



## 联合国儿童基金会

执行局

### 2023 年年会

2023 年 6 月 13 日至 16 日

临时议程\*项目 9

## 2022 年儿基会评价职能年度报告

### 摘要

本报告概述了 2022 年的儿基会评价职能，分析了业绩，着重介绍了一些敏捷性、创新性和适应性的例子，以及执行局先前提到的需要加快进展的两个领域的进展：影响评价和国家评价能力发展。

本报告还在可能的情况下确定了挑战和解决方案，借鉴了对儿基会评价职能的独立同行审查、关于儿基会实现儿基会《2022-2025 年战略计划》的定位的可评价性评估和形成性评价的结果和建议，并通过协商为审查和修订儿基会现行评价政策提供信息。

本报告最后提出了前瞻性的反思，包括关于支持儿基会日益重视在影响和结果层面为儿童确定成果而开展的新活动的反思。

第六节载列供执行局审议的决定要点。

\* E/ICEF/2023/9.

说明：本文件全文由儿基会负责处理。



## 一. 概述

1. 2022 年是评价职能演变的关键一年，涉及到《2022-2025 年战略计划》实施第一年的若干重大战略转变，以确保评价职能在本四年期及以后胜任其职。旨在打造“下一代”评价职能的一份愿景文件于 2022 年初定稿，当时还委托对评价职能进行独立同行审查，该审查为今年晚些时候对评价政策的审查和修订提供了依据，并将与本报告一起提交给执行局。关于儿基会实现儿基会《2022-2025 年战略计划》的定位的评价是联合国系统内首次开展此类评价，进一步推动了评价职能内部的变革。
2. 评价对儿基会监督和问责结构的贡献也得到了清楚阐明，并在创新评价方法的使用、国家评价能力发展和影响评价方面取得了重大进展——所有这些领域以前都被执行局列为需要加快进展的领域。
3. 在为评价职能设立的关键绩效指标方面，2022 年的进展情况参差不齐。一方面，这一年是有记录以来提交评价数量最多的一年，总体评价质量仍然很高，尽管由于选择了一家新的外部公司进行质量保证和采用了更严格的评估框架，质量有所下降。还实现了其他重要的里程碑，包括同时涵盖性别平等、人道主义行动和/或残疾包容问题的评价比例增加了；对人道主义行动的评价和联合评价的数量也增加了。
4. 另一方面，尽管在影响评价的总体战略方向上取得了进展，但影响评价的数量仍然很少。此外，将性别平等和增强妇女权能纳入评价分析范围的程度有所下降。最重要的是，在实现将总预算的 1%用于评价的目标方面取得的重大进展出现了明显逆转。

## 二. 背景和业务环境

5. 2019 冠状病毒病（COVID-19）疫情虽然仍然是儿基会业务环境的一个重要特征，但对 2022 年评价职能工作的制约较小。然而，就在一场全球危机开始消退之际，另一场危机进入了更严重的阶段：乌克兰危机在范围和规模上都有所扩大，要求儿基会采取紧急应对措施，同时对评价职能的工作产生了重大影响。阿富汗、埃塞俄比亚和也门正在发生的 3 级紧急情况也急需关注。
6. 在此背景下，儿基会开启了实施新的《2022-2025 年战略计划》的第一年。新战略计划的雄心是前所未有的，并高度重视在影响和结果层面为儿童取得结果，这反过来要求儿基会能够衡量这些层面的进展。随后，执行局在其第 2021/3、2021/10、2022/5 和 2022/13 号决定中连续强调了衡量影响级和结果级成果的重要性。执行局更加关注的其他领域，如国家评价能力发展以及问责制和监督，也为 2022 年评价职能确定优先事项提供了指导。
7. 2022 年也是评价职能发生更广泛变化的关键一年。独立同行审查在政策领域内外为评价职能提供了重要的见解。对评价政策的审查在修订过程中考虑到了这些见解和其他意见，修订后的政策草案将在本届年会上提交给执行局。
8. 影响 2022 年评价职能部门工作的另一个重要因素是总体组织资源环境，这将在第四章第 E 节讨论。

### 三. 评价职能的主要发展

#### A. 2019 冠状病毒病

9. 随着 2019 冠状病毒病疫情逐渐消退，组织运作逐渐正常化，儿基会和合作伙伴得以从关于应对工作的深入评价中获得重要的见解和经验。关于儿基会对 2019 冠状病毒病疫情的三级响应的评价评估了儿基会应对第一次真正的全球人道主义危机的表现情况。儿基会还为对 2019 冠状病毒病人道主义响应的机构间人道主义评价做出了贡献，该评价旨在评估机构间常设委员会成员机构在满足疫情造成的人道主义需求时的集体防范和应对情况。这两项评价都有助于为当前应对疫情的行动提供信息，并帮助儿基会及其合作伙伴为应对未来的突发卫生事件做好准备。

#### B. 乌克兰

10. 随着乌克兰危机加剧，评价办公室将资源转用于支持国家办事处的评价需求。这包括临时远程部署一名工作人员，支持及时建立必要的信息管理系统，以监测、报告并最终评估邻国应对难民外流的情况。评价大量采用二级数据，并应用自然语言处理技术分析来自会议纪要和人力资源监测工具等未被利用的、不同机构数据源的信息，从而避免给参与应急响应的外地工作人员带来不必要的负担。通过这一过程开发了用于评价的新型国家数据生成器原型工具，并借助该工具制作了互动的视觉产品，目前正在加以调整，以便在整个组织内使用。

11. 此外，在乌克兰国家办事处设立了一个 P-5 级评价专家主任职位，加强了国家层面的评价能力。由此，在总部和区域办事处两级评价人员的支持下，对乌克兰的应对工作进行了业务审查。

#### C. 战略和愿景

12. 随着儿基会开始实施新的战略计划，评价办公室开展了关于儿基会实现儿基会《2022-2025 年战略计划》的定位的可评价性评估和形成性评价，其结果已提交给执行局 2023 年第一届常会（[E/ICEF/2023/3](#)）。这项工作确定了 12 项关键建议，以改善准备情况，从而最大限度地提高儿基会实现其宏伟目标和衡量其进展的可能性。儿基会在其管理层回应（[E/ICEF/2023/4](#)）中接受了所有建议，所有建议也反映在执行局第 2023/3 号决定中，该决定还要求儿基会进一步拟定一项行动计划来执行这 12 项建议。儿基会执行主任的 2022 年年度报告（[E/ICEF/2023/10](#)）进一步着重介绍了评价结论。

13. 儿基会关于“下一代”评价职能的愿景文件也于 2022 年初定稿。除了提出总体愿景，即“始终为儿基会及其合作伙伴利用严格的、按战略优先顺序排列的评价证据，以在‘行动十年’中实现每个儿童的权利”之外，该文件还概述了将这些原则付诸实践的初步计划。这一愿景的一个关键组成部分是确保整个评价职能更加协调一致。在这方面已经采取了重大措施，包括在评价职能的所有三个级别举行例行会议，讨论共同战略利益和潜在的联合工作规划等问题。自 2019 冠状病毒病疫情暴发以来的首次面对面全球评价会议也于 2022 年 11 月举行。

14. 在这一内部愿景规划工作的基础上，根据 2018 年经修订的评价政策（E/ICEF/2018/14）第 68 段的授权，评价办公室于 2022 年着手对儿基会评价职能进行了两次进一步的审查：对评价职能的独立同行审查以及对评价政策的审查和修订。

15. 同行审查由一个外部小组和辅助咨询小组领导，目的是向儿基会提供一个独立、公正和基于证据的评估，评估儿基会评价职能的现状，以及评价职能在多大程度上最适合在儿基会内发挥学习和问责作用。正如在向执行局 2023 年第一届常会提交的最新情况报告（E/ICEF/2023/6）所指出的那样，同行审查承认，自 2017 年上一次同行审查以来，儿基会取得了重大进展，并就如何在评价职能内和更广泛的组织内进行改进提供了具体建议。其中包括确保对作用和责任（包括供资和报告关系）的问责和监测，这可能会影响独立性、可信度和效用的持续实现。查明的其他风险涉及对评价的利用有限，以及评价职能和儿基会其他知识职能之间的互动有限。还确定评价办公室需要通过扶持环境得到更多赋权和支持，包括儿基会在其他职能内的领导和管理。最终报告和管理层回应的副本将在联合国评价小组和儿基会网站上公布。

16. 在进行同行审查的同时，对儿基会 2018 年评价政策的审查和修订也已开始。目的是批判性地评估现行政策在多大程度上帮助评价职能有效履行职责，以及评价政策是否依然适用。在评价办公室的领导下，经与主要利益攸关方协商，这一进程在有条不紊、务实和循证的方法指导下进行。此次修订虽然保留了 2018 年政策的核心要素，但是对一些要素进行了修订，以加强评价职能有效履行职责的能力。

17. 与这些评价相关，并且为了配合执行局对问责制和监督的日益重视，评价职能在儿基会最新的问责制度（E/ICEF/2022/24）中确定了其关键的监督作用，包括为系统遵守良好做法以及声称的成果和良好管理的有效性提供独立保证。评价办公室的问责作用由评价职能在整个组织范围内实施，包括为国家、区域和全球各级的监督进程作出贡献。还指出了一系列的评价作用和对儿基会学习目标的贡献。

#### D. 影响和结果层面的评价

18. 针对执行局关于在全球范围内加强对影响和结果层面评价的重视的决定，以及 2021 年多边组织业绩评估网（MOPAN）评估提出的“就儿基会方案编制的效率和可持续性提供更有力的评价证据”的建议，评价办公室最终确定了其 2022-2025 年期间的[影响评价战略和行动框架](#)。该战略是通过广泛的内部和外部磋商制定的，是儿基会在努力通过提供结果和影响层面的有力评价证据，为国家计划和以儿童为重点的政策提供信息时的实用参考。该战略概述了相互关联的战略支柱，目的是(a)加强影响评价的启动并增加其覆盖面；(b)使方法多样化并进行创新；(c)改善学习和方案的协同作用，以及伙伴关系和资源需求。

#### E. 伙伴关系

19. 注意到执行局对加强机构间合作的重视（E/ICEF/2021/3），相关工作已在 2022 年开始，以优先考虑战略评价伙伴关系，首先是制定一项专门战略。儿基会继续共

同主持 EVALSDG，<sup>1</sup>积极配合机构间常设委员会关于机构间人道主义评价的工作，<sup>2</sup>并联合开展人道主义部门以外的许多其他机构间评价。2022 年，儿基会还探讨了成为全球评价倡议<sup>3</sup>正式合作伙伴的可能性，并将在 2023 年进一步确定其可为其他国际伙伴关系创造的附加价值，以促进国家评价能力发展和证据综合活动。

20. 得益于德国联邦经济合作与发展部对儿基会评价职能的长期支持，评价办公室与方案小组（目标领域 5）、儿基会因诺琴蒂-全球研究和展望办公室以及数据、分析、规划和监测司合作，在公共伙伴关系司的支持下，制定并签署了一项 900 万美元的协议，以发起关于脆弱环境中儿童适应性社会保护的全面证据生成伙伴关系。

21. 儿基会还继续参与更广泛的证据生态系统，加强评价以外的伙伴关系。

## F. 国家评价能力发展

22. 为响应 MOPAN 的建议和执行局最近呼吁加快加强国家评价能力的决定<sup>4</sup>，评价办公室在 2022 年推出并恢复了几项重要举措。这些举措包括推出评价领导人执行课程（ExCEL），这是一门由新加坡国立大学讲授的、为期两周的面对面课程。2022 年在新加坡举办了二期培训班，2023 年 3 月与欧洲大学学院跨国治理学院联合举办了第三期培训班。这些培训班汇集了来自 29 个国家的代表团，共有 101 名参与者，其中 65 名是高级别政府合作伙伴。每个代表团都制定了自己的行动计划，其中包括国家评价法案、法律、政策和计划方面的倡议；制定或修订国家评价指南；开展提高认识运动；开展新评价的计划；以及与国家评价能力发展有关的后续活动。

23. ExCEL 旨在加强对评价概念和方法的理解，在高级政府官员中树立使用评价的信心并打造评价倡导者社区，并加强国家内部和国家之间的评价伙伴关系。作为培训后支持的一部分，将在多个领域举办 70 场面向 ExCEL 毕业生及其国内同事的 ExCEL 深度网络研讨会，以确保课程产生最大的实际积极影响。ExCEL 的势头将在 2023 年继续，旨在加强评价系统的在线中级引导式课程的势头也将继续。

24. 创新的国家评价能力发展举措也在区域一级涌现。例如，西非及中非区域办事处评价小组和当地合作伙伴共同组建了一个由非洲法语区 26 所大学组成的联合会，合作伙伴通过该联合会共同开设了一门为期一年的区域评价硕士课程。院长、评价教授和高年级学生合作进行了为期六个月的对话，最终开发了 18 个评价模块。计划在 2023 年中期推出该课程，目标是培训 1000 多名学生。该举措还激发非洲教授和学生纷纷撰写关于评价相关问题的文章和博客，使目前关于国家评价能力发展的讨论有了更多样化的声音。

<sup>1</sup>见 <https://evalsdgs.org/>.

<sup>2</sup>见 <https://interagencystandingcommittee.org/inter-agency-humanitarian-evaluations>.

<sup>3</sup>见 [www.globalevaluationinitiative.org/](http://www.globalevaluationinitiative.org/).

<sup>4</sup>见 E/ICEF/2022/27.

## G. 面向下一代评价职能的创新

25. 持续的技术和方法进步也影响了 2022 年评价职能的工作。除了在乌克兰使用实时数据方面的创新之外，欧洲和中亚评价小组与儿基会创新办公室合作，启动了几项数字化随机对照试验，以评估数字化解决方案对区域方案交付的影响。该小组还探索了心理健康方面的大数据，应用行为洞察力模型来评价干预措施。

26. 评价职能还开始利用纵向二级数据源来审视关于儿基会在阿富汗的三级人道主义响应的评价结果，开始探索利用人工智能来合成专题学习，并利用大数据进行社会倾听分析。此外，由评价办公室牵头的、通过移动电话调查开发实时高频数据这一工作已在各区域展开，以协助快速生成方案洞察力和证据，用于跨部门方案编制。

27. 针对执行局和 MOPAN 提出的做更多工作以“促进成果层面的转型和系统性变革”的建议，<sup>5</sup>相关工作已经开始，旨在对儿基会《2018-2021 年战略计划》和《2022-2025 年战略计划》第一年期间开展的各类儿基会评价产生的证据进行系统化的综合集成。其目的是概述儿基会支持的儿童和青少年干预措施的成果和影响，确定和分析促成或阻碍实现组织业绩的关键因素，并确定需要组织采取行动的经常性建议。该综合集成使用应用型人工智能来标记和搜索评估报告中的术语。

28. 与此同时，作为进一步加强组织评价能力的持续努力的一部分，评价办公室开发了一系列方法简报会和网络研讨会，涵盖过程跟踪和贡献分析等主题。在分散层面，西非及中非区域办事处评价小组与其他区域合作，制定了参与式儿童工作评价方法的方法指南，包括创新的照片-语音方法。正在通过评价办公室内部 SharePoint 门户网站上的“方法库”整合资源。

## H. 人力资源

29. 2022 年还招聘了关键工作人员，包括最近设立的首席顾问评价职位和新的机构效力评价主任。此外，为了加强跨职能流动，包括跨证据职能的流动，评价办公室主持并参加了几次延展性任务。

30. 评价办公室还与 24 个组织最终确定了一项新的长期协议，大幅扩大了来自全球南方的优质评价公司库，并带来了影响评价和可评价性评估等领域的具体技能。

## 四. 业绩概览：评价职能的持续进展和主要经验教训

### A. 提交的评价数量和地理覆盖范围

31. 2022 年，提交的评价数量增加了约 12%，而地理覆盖范围与 2021 年大致持平。在前几年势头的基础上，评价职能提交了有史以来数量最多的评价报告。

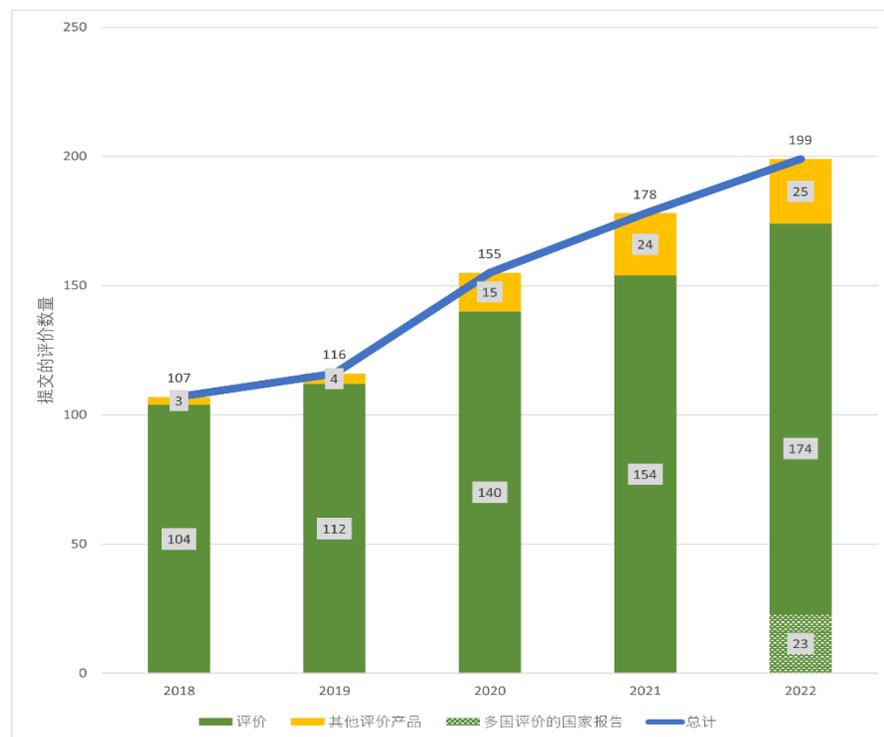
32. 评价工作在儿基会的三个层面开展，其中大部分在国家层面开展。在 2022 年提交的 199 个评价产品中，<sup>6</sup>有 5 个在总部层面开展，194 个在分散层面开展。

<sup>5</sup>多边组织业绩评估网，“2021 年年度报告”，MOPAN，巴黎，2022。

<sup>6</sup>另一项由总部牵头的机构间评价已于 2022 年完成，但在起草本报告时尚未上传到 EISI 系统或进行质量评估。这一点在有关 2022 年完成的评价数量的后续数字或其他引用中没有得到说明。

33. 如图 1 所示，2022 年产生的评价产品比 2021 年多 21 个。这一增长部分是因为全球疫情期间启动的 39 个评价产品在 2022 年才完成，部分是因为 2022 年开展了更多的多国评价，产生了针对具体国家的评价报告。

图 1  
2018-2022 年提交的评价数量



资料来源：证据信息系统集成（EISI）。

34. 这一增长也是由于其他评价产品的数量略有上升。2022 年进行了 10 次可评价性评估，而 2021 年为 6 次。这些评估力求通过衡量方案的准备程度来加强方案的规划和设计，以便在未来进行有意义的评价。2022 年没有进行任何审查。提交了 15 份实时评估，比 2021 年多了一份。其中 11 项与 2019 冠状病毒病有关，其他几项专门涉及中美洲飓风应对、少女怀孕、感染艾滋病毒/艾滋病的年轻人和就业能力。

35. 如表 1 和图 2 所示，2022 年，评价职能各级都出现了这两种趋势——提交的评价数量增加和评价产品更加多样化。

表 1  
2021 年和 2022 年提交的评价和其他评价产品的数量，按地点分列

地点	评价		其中包含 的多国评 价国家报 告的数量	可评价性评估		审查		实时评估		评价产品总数	
	2021	2022	2022	2021	2022	2021	2022	2021	2022	2021	2022
EAP	12	23	1	1	2	0	0	2	3	15	28
ECA	19	16	3	1	3	0	0	2	3	22	22
ESA	23	24	0	0	0	0	0	5	0	28	24
HQ <sup>a</sup>	7	4	0	2	1	2	0	3	0	14	5
LAC	22	27	7	0	0	0	0	1	9	23	36
MENA	20	21	0	0	0	0	0	1	0	21	21
SA	25	18	0	1	2	2	0	0	0	28	20
WCA	26	41	12	1	2	0	0	0	0	27	43
总计	154	174	23	6	10	4	0	14	15	178	199

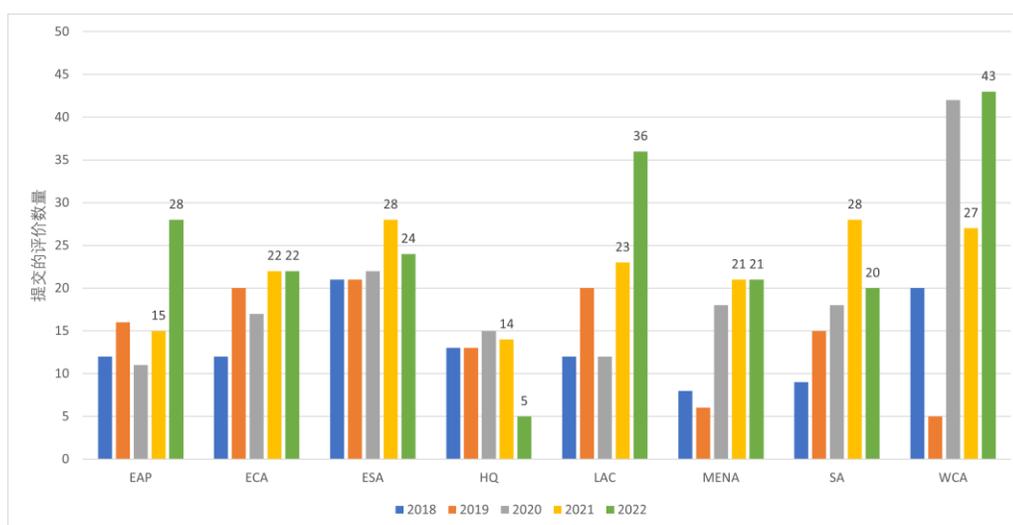
<sup>a</sup> 另一项由总部牵头的机构间评价已于 2022 年完成，但在起草本报告时尚未上传到 EISI 系统或进行质量评估。这一点在有关 2022 年完成的评价数量的后续数字或其他引用中没有得到说明。

图例：EAP：东亚和太平洋地区；ECA：欧洲和中亚地区；ESA：东部和南部非洲地区；HQ：总部；LAC：拉丁美洲和加勒比地区；MENA：中东和北非地区；SA：南亚地区；WCA：西部和中部非洲地区。

资料来源：EISI。

36. 东亚和太平洋地区、拉丁美洲和加勒比地区以及西部和中部非洲地区的提交量大幅增加。然而，在西部和中部非洲地区，有几份多国评价的国家报告——在拉丁美洲和加勒比地区也是如此，但程度较轻。虽然中东和北非地区以及欧洲和中亚地区提交的评价产品数量与去年相同，但东部和南部非洲地区略有下降，南亚地区以及总部则大幅减少。第五节探讨了总部一级提交量减少的可能原因。

图 2  
2018-2022 年在分散层面提交的评价产品数量，按地点分列

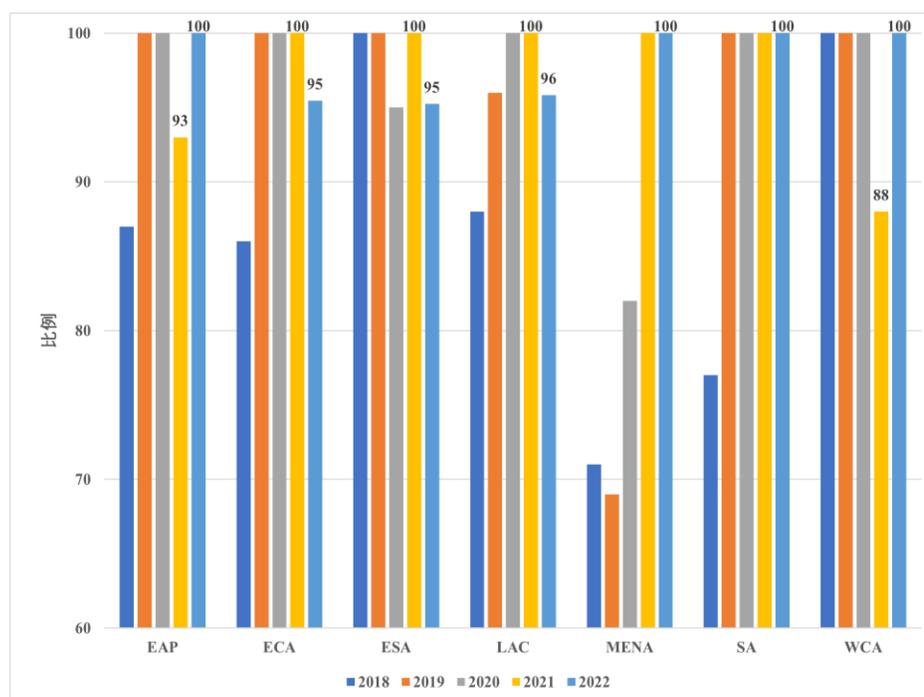


图例：EAP：东亚和太平洋地区；ECA：欧洲和中亚地区；ESA：东部和南部非洲地区；HQ：总部；LAC：拉丁美洲和加勒比地区；MENA：中东和北非地区；SA：南亚地区；WCA：西部和中部非洲地区。

资料来源：EISI。

37. 如图 3 所示，2022 年（98%）的总体地理覆盖率（指过去三年中进行过评价的国家和区域办事处）与 2021 年（97%）几乎相同。这要归功于多国评价（即各办事处汇集财政和人力资源开展评价），使评价工作得以在资源有限的情况下进行。在七个地区中，有两个地区的覆盖率仍为 100%（中东和北非以及南亚），而两个地区的覆盖率在 2022 年提高到 100%（东亚和太平洋地区，之前为 93%，西部和中部非洲地区，之前为 88%）。在三个地区，覆盖率从 100% 略微下降：欧洲和中亚地区以及东部和南部非洲地区降至 95%，拉丁美洲和加勒比地区降至 96%。

图 3  
2018-2022 年各区域的评价覆盖率

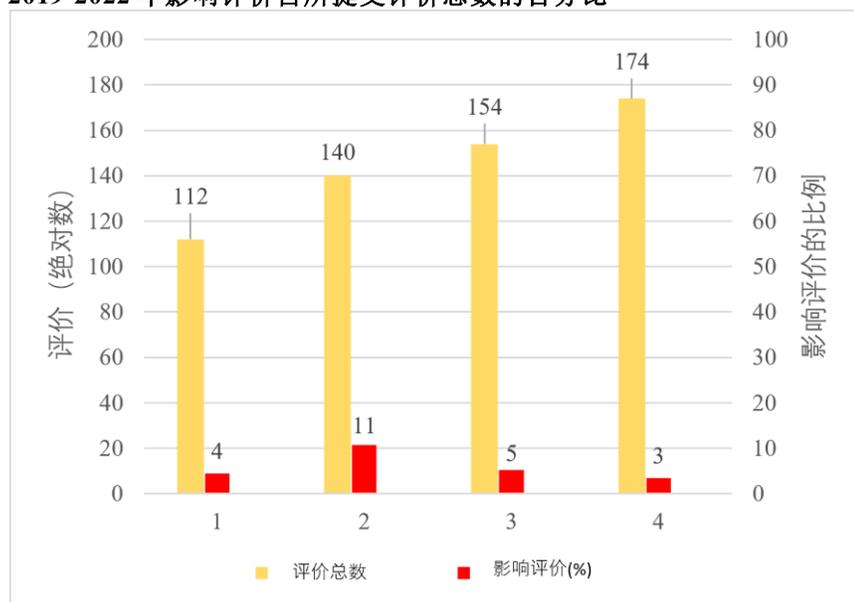


图例：EAP：东亚和太平洋地区；ECA：欧洲和中亚地区；ESA：东部和南部非洲地区；HQ：总部；LAC：拉丁美洲和加勒比地区；MENA：中东和北非地区；SA：南亚地区；WCA：西部和中部非洲地区。  
资料来源：EISI。

## B. 评价对可持续发展目标的影响和集体贡献

38. 如图 4 所示，对儿基会方案干预措施进行影响的总体覆盖率仍然很低，影响评价严格地评估了创新型、可扩展型干预措施的有效性。2022 年，总共有 3% 的评价采用了可信的反事实设计。过去四年的观察结果表明，社会保护领域的影响评价证据继续在 2022 年的主题分布中占主导地位，三分之二的的影响评价是在这个方案领域开展的。社会保护影响评价的结果在扩大有效的儿童补助金方案方面产生了影响。整个评价职能部门正在加强能力以开展更强有力的影响评价，以帮助在包括儿童保护和营养在内的其他方案领域打造严格的评价证据基础，但还需要投入更多的资源。

图 4  
2019-2022 年影响评价占所提交评价总数的百分比



资料来源：全球评价报告监督系统(GEROS)。

39. 2022 年产生的评价继续很好地结合了回顾过去（总结性）评价和展望未来（形成性）评价，尽管纯形成性评价较少。近一半（47%）的评价既是总结性的，又是形成性的，而大约三分之一（32%）的评价是纯形成性的。五分之一（20%）是纯总结性评价，还有一次元评价。

表 2

2021 年和 2022 年开展的评价，按类型分列

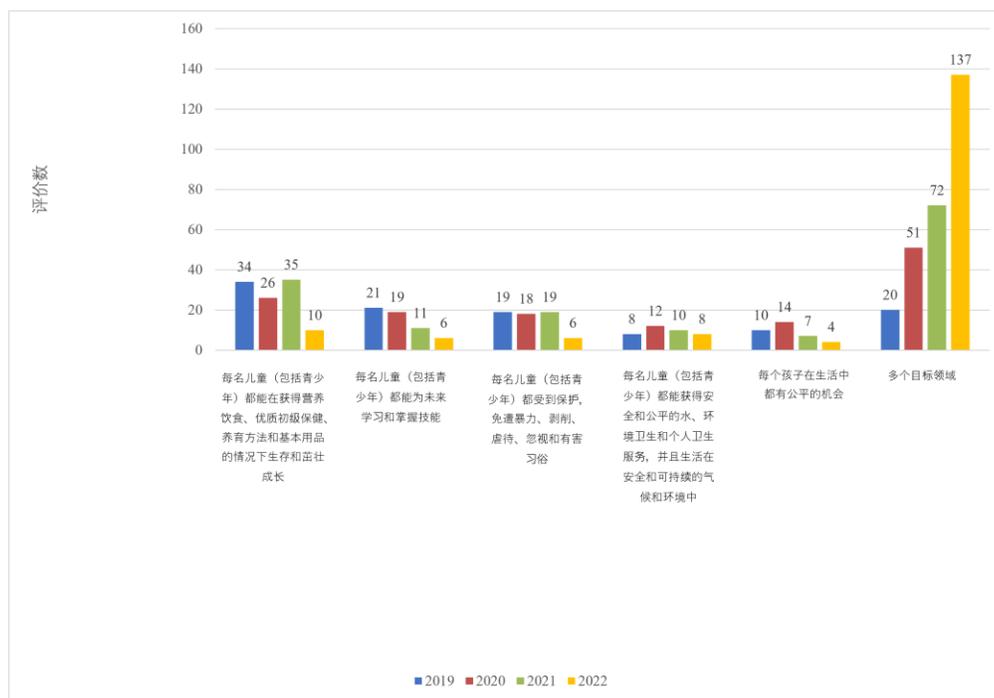
评价类型	评价数		占有所有评价的比例	
	2021	2022	2021	2022
形成性	53	56	34	32
总结性	19	35	12	20
总结性和形成性	81	82	53	47
元评价	1	1	1	1

资料来源：GEROS。

## 1. 目标领域分布

40. 自 2019 年以来，随着区域和国家办事处转向在分散层面进行更具战略意义的评价（这些评价反映了旨在确保广泛覆盖所有目标领域的 2022-2025 年全球评价计划（E/ICEF/2022/3）），覆盖儿基会战略计划多个目标领域的评价数量逐年稳步增加（从 2019 年的 20 个增加到 2022 年的 137 个）。

图 5  
2019-2022 年评价的专题覆盖面<sup>a</sup>



<sup>a</sup> 儿基会先前战略计划中的目标领域略有不同：1. 每个孩子都能生存并茁壮成长；2. 每个孩子都能学习；3. 每个孩子都受到保护，免遭暴力和剥削；4. 每个孩子都生活在安全和清洁的环境中；5. 每个孩子在生活中都有公平的机会

## 2. 联合评价和机构间评价

41. 2022 年，全组织共有 36 名儿基会评价工作人员积极参加了 20 个联合国评价小组工作组和兴趣小组，涵盖了众多专题和部门。儿基会继续致力于参与联合国评价小组，如支持制定新指导和更新指导，推动学习，并与更广泛的联合国评价界分享良好的评价做法。

42. 此外，自 2018 年以来，与一个或多个联合国机构联合管理的评价数量有所增加。在 2022 年与联合国其他机构联合管理的 14 项评价（2021 年为 10 项）中，东亚和太平洋地区以及欧洲和中亚地区各 4 项，东部和南部非洲地区 3 项，拉丁美洲和加勒比地区以及中东和北非地区各 1 项，全球一级有 1 项。西部和中部非洲地区以及南亚地区在 2022 年没有开展任何联合评价。

43. 2022 年，儿基会继续支持各国政府在其自愿国别评估中使用评价证据，并支持在国家主导的可持续发展目标评价中使用评价证据。例如，儿基会支持尼日利亚政府在尼日利亚开展由国家主导的、关于可持续发展目标 3 和 4 的有效性和影响的评价。评价办公室作为多利益攸关方 EvalPartners 工作队的成员，应尼日利亚常驻联合国代表团的要求，协助起草了一份关于通过评价加强自愿国别评估的联合国大会决议。评价办公室还与世界粮食计划署和 EvalPartners 工作队的其他成员在 2022 年可持续发展问题高级别政治论坛上共同主办了一个虚拟自愿国别评估实验室，主题是“国家主导的评价如何为可持续发展目标提供循证自愿国别评估”。2022 年 6 月，儿基会连续第二年与芬兰外交部、国际环境与发展研究所、国际发展评价协会、德国发展评价研究所和 EVALSDG 一起，举办了一次主题为“将国家优先事项与可持续发

展目标联系起来的评价”的在线研讨会。今年的研讨会是为亚洲各地的政府官员量身定制的。

### 3. 评价对残疾、性别平等问题和人道主义行动的覆盖

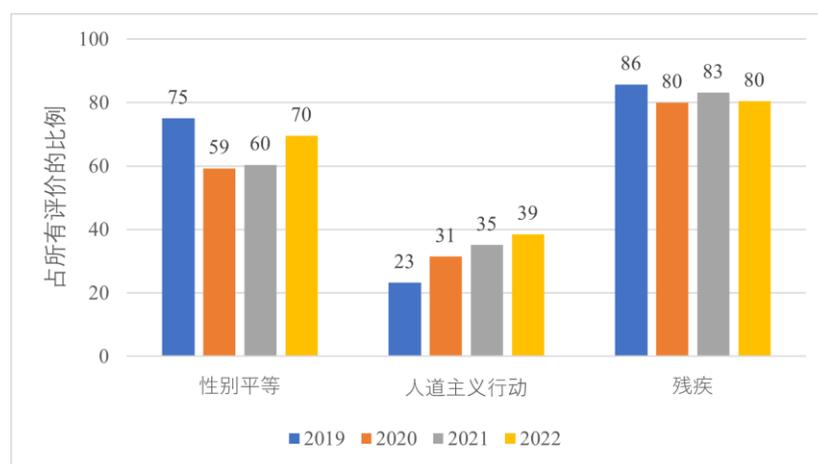
44. 儿基会在其评价中特别注意确保对残疾、性别平等问题和人道主义行动的覆盖。自 2019 年以来，残疾覆盖率基本稳定在 80% 至 86% 左右，与 2021 年相比，2022 年小幅下降了 3%。儿基会继续向各评价小组简要介绍评价设计、结论和建议必须如何在所有政策和方案中考虑残疾问题。要求评价人员与残疾人和/或其家人进行焦点小组讨论，并编制按残疾分类的数据，以确保没有任何一个人掉队。

45. 2022 年，儿基会在实现联合国残疾包容战略指标 10 方面取得了显著进展，该指标衡量在评价的所有阶段考虑残疾包容的程度。评价办公室对 2016 年至 2021 年的发展和人道主义评价中的残疾包容情况进行了综合集成，并在整个评价过程中为评价职能制定了兼顾残疾人的方法指南。因此，儿基会现在是指标 10 评级从“接近要求”上升到“符合要求”的极少数机构之一。该指南在内部和外部通过全球残疾问题行动网络和社交媒体进行了广泛介绍，得到了好评。这些努力有望纠正 2022 年出现的小幅下降。

46. 如图 6 所示，虽然涵盖残疾问题的评价比例略有下降，回到了 2020 年的水平，但 2022 年涵盖性别平等问题的评价比例增加了 10%，涵盖人道主义行动这一跨领域主题的评价比例增加了 4%。评价职能继续恪守评价小组的行为标准，让受影响的人群参与进来，告知他们自己的权利和应享待遇，并让他们作为主要利益攸关方参与制定建议。

图 6

2019-2022 年涵盖残疾、性别平等问题和人道主义行动的评价比例



资料来源：GEROS。

47. 虽然在过去几年中，在将性别平等和增强妇女权能纳入评价分析范围（评价目标、标准、评价框架中的问题/指标）方面实现了持续改进，但在 2022 年，儿基会在《联合国全系统性别平等和增强妇女权能行动计划》（UN-SWAP）中的表现有所下滑。得益于对儿童基金会《2018-2021 年性别行动计划》的评价，评价工作的总体业

绩在 2019 年大幅增长，并在 2021 年前“超过要求”。然而，如下图 7 所示，2022 年评价组合的总体业绩却略低于“符合要求”。

图 7

### 《2018-2022 年联合国全系统性别平等和增强妇女权能行动计划》中的评价业绩<sup>a</sup>

<sup>a</sup> 评估报告的 UN-SWAP 最低得分必须为 9 分，才能“超过要求”，或 6.5 分才能“符合要求”。

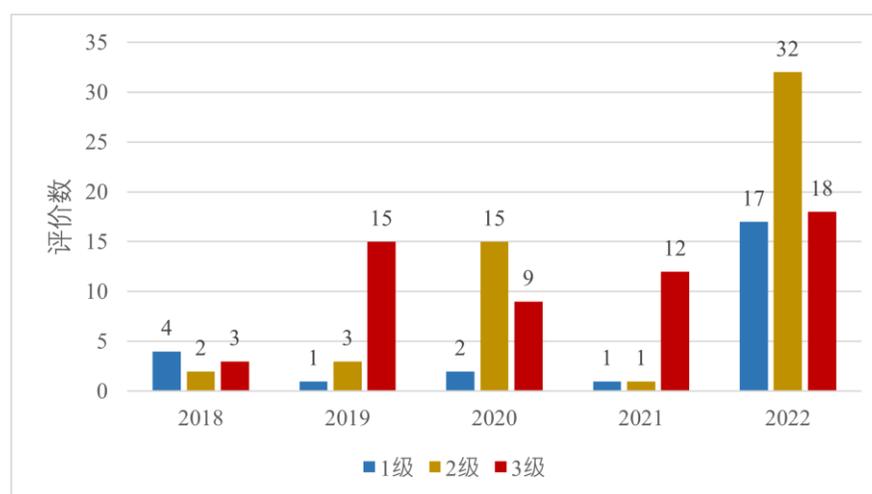


资料来源：GEROS。

48. 评价职能在 2022 年仍然侧重于对已宣布的紧急情况的评价。如图 8 所示，对人道主义行动的评价比往年多得多，包括在目前参与应对三级紧急情况的国家进行了 18 次评价，而 2021 年为 12 次。对二级紧急情况的评价也大幅增加（33 次，而 2021 年为 1 次，2020 年为 15 次）。

图 8

### 2018-2022 年进行的人道主义评价数量，按紧急情况级别分列

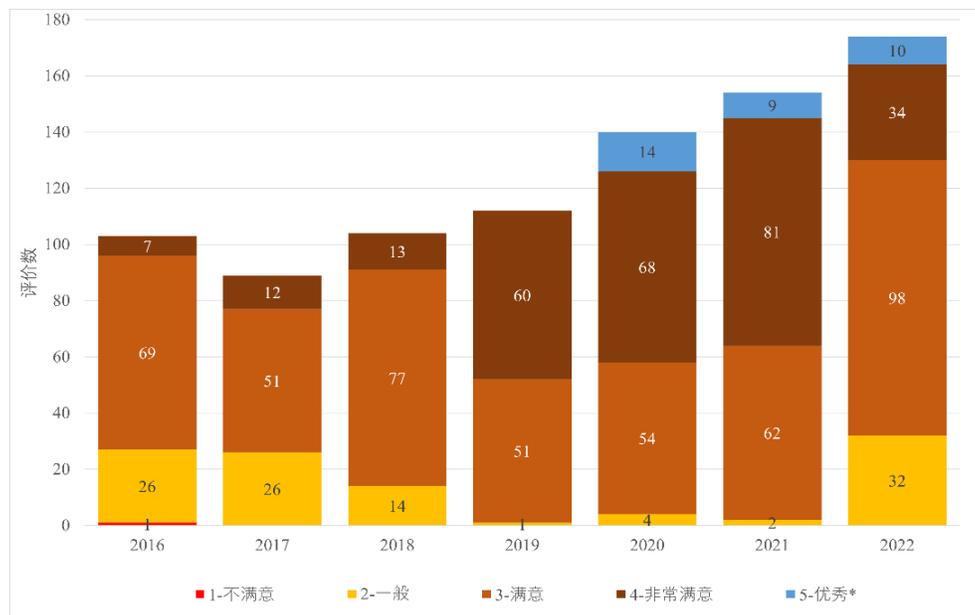


资料来源：GEROS。

## C. 评价质量

49. 虽然评价质量在 2019-2021 年期间一直保持较高水平，但在 2022 年却出现了下降。在外部公司利用全球评价报告监督系统（GEROS）独立评估的 174 项评价中，3 项被评为“优秀”，而 2021 年为 9 项；41 项为“非常满意”，而 2021 年为 81 项；100 项为“满意”，而 2021 年为 62 项。30 项评价被评为“一般”，而 2021 年只有 2 项。没有任何评价得到“不满意”的评级，这一积极趋势自 2017 年以来一直持续至今。

图 9

2016-2022 年总体评价质量评级<sup>a</sup>

<sup>a</sup> “优秀”类别是在全球评价报告监督系统审查后于 2020 年周期推出的。  
资料来源：GEROS。

表 3

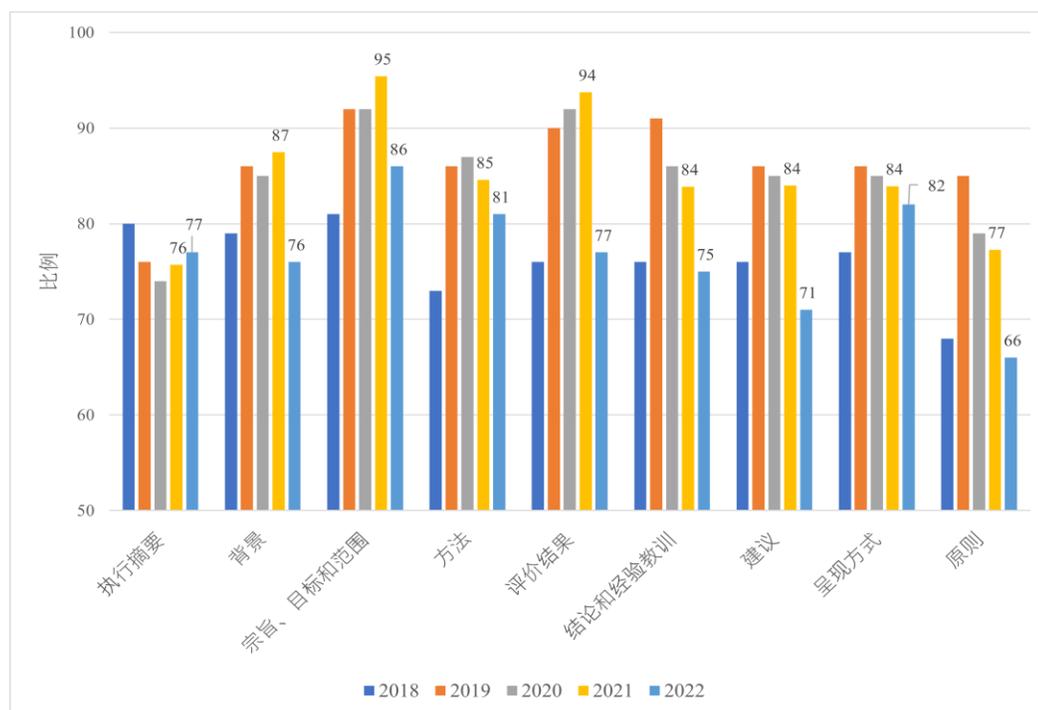
2016-2022 年评价质量评级的平均分和中位数得分

年份	2016	2017	2018	2019	2020	2021	2022
平均得分	2.80	2.84	2.92	3.53	3.66	3.63	3.13
中位数得分	3.00	3.00	3.00	4.00	4.00	4.00	3.00

50. 这一质量下降的趋势主要是由于 2022 年推出了更严格的评估框架，并由签约评估儿基会评价报告的一家新的外部公司进行了试点。更严格的质量评估将有助于确保所有评价逐步达到更高的标准，评价职能继续得到改善。

51. 下图 10 中更详细的分类显示，更严格的评估令评价报告的所有关键要素得分都有所下降。

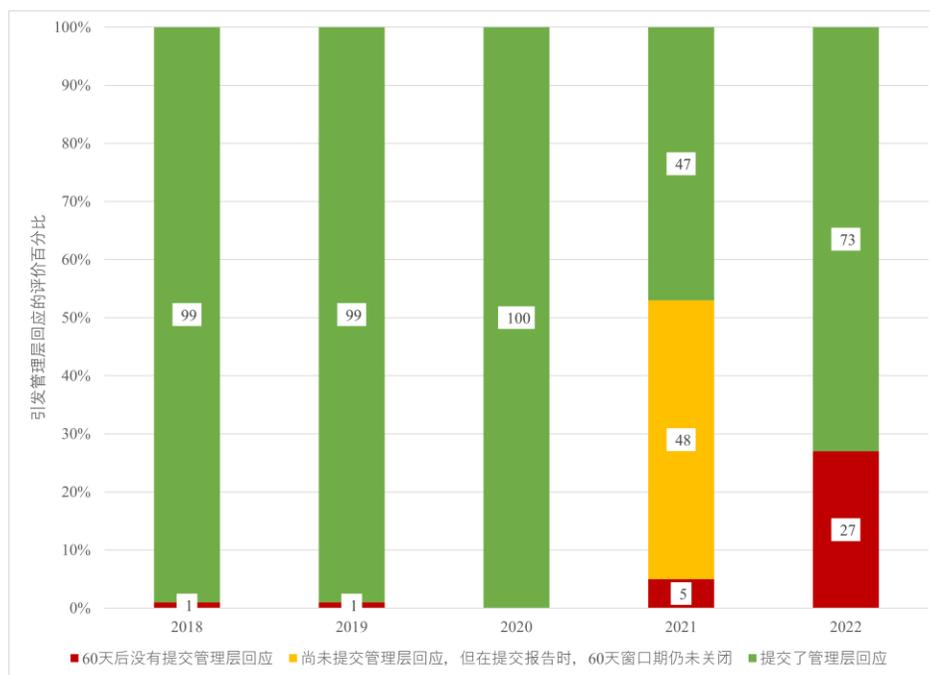
图 10  
2018-2021 年评价质量评级平均得分，按报告章节分列



#### D. 评价管理层回应的提交和执行

52. 在 2022 年进行的需要管理层回应的 165 项评价中，121 项按时提交（73%）。有 44 份管理层回应（27%）已经超过了 60 天期限，现已逾期，比 2021 年的 5% 大幅增加。然而，较早年份（2018-2020 年）的趋势显示，几乎所有的评价最终都会得到管理层回应。

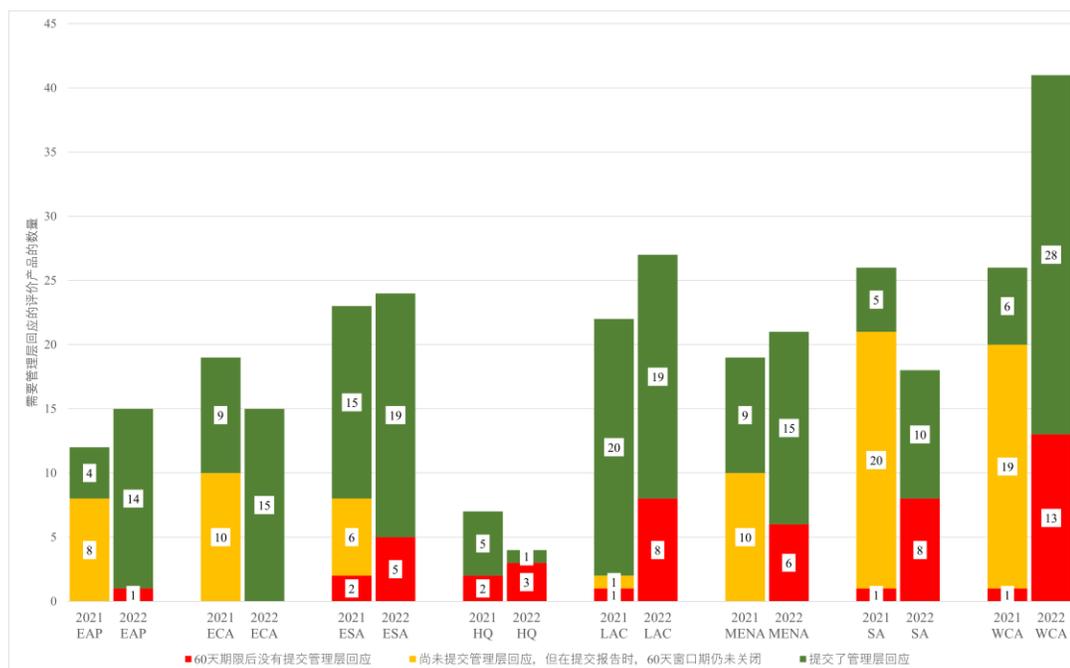
图 11  
2018-2022 年遵守管理层回应要求的及时性<sup>a</sup>



<sup>a</sup> 由于 2019 冠状病毒病疫情暴发，2020 年和 2021 年提交管理层回应的时限暂时延长到 90 天。  
资料来源: EISI。

53. 如下图 12 所示，有一个区域在报告之时没有逾期的管理层回应：欧洲和中亚地区。逾期未作出管理层回应的数量最多的是西部和中部非洲地区（13 份），其次是南亚地区和拉丁美洲及加勒比地区，各有 8 份，中东和北非地区有 6 份，东部和南部非洲地区有 5 份。总部有 3 份逾期的管理层回应，东亚和太平洋地区有 1 份逾期未作出的管理层回应。

图 12  
2021-2022 年遵守管理层回应要求的及时性，按地点分列<sup>a</sup>

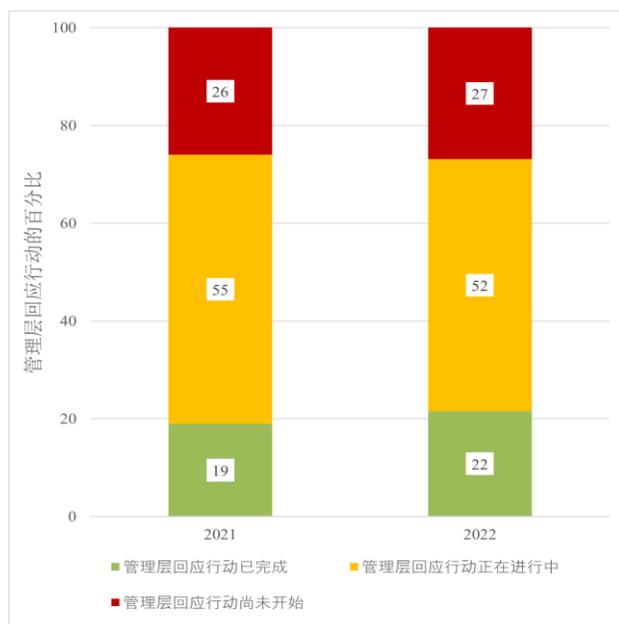


<sup>a</sup> 由于 2019 冠状病毒病疫情暴发，2020 年和 2021 年提交管理层回应的时限暂时延长到 90 天。

图例：EAP：东亚和太平洋地区；ECA：欧洲和中亚地区；ESA：东部和南部非洲地区；HQ：总部；LAC：拉丁美洲和加勒比地区；MENA：中东和北非地区；SA：南亚地区；WCA：西部和中部非洲地区。

资料来源：EISI。

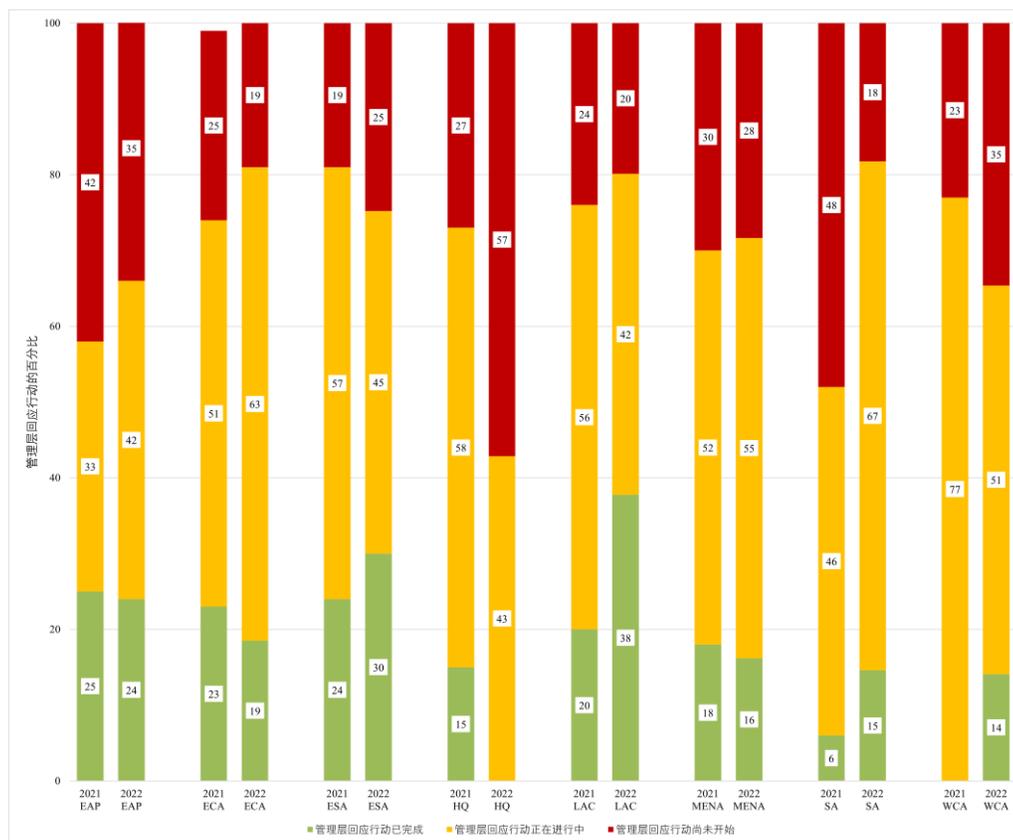
图 13  
2021-2022 年管理层回应中确定的行动的执行状态



资料来源：EISI。

54. 如下图 14 所示，从区域来看，2021 年至 2022 年期间，管理层回应行动的执行速度有所不同。

图 14  
2021-2022 年管理层回应中确定的行动的执行状态，按地点分列



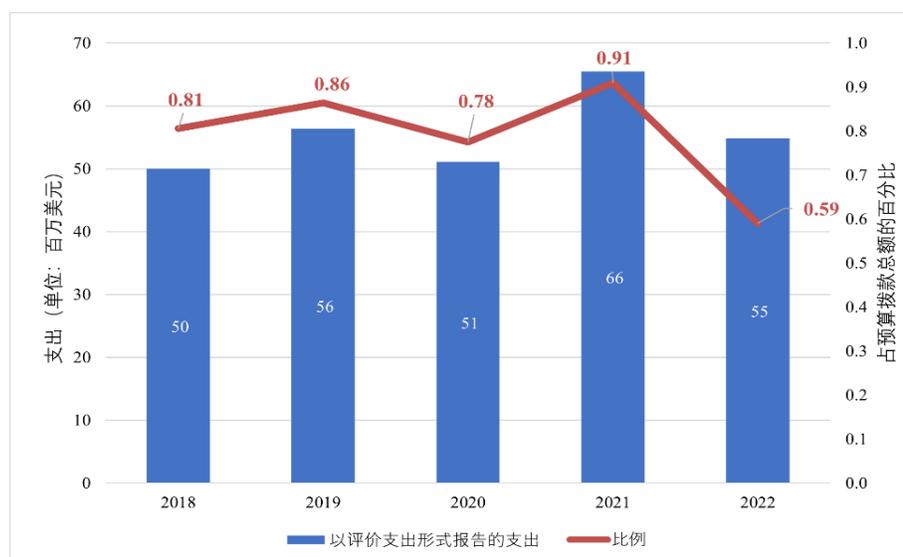
图例：EAP：东亚和太平洋地区；ECA：欧洲和中亚地区；ESA：东部和南部非洲地区；HQ：总部；LAC：拉丁美洲和加勒比地区；MENA：中东和北非地区；SA：南亚地区；WCA：西部和中部非洲地区。

资料来源：EISI。

## E. 以评价支出形式报告的财政资源和支出

55. 图 15 显示了这一期间使用现行公式计算的、以评价支出形式报告的总支出趋势，包括以评价支出形式报告的所有支出（实际支出和承付款）占预算拨款总额的百分比。

图 15  
2018-2022 年以评价支出形式报告的儿基金会支出（单位：百万美元）占预算拨款总额的百分比<sup>a</sup>



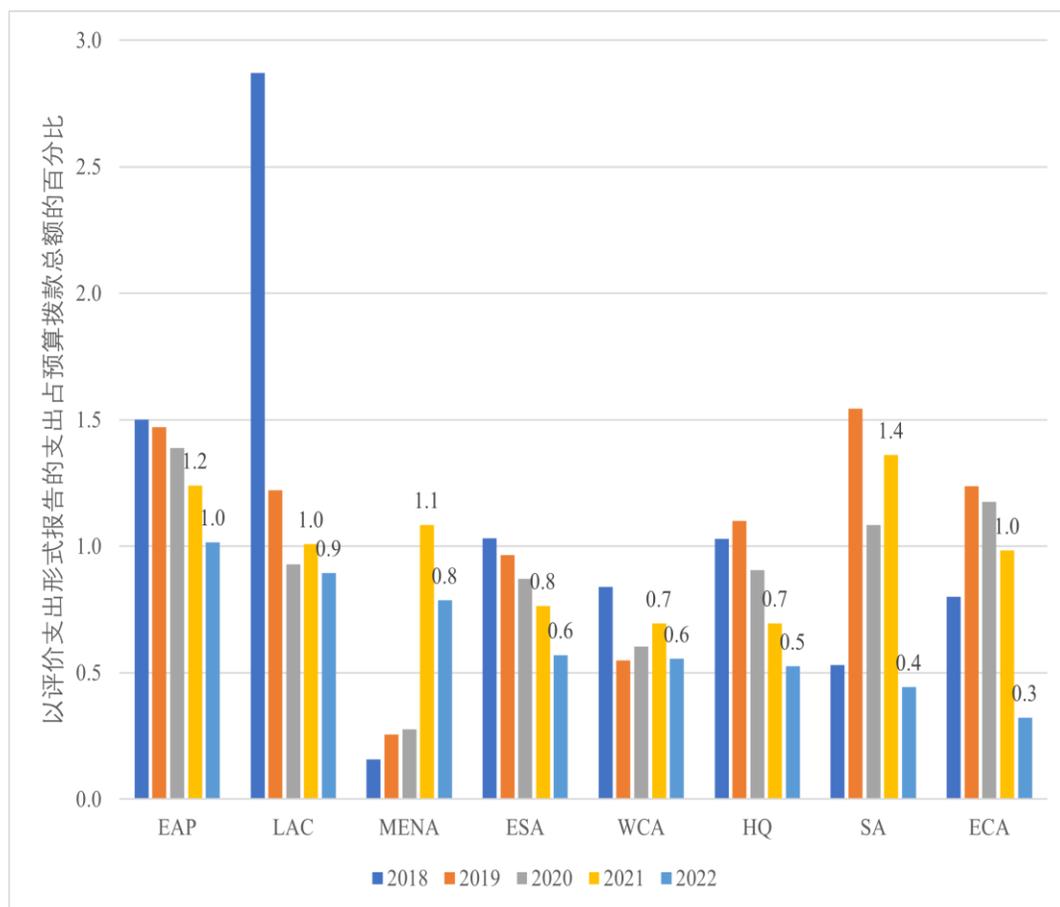
<sup>a</sup> 柱状图和里面的金额显示了每年以评价支出形式报告的支出（单位：百万美元）。红线和数字表示每年以评价支出形式报告的支出占预算拨款总额的百分比。

资料来源：分子数据：2022 年 12 月 31 日检索的以评价支出形式报告的活动（实际支出和承付款）；分母数据：2023 年 1 月 31 日检索的预算拨款总额。

56. 与 2021 年相比，2022 年以评价支出形式报告的支出业绩大幅下降，降至预算拨款总额的 0.59%。这一信息基于新年伊始记录的内部数据。最近的情况是，财务报表数字已于今年早些时候公布，因此可用于更准确地计算这一百分比。如果将 2022 年的财务报表用于计算方案支出，这一比例将升至 0.66%。由于儿基金会的财务报表现已在年初提供，因此可以修订这一评价支出业绩指标的公式，以提供从 2023 年开始的、截至报告时的最新数字。

57. 下图 16 显示了分散层面的趋势。有一个地区，即东亚和太平洋地区，达到了 1% 的目标，而拉丁美洲和加勒比地区以及中东和北非地区则接近目标，分别有 0.9% 和 0.8% 的预算拨款总额用于以评价支出形式报告的活动。包括总部在内的所有其他地区都在 0.5% 左右。

图 16  
2018-2022 年以评价支出形式报告的支出占预算拨款的百分比，按区域分列



图例：EAP：东亚和太平洋地区；ECA：欧洲和中亚地区；ESA：东部和南部非洲地区；HQ：总部；LAC：拉丁美洲和加勒比地区；MENA：中东和北非地区；SA：南亚地区；WCA：西部和中部非洲地区。  
资料来源：分子数据：2022 年 12 月 31 日检索的以评价支出形式报告的活动（实际支出和承付款）；分母数据：2023 年 1 月 31 日检索的预算拨款总额。

58. 然而，下文表 4 提供了对作为评价活动而报告的活动支出的分析。这项分析显示，2022 年，虽然所报告的 70% 评价支出实际用于评价和其他评价产品和活动，但有相当大的比例（30%）用于其他类型的活动。

表 4  
2022 年以评价支出形式报告的活动，按活动类型分列

活动	支出 (单位: 美元)	比例
<b>与评价直接相关的活动</b>	<b>38 073 275</b>	<b>70</b>
数据和调查	3 505 294	6
规划和方案审查	1 755 344	3
监测	1 294 542	2
研究	264 703	1
证据生成	612 357	1
系统强化	4 187 139	8
其他	5 128 126	9
<b>总计</b>	<b>54 820 780</b>	<b>100</b>

资料来源: 2022 年 12 月 31 日检索的、以评价支出形式报告的活动 (实际支出和承付款)。

59. 如上所述, 2022 年出现的、关于以评价支出形式报告的支出占预算拨款百分比的一些趋势, 可能是由于近年来预算拨款的总体增加所致。这一因素可能会扭曲以评价支出形式报告的支出分配, 为了消除这一因素, 表 5 提供了更详细的细目。

表 5  
2022 年以评价支出形式报告的活动, 不包括总部和预算最高的办事处

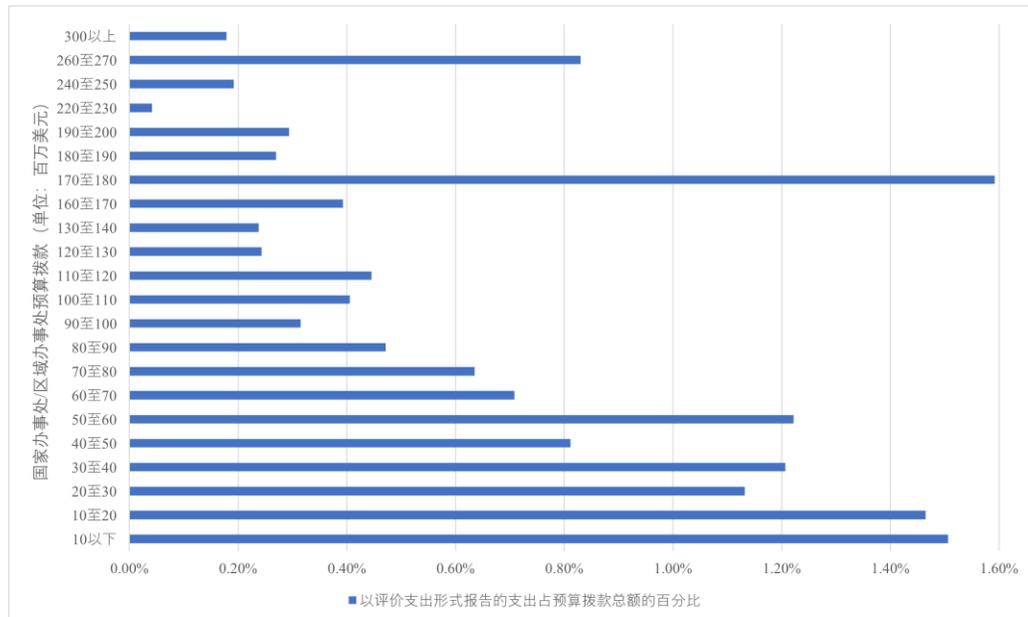
办事处	以评价支出形式报告的支出 (单位: 百万美元)	占预算拨款总额的百分比	与评价直接相关的活动支出 (单位: 百万美元)	占预算拨款总额的百分比	其他活动支出	占预算拨款总额的百分比
所有国家办事处和区域办事处 (不包括总部)	46.3	0.60	30.9	0.40	15.4	0.20
国家办事处和区域办事处 (不包括方案支出超过 2.5 亿美元的六个办事处)	36.5	0.71	24.8	0.48	11.7	0.23

资料来源: 分子数据: 2022 年 12 月 31 日检索的以评价支出形式报告的活动 (实际支出和承付款); 分母数据: 2023 年 1 月 31 日检索的预算拨款总额。

60. 上表第一行的数字不包括总部办事处, 仅反映国家办事处和区域办事处的支出。在这种情况下, 以评价支出形式报告的支出百分比略有提高, 从 0.59% 增至 0.60%。与评价直接相关的支出比例为 0.40%。

61. 第二行不包括预算拨款超过 2.5 亿美元的六个最大的国家办事处, 它们往往在复杂的方案环境中运作 (在被排除的六个国家办事处中, 有三个办事处参与了三级应急响应)。在这种情况下, 以评价支出形式报告的支出上升到 0.71%, 与评价直接相关的活动支出占预算拨款总额的 0.48%。

图 17  
2022 年以评价支出形式报告的活动，按预算拨款总额规模分列

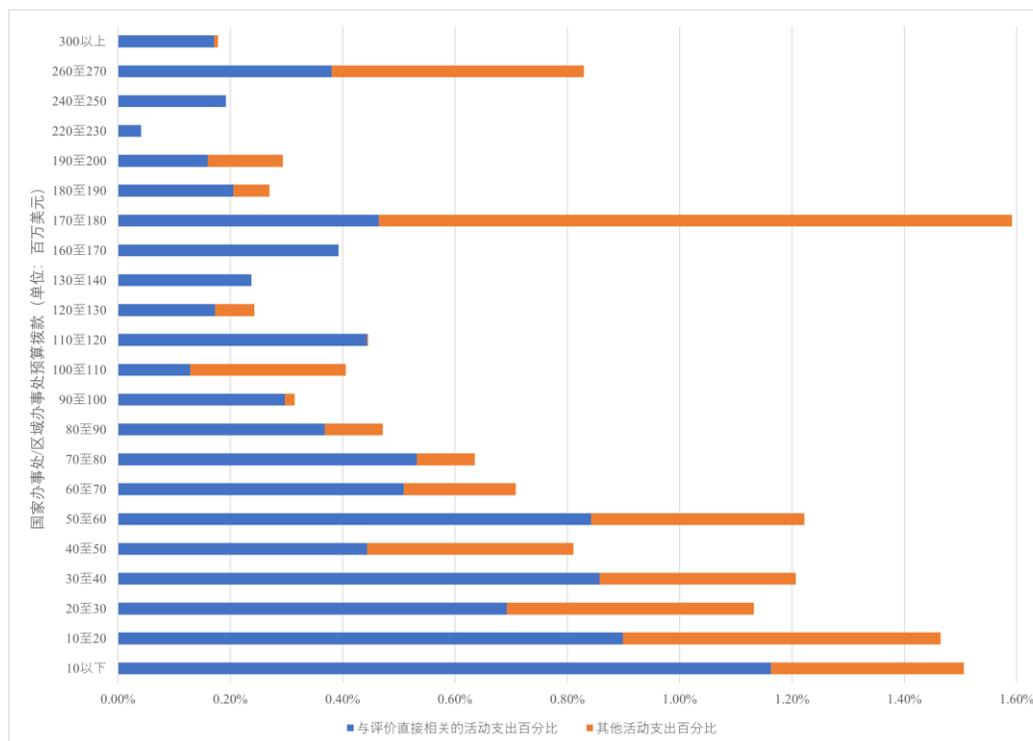


资料来源: 分子数据: 2022 年 12 月 31 日检索的以评价支出形式报告的活动 (实际支出和承付款); 分母数据: 2023 年 1 月 31 日检索的预算拨款总额。

62. 图 17 显示, 预算拨款总额超过 8000 万美元的区域办事处和国家办事处 (不包括总部, 以避免对结果产生负面影响的风险) 将不到 0.5% 的资金用于以评价支出形式报告的活动。

63. 上图中 2.60 亿美元至 2.70 亿美元和 1.70 亿美元至 1.80 亿美元这两个括号中的异常值是由于在与评价没有直接关系的活动中支出较高的办事处造成的。

图 18  
2022 年以评价支出形式报告的活动支出占预算拨款总额的百分比



资料来源：分子数据：2022 年 12 月 31 日检索的作为评价支出而报告的活动（实际支出和承付款）；分母数据：2023 年 1 月 31 日检索的方案支出总额（仅实际支出）。

64. 图 18 显示，两类预算拨款高的办事处（2.6 亿至 2.7 亿美元和 1.7 亿至 1.8 亿美元）以评价支出形式报告的活动支出多，原因是以评价支出形式报告的其他（非评价）活动比例高。预算拨款总额超过 8000 万美元的大多数办事处将大约 0.4% 的资金用于与评价直接相关的活动。

## 五. 结论和未来的方向

65. 2022 年是评价职能演变的关键一年，涉及到《2022-2025 年战略计划》实施第一年的多项战略转变，以确保评价职能在本四年期及以后胜任其职。评价职能于今年年初最终确定了战略愿景，随后采取了重大措施，将其宏大目标转化为具体行动。这些措施包括更有重点地确定评价职能各种评价伙伴关系的优先次序，促进在儿基会的三个层面实现更加一体化和更加统一协调的评价职能（包括在实现更一体化的评价规划和实施方面取得初步进展）；不断提高对评价综合的重视程度，以提供对什么可行和什么不可行（以及方式和原因）的综合见解；改进评价质量评估框架；填补长期存在的人力资源差距，以及使评价办公室的业绩预期标准化，以便使其符合愿景。

66. 愿景中包含的并在执行局决定中提到的需要加快进展的另外三个领域取得了特别显著的进展：国家评价能力发展、更多的方法创新以及加强影响评价能力。在评价范围内与影响评价相反的一面取得了重大进展，以进一步融入应用可评价性评估和其他前瞻性评价的新方法和创新型方法，帮助儿基会及其合作伙伴（包括执行局）

识别和管理风险，抓住机遇，并尽早和灵活地改变方向。这些评价中最主要的评价包括关于儿基金会实现《2022-2025 年战略计划》的定位的可评价性评估和形成性评价，这是有史以来第一次为整个儿基金会提供关于其在新战略计划前六个月的优势和差距的早期看法。

67. 总之，这些进展标志着评价职能在保持对其作用至关重要的独立性的同时，正在努力确保与儿基金会、其业务环境及其学习和问责需求同步发展。除了这些职能推动的举措之外，还启动了涉及更广泛组织的另外两项工作，以帮助确保评价职能能够最好地发挥其学习和问责作用，从而帮助儿基金会及其合作伙伴（包括执行局）为儿童取得尽可能好的结果。其中包括对儿基金会评价职能的独立同行审查以及对儿基金会评价政策的内部审查和修订，这两项工作都是儿基金会 2018 年现行评价政策所要求的。政策审查和修订过程在很大程度上依赖于独立同行审查的结果和建议，以及一系列其他投入，包括磋商、案头审查、工作人员调查和对其他机构评价政策的比较分析。独立同行审查提出了改进评价职能本身及其在儿基金会中的定位的建议，其中一些建议对评价政策有影响，一些建议对儿基金会的评价做法和文化有影响。

68. 除了这些广泛的进展领域之外，评价状况可谓喜忧参半。今年提交的评价数量创下新高，并延续了近年来所实现的高地理覆盖率，包括所开展的人道主义评价数量、联合评价和机构间评价数量以及涵盖性别平等的评价数量都有所增加。与此同时，这一数字增长只是评价职能总体成熟度的一个指标：更重要的是确保所有评价都以涵盖儿基金会面临的最具战略重要性的问题为最佳目标，并且具有很高的可信度和效用。影响和结果层面的评价比例也仍然很低。虽然评价质量略有下降，但相对而言质量仍然较高——考虑到推出了更严格的新质量评估框架，这是一个特别令人鼓舞的结果。评价指南和有具体的评价供资目标仍然是关键的优势领域，在独立同行审查中受到了称赞。

69. 需要继续改进的领域包括为评价职能获取可预测的充足资源，包括用于影响评价的资源，这是执行局在第 2022/5 号决定中强调的一个优先事项。在评价职能的资源配置方面，2018 年至 2021 年的积极趋势发生了明显的逆转。独立同行审查强调需要更加关注这一领域，并强调管理层有责任确保为评价提供可预测的充足资金。独立同行审查还建议，应该明确 1% 的目标，包括计算方式以及包含的内容。这一领域取得进展的一个重要标志是，执行主任办公室、财务和行政管理司和评价办公室在 2023 年初进行的讨论和达成的协议在最终解决这一持续挑战方面取得了重大进展。经修订的政策草案及其配套的实施指南将巩固这一进展，在 2023 年，将修订公式以确保一致性和准确性，并加强对哪些组织支出可以合法地列为评价支出的指导。

70. 为了恢复以前在实现 1% 支出目标方面取得进展的积极趋势，评价集合基金也必须补充到上一个四年期开始时的水平。这将大大扭转总部支出下降的趋势，并提高供资的可预测性和及时性。还需要在分散层面对支出进行适当监测，确保管理层回应得到有意义的落实，建立更强有力的问责机制，以及明确国家办事处的评价作用和责任。

71. 依然存在更多的挑战。独立同行审查发现，在确保问责制以及作用和责任方面需要更加明确，包括供资和报告关系，这可能会限制独立性和可信度的一致性。独

立同行审查还查明了其他风险，包括管理层对评价的使用受到限制，以及评价职能与儿基会其他证据和知识职能之间的互动有限。报告的结论是，虽然评价办公室在解决这些限制方面发挥了作用，但它必须得到扶持环境中其他因素（包括儿基会领导层和其他职能）的赋权和支持，才能取得切实进展。

72. 其他几个领域确实属于评价职能的职权范围。例如，涵盖残疾问题的评价数量在 2022 年略有下降。虽然 2022 年涵盖性别平等的评价数量有所增加，但将性别平等和增强妇女权能纳入评价分析范围的质量却有所下降。需要仔细监测这些问题，以便在 2023 年扭转这些趋势。

73. 同样令人关切的是，在 2022-2025 年全球评价计划中指出应于 2022 年完成的 13 项评价中，只有 4 项<sup>7</sup>按时交付。所有这些评价现在都将在 2023 年完成，但若有延迟，就可能会对 2023 年评价的交付产生连带影响。影响这些延迟的一个关键因素是资金支付的及时性和充足性，资金问题令一些评价的启动出现延迟（从而使完成时间推迟）。其他因素包括针对组织优先事项、长期人力资源缺口和工作人员轮调，在执行快速增援和延展性任务时临时部署关键工作人员。

74. 展望未来，正如同行审查（以及 2021 年 MOPAN 评估）所建议的那样，评价办公室将在 2021 年和 2022 年的基础上再接再厉，积极加强与其他证据职能部门的协调与合作，包括通过在证据生成的道德伦理方面的合作、联合证据差距摸底调查和证据战略、联合专业发展倡议、资源调动支持等。

75. 在 2022 年其他领域取得的进展将会继续。儿基会不久将启动影响催化剂基金，这是一项战略举措，旨在支持儿基会优先专题领域的高质量影响评价，同时继续 2022 年开始的具体影响评价工作，包括童婚、精神卫生和社会心理支持以及儿童营养不良等领域的工作。此外，2022 年取得的国家评价能力发展强劲势头将持续下去。

76. 本年度报告中确定的趋势并不令人惊讶，因为它们只是为独立同行审查中确定的问题增加了客观数据，并在与审查和修订评价政策有关的组织磋商中得到进一步附和。在提交给执行局本届会议的评价政策修订草案中，将反映从政策角度解决这些问题的努力。在未来几个月，评价办公室期待与儿基会管理层合作落实同行审查的建议，并与执行局一起最终确定经修订的儿基会评价政策，以解决悬而未决的挑战，从而帮助评价职能和儿基会在未来几年尽可能产生积极影响，造福儿童。

## 六. 决定草案

### 执行局

1. 注意到关于儿基会评价职能的 2022 年年度报告（[E/ICEF/2023/18](#)）及其管理层的回应（[E/ICEF/2023/19](#)）；
2. 还注意到关于儿基会对全球 2019 冠状病毒病（COVID-19）疫情的三级响应的评价、其摘要（[E/ICEF/2023/20](#)）及其管理层的回应（[E/ICEF/2023/21](#)）。

<sup>7</sup> 另一项由总部牵头的机构间评价已于 2022 年完成，但在起草本报告时尚未上传到 EISI 系统或进行质量评估。