



经济及社会理事会

Distr.: General
14 April 2025
Chinese
Original: English

供作决定

联合国儿童基金会

执行局

2025 年年度会议

2025 年 6 月 10 日至 13 日

临时议程* 项目 12

2024 年儿基会评价职能年度报告

摘要

本报告概述了 2024 年儿基会评价职能。报告概述了背景情况，重点介绍主要成就和挑战，最后对今后几年进一步加强该职能及其有利环境的优先事项进行了思考。

供执行局审议的决定要点载于第五节。

* E/ICEF/2025/11。



一. 概述

1. 今年，无论是在评价职能方面，还是在其有利环境方面，总体上都出现了积极势头。儿基会经修订的评价政策(E/ICEF/2023/27 和 E/ICEF/2023/27/Corr.1) 于 2024 年 1 月 1 日生效，标志着朝向进一步加强评价职能的一系列战略转移，以便通过评价职能的学习和问责等双重作用，使其能最佳地帮助儿基会及其合作伙伴在今后关键的几年里加快儿童工作的进展。在实行修订后政策的第一年，评估职能部门和管理层共同努力，开始实施“硬连接”和“软连接”要素，这将有助于确保在整个儿基会内始终有力地实施该政策。
2. 在变革的第一年，评价职能继续开展其核心工作，即提供及时、相关、可信和有影响力的证据，为儿基会全年的战略方向提供信息。儿基会对关键决策进程的贡献最终体现在《对儿基会 2022-2025 年战略计划的评价》(E/ICEF/2025/3)中，该评价已提交执行局 2025 年第一届常会，儿基会在制定《儿基会 2026-2029 年战略计划》时广泛利用了该评价。
3. 在为评价工作制定的主要绩效指标方面继续取得进展，出现了若干值得注意的积极进展。与前一年相比，评价数量趋于稳定，部分原因是通过优先考虑最具战略意义的主题，为减少这一数量作出了协调一致的努力。区域和国家两级评价数量的减少被总部一级评价数量的大幅增加所抵消，特别是在机构效力和人道主义行动领域。
4. 尽管横跨整个评价职能评价质量之间差异很大，但总体质量仍令人满意或非常满意。然而，尽管评价更系统地纳入了性别平等和残疾视角，但有意义地纳入性别平等关切问题仍然少得令人无法接受，尽管为提高这方面的绩效作出了巨大努力。
5. 一个特别令人鼓舞的发展情况是，评价支出的绝对值和占方案支出总额的百分比都有所增加。总部的支出超过了 1%的组织目标，原因是从评价集合基金预支了原本分配给 2025 年的资源，但后来被用于支付 2024 年的评价优先事项。此外，从 2023 年起，除一个区域外，其他所有区域的支出都有所增加，其中一个区域(东亚和太平洋区域)接近 1%的目标。之所以能够扭转上一年的下降趋势，直接原因是评价职能(全球和分散两级)、区域主任和许多驻国家代表之间的大力协作。然而，这一进展需要持续下去；在大多数区域，评价支出仍未达到下降前的水平，在一些国家，与国家大小和风险状况的其他方面相比，支出仍然低得令人无法接受。
6. 在儿基会内部开展这些努力的同时，儿基会继续加速推动更广泛的全球学习和问责运动，以加快儿童工作的进展。儿基会的国家评价能力发展工作显著扩大：通过其针对高级政策制定者的创新旗舰课程，到 2024 年底，儿基会已为来自 70 个国家的 224 名高级政府官员提供了实质性参与评价的知识和技能，促进更知情的决策过程，造福儿童。儿基会加大了工作力度，将影响评价作为一种工具，促进就在何处以及如何为儿童实现最高投资回报做出正确的战略决策。儿基会与世界粮食计划署(粮食署)一道，在联合国主办了 2024 年全球影响评价

论坛，论坛促进了政府间、非政府、私营、公共和慈善部门各组织之间的战略伙伴关系。最后，儿基会在促进有关评价综合方面的全球伙伴关系中发挥了领导作用，这是在日益复杂的决策环境中的一个战略优先事项。作为全球可持续发展目标综合联盟的共同主席，儿基会在证据综合基础设施协作的治理方面发挥了核心作用，为评价综合筹集了 7 400 万美元。

7. 利用创新性和创造性的数据驱动解决方案，最大限度地提高评价的及时性、战略相关性、可信度、效用、使用率、效率和成本效益，这是 2024 年取得进展的另一个标志。这些解决方案包括从流程改进(如通过数据驱动的法证分析，了解绩效不佳的根源，开发基于风险的评价规划工具，让由青年人组成的评价咨询小组参与其中)到数据系统的简化(如更多地使用行政数据系统，以用户为中心改进儿基会的评价数据库)，到方法上的进步(如预测未来绩效情况的建模工具)和对新技术的使用(如人工智能和自然语言处理)等。还发起一项举措，以加强评价职能的战略传播能力，最大限度地扩大评价影响。

8. 在这些重大进展中，2024 年仍存在几项挑战。除了整个职能部门在关键指标方面的绩效持续存在差异外，由于缺乏强有力的成果和影响层面的数据，在衡量产出和活动之外的组织成就方面继续面临挑战。这一差距延伸到了评价职能本身：除了传闻之外，几乎没有强有力的证据表明评价对儿基会工作的更广泛影响的范围或规模。最严重的不足是一些战略举措的延迟：由于能力有限，无法应对评价职能面临的多重挑战和机遇，2024 年开始编写补充经修订的评价政策的三份关键文件，即评价政策程序文件、政策实施战略和评价手册，但其完成时间被推迟到 2025 年。

9. 2025 年及以后，儿基会将继续努力应对持续存在的挑战，同时保持 2024 年取得的成果。这些努力将在儿基会组织现实不断变化的背景下进行。在这一变革时期，评价职能的首要重点将有两个方面：(a) 通过相关的评价证据，支持儿基会驾驭复杂和不确定的局面；(b) 保持评价的公正性，使其成为一个强有力的独立职能，拥有适当的架构及充足、可预测的资源，能够发挥学习和问责的双重作用，从而在重塑儿基会的更广泛的潮流中保持适应未来的能力。

二. 评价职能的主要发展

A. 2024 年的背景和优先事项

10. 2024 年之前的几年是评价职能的变动和过渡期。在 2018 年发布修订后的评价政策以及通过建立分散一级职能(借助评价集合基金的设立)从 2018 年到 2019 年相应扩评价大职能后，COVID-19 大流行才显现出来。在此期间，评价主任的职位空缺，使新成立的、地理位置分散的、远程运作的全球团队在近一年的时间里没有领导。

11. 正是在此背景下，2021 年底，随着评价主任的到任，关键优先事项是在技术和关系方面(重新)建立团队，并根据联合国评价小组的规范和标准，建立独立评价职能良好做法的一致性。除这一优先事项外，还有另一个更高阶的战略优

先事项：在 2023 年对评价政策进行政策授权的审查和修订，在此之前，2022 年对评价职能进行独立同行审议。

12. 因此，在 2024 年之前的两年中，优先事项是采取系统性、包容性的循证办法来完成这些进程，并制定经修订的儿基会评价政策(E/ICEF/2023/27 和 E/ICEF/2023/27/Corr.1)，该评价政策已获执行局 2023 年第二届常会批准，并于 2024 年 1 月 1 日生效。因此，持续、有力和有意义地实施修订后的政策是 2024 年的首要优先事项，与此同时，要继续建立综合协调的全球评价职能，确保统一的高绩效，加强有利的环境和评价文化，实施雄心勃勃的评价工作计划，所有这一切都在这样一种组织结构中进行：在这种结构中，分散一级的独立职能工作人员不向评价主任报告。

13. 经修订的政策体现了一系列渐进但重要的变化，使评价职能及其有利环境更加成熟，从而优化评价职能，帮助儿基会及其合作伙伴在未来几年为儿童取得尽可能好的成果。经修订的政策还划分了具体的战略优先事项，以适应未来的职能，对儿童产生最大的影响。其中包括国家评价能力发展、影响评价、评价综合、战略评价伙伴关系和战略传播。

14. 确保在儿基会的分散一级结构中始终有力地执行该政策的广泛战略沿着两条平行的轨道落实，大致分为行动的“硬连接”和“软连接”两方面。“硬连接”要素包括实际产出，以更加具体和细化的细节落实该政策的广泛目标。其中包括三个关键文件：(a) 政策实施战略，为履行该政策中所列承诺提供总体路线图；(b) 政策程序文件，其中规定了分配给负责实施的具体个人的行动和期望；(c) 评价手册，为评价工作人员和其他参与评价的人员提供详细的技术性“操作方法”，以便作出始终相关、及时、可信和实用的评价。所有这三份文件都是 2024 年启动的，但由于能力所限，其定稿推迟到 2025 年。

15. 与此形成对比关系的是，“软连接”要素侧重于建立和深化关系，这对于政策的内部化——尤其是支撑政策的基本原则——以及确保政策的有效实施至关重要。这一要素包括在儿基会的三个层面之间和内部建立关系。由于评价职能(评价办公室和区域评价顾问)与管理层(区域主任、司长和执行主任办公室)之间的密切合作，在这一领域取得了相当大的进展。该政策的推出已列入大多数区域管理小组会议的议程，并在区域主任(或副区域主任)、区域评价顾问和评价主任之间建立了三方会议，以同时加强评价职能的独立性、其与管理及决策进程的紧密结合及其在分散一级的总体绩效。

B. 评价对成果的突出贡献

16. 在这些总体战略潮流中，儿基会继续开展评价工作，以促进儿基会在为儿童取得成果方面的绩效。

17. 《对儿基会 2022-2025 年战略计划的评价》(E/ICEF/2025/3)是一项特别值得注意的全球评价，原因有二，它利用了创造性的解决方案，确保对整个四年期进行及时、彻底和强有力的评估，同时又对组织决策进程产生了积极影响。该评价所依赖的创造性解决方案以方法论为中心，不仅能够在短短 5 个月内完成高

质量的评价，而且还使评价小组能够预测到 2025 年底可能观察到的成果。因此，儿基会广泛而周到地利用了此次评价的结果，其中最突出的是在制定《儿基会 2026-2029 年战略计划草案》的过程中利用了此次评价结果。

18. 对儿基会广泛战略办法的其他范围更窄、重点更突出的全球评价也证明具有同样的影响力。《关于儿基会性别平等政策和性别平等行动计划执行情况的评价》(性别平等行动计划 2 和性别平等行动计划 3)确定了在人道主义、发展与和平背景下将性别平等纳入主流的领域，并为下一个性别平等行动计划以及将其纳入下一个战略计划提供了参考。对合理便利基金的评价发现，该基金通过消除残疾员工在工作场所遇到的障碍，有效促进了更具包容性的工作环境，同时该评价还提出了一系列改进建议。¹与此同时，《对儿基会内部沟通和工作人员参与职能的发展评价》建议改变单向沟通的做法，转而采取更具活力的协作办法，促进员工的参与和归属感。《对儿基会在促进减少儿童贫困、包括在人道主义和脆弱环境中的儿童贫困工作的评价》同样具有深刻的见解。在所有这些评价中，儿基会管理层都积极、正面和建设性地参与，并一直在根据评价建议采取行动。

19. 伴随着儿基会在全球紧急情况中发挥的作用不断增强，对人道主义行动的评价数量也同时增加。评价办公室总共进行了 12 次应急反应评价，比 2023 年的评价数量增加一倍多，提供了战略和业务方面的相关证据，为儿基会的工作提供了参考。《对儿基会乌克兰三级应对措施的评价》强调了一些良好做法，如转向有针对性的、基于需求的实施模式，以及通过利用当地能力将短期援助与长期复苏工作相结合。同样，《对儿基会也门三级应对措施的评价》指出了 2019 至 2022 年所取得的进展，但建议改进人道主义工具，加强方案一致性，重新评估应急程序，利用实地经验，加强监督和伙伴关系，以更好地确定弱势群体目标，并确保采取可持续的干预措施。评价办公室在拟订第一份《联合国评价小组关于将人道主义原则纳入人道主义行动评价的指导意见》方面发挥了关键作用。

20. 在分散一级开展了数量庞大的评价，为儿基会实地业务行动的决策过程提供了帮助，其数量之多，无法一一概述。突出的例子包括《对儿基会应对马拉维二级霍乱和弗雷迪飓风紧急情况的评价》，其中建议在今后的紧急情况中实现供应链管理和采购战略的区域化及强化，以建设地方复原力，并加强人道主义-发展-和平之间的联系。《对欧盟全球促进移民儿童最佳做法方案的总结性评价》提高了对这一专题的关注，并探讨了评价对今后方案拟订的影响。《对埃塞俄比亚北部和阿富汗危机应对措施的机构间人道主义评价》评估了整个联合国应对措施的附加值，从而为两国人道主义协调小组的决策提供了参考。《对儿基会 2018-2023 年太平洋儿童保护方案的评价：基里巴斯、萨摩亚、所罗门群岛和瓦努阿图》促使 2024 年年中赠款大幅延期。

¹ 本报告中的所有评价均可查阅 <https://www.unicef.org/evaluation/reports#/>。

21. 除了这些单独的评价之外，儿基会还在 2024 年建立了评价综合和元分析做法，以此提请利益攸关方注意评价中反复出现的问题。《全球评价证据综合：儿基会 2018-2022 年为儿童取得成就的评价证据》分析了儿基会 2018-2022 年的工作成果，利用人工智能按目标领域、变革战略、能动因素以及《儿基会 2022-2025 年战略计划》的其他要素绘制评价结果。其中确定的关键证据差距将直接为未来的评价规划进程提供参考。在方案层面，《南亚童婚证据元综合与元分析》审查了 65 项关于南亚童婚问题的研究——包括儿基会本身以外的对干预措施的研究，并确定了减少该区域童婚的有效战略。此后，这一分析为南亚以外区域关于循证方案拟订和干预措施有效性的全球讨论提供了参考。

22. 尽管这些证据确凿的积极成果值得注意，但 2024 年围绕儿基会评价整体效果的总体范围和规模继续存在长期的知识差距。除了这些高质量、高使用率评价的个别例子外，由于缺乏监测系统和能力，评价职能对儿基会工作的影响程度的总体情况仍不清楚。

C. 影响评价

23. 影响评价提供了强有力的证据，表明政府、捐助方和执行伙伴需要进行最有效和最具成本效益的投资，从而为儿童取得最大成果。影响评价有助于更好地决定一项干预措施(或干预措施组合)是应该“按现状”或以经修改的形式继续实施并扩大规模，还是停止实施。最重要的是，它可以帮助决策者了解哪种干预措施(或干预措施组合)会带来最高的投资回报率。随着儿基会和许多其他组织在资源日益紧张的环境中努力提供更多更好的服务，影响评价比以往任何时候都更加重要。正因如此，影响评价模式成为修订后评价政策的核心。

24. 与往年相比，2024 年完成了更多基于影响方法的评价(5 项)和评估影响水平结果的评价(18 项)。对格鲁吉亚有针对性的社会援助的影响评价表明，现金支付的好处优于食品券，促成方案覆盖范围扩大了 30%。同样，对尼泊尔多部门营养计划的影响评价提供了强有力的证据，为扩大规模的决定提供了依据，该计划涵盖 352 000 名儿童，促进了不同利益攸关方之间的持续合作。对秘鲁向儿童早期教育方案进行转移支付的影响评价得出了积极的结论，政府因此追加拨款 860 万美元，以扩大方案覆盖面，使更多儿童受益。

25. 影响催化基金是评价办公室的一项旗舰举措，事实证明它在激励国家层面使用影响评价方面发挥了关键作用。布基纳法索、加纳、印度和土耳其收到了影响催化基金的对等赠款，用于设计和开展以童婚和社会规范为重点的严格影响评价。目前共有 10 个国家办事处被指定在两个专题窗口下进行影响评价：童婚和社会规范以及适应性社会保护。

26. 儿基会在推动影响评价以造福儿童方面的思想领导力，最终于 12 月在纽约由儿基会和粮食署共同主办的联合国 2024 年全球影响评价论坛上达到顶峰。来自 30 个国家的 80 个组织参加了论坛，其中包括会员国代表团、双边和多边捐助方、联合国 12 个机构的高级官员、非政府组织、私营部门和慈善机构。儿基会执行主任和负责伙伴关系的副执行主任参加了论坛，强调加快共同努力进行影

响评价的道义责任。这次活动为指导影响评价的未来提供了可行的思路，并促进了与会者之间的一系列合作。

D. 国家评价能力发展

27. 影响评价通过确保对高收益干预措施进行目标明确、以证据为导向的投资，有助于实现规模化成果，而国家评价能力发展则通过将儿基会以外的其他关键部门和行为体纳入评价领域，帮助取得系统一级的大规模成果。由于儿基会只是为儿童取得成果的众多贡献者之一，只有通过这些其他利益攸关方——特别是高级决策者——对评价的参与，学习和问责才能为儿童带来广泛而切实的惠益，因为政府是儿童权利的主要责任承担者。正因为如此，国家评价能力发展与影响评价一样，被确定为订正后评价政策的一个主要优先事项。

28. 与影响评价一样，2024年，儿基会在国家评价能力发展方面取得了重大进展。继续提供“加强评价系统的中级引导式课程”这一在线课程，2024年有来自67个国家的1187名政府官员、合作伙伴、联合国工作人员以及其他技术评价从业人员和管理人员接受了培训。儿基会还扩大了评价领导者执行课程的规模，这是与新加坡国立大学合作为高级决策者提供的创新方案，重点是帮助这些官员及其工作人员从评价角度寻找解决本国儿童问题的办法。截至2024年底，即第一批学员接受培训刚刚两年多的时候，评价领导者执行课程已接待了来自儿基会所服务的所有区域的70个国家的101个代表团。共有224名高级决策者和其他政府伙伴(如司长、副部长、高级顾问)以及陪同他们参加课程培训的107名儿基会工作人员从评价领导者执行课程毕业。这些共计331名的评价领导者执行课程校友组成了一个评价倡导者网络，他们继续促进国家、国家间、区域和全球的经验交流。

29. 此外，评价领导者执行课程促进了政府间评价合作的更广泛运动。例如，南非规划、监测和评价部利用评价领导者执行课程网络，邀请其他国家政府加入其新设立的国家评价系统资源中心，该中心本身就是因评价领导者执行课程而设立的。在参加了评价领导者执行课程之后，孟加拉国政府派出了一个代表团与斯里兰卡政府交流经验，从而与儿基会达成了加强评价能力促进发展的双边协定。在土库曼斯坦，参加评价领导者执行课程促成举办深入讲习班，并将评价能力建设纳入业经总统批准的《2023-2028年实现儿童权利国家行动计划》。在柬埔寨，儿基会利用评价领导者执行课程的几位校友获得的知识，通过国立管理大学为评价从业人员和管理人员开发了一个专业认证课程。

30. 儿基会还继续在国家一级满足更加个性化的国家评价能力发展需求，以加强评价系统和评价在决策中的作用。实例包括：(a) 贝宁、喀麦隆和多哥的国家评价周；(b) 向塞拉利昂政府提供有针对性的支助；(c) 刚果国家能力诊断；(d) 在布隆迪实施能力建设方案，以提高监测和评价部长级技术小组进行快速评价的技能，并建立一个由布隆迪大学主办的评价同业交流圈。2024年的新举措包括：(a) 也门的年轻评价人员和新评价人员学习方案；(b) 与泰国公务员制度委员会办公室共同举办了一个讲习班，为未来的公务员开发课程；(c) 与越南规划和投资部一起举办一个培训课。儿基会还促进举办一系列区域学习网络研讨会，

并参加了众多全球和区域会议，包括在菲律宾举行的亚太评价协会会议、在摩洛哥举行的评价伙伴(EvalPartners)全球评价论坛、在中国举行的国家评估能力会议以及在斯里兰卡举行的监测和评价教育学术专题讨论会。

E. 创新和其他创造性解决方案

31. 经修订的评价政策确定，创新和创造性地解决问题是最大限度地提高评价的相关性、及时性、质量、可信度、效用和效率的一种手段。2024 年，在评价规划、管理和青年参与等领域取得了众多进展。

32. 在拟订儿基会 2026-2029 年全球评价计划的过程中，儿基会着手开发一个基于风险的规划工具，该工具将在确定下一个四年期拟予以评价的主题方面带来更大的一致性、严谨性、客观性和透明度。这一工具将于 2025 年试行，其中整合了广泛的数据和信息来源，并将成为评价职能内部进行讨论以及与内部和外部利益攸关方开展协商的基础。除了简化规划过程外，这一工具还有助于确保优先考虑最具战略意义的评价专题，并进而合理分配资源。它还将促进评价职能的三个层级之间采取更加协调一致的评价规划方法，同时通过提升系统、数据驱动和循证分析在规划过程中的作用来加强独立性。

33. 2024 年，随着人工智能驱动的自动化和新兴技术的不断发展，儿基会推进了人工智能的使用，以改善其评价流程和行为。通过使用人工智能改进数据提取、处理、分析和共享，提高了效率和准确性，并能从不同的数据集中制作更有力的证据综合，从而有助于实时决策。此外，对儿基会证据信息系统集成平台搜索引擎的人工智能增强，提高了评价报告的可搜索性，用户搜索量激增。

34. 儿基会还将人工智能提升为组织内外的总体战略评价优先事项。取得的进展包括发表了“评价中机器学习的操作框架”以及两篇技术论文和《社会科学及医学》上的一篇期刊文章。作为联合国评价小组数据和人工智能工作组的主席，儿基会分享了可能与其他组织相关的人工智能使用案例和数据科学应用。该工作组最后敲定了一份讨论文件和建议，以及与人工智能在证据生成活动中的伦理问题咨询小组一起编写的一份技术文件和建议，内容涉及人工智能在证据生成中的伦理问题。

35. 这些“高科技”解决方案得到了一系列同样重要的“低科技”进步的补充。为了让更多人可查阅应急情况下的评价结果、结论和建议，儿基会改进了对视觉辅助工具的使用，以传播主要结论和经验教训，并测试了让利益攸关方参与评价过程的创新性替代办法，从而增强他们的认同感。例如，在选定的人道主义评价中采用了“客户体验”方法，以更好地陈述受影响民众在与应对紧急情况的组织互动之前、期间和之后的实际经历。此外，儿基会采取了更系统的办法，通过设立青年咨询委员会，提高其服务对象——特别是青年人——的声音，该委员会在对战略计划和性别平等行动计划的全球评价中发挥了积极作用。由于儿基会在有意义地让青年人参与评价方面作出了杰出贡献，它被 Eval4Action 运动确认为联合国各机构中的“超级冠军”，该运动推动评价成为实现可持续发展目标的重要和关键加速器。

36. 分散一级也探索了许多创造性的解决方案。秘鲁和厄瓜多尔的小组进行了测试，将行政数据用作提高影响评价及时性的一种成本效益高的办法。在南亚区域，采用人工智能驱动的分析与人工审查相结合的方式，对过去3年来儿基会在该区域开展的所有评价进行了全面评估，为区域办事处管理计划提供了有据可依的分析。同样，在东亚和太平洋区域，以“共同试点”工具的形式部署了机器学习，该工具经过培训，可从评价报告中提取关键词和短语，为评价综合提供信息。在欧洲和中亚区域，一个经过青少年训练的人工智能模型被用来评估政府的精神健康服务是否满足不断变化的客户需求。东部和南部非洲区域利用人工智能工具编写了两份关于人道主义行动中儿童保护问题的微型综合报告：一份报告为儿基会加强东部和南部非洲民事登记和生命统计战略方案框架提供了证据基础；另一份报告有助于从多项评价中快速提取关键结论，以支持关于儿基会应急工作的综合报告。

F. 伙伴关系

37. 伙伴关系与创新和创造性地解决问题一道，构成了履行经修订的评价政策中提出的雄心勃勃的承诺的关键战略支柱——进而将驾驭评价作为促进儿童进步的加速器。儿基会与粮食署在联合国共同主办的2024年全球影响评价论坛催生了与会者之间的伙伴关系，这是一项具有试金石意义的成就。此外，儿基会建立其影响评价能力和做法的能力是与德国联邦经济合作和发展部(联邦经济合作部)就“适应性社会保障：儿童成果证据”项目建立创造性伙伴关系的直接结果。

38. 伙伴关系在儿基会的国家评价能力发展努力中也很突出，特别是在设计和交付以及这些努力在参与者之间产生的关系方面。全球一级的试金石活动——评价领导者执行课程和加强评价系统的中级引导式在线课程——得到了区域定制举措的补充。例如，中西部非洲区域办事处与非洲高级管理研究中心正式建立了多年伙伴关系，以便将评价能力嵌入该区域的公共机构，使各国政府具备先进的技能和系统，开展和利用评价来改进政策和方案。在斯里兰卡，儿基会与粮食署和德国发展评价研究所建立了伙伴关系，就评价能力进行国家诊断，以全面评估该国的评价生态系统和支持循证决策的能力。东部和南部非洲区域办事处与联合国合作伙伴一道，举办了一个为期五周的培训课程，以建设各国政府和联合国工作人员在评价框架、规划和国家评价系统发展方面的机构知识和能力。

39. 在评价综合领域取得的重大成果同样也是与联合国其他机构及其他方面建立牢固伙伴关系的结果。儿基会与联合国姊妹机构²合作，汇总了所有与可持续发展目标5(实现性别平等，增强所有妇女和女童的权能)相关的联合国系统评价的经验教训。此外，儿基会和联合国开发计划署(开发署)共同主持了由超过46个联合国机构组成的全球可持续发展目标综合联盟。儿基会还在一项由韦尔科姆信托基金资助的新倡议即证据合成基础设施协作倡议的治理中发挥了关键作

² 联合国促进性别平等和增强妇女权能署(妇女署)、联合国开发计划署(开发署)、联合国人口基金(人口基金)和粮食署。

用，该倡议旨在通过扩大使用评价证据、安全和负责任地使用人工智能以及增加数据共享和再利用来应对社会挑战。在很大程度上，由于这一伙伴关系的成功，并通过联合王国研究和创新以及韦尔科姆信托基金的初始投资，为支持全球可持续发展目标综合联盟和证据合成基础设施协作倡议筹集了 7 400 万美元，不是用于儿基会本身，而是用于更广泛的全球努力。

40. 联合、机构间和全系统协作是另一个取得显著进展的领域。与其他机构联合管理了 10 项评价，包括对 COVID-19 疫苗全球获取倡议、³《人人享有健康生活和福祉全球行动计划》⁴ 的评价，以及世卫组织/儿基会对全球霍乱控制专题小组的联合评价。儿基会向 2024 年全面运作的联合国可持续发展集团全系统评价办公室提供支助，并确定协调与协作的领域。这包括为联合国系统发展方面业务活动四年度全面政策审查(四年度全面政策审查)的状况报告制作 5 份联合国评价证据摘要以及利用人工智能生成自 2021 年以来全球、区域和国家层面所有联合国评价报告的互动证据图。这些联合国系统评价证据图和证据摘要提供了有关几个专题的证据，即：(a) 驻地协调员系统；(b) 联合国发展系统区域架构；(c) 可持续粮食体系；(d) 人道主义-发展-和平的联系；(e) 供资质量。这些证据图和摘要引起了会员国的兴趣并获得了积极反馈，为参与 2024 年四年度全面政策审查的利益攸关方提供了关键信息。

41. 儿基会继续积极参与全球和区域评价网络。通过参与联合国评价小组各工作组的工作，编写了几份关于将人道主义原则、人权和性别平等纳入评价工作的指导文件。在以下方面也作出了其他重要贡献：评价工作中的道德操守；联合国评价小组评价综合报告；环境和社会影响；方法；数据和人工智能；性别、残疾和人权；建设和平；影响评价；分散评价；年轻评价人员和新评价人员；人道主义评价；全系统的联合国残疾包容战略工作组。儿基会与联合国人口基金(人口基金)合作，为“联合国青年战略：优先领域教育和就业”的综合报告作出了贡献。联合国亚洲及太平洋评价发展工作组是一个区域机构间网络，儿基会是该网络的核心成员，该工作组与联合国发展协调办公室密切合作，为已安排对联合国可持续发展合作框架进行评价的国家的政府和联合国国家工作队提供培训。

三. 绩效概览：评价职能的持续进展和主要经验教训

42. 本节概述了 2024 年在有关评价的主要绩效指标方面观察到的主要趋势。根据经修订的评价政策中的一项承诺，儿基会《2026-2029 年战略计划》将为制定更多的衡量标准提供机会，以全面反映职能及其有利环境在质量上的健康和成熟程度，作为对现有定量指标的补充。

³ 与全球疫苗免疫联盟和世界卫生组织(世卫组织)一道。

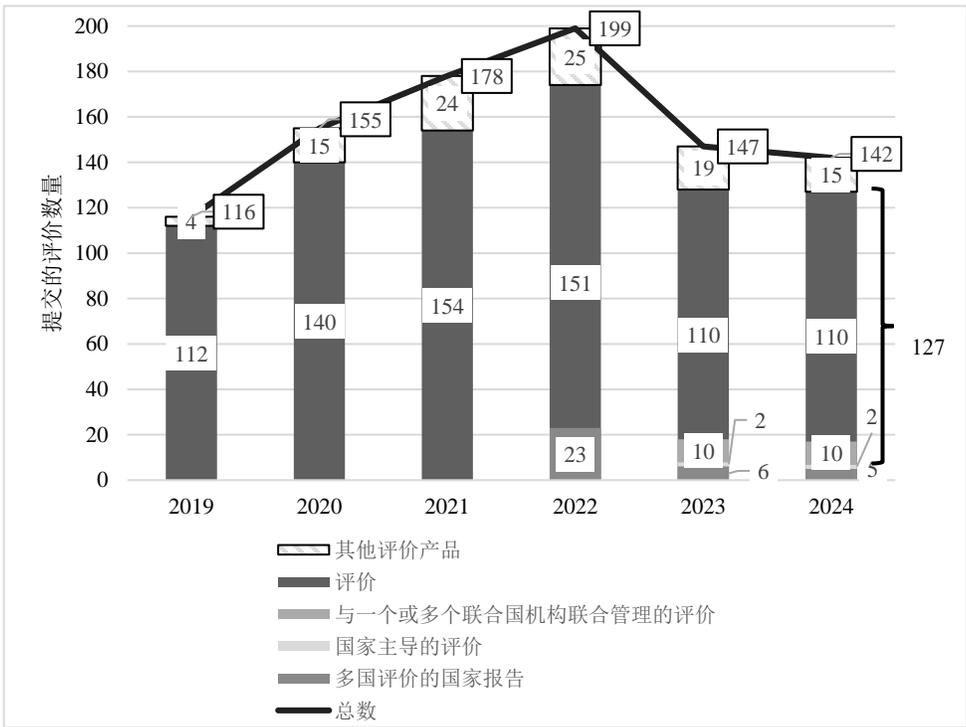
⁴ 与世卫组织牵头的 13 个组织组成的全球联盟一道。

A. 提交的评价数量和地理覆盖面

43. 如图一所示，2024 年共记录了 142 项评价和其他评价产出。这些包括 127 项评价——其中 10 项是与联合国其他机构联合管理的评价，5 项多国评价和 2 项国家主导的评价——以及 15 项其他评价产品。

44. 值得注意的是，评价数量持续稳定在与疫情前相当的水平。这种稳定可能是多种因素共同作用的结果，其中包括：(a) 2022 年推出将一项活动归类为评价的更明确标准，2023 年和 2024 年又以核查活动加以补充，以确认所提交资料的准确性；(b) 经修订的评价政策提出了广泛的战略诉求，将及时性、战略相关性、质量和可信度置于数量之上；(c) 由于资源限制、评价和其他地方决策进程资源不足，几个区域的评价数量有所减少。

图一
2019-2024 年提交的评价数量



注：2023 年出台的评价类型详细分类。

资料来源：证据信息系统集成平台数据库。

45. 导致数量下降的另一个因素是独立的可评价性评估数量减少。虽然数量有所减少，但 2024 年将一些早期、快速和面向未来的活动纳入标准评价的情况有所增加，这表明出现一种积极的趋势，即进行更加综合、全面和平衡的评价，同时满足眼前的前瞻性学习需求，并对绩效进行更加全面和深入的评估。

表 1

2022-2024 年提交的按地点分列的评价和其他评价产品的数量

地点	提交的产品																
	评价			其中由国家主导的评价数量	其中与一个或多个联合国机构联合管理的评价数量	其中包含的多国评价国家报告的数量	可评价性评估			实时评估			其他(包括基线评估、综合、审查)			共计	
	2022	2023	2024	2024	2024	2024	2022	2023	2024	2022	2023	2024	2022	2023	2024	2023	2024
EAP	23	13	10	1	0	0	2	5	3	3	1	0	0	0	1	19	14
ECA	16	11	23	1	2	5	3	2	0	3	0	0	0	0	0	13	23
ESA	24	24	25	0	2	0	0	0	0	0	0	0	0	1	0	25	25
HQ	4	11	14	0	0	0	1	1	0	0	0	0	0	1	4	13	18
LAC	27	17	16	0	0	0	0	1	1	9	0	0	0	0	0	18	17
MENA	21	21	11	0	1	0	0	2	2	0	0	0	0	1	0	24	13
SA	18	14	7	0	2	0	2	1	0	0	0	0	0	0	2	15	9
WCA	41	17	21	0	3	0	2	3	1	0	0	0	0	0	1	20	23
共计	174	128	127	2	10	5	10	15	7	15	1	0	0	3	8	147	142

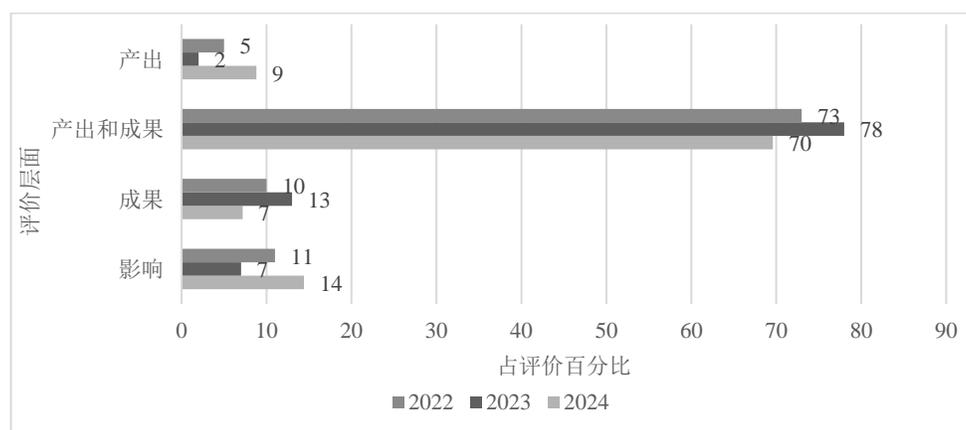
说明：EAP：东亚和太平洋区域；ECA：欧洲和中亚；ESA：东部和南部非洲；HQ：总部；LAC：拉丁美洲和加勒比区域；MENA：中东和北非；SA：南亚；WCA：西部和中部非洲区域。

资料来源：证据信息系统集成平台数据库。

46. 2024 年的评价延续了之前的趋势，即大部分评价侧重于更高层次的成果和对儿童生活的影响，这进一步反映了注重具有战略意义的评价而不是增加数量的呼吁。对图二所示数据进行的次级分析进一步表明，在 2024 年进行产出层面评价的 11 个国家办事处中，7 个还进行了影响和成果层面评价。这一模式表明，这些办事处普遍采用了全面和平衡的评价视角，而不是一味地进行不那么雄心勃勃的产出层面评价。

图二

2022-2024 年按层面和年份分列的评价



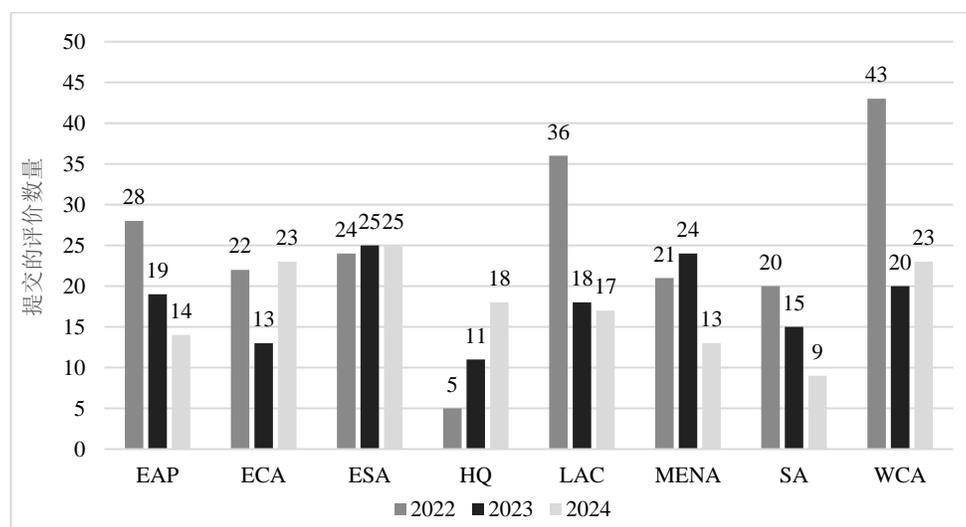
资料来源：全球评价报告监督系统。

47. 图三汇总了按地点分列的提交评价报告的趋势。在数量持续减少的大趋势下，整个组织的产出差异很大。总部提交的报告是原来的3倍多，原因是需要弥补战略关键性评价覆盖面存在的差距，加上对该组织这一级评价的需求不断增加。这种需求的增加可以说是一个积极的迹象，表明人们越来越重视评价工作的价值；这也证明评价职能愿意满足各种要求和需求，而且由于采取了各种创新和创造性措施来提高效率，评价职能有能力按时交付，并始终保持高质量。然而，这一吸收能力在2024年达到了极限。

48. 各区域的产出水平差别很大。欧洲和中亚以及西非和中非超过了2023年的产出水平。总体而言，东部和南部非洲提交的评价最多，南亚(仅涵盖8个国家)提交的评价最少。中东和北非、南亚以及东亚和太平洋出现了大幅下降。中东和北非提交评价报告出现波动主要与具有挑战性的方案环境有关，特别是在二级和三级紧急情况下。在南亚和东亚及太平洋区域，国家一级的评价更多地融入了全球或专题评价，而且这些区域参与影响评价的程度较高，因而持续时间较长，这抵消了数量下降的影响。值得注意的是，虽然这些区域的数量有所下降，但全球评价报告监督系统的评级却有所提高，这说明从数量到质量的转变可能带来的收益。

图三

2022-2024年按地点分列的评价报告提交情况



说明：EAP，东亚和太平洋区域；ECA，欧洲和中亚；ESA，东部和南部非洲；HQ，总部；LAC，拉丁美洲和加勒比区域；MENA，中东和北非；SA，南亚；WCA，西部和中部非洲区域。

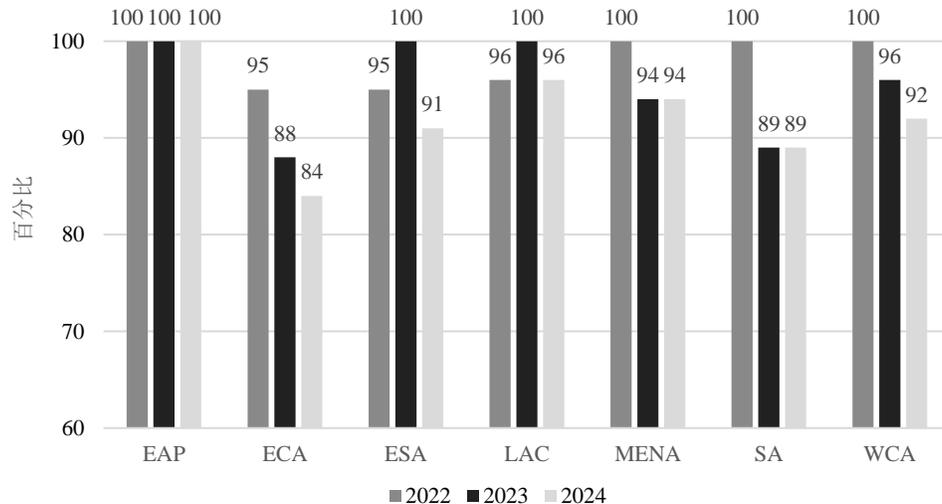
资料来源：证据信息系统集成平台数据库。

49. 如图四所示，评价覆盖面不均，与2023年相比，大多数区域2024年的评价覆盖面明显下降。这是因为各区域有多达24个国家在过去3年中未提交评价报告。东亚和太平洋区域尽管报告提交总量有所下降，但仍保持全面覆盖，因为前两年没有进行评价的办事处在2024年进行了评价。与此相反，欧洲和中亚区域的覆盖面急剧下降，尽管该区域提交的评价报告数量大幅增加，因为一些办事处开展了多项评价。2024年，南亚以及中东和北非的覆盖面趋于平稳，但除东亚和太平洋区域外，其他区域的覆盖面大幅下降。

50. 2024年最令人担忧的观察结果是，30多个国家完全缺乏评价活动，而这些国家加在一起约占儿基会所有方案的四分之一。在这30个办事处中，只有5个在二级或三级紧急情况的国家开展业务，而且大多数办事处管理着大量资源：7个办事处的预算在2500万至5000万美元之间，3个办事处的预算在5100万至1亿美元之间，3个办事处的预算超过1亿美元。尽管评价职能和管理层采用“软连接”策略鼓励国家办事处履行义务，但这些差距依然存在。

图四

2022-2024年各区域的评价覆盖率



说明：EAP，东亚和太平洋区域；ECA，欧洲和中亚；ESA，东部和南部非洲；LAC，拉丁美洲和加勒比区域；MENA，中东和北非；SA，南亚；WCA，西部和中部非洲区域。

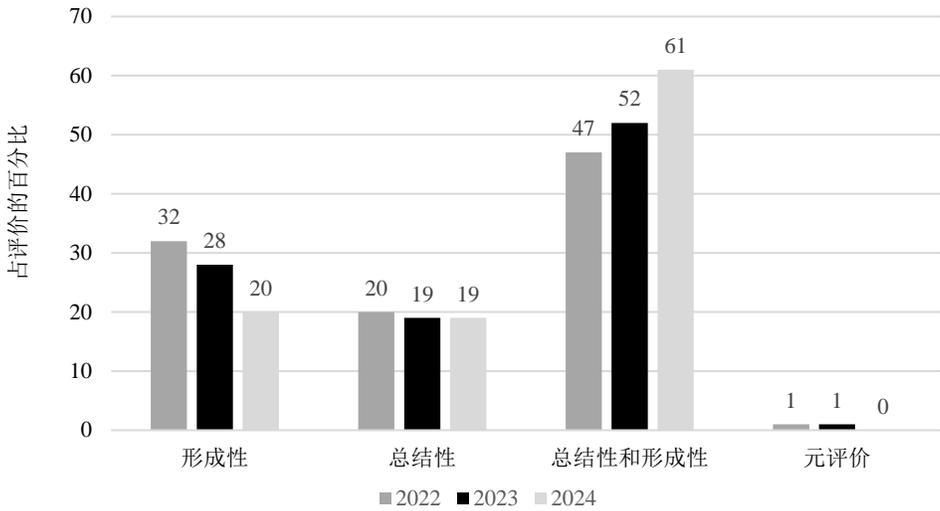
资料来源：证据信息系统集成平台数据库。

B. 评价概况

1. 兼顾目标和专题重点

51. 有些评价主要是形成性的，目的是在方案周期较早的时候向利益攸关方提供“前馈”，另一些评价主要是总结性的，目的是在方案周期较晚的时候对绩效进行更明确的评估，并提供详细的反馈。两种形式的平衡混合是理想的，许多评价提供了两种类型的见解。如图五所示，2024年的评价概况继续保持良好平衡。

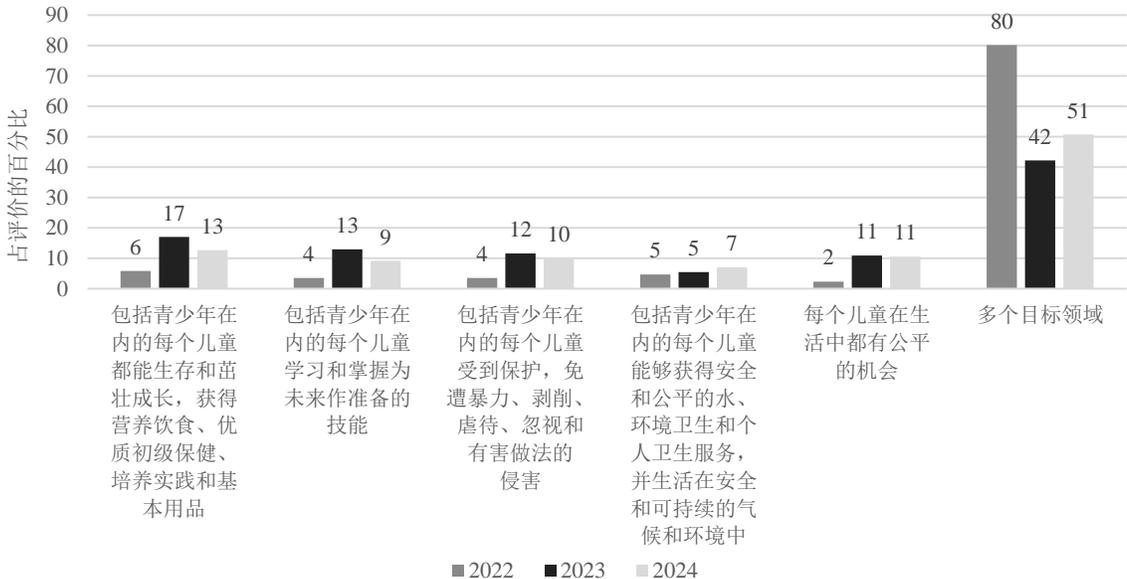
图五
2022-2024 年评价的主要目的



资料来源：全球评价报告监督系统。

52. 儿基会目标领域的覆盖面最好也能保持平衡。由于儿童发展的整体性——以及儿童所面临风险的交叉性，多种脆弱性来源导致风险增加，需要相应采取综合方案办法，因此，评价职能有责任确保评价以综合方式审查绩效，而不是延续各自为政的做法。2024 年，在扭转前一年多部门评价覆盖面下降的趋势方面取得了一定进展。

图六
2022-2024 年按儿基会目标领域分列的评价重点



资料来源：证据信息系统集成平台数据库。

2. 联合评价、机构间评价和国家主导的评价

53. 在确保对儿基会本身及其与合作伙伴的工作进行评价时，也最好保持平衡。2024 年，儿基会通过联合评价、机构间评价、全系统评价和国家主导的评价，增加对后一类评价的参与。与前一年类似，与世卫组织、人口基金、粮食署、联合国难民事务高级专员公署、驻地协调员办事处、国际移民组织以及联合国毒品和犯罪问题办公室等其他联合国机构在不同国家联合管理共计 12 项评价，其中包括机构间人道主义评价，涉及的专题领域包括儿童保护、童婚、女童教育和全球卫生等。除了对全系统评价办公室的一般性支助外，儿基会还支持具体的全系统评价，包括对致力于消除暴力侵害妇女和女童行为的聚光灯倡议和目前正在进行的联合国残疾包容战略的评价。

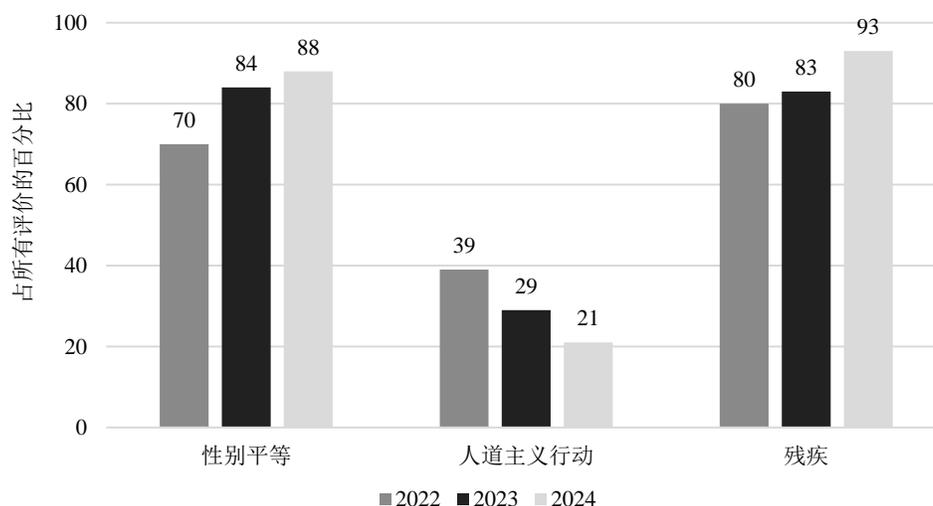
54. 2024 年，儿基会还支持了两项国家主导的评价：对黑山 2019-2023 年行使儿童权利战略的评价；对东帝汶 2019-2023 年营养方案的评价。

3. 评价对残疾、性别平等问题和人道主义行动的覆盖

55. 如图七所示，2024 年，系统纳入性别平等和包容残疾人情况的评价比例稳步提高。

图七

2022-2024 年覆盖残疾、性别平等和人道主义行动的评价

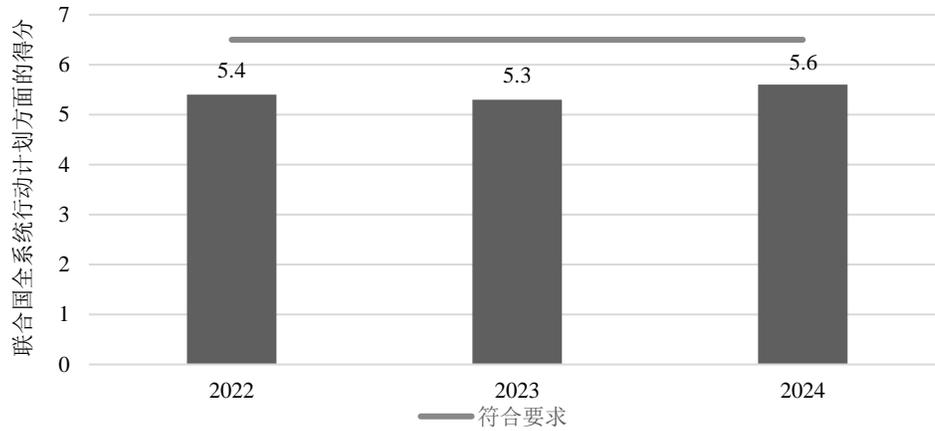


资料来源：全球评价报告监督系统。

56. 相比之下，在将性别平等考虑因素有意义地纳入儿基会评价方面取得的进展不足，在纳入《联合国全系统性别平等和增强妇女权能行动计划》（《联合国全系统行动计划》）方面，2024 年的得分仅略有增加，达到 5.6 分。在 2023 和 2024 年投入大量时间和精力来改善这一领域的绩效(如通过重点培训和网络研讨会以及在整个评价职能设立性别平等问题协调中心)之后，这一进展平淡无奇。2025 年及以后将继续努力，直至取得实质性和持续的改善。

图八

2022-2024 年评价在纳入《联合国全系统性别平等和增强妇女权能行动计划》方面的绩效



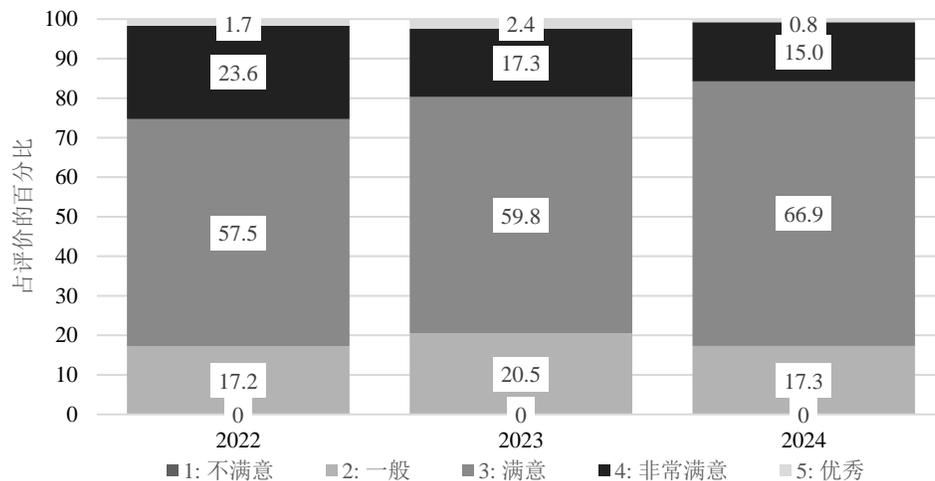
资料来源：全球评价报告监督系统。

C. 评价质量

57. 总体上质量保持令人满意至非常满意的水平，2024 年略有提高。被评为“非常满意”或“满意”的评价大幅增加，被评为“一般”的评价比例下降，没有评价被评为“不满意”。在这些广泛的模式中，质量仍然参差不齐。2024 年，各区域的全球评价报告监督系统的总体评分从 68.9 分到 83 分不等(满分 100 分)，按照 1 到 5 分的评分标准，质量评分中位数为 3 分(满意)。各区域的平均得分也各不相同，从最低的 2.6 分到最高的 3.3 分不等。这些差异突显需要有工具来确保评价做法的一致性，更加重视质量而不是数量，更加突出的优先次序，在人才和绩效管理过程中更好地利用全球评价报告监督系统的数据，并加强质量控制和质量保证过程。

图九

2022-2024 年总体评价质量评分



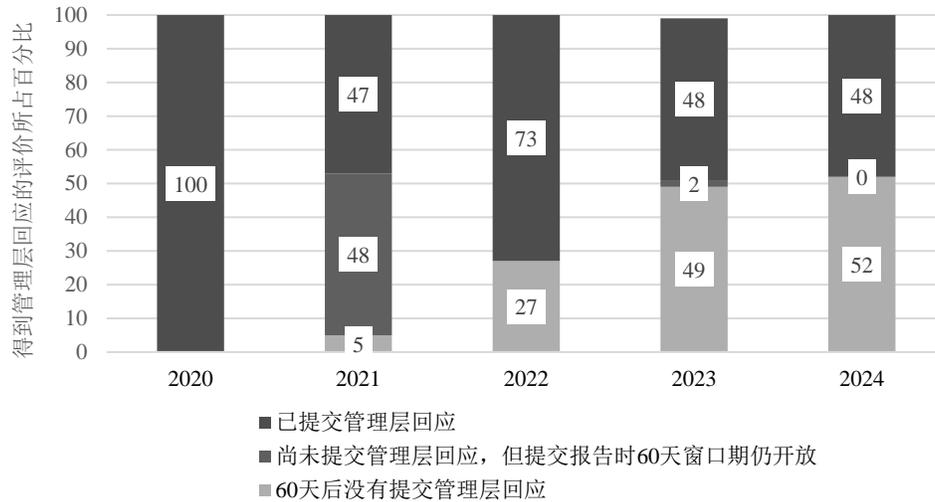
资料来源：全球评价报告监督系统。

D. 管理层提交和执行对评价的回应

58. 如图十所示，在过去两年中，管理层提交回应的及时性停滞在评价报告的一半左右，与 2022 年的近三分之二相比已大幅下降。评价职能将继续支持儿基会各级管理层按照评价政策的要求，在 60 天窗口期内提交管理层回应。

图十

2020-2024 年遵守管理层回应要求的及时性

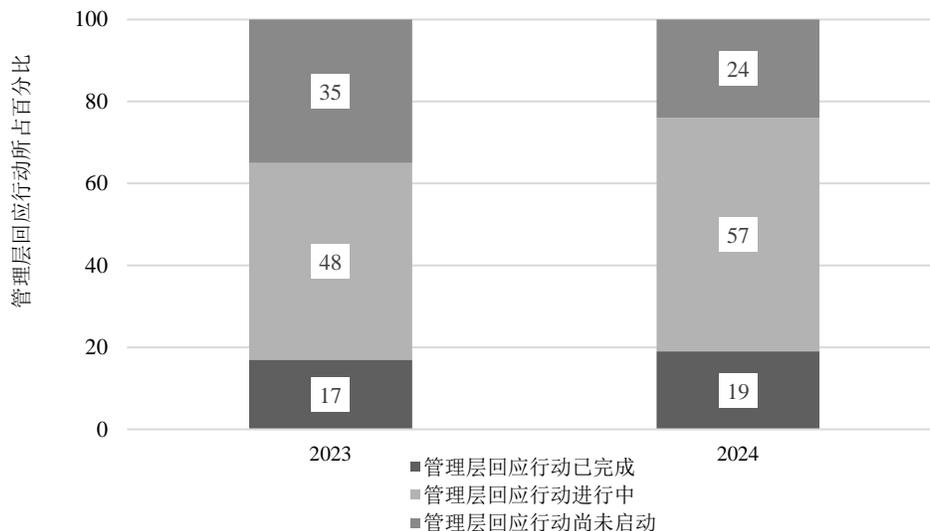


资料来源：证据信息系统集成平台数据库。

59. 相比之下，在提交的管理层回应中，管理层在采取行动和跟踪落实行动方面呈现积极趋势。此外，根本原因分析表明，未启动行动或行动被取消的主要原因是由于在与政府合作伙伴和其他利益攸关方进行有意义的接触时，在广泛协调、长期承诺和外部协作方面遇到挑战。

图十一

2023-2024 年管理层回应行动和执行情况



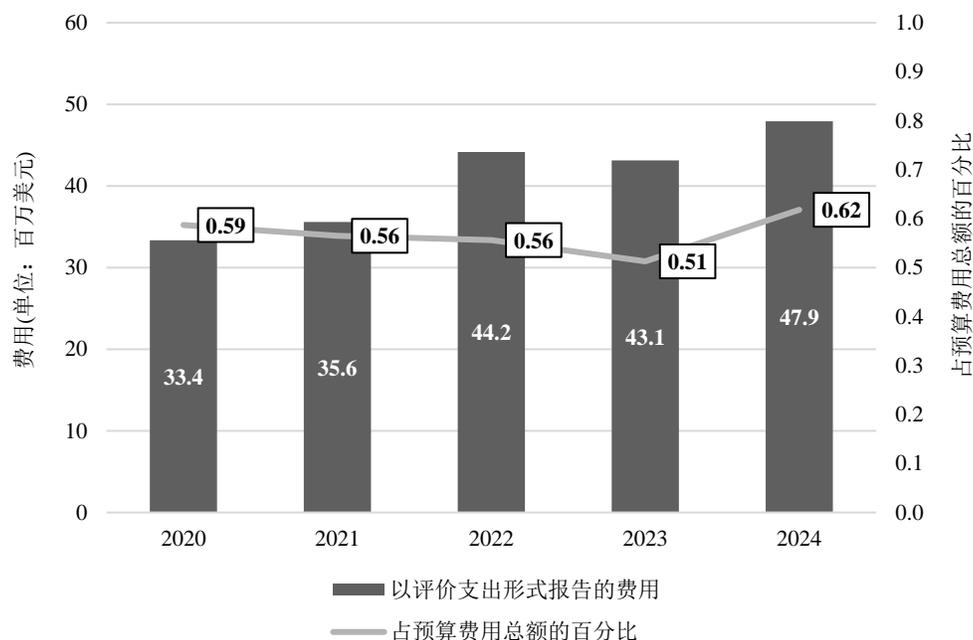
资料来源：证据信息系统集成平台数据库。

E. 以评价支出形式报告的财政资源和支出

60. 2024 年用于评价的方案支出在绝对值(归功于执行局批准提高评价集合基金的上限, 以及管理层随后向该基金追加拨款)和占方案支出总额的百分比(由于方案支出总额从 85 亿美元降至 78 亿美元)方面都出现了正增长。

图十二

2020-2024 年以评价支出形式报告的儿基会支出(单位: 百万美元)和占费用总额的百分比



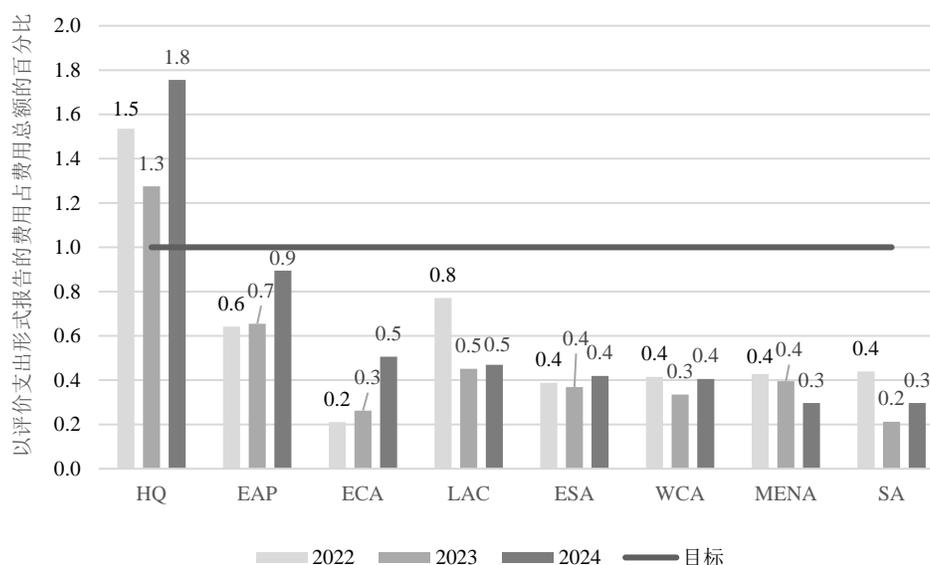
注：柱状图及其中的金额显示了每年以评价支出形式报告的费用(单位: 百万美元)。线条和数字表示每年以评价支出形式报告的费用占费用总额的百分比。

资料来源：分子数据：以评价支出形式报告的费用；分母数据：方案总费用净额；摘自虚拟集成信息系统——企业资源一体规划，2025 年 3 月 28 日。

61. 评价办公室是唯一一个评价支出超过 1%这一组织目标的单位，原因是为了履行 2024 年工作计划承诺而支取了分拨给 2025 年的评价集合基金资源。重要的是，这些拨款的很大一部分投入到分散一级，即通过加强政府能力（即通过评价领导者执行课程和加强评价系统的中级引导式在线课程）以及为区域和国家两级的 16 个职位提供资金。在分散一级，几乎所有区域都成功扭转了 2022 至 2023 年期间出现的支出下降趋势，这在很大程度上归功于政策实施的“软连接”要素，即评价领导层与区域和国家领导层在这一年中共同努力解决投入不足的问题。然而，尽管作出了这些努力，只有一个区域(东亚和太平洋)接近实现 1% 的目标，其他区域的支出水平仍远低于目标(即 0.3%至 0.5%)；虽然两个区域(东亚和太平洋以及欧洲和中亚)的投资随着时间的推移稳步增加，但其他区域的投资仍相对停滞。

图十三

2022-2024 年按区域分列以评价支出形式报告的支出占费用总额的百分比



说明：EAP，东亚和太平洋区域；ECA，欧洲和中亚；ESA，东部和南部非洲；HQ，总部；LAC，拉丁美洲和加勒比区域；MENA，中东和北非；SA，南亚；WCA，西部和中部非洲区域。

资料来源：虚拟集成信息系统——企业资源一体规划，2024年3月5日提取。

62. 与评价没有直接关系的所有支出的准确性均有所提高，但规划和方案审查除外，这两项支出的“不准确”编码百分比和美元金额都增加了一倍。评价办公室就活动编码开展了持续培训和监测，从而大大提高了准确性。尽管如此，错误编码仍占支出总额的 12.2%。

表 2

2023 和 2024 年以评价支出形式报告的活动，按活动类型分列

活动	2023 年支出 (单位：美元)	百分比	2024 年支出 (单位：美元)	百分比
与评价直接相关的活动	35 163 181	81.5%	42 071 555	87.8%
数据和调查	1 546 505	3.6%	221 046	0.5%
规划和方案审查	580 956	1.3%	1 259 217	2.6%
监测	935 320	2.2%	864 701	1.8%
研究	185 091	0.4%	158 623	0.3%
证据生成	372 498	0.9%	364 328	0.8%
系统强化	2 417 720	5.6%	1 829 042	3.8%
其他	1 945 589	4.5%	1 154 821	2.4%
共计	43 146 861	100.0%	47 923 330	100.0%

资料来源：2025 年 3 月 6 日从虚拟集成信息系统——企业资源一体规划中检索到的以评价支出形式报告的活动。

四. 结论和前进方向

63. 2024 年出现了一系列积极变化，首先是推出了经修订的评价政策及其众多措施，以优化评价职能，对儿基金会的工作产生最大的积极影响。此外，2024 年，几项关键绩效指标出现可喜的稳定或逆转趋势；这一进展归功于评价职能和儿基金会管理层从 2022 年开始并在 2023 和 2024 年继续作出协同努力，共同致力于需要改进的领域。

64. 这一年还出现了说明评价职能对儿童工作产生了重大和切实影响的许多实例。在一个层面上，儿基金会各级的一系列试金石评价展示出很高的价值——包括对儿基金会战略计划执行情况的评价，评价结果和建议被管理层系统和有意义地纳入即将出台的战略计划草案。在另一个层面上，2024 年，具有战略意义的重要举措——影响评价、国家评价能力发展和评价综合——得到了显著扩大，这些举措共同努力在系统层面上为儿童“带来积极影响”，远远超出儿基金会本身的影响范围。

65. 总之，2024 年的进步突出了评价作为一项投资的独特价值主张——这是一项小投资，目标仅占儿基金会方案预算总额的 1%，有助于儿基金会适应未来，实现最大绩效，为儿童实现更高的投资回报率，管理其风险，并增强合作伙伴对儿基金会的信任。至关重要的是，这些成果并不只是评价职能努力的结果，而是评价职能与儿基金会管理层密切协作的结果。

66. 在 2025 年及以后，需要继续并扩大这项工作，以保持 2024 年的积极势头并再接再厉。将扩大评价职能各级之间以及评价职能与管理层之间的协作，以确保继续加强评价职能及其有利环境。将特别关注提高整个评价职能绩效的一致性，更加注重成果层面的变化，并利用战略传播来加强对评价的使用和建立评价文化。将完成补充这些工作的三个具体工具，即政策实施战略、政策程序和评价手册。

67. 还将制定 2026-2029 年全球评价计划草案，供执行局在 2025 年审议。为此，将建立一个更强有力的整个评价职能一体联动和更健全的基于风险的系统，以加强独立性和公正性在评价规划中的作用，帮助确定最关键评价专题的优先次序，并最大限度地提高儿基金会的三个层面的评价工作的效率。

68. 所有这些努力都将在儿基金会面临前所未有的不确定性和变革的大背景下进行，首先是在 2025 年上半年进行一次全组织范围的全球审查。为适应这些挑战，评价职能还将侧重于：

(a) 通过向管理层提供相关评价意见，以纳入变革进程，协助儿基金会驾驭这些变革；

(b) 维护评价职能的公正性，确保其保持活力，保持其独立性(并根据执行局以往的决定予以加强)，并确保其作为具有学习和问责双重作用的监督职能的独特地位保持不变；

(c) 继续努力确保该职能在当前财政限制的情况下，能够获得充足且可预测的资源，包括通过补充评价集合基金和保留所有三个层面的人能；

(d) 重组评价职能本身，确定如何在更广泛的组织变革过程中提高效率和成本效益。

69. 根据执行局第 2023/20 号决定，将于 2025 年底委托进行一次独立评价，以评估经修订的评价政策，特别是其加强独立性的措施的实施情况。

五. 决定草案

执行局

注意到 2024 年儿基会评价职能年度报告(E/ICEF/2025/23)及其管理层的回应(E/ICEF/2025/24)。