



供采取行动

联合国儿童基金会

执行局

2017 年年度会议

2016 年 6 月 13-16 日

临时议程* 项目 7

2016 年关于儿基会评价职能的年度报告

摘要

本报告记载了儿基会在 2016 年开展的评价活动。报告审查了儿基会评价职能的现况，并报告了订正评价政策(E/ICEF/2013/14)的执行进展情况。报告还涉及儿基会评价职能的管理情况以及旨在促进联合国系统和更大范围内评价一致性的努力。报告提供有关绩效和成果的最新情况、最近有关评价职能的评估结果摘要以及详述 2016 年全球专题评价计划拟议评价状况的附件。报告还载有供执行局审议的决定要点。

* E/ICEF/2017/5。



一. 引言

1. 到 2016 年底，大会选举安东尼奥·古特雷斯先生为现任联合国秘书长。在获任命前，他在一份全面陈述中阐明了他对本组织的愿景，其中强调透明度、问责制和监督的重要性。2016 年 4 月向各国代表发言时，他进一步指出，在联合国，“我们需要一种评价文化——完全透明地进行独立和实时评价”。¹

2. 儿基会致力于透明度，问责制和有效监督，儿基会的评价职能有助于实现这一点。按照其订正评价政策(E/ICEF/2013/14)，儿基会利用评价来展示成果，促进建立问责制，并为制订更好的政策和方案提供证据。因此，评价直接促进实现儿基会、甚至更广泛的联合国系统的使命、任务和优先事项。

3. 本报告记载了儿基会在 2016 年开展的评价活动。其中报告了订正评价政策的执行进展情况，并提供有关绩效和成果的最新情况。同前几年一样，本报告审议了对儿基会职能的管理情况以及旨在促进联合国系统内和更大范围内评价一致性的努力。报告还概述了最近对评价职能进行评估的结果。第八节载有供执行局审议的决定要点。

二. 对评价职能的管理

4. 执行局保持对儿基会评价职能进行监督，并支持评价职能在儿基会发挥的核心作用。在 2016 年 6 月举行的执行局年度会议上，评价局提出了 2015 年关于评价职能的年度报告，并由高级管理人员提出管理观点加以补充。在对年度报告作出回应时，执行局成员就这一职能提出相关问题，并发布第 2016/7 号决定，其中涉及以下问题：(a) 评价所覆盖的地域和专题范围；(b) 在实现将至少 1% 的总体方案支出用于评价的评价政策目标方面取得进展；(c) 及时提供管理层的回应及其落实情况；(d) 国家评价能力建设。应执行局的要求，管理层在 2016 年第二届常会上向执行局提出一项有关实现 1% 政策目标的行动计划。实现目标进展情况载于第五节。

5. 执行局还要求对评价职能进行一次外部同行审查，以满足订正评价政策的要求。第六节报告了这方面的安排。

6. 在 2016 年执行局所有三届会议上提交的评价报告分别都附有管理层的回应，并在每届会议之前举行的非正式简报会上进行更广泛的介绍和讨论。提交以下文件：

(a) 《2014-2017 年全球专题评价计划：供 2016-2017 年审查和更新》(E/ICEF/2016/3)，其中包括原计划所列一些评价的时间调整安排；

(b) “保护儿童免遭暴力侵害：对儿基会战略和方案绩效的全面评价”；

¹ 联合国评价小组，“联合国指定秘书长呼吁建立强有力的评价文化”(www.unevaluation.org/mediacenter/newscenter/newsdetail/121)。

(c) “水、环境卫生和个人卫生(水卫项目)规划的公平性、可扩张性和可持续性：儿基会 2007-2015 年评价的证据，即评价综合报告”；

(d) “扩大中东欧/独联体区域接受早期幼儿教育的机会并增进其公平性”，区域办事处委托进行和管理的独立多国评价。

7. 作为监督评价问题的内部机构的全球评价委员会于 2016 年 6 月举行会议。各成员核准了委员会的成员组成和运作模式，审议了关于加强评价覆盖面的提议以及评价管理层回应中商定行动的执行情况，并注意到有关外部同行审查的安排。

8. 外部审计咨询委员会将评价视为监督系统的重要部分，并监测评价职能的绩效。委员会在 2016 年提交给执行局的 2015 年年度报告中特别注意到，管理层需要继续改进其对评价建议的后续落实，并要求了解后续落实方面取得的进展。²

三. 促进联合国系统内的评价一致性

9. 大会 2016 年 12 月 21 日通过的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)³ 为包括评价在内的联合国所有发展活动提供重要的方向。四年度全面政策审查强调，需要通过全系统战略规划、执行、报告和评价，更好地支持执行《2030 年可持续发展议程》。2016 年，儿基会继续促进建立全系统独立评价机制。

10. 在 2016 年四年度全面政策审查中，大会指出，需要加强支持国家机构在规划、管理和评价方面进行能力建设。儿基会 2016 年支持国家评价能力发展的情况载于第四节。

11. 大会还敦促联合国系统各组织继续努力统一其在报告、监测和评价方面的要求，以减少交易成本，提高效率。儿基会确认有效伙伴关系的价值，并促进努力协调各项评价活动和进行能力建设，特别是通过制订商定全球标准、方法和议程。主要举措包括联合国评价小组(评价小组)开展的活动、联合评价活动以及支持全系统独立评价机制。

A. 联合国评价小组

12. 评价小组汇集整个联合国系统各评价办公室开展联合活动，以便加强和协调本组织内部的评价活动，并在联合国系统之外建立更强有力的评价伙伴关系。2016 年，儿基会继续在评价小组中发挥积极作用。评价局局长继续担任评价小组 4 名副主席中负责伙伴关系的副主席，儿基会工作人员也参加了若干工作组。

13. 2016 年，评价小组集中努力促使评价职能进一步专业化。主要产出包括出版评价小组的订正最新版本评价规范和标准⁴ 以及评价小组的评价胜任能力框架。⁵

² 见 www.unicef.org/about/execboard/files/UNICEF_Audit_Advisory_Committee_2015_AR-13Apr16-FINAL.pdf。

³ 大会第 71/243 号决议。

⁴ www.uneval.org/document/detail/1914。

⁵ www.uneval.org/document/detail/1915。

评价局工作人员对评价使用 and 专业化方面各工作组以及评价小组人道主义评价关注小组作出贡献，后者于 2016 年发表了题为“在评价中反映人道主义原则”的评价做法整合分析。⁶

14. 儿基会还支持评价小组就评价问题作出的倡导，其中根据大会 2015 年通过的具有里程碑意义的第 69/237 号决议，呼吁联合国支持会员国建立评价国家一级发展活动的的能力。在国家开展旨在实现可持续发展目标的活动方面，支持加强国家评价系统能力具有特别重要的意义。在这方面，儿基会 2016 年与姐妹机构一起合作主办了两次广受欢迎的活动，重点讨论评价和各项目标，许多国家及联合国的高级别发言者也有参加。这些活动是分别作为妇女地位委员会第六十届会议⁷以及经济及社会理事会主持下举行的可持续发展问题高级别政治论坛会议⁸的附带会议安排的。

B. 联合国发展方面业务活动全系统独立评价

15. 在 2012 年四年度全面政策审查(第 67/226 号决议)中，大会规定设立全系统独立评价机制，并在 2016 年四年度全面政策审查中再次要求建立这一机制，特别指出“必须提高发展方面业务活动全系统评价机制的质量、独立性和公正性，加强联合国发展系统全面评价架构的协调一致和相互依存，……利用评价机制的调查结果和建议改善发展系统的运作”。

16. 儿基会从一开始就支持有关该机制的工作，提供资金，并通过参加为两项试点全系统独立评价而设立的评价管理小组提供大量技术捐助。2016 年顺利完成两项评价：

(a) “以消除贫穷为特别重点的联合国发展援助框架元评价和评价汇总”。⁹

评价结果要求提高联发援框架评价的覆盖面、质量和资源配置；

(b) “联合国发展系统为了支持实现《千年发展目标》和其他国际商定发展目标而对加强国家统计分析和数据收集能力所作贡献的评价”。¹⁰

有鉴于《2030 年可持续发展议程》，评价报告要求加强整个联合国系统的联合数据收集和统计能力发展。

17. 2016 年底独立审查了全系统独立评价机制，审查报告即将发表。

⁶ www.uneval.org/document/detail/1862。

⁷ www.evalpartners.org/evalgender/evaluating-SDGs-with-an-equity-focused-and-gender-responsive-lens。

⁸ <https://sustainabledevelopment.un.org/index.php?page=view&type=20000&nr=480&menu=2993>。

⁹ www.unjui.org/en/reports-notes/JIU%20Products/JIU_REP_2016_6_Final_English.pdf。

¹⁰ www.unjui.org/en/reports-notes/JIU%20Products/JIU_REP_2016_5_Final_English.pdf。

C. 机构间人道主义评价指导小组

18. 机构间人道主义评价指导小组由各联合国机构和各人道主义非政府组织组成，继续开展联合评价活动。2016年，该小组散发了南苏丹危机联合国应对措施的评价和评价性研究的机构间评价结果(2015年完成)。¹¹ 它还完成以下三个报告：

(a) “叙利亚协调问责制和吸取经验评价综合分析和差距分析”，¹² 其中涉及有关国际社会应对2012至2015年期间叙利亚危机的24项公布的评价和评价性研究中突出强调的主要问题；

(b) “中非共和国危机应对措施的机构间人道主义评价”¹³ 得出结论指出，虽然覆盖面仍是一个重大挑战，但机构间应对措施产生了很大的积极影响；

(c) “菲律宾(台风“海燕”)、南苏丹和中非共和国危机方面国际社会应对措施的机构间人道主义评价的主要结果的综合报告”。¹⁴

该主要结果综合报告吸取了以下方面的一系列经验教训：需要评估和优先次序；战略和业务规划；衡量效力；领导和问责；利益攸关方参与；有关需要以对背景保持敏感的方式实行人道主义方案周期的问题。

四. 促进全球评价一致性

A. 发展网络

19. 儿基会主要通过一项全球运动即 EvalPartners 参与促进超越联合国网络的全球一级评价一致性，以便通过创新伙伴关系和建立联系加强民间社会组织影响决策者和参与国家评价进程的能力。¹⁵ Evalpartners 是联系评价小组与国际评价合作组织的伙伴关系，是由国家和区域评价协会组成的网络。儿基会2012年共同创立了 EvalPartners；该伙伴关系此后变得越来越积极和有影响力。通过协商进程拟订“2016-2020年全球评价议程”是其一项主要产出，重点放在评价能力建设和专业化。全球评价议程的发表和传播获得儿基会的支持。该组织还在2016年大力促进完成 EvalPartners 的任务和目标。在这一年中，儿基会评价局局长以评价小组副主席(伙伴关系)的身份担任 EvalPartners 的共同主席。

20. Evalpartners 还建立一整套全球网络在其旗下运作。儿基会在 EvalSDGs 发挥主导作用，后者是为支持将评价纳入国家、区域和全球系统而组成的网络，以便规划和审查实现可持续发展目标的进展情况。与国际环境和发展学会建立的伙伴关系为发表一系列目标评价简报文件提供渠道。2016年编制了5份文件，其中3份是该研究所2016年下载次数最多的出版物。

¹¹ www.unicef.org/evaldatabase/index_92844.html。

¹² www.interagencystandingcommittee.org/system/files/ocha_syria_web_interactive_2.pdf。

¹³ www.interagencystandingcommittee.org/system/files/print_final.pdf。

¹⁴ www.interagencystandingcommittee.org/system/files/web_interactive_0.pdf。

¹⁵ 见 www.evalpartners.org。

21. 儿基会还向全球议员评价论坛指导委员会提供技术支持，并在 2016 年协助建立欧亚区域议员评价论坛。儿基会吉尔吉斯斯坦国家办事处与国家当局密切合作，发挥重要的支助作用。

B. 人道主义网络

22. 在人道主义评价方面，儿基会继续与人道主义行动问责和绩效动态学习网络合作。2016 年，该网络发表了《人道主义行动评价指南》。根据该指南，儿基会就开展评价工作为儿基会各国家办事处和政府伙伴拟订了一项实践能力建设举措(见第五节)，还于 2016 年在摩洛哥马拉喀什举行的第二届国际法语国家评价论坛上为民间社会组织提供培训。

五. 评价职能的绩效

23. 儿基会关于比照 7 项核心指标的评价职能绩效报告：

- (a) 管理和提交给评价和研究数据库的评价数目；
- (b) 主题分配；
- (c) 进行评价的类型；
- (d) 评价质量；
- (e) 评价利用；
- (f) 评价的机构开支；
- (g) 机构评价生产。

下文将讨论各项指标。呈现的总体情况是不断改进的情况，确认了以往的年度报告中所提出的总体情况。但是，这一总体上积极的情况也因时间和区域的不同带来的绩效差异而令人感到不安。显然需要加强一致性。

24. 主要结果如下：

(a) 提交给评价和研究数据库的评价数目已恢复到可接受的水平，2015 年达到 90 项评价，2016 年达到 102 项评价；

(b) 但儿基会评价的地域覆盖范围在儿基会开展业务的所有国家中仍低于 80%，显示出各区域之间存在巨大差异；

(c) 专题覆盖面显示儿童生存、教育和儿童保护之间的专题分布平衡，但社会包容、人道主义行动和性别平等方面的跨领域问题覆盖面则很小；

(d) 与前几年相比，2015 年提交的评价质量大幅下降，但 2016 年恢复了。分析将质量下降与评价组合多样性增加联系起来；

(e) 提交管理层回应方面有了很大的改进，在 2015 年提交的评价中，95% 都有管理层回应跟进。还需注意促进落实管理层的回应；

(f) 2016 年用于评价的预算超过 5 000 万美元，比 2015 年增加了，当时的相应数字还没有超过 4 500 万美元。但这些数字只占每年总预算使用的 0.8%。要实现 1% 的政策目标，还需要取得进一步的进展。

25. 本节的这些指标借鉴了儿基会全球评价报告监督系统提供的数据。提交给评价和研究数据库的评价通常由一家外部公司进行评估，并每年发布其结果的元评价。2016 年对 2015 年提交给该数据库的评价进行评估和质量评定，并公布其结果。¹⁶

指标 1：管理和提交给评价和研究数据库的评价数目

26. 在其第 2016/7 号决定¹⁷ 中，执行局要求儿基会在关于评价职能的下一份年度报告中列入关于评价覆盖面的地域和专题概览。关于指标 1 的这一部分以及关于指标 2(专题分配)的部分载有该概览。

27. 2015 年提交给评价和研究数据库的评价报告达 90 份，比 2014 年共提交 80 项评价增加了 13%。2016 年提交的评价报告数量显示继续增加 102 份报告。

28. 尽管提交的报告总数有波动，但 2010-2015 年期间地域覆盖面几乎没什么改变，如图一所示。其中地域覆盖面被认为是在本报告年度或前两个年度向该数据库提交至少一份评价报告的国家办事处中的所占比例。地域覆盖面在这一期间仍保持在 75% 左右。在 2013-2015 年期间，103 个国家(占 76%)提交了至少一份评价报告，另有 32 个国家办事处(占 24%)没有提交。但在 2016 年，32 个办事处中有 9 个提交了评价报告。

29. 鉴于最后一点，可以预计 2016 年的全球覆盖面将会有所改善。关于 2014-2016 年期间地域覆盖面的分析表明：虽然 107 个国家办事处在有关期间提交了评价，与前几年相比有所改进，但仍有多达 28 个国家办事处没有提交。因此，2016 年全球覆盖面与往年一样，仍低于 80%。

¹⁶ “儿基会全球评价报告监督系统 2015 年元分析：关于儿基会评价报告质量和趋势的独立审查，2009-2015 年” (https://www.unicef.org/evaldatabase/index_92693.html)。

¹⁷ 见 E/ICEF/2016/15。

图一
2010-2016 年评价覆盖面趋势



资料来源：儿基会全球评价报告监督系统。

30. 区域分析说明了这些问题。2015 年提交的评价数目因区域而异，中欧和东欧及独立国家联合体(中东欧/独联体)、东亚和太平洋、东非和南部非洲(东南非区域)、南亚以及西非和中部非洲分别提交了 12 份或更多的评价，而其他区域提交的评价较少(见表 1)。

表 1
2012-2016 年区域评价提交率

年度	区域							共计
	中东欧/独联体	东亚和太平洋	东南非区域	拉丁美洲和加勒比	中东和北非区域	南亚	西非和中部非洲	
2012	16	8	29	10	10	7	34	114
2013	17	7	36	17	7	4	24	112
2014	15	6	13	8	4	14	12	72
2015	14	14	17	7	8	12	15	87
2016	9	12	26	13	6	13	13	92

简称：中欧和东欧及独立国家联合体(中东欧/独联体)；东非和南部非洲(东南非区域)。

资料来源：全球评价报告监督系统的。

31. 但审议各区域的覆盖面和趋势更有意义。在 2013-2015 年期间，南亚所有国家至少提交了一份评价，促使地域覆盖率达到 100%。东南非区域紧跟其后，覆盖率达 95%，中东欧/独联体的表现也高于平均水平，达到 82%，而其他区域则落在其后。2014-2016 年期间的数字显示各区域加强一致性(见表 2)。东亚和太平洋的覆盖面大幅提高尤其值得注意。

表 2
2013-2016 年各区域的地域评价覆盖面

期间	区域(百分比)						
	中东欧/ 独联体	东亚和 太平洋	东南非 区域	拉丁美洲 和加勒比	中东和 北非	南亚	西非和中部 非洲区域
2013-2015	82	73	95	56	65	100	76
2014-2016	86	93	82	64	65	100	80

资料来源：全球评价报告监督系统。

32. 评价覆盖面取决于不同国家和区域的条件以及区域和国家办事处的可用资源。勉强开展评价的国家往往其方案规模小，资源也有限。创造性评价规划顺利地处理了中东欧/独联体区域的这一问题。该区域办事处几年来利用汇集的资源，在协调和管理多国评价方面发挥关键作用。这种办法可适用于其他区域。

33. 政治情况对在一些国家进行评价构成重大制约因素，经历冲突的国家面临很大的挑战。但在中东和北非区域，例如，尽管该区域受旷日持久的武装冲突和巨大挑战的影响，但一些儿基会办事处还是设法开展有益的评价。

34. 最后，对今后一段时间的区域生产力进行分析带来了一些意外。在两个区域，如表 1 所示，已完成的评价数目比前几年增加了一倍。南亚的生产从 2013 年 4 项评价增加到 2014 年的 14 项，东亚和太平洋区域的评价数目从 2014 年的 6 项增加到 2015 年的 14 项。其他区域的评价数目减少了一半。东南非区域以及西非和中部非洲区域 2012 年和 2013 年提交多项评价，但随后两年产出却减少约 50%。高生产率的情况表明什么是可能的，而普遍波动的情况则表明必须持续努力和密切监督。

指标 2：主题分配

35. 这一部分涉及的专题覆盖面，是执行局已要求了解的问题。专题覆盖面是通过根据专题分布的主要业绩指标提出年度报告来处理的，其中根据 2014-2017 年儿基会战略计划所载的各种成果领域和贯穿各领域的问题或专题对评价组合进行审查。

36. 为本分析，已审议 2015 年提交的 87 份报告。¹⁸ 结果表明儿童生存、教育和儿童保护之间的专题分配平衡，其中儿童生存(包括保健、艾滋病毒/艾滋病、营养和水卫项目)成为 25 份报告(占 29%)的主题；教育成为 24 份报告(占 28%)的主题；儿童保护成为 22 份报告(占 25%)的主题。¹⁹

37. 只有少数经审查的 87 项评价特别侧重于社会包容、人道主义行动和性别等贯穿各领域的主题，虽然这些领域非常重要。社会包容是 2014-2017 年战略计划下一个主要成果领域，但就这一专题提交的评价报告只有两份。虽然涉及这些问

¹⁸ 三份报告无法归入某一特定专题，因而排除在分析之外。

¹⁹ 2015 年儿基会全球评价报告监督系统元分析。

题似乎较少，但主要侧重于部门问题的许多评价的确在某种程度上涉及社会包容、人道主义行动和性别平等问题。

38. 人道主义行动作为重要的交叉问题被列入战略计划，但所提交的特别侧重于人道主义问题的评价报告只有两份。近年来，人道主义行动已成为儿基会不断发展的主要工作领域。虽然儿基会开展了各种未公布的审查和吸取经验教训活动，但整个儿基会有必要对人道主义行动进行系统评价，并采取必要的后续行动。

39. 认识到有限的评价“专门技能”在人道主义紧急情况下可能是一个制约因素，但评价局启动了一个上手培训方案，并提供后续支助，以协助国家办事处开展评价工作。该方案旨在建立国家办事处的能力，使其不仅能评价人道主义行动，而且还能编制一整套评价。预计该方案将在今后几年促进扩大覆盖面。

40. 2016年，该方案支持在包括一些处境非常艰难的国家在内的17个国家开展有关人道主义活动的评价工作。其中3项评价于2016年完成并提交，预计2017年还将报告十几项评价。此外，评价局将就现有有关人道主义应对措施的评价编写一份综合报告，应于2017年9月提交给执行局。

41. 虽然性别平等是战略计划下一个重要的交叉问题，但只有一份经审查的评价报告有具体侧重涉及。但可以指出的是，在2012至2015年期间，部门专题评价所占比例有所增加，其中也涉及性别平等、公平和人权等关键主题。2016年委托进行的“联合国儿童基金会(儿基会)：审查发展实效，2012-2015年”也强有力地表明，儿基会通过将评价重新关注这些主题作出适当反应。²⁰ 审查表明，与2012年开展的类似活动相比，涉及以下主要内容的所占比例增加了：性别平等评价(66项中有58项)；公平评价(66项中有54项)；采取基于人权的办法拟订方案评价(66项中有39项)。大多数评价将这些专题列入其中减少了开展具体侧重于这些专题的评价的必要性，但也引起争议。

42. 对今后一段时间的专题分配进行分析表明出现一些重大转变。2015年，针对具体部门的评价的所占百分比增加了，其中84%的评价侧重于特定部门。这一增加分别始于2014年和2012年所报的66%和41%。相反，正如上文所述，具体涉及交叉问题的评价报告所占比例却减少了：从2014年9%减少至2015年仅5%。侧重于儿童保护的评价报告所占百分比增加是一个显著特征，即从2014年的6%增加到2015年的25%。

²⁰ www.unicef.org/evaldatabase/index_94430.html。

表 3
2010-2015 年评价报告主题分配

主题	百分比			
	2010	2012	2014	2015
针对具体部门的主题,按《2014-2017年战略计划》关键成果领域	66	41	66	84
儿童生存和发展		26	23	22
保健			13	15
营养			0	1
水、环境卫生和个人卫生			10	6
教育	15	13	32	28
保护儿童	15	8	6	25
艾滋病毒/艾滋病	10	5	4	7
社会包容	1	1	1	2
覆盖若干部门的方案	34	43	16	13
贯穿各领域的主题	—	10	9	5
组织绩效/其他	—	6	9	3

资料来源：全球评价报告监督系统。

指标 3：进行评价的类型

43. 2015 年根据受评估的结果层面提交的评价报告分析明确将重点防止较高层面的结果。大多数评价定在成果层面(61%)和影响层面(29%)，而不是定在产出层面(10%)。这种注重较高层面结果表明，应适当利用评价来理解儿基会的贡献和成就。

44. 进一步的分析(见表 4)表明，大多数结果层面评价被评为“高度满意”或“出色”，没有一项评价被评为“不满意”。相比之下，超过一半受审查的影响层面评价显示出一些不足之处。这表明，儿基会在其实践经验最多的领域能力最强，如制订和执行成果层面的方案和评价。这对评价组合的总体质量具有重大意义，正如有关儿基会评价报告质量的部分所审查的。

表 4
2015 年评价重点和质量

重点	质量				共计
	出色	高度满意	非常满意	不满意	
影响		10	15	1	26
成果	3	31	21	—	55
产出	1	3	3	2	9
共计	4	44	39	3	90

资料来源：全球评价报告监督系统。

45. 2015 年提交的评价报告分析表明,大多数评价是由儿基会各办事处直接管理的。儿基会管理的评价占评价组合的 72%,而联合国联合管理的评价、国家联合管理的评价以及国家主导的评价分别占 18%、3%和 1%。6%的评价报告管理无法确定。

46. 2015 年提交的评价报告分析表明,评价目的大致分为形成性评价(前瞻性和注重学习)、总结性评价(回顾性,通常十分注重问责)以及总结性和形成性相结合的评价(见表 5)。这表明,总体而言,适当部署评价工作,不仅有助于“实时”学习,而且有助于问责制和今后的改进。

表 5
2015 年评价目的

目的	数目	百分比
形成性	23	26
总结性	30	33
总结性和形成性	37	41

资料来源:全球评价报告监督系统。

指标 4: 儿基会评价报告的质量

47. 下列评价结果是根据对全球评价报告监督系统评价质量保证和联合国性别平等和增强妇女权能全系统行动计划(全系统计划)审查 2015 年提交给评价和研究数据库的 90 份儿基会评价报告提供的数据进行定量和定性分析所作出的。

48. 2015 年,大多数评价报告获评“高度满意”(49%)和“非常满意”(43%)几乎比例相当(见表 6)。4 份报告(4%)获评“出色”,3 份报告获评“不满意”。换言之,提交的报告中 53%被视为“良好”,即“高度满意”或较好,只有 3%被认为“不满意”。然而,受评估的评价中 43%显示出一些不足之处,下文的进一步分析探讨为什么会是这样的情况。

49. 评价报告的质量因部门不同而有很大的不同。大多数关于人道主义行动和水卫项目的评价报告被认为“高度满意”或较好(虽然样本数量少)。大多数涉及儿童保护、艾滋病毒/艾滋病和社会包容领域的评价报告以及多部门评价也被认为是“高度满意”。然而,24 份教育部门评价报告中只有不到一半获评“满意”,13 份卫生部门评价报告中只有 4 份获评“满意”。

50. 总体而言,评级表明,与 2014 年提交的评价报告相比,2015 年提交的评价报告质量急剧下降。虽然 2014 年提交的儿基会评价中 74%获评“高度满意”或“出色”,但 2015 年提交的评价中只有 53%获得这样的评级。与此同时,获评“非常满意”的评价报告从 2014 年的 23%增加到 2015 年的 43%,而获评“不满意”的评价所占百分比仍很低,占 3%。

51. 但 2016 年提交的评价结果显示更积极的情况。在被评级的 101 项评价中，74% 获评“高度满意”或“出色”，只有一项获评“不满意”。这表明，2015 年提交的评价质量下降是不正常现象，过去几年报告的改善趋势得到恢复。

表 6
按百分比分列的 2009–2014 年已完成的评价的质量

质量评级	年度				
	2009	2012	2014	2015	2016
出色	4	3	6	4	6
高度满意	32	59	68	49	68
非常满意	34	30	23	43	25
不满意	30	8	3	3	1

资料来源：全球评价报告监督系统。

52. 然而，2015 年评价报告质量严重下降，要求给予说明。此外，为扭转这一下降趋势，不仅必须了解可能的原因，而且必须核实下降是否实际发生或是否为计量不一致造成的影响。

53. 2015 年质量下降正好碰上更换为儿基会评价报告提供质量保证的外部公司。这就提出这样一个问题：评估标准是否变得更加严格，以便进行更严格的评级。评价局采取步骤管理这一风险，办法是：向新公司提供强化培训，以便适用同样的标准，并密切监测质量审查过程。评价局相信适用统一标准。

54. 有关评级的元分析揭示了一个更重大的问题。获评符合儿基会标准的报告所占百分比似乎与成果层面评价报告所占百分比是相称的。在 2015 年提交的评价报告中，62% 的成果评价报告获评“高度满意”或较好，只有 38% 的影响评价报告获得此种评级。此外，有关影响评价的报告所占比例高于前几年。总之，这些因素似乎促使 2015 年出现质量全面下降的情况。这意味着，评价类型组合越多样化，就越难一致提供符合规定质量标准的报告。这还意味着，需要提供更多的支持，以确保所有类型的评价(及相关评价报告)达到规定的质量水平。需要进一步监测来确认这一假设。

55. 对获得较低评级的评价进行审查也为以下主张提供更多的支持：在处理复杂专题或采用先进方法时，可能难以维持组合的质量。在 2015 年提交获评“非常满意”的评价中，似乎有一组共 6 项关于加强保健系统(儿基会一个具有挑战性的新工作领域)的评价。其中得出这样的结论：虽然必须保持质量，但也必须为创新和学习保留空间。评级和报告是支持学习的重要机制。因此，就每项评价质量评估而言，应向提交评价的办公室提供执行反馈意见。

联合国性别平等和增强妇女权能全系统行动计划

56. 儿基会根据全系统计划规定的主要指标监测和报告其评价绩效。就 2015 年提交的评价而言，儿基会的累计平均得分为 6.36，属于“接近要求”。这表明年同比有所提高，儿基会在 2014 年周期报告评级为 6，即“接近要求”。这一评级

与包括联合国开发计划署、妇女署以及联合国教育、科学及文化组织在内的类似联合国实体的情况一致。

57. 关于将性别纳入评价的范围、指标、标准和问题的报告获评“满意”，但利用性别分析作出知情评价结果、结论和建议方面则获评较低，需要作出改进。

指标 5：评价的使用，包括管理层回应

58. 管理层回应是评价周期的主要特点。在其第 2016/7 号决定中，执行局敦促儿基会优先及时提供管理层回应，并提高已同意的评价建议的执行率。为此，儿基会已按照评价政策采取行动，以确保上载到评价和研究数据库的所有评价报告都附有管理层的回应。在 2015 年提交的评价报告中，86 份报告(占 95%)有管理层回应跟进，比 2014 年提交的评价所报告的 79%增加了 17%。

59. 管理层的回应行动应在一年内完成。就 2015 年提交的评价而言，据报管理层回应中所列商定行动中 76%在 2016 年底之前已完成(33%)或正在执行中(43%)。这与 2014 年提交的评价执行率相同。

60. 按区域进行的分析表明，在完成和执行商定行动方面存在相当大的差异。就 2015 年提交的评价而言，中东和北非区域报告取得良好进展，商定行动已完成 85%，9%正在执行中，只有 6%尚未开始执行。东南非区域也取得良好进展，报告商定行动已完成 62%，33%正在执行中，只有 6%尚未开始执行。东亚和太平洋区域也报告其绩效超过平均水平(56%已完成，31%正在执行中)。但其他区域则在不同程度上落后了。一些区域的良好绩效再次表明什么是可能的，同时指出必须加强各区域的一致性。

指标 6：评价的机构开支

61. 儿基会评价政策规定评价支出目标为方案支出的 1%。执行局一直在密切监测这一点，并敦促儿基会在实现这一目标方面加快取得进展。

62. 在 2016 年底，据报告，评价预算使用已达到 50 121 716 美元，占总体方案预算使用 0.8%。²¹ 2015 年底的相应数字为 44 959 817 美元，也是占总体方案预算使用 0.8%。虽然这些数字显示评价预算使用大幅增加，但在在这一年实现 1% 的政策目标方面显然没有取得任何进展。

63. 这些数字来自虚拟集成信息系统即儿基会的企业资源管理平台。2016 年审查了用来计算评价预算使用的方法，发现重要支出领域遗漏了。因此，评价局修订了该方法，将在儿基会各系统可随时追踪的所有评价支出都纳入其中。²² 上文所提 2015 年和 2016 年的数字也采用同样的计算方法。

²¹ 正如虚拟集成信息系统即儿基会的企业资源管理平台所记录的，预算使用包括实际支出和承付开支。

²² 以前，评价预算使用估算是根据经常资源和其他资源(经常)计算的。为了对评价预算使用进行更全面的估算，现在的计算包括在一些其他预算项目中入账的评价费用。

指标 7：机构一级评价

64. 本指标下审议的活动为评价局开展的机构一级评价。其中包括 2014-2017 年全球专题评价计划(E/ICEF/2014/3)以及 2016-2017 年的审查和更新(E/ICEF/2016/3)下所列的评价。

65. 2016 年，评价局完成并公布下列评价及相关研究：

(a) 全球专题评价：²³

(一) “评价儿基会预防母婴传播/儿科艾滋病毒护理和治疗方案”；²⁴

(二) “交流促进发展：评价儿基会的能力和行动”；²⁵

(b) 评价儿基会人道主义行动：

(一) “评价儿基会应对西非埃博拉病毒的工作”；²⁶

(二) “儿基会应对中非共和国危机的措施”；²⁷

(c) 根据现有评价的元分析评价综合研究报告：

(一) “儿基会水卫项目方案拟订方面的公平、可持续性和可扩放性：2007-2015 年儿基会评价提出的证据”；²⁸

(二) “联合国儿童基金会(儿基会)：审查发展实效，2012-2015 年”；²⁹

(d) 关于儿基会的建设和平、教育和宣传方案在埃塞俄比亚³⁰ 和缅甸³¹ 开展发展评价³² 活动的国家报告。

高级管理层已对这些评价提出的建议拟订管理层应对措施。

66. 评价局还参与若干联合评价的管理，正如本报告其他部分所指出的。协同联合国人口基金评价办公室继续开展卫生部门联合评价工作：2011-2016 年保健 4 组织伙伴关系联合支助底线评价，以促进妇女和儿童的健康。

²³ 全球专题评价活动范围广泛，宗旨是综合审视儿基会就特定议题领域或专题进行的工作，评价工作的目的是提供与每一个案有关并对本组织具有战略意义的调查结果、结论和建议。

²⁴ www.unicef.org/evaldatabase/index_95015.html。

²⁵ www.unicef.org/evaldatabase/index_C4D.html。

²⁶ www.unicef.org/evaldatabase/index_95016.html。

²⁷ www.unicef.org/evaldatabase/index_92843.html。

²⁸ www.unicef.org/evaldatabase/index_94431.html。

²⁹ www.unicef.org/evaldatabase/index_94430.html。

³⁰ www.unicef.org/evaldatabase/files/UNICEF-PBEA_LR_DE-Ethiopia.pdf。

³¹ www.unicef.org/myanmar/PBEA_Developmental_Evaluation_Myanmar.pdf。

³² 发展评价是了解在动态、新的或复杂的环境中实施干预的活动的办法。

67. 全球专题评价计划所列评价交付进展情况载于附件中。

六. 评价职能评估

68. 第六节介绍了儿基会已完成的若干项评价职能评估。此外，外部同行审查也在进行之中。

A. 联合国评价小组/发展援助委员会同行审查

69. 执行局要求按照订正评价政策的规定，对评价职能进行外部同行审查。经济发展组织-发展援助委员会(经合组织-发援会)、发展评价网络(评价网)和评价小组维持一个联合工作队，以协助对联合国各组织的评价职能进行专业同行审查。每次同行审查都旨在查明良好做法和机会，以进一步加强所审机构的评价职能，促进改善业绩和成果。儿基会是 2006 年首批接受同行审查的机构之一，进而拟订儿基会第一项评价政策。迄今为止，已进行了 16 次评价小组/发援会同行审查。

70. 到 2016 年底，已为评价小组/发援会同行审查做好安排，将由联合国环境规划署评价主任主持的小组进行。已商定评估儿基会评价职能的独立性、公信力和效用的职权范围。职权范围注重以下方面：评价职能的独立性和公信力；整个组织评价的质量、使用和后续落实，以促进问责制、学习和改进；儿基会评价职能表现出领导能力和远见，包括适应不断变化的环境的能力。

71. 已为该小组 2017 年访问儿基会若干地点做好安排，以期在本年度里向儿基会管理层和执行局提交报告。值得注意的是，本次同行审查将首次使用 2016 年核准的评价小组订正评价规范和标准。

B. 儿基会评价职能的自我评估

72. 2016 年，评价局着手对儿基会评价职能进行自我评估，目的是：提供证据供预定的外部同行审查参考；评估儿基会如何妥善应对评价小组各成员正在进行的努力，以加强专业精神；供今后可能修订儿基会评价政策参考。

73. 评价局协调进行的自我评估所依据的方法包括整理和审查有关文件、在线调查和各种参与性协商。回顾评价是一项共同职能的同时，不仅要特别重视整个儿基会的监测和评价工作人员的作用和看法，而且要重视评价证据用户的意见，包括高级管理层的意见。

74. 跟岗实习外部同行审查的办法，自我评估的重点放在评价小组的规范和标准，特别是涉及这一职能的独立性、公信力和效用方面的规范和标准。与工作人员和管理人员以及入驻的外部同行审查小组广泛交流评估结果。

75. 主要评估结果如下：

(a) 整个组织的工作人员和管理人员认为儿基会评价职能是可信的；

(b) 订正评价政策广为人知，并被公认为指导儿基会评价做法的框架。但有些方面可以加以调整和更新；

(c) 高度分散评价模式被认为是该组织的一项资产。然而，三层结构对这一职能的协调管理带来挑战；

(d) 虽然评价职能被认为享有相当高度的独立性，但涉及独立性的一些相关问题值得进一步的思考。认识到以下方面：很难做到分开承担小型国家办事处评价的责任；在没有适当保障的情况下依赖于方案预算可能会危及评价工作的独立性；在机构一级，各部门小组也需要有适当的职责分工；

(e) 一般认为评价是有用的，但所采取的办法必须是更方便用户，且不过于技术官僚色彩和死扣规定。及时性、知识管理和建议的质量都是需要注意的领域；

(f) 监测和评价工作人员普遍关切的是，他们没有足够的时间用于评价工作：对于小型办事处，他们报告在评价事项上花费不到 15% 的时间。在许多情况下，同一批工作人员不仅要处理评价，而且还要处理业绩监测和报告，并监测儿童权利及做调研，有时还涉及社会政策。虽然这已经是不现实的，但成果管理制度规定方面的需求日益增多，更加重了负担；

(g) 工作人员的能力各不相同，需要注意加强包括评价证据的用户在内各级的技能、知识和系统。

76. 总体而言，虽然儿基会的评价体制环境被认为是有助益的，对评价的需求被认为很高，但应对需求的能力和资源却受到制约。

C. 联合国儿童基金会(儿基会)：审查发展实效，2012–2015 年³³

77. 2016 年，评价局委托进行一项审查，审议儿基会一直如何有效实现儿童方面的成果，通过对抽样的 2012-2015 年期间 66 项评价报告进行审慎和有条理的审查加以评估。审查包括有系统地关注监测和评价对该组织的业绩和成果的贡献。

78. 关于利用评价证据提高效力的审查结果“十分正面”，80% 的样本报告审查结果获评“满意”(21%)或“高度满意”(59%)(见第 47 页)。

79. 审查列出了有助于利用评价提高效力的因素，并提出意见指出，评价报告“往往指出……利用以往项目或方案的评价得出的经验教训来加强制订当前的方案”(第 48 页)。若干项评价确认以下积极因素促进利用评价结果提高效率：

(a) 强有力的基准收集工作；

(b) 利用外部评价人员和外部评价来支持学习；

(c) 对评价采取系统的办法，承诺定期持续开展评价，并涵盖多数方案；

(d) 承诺利用各种评价(形成性、中期、方案或项目终了评价)的结果来改进方案设计。

80. 也从被审查的评价中找出消极因素。这些因素“侧重于似乎不是定期评价对象的项目和方案”(第 49 页)，特别是：

³³ www.unicef.org/evaldatabase/index_94430.html。

- (a) 主要重点放在项目监测上，而儿基会和执行伙伴对评价的关注不足；
- (b) 对评价经常方案没有作出任何承诺或提及，因而方案中没有体现评价文化；
- (c) 未经事先评价而延长方案；
- (d) 注重报告产出，忽视学习和问责。

81. 审查还将评估了儿基会各系统和评价流程的效力。在这方面，接受审查的评价中 70% 报告关于这一问题的积极结果。审查还指出，在利用评价促进业绩和成果方面，与 2012 年对类似活动进行评价审查相比，2016 年审查的评价更有可能报告积极成果。³⁴ 审查指出，“2016 年审查的评价总的来说质量较高，并表示儿基会一直在有效提高过去 3 年所作评价的质量和效用”（第 47 页）。关于性别平等的具体审查结果见本报告第三节。

82. 这些意见证实提高效力和效用以及儿基会评价职能大致胜任其职的总体情况。

D. 2015–2016 年多边组织业绩评估网儿基会体制评估报告

83. 多边组织业绩评估网的审查对儿基会进行了详细的评估，详细研究该机构的组织和发展实效，包括评价职能。根据一系列微观指标进行评估，审查对评价职能的评级为“高度满意”，对一些其他因素的评级为“满意”或“高度满意”（见表 7）。但又认为儿基会在采用评价性证据方面的业绩是薄弱的。

表 7
采用循证规划和方案拟订

序号	微观指标	评级
8.1	存在一个机构独立评价职能	高度满意
8.2	一致、独立的结果评价(覆盖面)	满意
8.3	确保评价质量所采用的系统	高度满意
8.4	强制性展示制订新干预措施的证据基础	不满意
8.5	积极主动地查明、跟踪和处理表现不佳的干预措施	高度满意
8.6	明确的问责制确保对评价建议采取应对措施和后续落实及使用	满意
8.7	吸取评价提出的经验教训和最佳做法	满意

资料来源：多边组织业绩评估网报告：附件 1。

84. 这一弱点体现在以下结论：“业绩数据尚未总是在规划和决策中得到系统地利用，或作为方案调整的参考”。虽然确认“在质量保证系统强有力的支持下，评价质量总体上正在提高”，但也指出“保持覆盖层面存在挑战”，且“尚未系统或全面地吸取和利用评价方面的经验教训”（第 27 页）。

³⁴ 经济发展组织-发展援助委员会，“审查儿基会发展实效，2009-2011 年：最后报告”（2013 年）。

85. 尽管上文提到积极的评级以及评价职能不断改进，但本年度报告其他部分以及以往各项报告提供的资料证明了这些判断，并很有益地指出需要改进的领域。

七. 结论

86. 本报告所载资料确认，总体而言，儿基会拥有继续发展和改进的有效评价职能。该组织的高度分散结构提出了实际管理挑战：需要跨时间和跨区域加强业绩一致性。此外，随着发展议程变得更加复杂和儿基会作用的变化，有必要重新讨论各种方法，并为工作人员和发展伙伴提供适当的工具和方法，以便进行有益和富有远见的评价。一个基本上分散的结构有助于快速、局部和非正规的学习，但随着儿基会本身的规模和复杂性日增，并与更广泛的利益攸关方网络接触，也需要建立更正式的机制，以便接受和适用证据和经验教训。

87. 然而，最近的审查和评估已重申，儿基会评价职能的主要实力是在整个组织广泛传播评价问责制和责任制以及工作人员和管理人员在努力加深了解在迅速变化的世界中什么是对儿童有益的方面的参与和热情。

八. 决定草案

执行局，

注意到 2016 年关于儿基会评价职能的年度报告([E/ICEF/2017/11](#))。

附件

2016 年全球专题评价计划中提出的评价状况

1. 2014 年，执行局核准了 2014-2017 年全球专题评价计划(E/ICEF/2014/3)。经修订和更新，执行局 2016 年核可了该订正计划(E/ICEF/2016/3)。
2. 下表列出计划在 2016 年进行的评价状况。2016 年完成的所有评价清单载于本报告第五节(指标 7：机构一级评价)。

表 1
2016 年全球专题评价拟议专题

类型	专题	2016年底的状况和完成时间表
主要评价	加强国家卫生系统和能力	已完成的全面范围界定工作和已启动的主要评价 国家案例研究：2017 年第四季度 全面综合报告：2018 年第三季度
主要评价	女童教育和性别平等	进行中 最后报告：2017 年第二季度
主要评价	农村/小城镇水供应方案编制	进行中 最后报告：2017 年第四季度
主要评价	预防发育障碍：改善公平利用营养支助和护理	进行中 最后报告：2017 年第二季度
评价	预防艾滋病毒母婴传播	已完成
评价	保健 4+组织孕产妇、新生儿和儿童保健联合方案(联合)	进行中 最后报告：2017 年第二季度
评价综合报告	艾滋病毒和艾滋病领域的评价结果综合报告	进行中 最后报告：2017 年第二季度
评价综合报告	联合国儿童基金会(儿基会)：审查发展实效，2012-2015 年	已完成