂直基金外，大多数要求进行评价的被审查捐助方既不监测建议的执行情况，也不提供对评价结果的反馈，因此不清楚其如何使用联合国实体出具的评价报告。在访谈中，捐助方也表示它们更倾向于自己进行评估，而不是由接收方进行评估。总而言之，这些评价结果导致人们质疑，一些捐助方授权开展的评价除了其作为额外保证机制的作用是否还有其他价值。

D. 进行分散评价的效率

77. 如果不能满足组织学习、管理或问责的需求，那么分散评价就失去了价值：如果评价结果没有被利用，那么节省成本的意义就微乎其微。评价只有在范围明确、及时开展且得到高层管理人员支持的情况下才被认为是有价值的；如果主要是受合规或捐助方要求驱动，所谓的效用就会下降。终期项目评价就说明了这个问题，因为一旦项目结束，评价结果就鲜少会被采纳；然而，这些评价通常是强制性的，而被认为更有用的中期评价则往往是可选的，因此并不总是有资金支持。一般的评价结果和建议几乎没有增加价值，因此加剧了这一问题。它们反映了评价质量差且影响效率，因为投入到得出一般性结论的评价中的资源产生的回报有限，并且常常给委托单位留下其资源利用不善的印象。捐助方的要求不明确进一步加剧了压力，迫使各机构默认选择投入开展全面评价，而不是采取更相称的方案。检查专员建议加深与捐助方的接触，包括通过总部和国家层面的机构间论坛做到这一点，并在捐助方的要求不太可能产生及时或有用的证据时让中央评价实体尽早参与。

78. 调查受访者倾向于更灵活、更相称的方法：超过一半的受访机构认为，在某些情况下就采用更轻便的工具(审查、自我评估、监测和评价报告、实时评价等)就足够了，只为风险、规模或问责需求证明有必要开展全面评价的情况保留全面评价的选择。粮食署对不同证据生成产品进行区分是一种有益的做法。效率也受到体制安排的影响：组织内部和组织之间的协调不力导致重复、择时欠佳和协同效应缺失，而评价数量的不断增加则引发了人们对数量可能超过效用的担忧。检查专员鼓励在可行的情况下更多地使用替代评价产品，同时考虑到成本、效用和及时性。

79. 集群评价、组合评价和专题评价日益被确认为更有效的方法，可以减少评价疲劳并产生战略学习。然而，由于评价规划和供资通常与单个项目和捐助方要求挂钩，因此采用程度仍然有限。对于垂直基金而言尤其如此，因为强制性项目评价决定了数量和择时。有限的监测和评价能力、合规要求和协调需求进一步限制了采用。

80. 联合评价和机构间评价提供了类似的潜在好处，但仍主要局限于联合实施的项目。由于预算编制、合同签订、审批流程和质量标准的差异及其蕴含的更高协调成本，更广泛的机构间专题评价和组合评价极少开展。与集群评价一样，效率的提高取决于早期规划、强有力的领导和持续的机构间承诺。

81. 类似的挑战影响到可持续发展合作框架与国别方案评价之间的协调。虽然有些案例²⁶显示了早期协调和共享数据收集的好处，但这种情况仍然不常见。时间安排不一致、分类不同、具体组织的要求和不同的问责目的，往往导致重复让相同利益攸关方参与的平行评价，增加了成本，加剧了评价疲劳。

82. 由于行政限制、能力有限、采购瓶颈和难以征聘合格评价人员，延误仍然是一再出现的挑战。各组织已引入缓解措施，包括评价人员名册、框架协议、标准化指南、质量保证机制和数字跟踪系统，但这些措施的应用并不均匀。要实现持久的改进，就必须尽早将评价纳入方案周期、建设更强的能力、更系统地使用可用工具。一个更根本的挑战是对分散评价成本的理解有限。虽然经常跟踪咨询顾问费用，但很少记录工作人员投入时间、管理工作和建议跟进等更广泛的成本。这就限制了各组织评估效率以及就评价数量和设计做出循证决策的能力。检查专员强烈建议各组织收集已完成分散评价的成本数据，并认识到即使是粗略的估计也能支持更好的资源分配和规划。

83. 尽管证据仍然有限，但人工智能的使用正在开始提高效率。已报告的好处集中在数据处理、分析、证据检索、质量保证和起草方面，在工具是专门构建、已实现制度化且有训练有素工作人员支持的情况下尤其如此。然而，培训、开发和验证成本部分抵消了节省的费用。人工监督和来源核实仍然至关重要，事实证明，人工智能作为第一阶段的支持工具最为有效。

²⁶ 如可持续发展合作框架评价以及人口基金、儿基会和妇女署在东帝汶为2021-2025年方案拟订周期开展的国别方案评价。

六. 分散评价的未来展望

84. 在过去的十年里，整个联合国系统的评价职能通过更有力的政策、更明确的治理框架和更高的机构认可而得到加强。然而，这一进展造成了日益严重的紧张局面。在本次审查的重点——分散评价中，评价的扩展主要受覆盖范围规范和捐助方期望的驱动，而不是受评价证据的内在需求驱动，这导致本已面临相互竞争优先事项和日益减少资源的机构的评价数量不断增加。许多评价既不符合最紧迫的证据需求又没有做到适度 and 契合目的，从而在资源最受制约的地方造成了效率低下。指定用途资金加强了捐助方对评价需求的影响力，而覆盖范围规范将评价要求与供资门槛和其他标准而不是战略证据需求挂钩，这一挑战因此变得更加严峻。结果就是一个悖论：该系统正在产生更多的评价，但许多评价并没有解决各组织最需要回答的问题，而且，即使生成了相关证据，利用这些证据的能力也往往有限。

85. 这些挑战是在联合国系统面临越来越大压力之际出现的。对多边机构信心的下降提高了人们对联合国展现意义和成果的期望，使评价比以往任何时候都更加重要。与此同时，人工智能正在通过提高证据收集、分析、综合和报告的速度和规模来重塑评价，同时仍然需要人工监督以确保严谨性和可信度。人们也日益认识到，评价是公共产品的一个来源，可以为各组织以外的学习和决策提供参考。

86. 综合来看，这些趋势表明，目前的分散评价模式不再完全契合目的，而仅仅进行更多的评价并不能有效地表明其职能强大。不存在最佳的分散评价数量或支出水平；只有当评价的价值足以抵消投入的资源时，评价才是合理的。这就要求确保评价具有相关性、相称性并得到有效体制安排的支持。因此，效率与效用密不可分。应定期评价由财务门槛或其他政策要求驱动的评价而不仅仅是捐助方要求开展的评价的有用性。此类审查还应考虑的是，包括评估、研究、审查或有限范围评价以及专题、联合或集群评价在内的替代产品或方法是否可以更高效地产生更相关的证据。将分散评价的重点重新放在真正的证据需求上并有目的地适当应用分散评价，对于维持其可信度、可持续性和对组织效力和效率的贡献至关重要。²⁷

87. 以下建议旨在确保分散评价继续顺应各参加组织不断变化的需求，定期进行效用评估，并重新审视覆盖范围规范。评价可能并不是适合所有场合的正确工具，在某些情况下，其他更具成本效益且根据证据需求量身定制的评价工作可能更适合。因此，理事机构应了解各参加组织可采用的所有问责和证据生成选项。

²⁷ 关于进一步分析，见补充分析文件(JIU/REP/2026/5 [Expanded report])，第 S29 节。分散评价的未来展望。

建议 5

到 2029 年底，各参加组织的行政首长应通过适当的现有或专门机制委托对分散评价(包括政策要求驱动的评价)的效用和效率进行评估，以重新审视覆盖范围规范，确定哪些地方可能更适合开展更具成本效益的替代评价或其他工作，并为战略、资源配置和体制安排提供参考。

建议 6

到 2030 年底，联合国系统各组织的理事/立法机构应要求提供关于评估分散评价效用和效率的最新信息，包括分散评价的驱动因素、数量和费用，以及加强这一职能的方法。

附件

各参加组织根据联合检查组的建议应采取的行动概览

		预期影响	联合检查组参加组织																											
			联合国 ^a	艾滋病署	贸发会议	国贸中心	开发署	环境署	人口基金	人居署	难民署	儿基会	毒罪办	项目署	近东救济工程处	妇女署	粮食署	粮农组织	原子能机构	国际民航组织	劳工组织	海事组织	国际电联	教科文组织	工发组织	联合国世旅组织	万国邮联	世卫组织	知识产权组织	气象组织
报告	供采取行动		☒	☐	☐	☒	☒	☒	☒	☒	☒	☐	☒	☒	☒	☒	☒	☐	☐	☒	☐	☒	☒	☒	☐	☒	☒	☒	☒	
	供参考		☐	☒	☒	☐	☐	☐	☐	☐	☐	☒	☐	☐	☐	☐	☐	☐	☒	☐	☒	☐	☐	☐	☐	☒	☐	☐	☐	☐
建议 1		a				E		E	E	E		E		E	E	E		E		E	E		E	E	E		E	E	E	
建议 2		a				E		E		E				E	E	E	E	E		E	E		E	E	E		E	E	E	E
建议 3		c				E				E				E	E			E		E			E		E		E			E
建议 4		a						E		E	E			E	E			E					E	E	E		E		E	E
建议 5		h	E			E	E	E	E	E	E			E	E	E	E	E		E	E		E	E	E		E	E	E	E
建议 6		h	L			L	L	L	L	L	L			L	L	L	L	L		L	L		L	L	L		L	L	L	L

图例：

- L：供立法机关作出决定的建议
- E：供行政首长采取行动的建議
- ☐：不要求该组织采取行动的建議

预期影响：

- a：透明度和问责制得到加强；b：传播良好/最佳做法；c：协调与合作得到加强；d：一致性和统一性得到加强；e：控制与合规性得到加强；f：有效性得到提高；g：节省大量资金；h：效率得到提高；i：其他。

^a 虽然建议 1-4 与可持续发展合作框架评价的背景相关，但联检组了解正在进行有关其总体实施框架的讨论，因此向联合国秘书处发布这些建议仅供参考。但无论可持续发展合作框架评价方法未来可能发生何种变化，建议 5 都针对发展协调办公室。