



经济及社会理事会

Distr.: General
19 April 2013
Chinese
Original: English

供采取行动

联合国儿童基金会

执行局

2013 年年度会议

2013 年 6 月 18 日至 21 日

临时议程* 项目 8

关于评价职能和主要评价工作的年度报告

摘要

本年度报告是按照评价政策 (E/ICEF/2008/4) 和执行局有关决定编写的。报告介绍了评价工作的全球发展情况以及儿童基金会评价职能的现状, 包括在加强评价职能的权力下放方面的进展情况。

决定草案载于第八节。

* E/ICEF/2013/10。



一. 导言

1. 儿基会的评价工作是为了实现儿基会保护和促进儿童权利的宗旨，解决儿童的需要，帮助儿童在正常发展过程和处于危机时刻充分发挥潜能。评价促进了本组织的学习、问责制和透明度，以期加强绩效，为儿童提供更好的成果。评价就政策、方案和举措的执行和结果提供可靠的证据，用来支持决策和宣传。

2. 本报告介绍 2012 年儿基会评价职能的绩效。近几年来，年度评价报告也包括关于一年内主要评价工作的评论。执行局希望看到关于最近评价工作的深入专题分析，并要求就此另外提出报告。因此，本报告扼要介绍 2012 年评价职能的现状，另一份报告 (E/ICEF/2013/15) 则就儿基会人道主义行动的最近评价提出专题综合报告。

3. 儿基会的评价职能是按照执行局在 2008 年核准的评价政策行使的。近几年来，儿基会内部和广泛的发展背景都出现了重大变化。预期还会出现进一步变化，儿基会自己正在拟订新的中期战略计划；此外，制订可持续发展目标以指导今后发展活动的工作已经开始，并要改变人道主义议程，做出更有力的回应。处于这样不断变动的环境，执行局请儿基会更新评价政策。经修订的评价政策 (E/ICEF/2013/14) 已经提交执行局审议。

4. 本报告涵盖了儿基会评价职能的要点。报告介绍儿基会在全球一级支持评价工作的努力，特别是联合国系统内机构间活动；重点讨论儿基会自己的评价职能的绩效，根据一套关键指标来衡量；对分配给评价的人力资源 and 财政资源提出意见。报告接着概述为加强儿基会评价职能所采取的措施；提出关于加强国家和全球各级的评价能力所作努力的报告。在结论之后，报告提出一项决定草案供执行局审议。

二. 支持在全球一级的评价行动

5. 儿基会在全球一级进行了大量的评价活动，大部分是为了支持广泛的联合国系统内的评价工作。这些活动主要有三个方面：(a) 关于发展、传播和协调技术知识和良好做法的工作组和会议；(b) 机构间和联合评价以及有关活动；(c) 总体领导和治理进程。大部分工作都是通过联合国评价小组进行或与其协调后进行，也通过机构间常设委员会 (机构间常委会) 进行。本节概述在全球一级的主要工作及其对今后行动的影响。

协调技术知识和良好做法

6. 联合国评价小组从一开始就安排各个机构间技术工作组制订技术和政策指导，并累积完成了大量工作。过去一年来，儿基会主导的好几个工作组取得了进展：

(a) 联合评价工作队与联合国发展业务办公室合作，为联合国发展援助框架的评价职权范围以及管理部门对联合国发展援助框架评价的答复拟订准则。工作队正在拟订关于规划和管理联合评价的资源包，这方面的工作量和复杂性都在增加。这项工作的目的是改进联合国发展援助框架评价的质量和使用，因为普遍认为它在质量和更新资料方面比不上别的评价。

(b) 国家评价能力建设工作队编写了一本手册《如何加强国家评价系统的切实可行的建议》，作为技术和非技术工作人员的指导。此外，儿基会和其他联合国机构与联合国评价小组一起，支持一个重要的伙伴关系举措（‘EvalPartners’），旨在加强民间社会的评价能力（详情见第六节）。

(c) 人权和性别平等工作队完成了把人权和性别平等纳入评价的准则，填补了联合国评价小组指出的联合国评价工作的一个缺口。

(d) 影响评价工作队正在制订影响评价技术准则，特别提到联合国内部的规范工作和机构支持。联合国评价小组每年召开影响评价网网络会议，以便与经济合作与发展组织发展援助委员会（经合组织/发援会）和国际金融机构的评价伙伴讨论最近的发展情况。

7. 虽然大多数工作队都很成功，仍然有一些并无活动，还有一些需要与更多成员分享工作。联合国评价小组产出的利用和效果还没有经过系统的审查。为此，联合国评价小组已经委托外部人士来进行评价，以便于 2013 年在充分了解情况的基础上进行反思和采取行动。儿基会对此表示欢迎，希望联合国评价小组将重新调整重点和工作。

机构间评价和联合评价以及有关活动

8. 儿基会在 2012 年执行的联合评价和有关活动，比通常的数目多了一倍，发展活动和人道主义行动各占一半。这些联合评价包括以下几项：

(a) 联合国女孩教育倡议的评价。儿基会与联合国女孩教育倡议的各个合作机构合作，完成了这一重要的评价，旨在审查联合国女孩教育倡议的治理结构和管理组成，找出支持建立健全伙伴关系以促进女孩教育的因素。在埃及、尼泊尔、尼日利亚和乌干达进行了国家个案研究，对东亚—太平洋伙伴关系进行了区域个案研究。全球咨询委员会代表所有联合国女孩教育倡议的伙伴提出了联合管理部门的答复。这个评价加上管理部门的答复阐明了联合国女孩教育倡议的新战略方向，是在乌干达召开的全球会议上商定的。儿基会自己另外提出了管理部门答复，主要是针对它作为牵头机构和联合国女孩教育倡议秘书处主管的建议。

(b) 联合国系统性别平等联合方案的联合评价。这是由联合国促进性别平等和增强妇女权能署（妇女署）牵头的活动，参加的还有儿基会、联合国开发计划署（开发署）、联合国人口基金（人口基金）、实现千年发展目标基金、以及挪威政府和西班牙政府。评价审查联合性别方案对于改进联合国系统的一致性做出什么贡

献，确定该方案把促进性别平等和增强妇女权能实现到什么程度。该评价在 2012 年开始实施，将在 2013 年年中结束。

(c) 人口基金-儿基会切割女性生殖器官联合方案的联合评价。该评价由儿基会和人口基金联合管理，正在联系主要的国家对应机构，评估联合方案在 2008-2012 年期间使得切割女性生殖器官现象在什么情况下不再发生，以及进行到什么程度。该评价将在 2013 年 7 月结束，重点是执行联合方案的 13 个国家。在四个国家(布基纳法索、肯尼亚、塞内加尔、苏丹)深入进行个案研究。

(d) 关于武装冲突中严重侵害儿童行为的监测和报告机制的机构间研究。该研究由儿基会牵头，与联合国维持和平行动部和负责儿童与武装冲突问题秘书长特别代表办公室合作，将填补知识上的缺口，报告关于监测和报告机制的良好做法。

9. 机构间联合评价的管理通常比较复杂，执行所需时间比单独一个机构的长，但儿基会的客户普遍认为投资取得了很好的回报。这些经过仔细规划并获得充分资源的评价的质量一直很好。另一方面，关于应对人道主义紧急情况的机构间联合评价的质量却参差不齐，尽管参加的机构做了相当大的努力。¹ 儿基会对于修订各种联合评价人道主义情况方法的机构间努力做出了重要贡献，包括实时评价，以期改进质量和使用。参加联合评价——“统一评价”——是全面评价联合方案的重要机制，特别是在缺乏联合国全系统评价的有效安排的情况下。不过，鉴于整个联合国的评价能力有限，联合评价与个别机构内部评价的需要之间谁先谁后，必须取得适当的平衡。

总体领导和治理进程

10. 常务副秘书长 2012 年委托进行的对于全系统评价安排的审查结论认为，² 整个联合国关于机构间事务的评价缺乏统一的领导，治理安排薄弱。联合国评价小组是一个专业网络，但在联合国系统内没有指导或管理评价活动的正式职权，尽管它为联合国内部的评价专家提供了有用的协调、咨询和分享看法的论坛。

11. 关于一体行动的独立评价。联合国评价小组对于一体行动的重要研究提供咨询支持，向会员国说明“一个联合国”议程的进展情况。联合国评价小组也是就有关评价问题提出意见的渠道，由联合国系统内发展方面业务活动四年期全面政策审查加以审议，包括上述审查联合国全系统的评价安排。

12. 联合国评价小组是一个有用的征询意见的地方，并提供技术咨询；它可以为联合国关心的评价问题提供咨询和宣传的重要平台。重要的问题包括：评价可以

¹ 见 E/ICEF/2013/15。

² Angela Bester and Charles Lusthaus, *Comprehensive review of the existing institutional framework for the system-wide evaluation of operational activities for development of the United Nations system. Final Report*, March 2012.

提供相关投入，以提高效率、效用和性价比；在驻地协调员办公室和广泛的联合国国家工作队建设监测和评价的能力，支持在国家一级充分评估联合国的绩效和结果。通过联合国评价小组，儿基会与其他成员合作，就评价问题提供更为清楚、有力的专业意见。

三. 儿基会的评价职能：绩效和结果

导言和概述

13. 儿基会在 2006 年第一次向执行局提出关于评价职能绩效的报告，此后每年提出报告。随着信息协调改进，内容也逐渐增加。

14. 六组关键绩效指标概述了绩效情况。³ 根据现有的数据可以进行一些趋势分析。根据关于趋势的数据，得出四个主要结论：

(a) 近几年来，进行评价的数目和专题的范围都低于最低必要水平；这两方面都必须加强；

(b) 影响评价和累积性评价在整个评价活动中正在减少。虽然这可能是因为实施新方案而出现的暂时趋势，但如果是表示儿基会忽视成果和影响，则值得关注；

(c) 自 2000 年代中期以来，评价质量的改进有限；

(d) 儿基会成功地建立了一种文化，管理部门对评价做出答复已成为常规。

指标 1：接受管理评价的范围以及提交全球评价数据库的数量

15. 儿基会四年前开始实施比较有效的监督制度，在此之前，各个办公室声称每年进行的评价多达 250 项。但是提出报告的只略多于该数字的一半，而且许多报告很迟才提出。管理部门的答复为数极少。因此，评价办公室建立了全球评价报告监督系统，强调加强关于评价原则的培训，向中央数据库及时提交评价报告和管理部门答复，以及在儿基会外部网站公布评价报告。全球评价报告监督系统提供关于报告的及时评估和评级，提交给办公室关于报告质量的回馈意见，以及把绩效信息编入管理部门信息表。

16. 这些内部能力建设和监督措施大致很有成效。在评价办公室与各区域办事处和国家工作队的密切合作下，2012 年对全球评价数据库的提交率达到 98%。而且，所有报告和质量评估都在儿基会网站上免费公布，很容易通过专题、国家、区域或时间等标题进行检索。⁴

³ 在大多数情况下，可以根据过去三年的情况确定趋势。对于有些问题，要根据过去十年才能确定可靠的趋势。

⁴ <http://www.unicef.org/evaldatabase/index.html>。

17. 不过，提交报告的数目从 2010 年的 140 件下降到 2011 年的 99 件。遗憾的是，在 2009-2011 年期间，有 38 个国家办事处，其中包括好几个管理大型方案的办事处，似乎从来没有进行过评价。还有几个只进行了一个或两个评价，尽管 2008 年的评价政策规定，所有方案构成部分应该在方案周期内至少评价一次。不仅如此，虽然人道主义回应的数目和资源都已经增加，儿基会对 2008 至 2011 年期间的 1 025 个紧急情况做出了回应，却只评价了 34 个。重要的是，绝大多数的评价是对 2 级和 3 级紧急情况，对儿基会最常见的 1 级人道主义情况却大都不予理会。有一些评价虽然进行，却可能没有提交中央数据库。不过，如果全球评价报告监督系统的数字准确无误，许多国家办事处以及整个组织都没有掌握机会，创造重要的知识来进行管理、学习和问责。这里需要一个关于评价范围的新绩效指标，需要对按照该指标收集的信息核实其准确度，并找出评价范围始终不佳的原因。

指标 2：主题分配

18. 2011 年的结果显示，评价范围涵盖所有战略重点领域。评价主题分配的数字载于下文表 1。

19. 随着时间演变，儿基会经常方案每个部门或重点领域的评价数目变动很大。最明显的趋势是多部门评价倍增，2009 年是 12%，2011 年增加到 25%。这可能是好几个变化的副产品：走向性别平等目标的综合方案编制；走向“上游”，在那里的政策引导举措(如现金转移)影响好几个部门的结果。因此，这似乎是个积极的变动。同样，艾滋病毒/艾滋病评价减少很多，从以前的 10%下降到 2012 年的 4%。评价办公室将继续关注这些趋势。

表 1

2011 年评价报告的主题分配

主题	百分比*
多部门	25
教育和性别	25
儿童幸存和发展问题	13
儿童保护	10
政策宣传和伙伴关系	4
艾滋病毒/艾滋病	4
各种交叉专题和组织绩效	20
未分类	3

资料来源：2012 年全球评价报告监督系统的报告。

* 总数超过百分之百，因为有些报告涉及几个主题。

指标 3：评价的类型

20. 评价评估不同方案层面的结果；这表明机构的关注所在。例如，强烈面向结果表明希望对影响和成果、而不是对产出进行评价。2009–2011 年的数据显示出矛盾的趋势（见表 2）。在积极的一面，产出层面的评价从 2009 年的 33%减少到 2011 年的 27%。然而，影响层面的评价减少更多，从 2009 年的 43%减少到 2011 年的 27%。另一方面，成果层面的评价从 2009 年的 24%增加到 2011 年的 46%。

表 2

评价的类型：2009–2011 年趋势

指标	2009 (%)	2011 (%)	趋势走向
产出层面评价	33	27	减少：可取
影响层面评价	43	27	减少：不可取
构成层面评价	45	59	增加：原因？
累积层面评价	55	41	减少：不可取

21. 这些趋势可能反映出机构战略的演变状况。成果评价增加可能是因为更加注重加强国家能力，而准确的模式增加和更多使用调查数据导致可以用监测方式来衡量影响，不必再用评价的方式。

22. 另一个趋势表现于在方案周期的哪一个阶段进行评价。在周期初期阶段进行的构成评价—评估方案是否按照计划编制，以便及时提供证据来指导调整—几乎增加了三分之一，从 2009 年的 45%增加到 2011 年的 59%。在方案周期末期进行的累积评价调查累积的结果。这类评价的百分比从 2009 年的 55%下降到 2011 年的 41%。重新注重公正可能是促成这一转变的原因，因为方案管理人要确保调整结构后的方案战略能够覆盖处境最不利的群体。不过，现在还不能确定，而缩小评价范围和减少累积评价则有可能扩大关于成果和影响的证据缺口。

指标 4：儿基会评价的质量

23. 评价质量的重要性不言而喻。高质量可以保证评价的结果可靠，可以放心使用，并保证值得对评价投资。

24. 应该指出，良好或杰出的评价报告几年来略有增加。相当大一部分的评价报告属于“几乎令人满意”或“几乎有信心采取行动”的类别。鉴于评级过程很严格，这类报告通常都是因为提出时的一些技术问题或缺漏使得评估低于“满意”的级别。不过，这种情况的评价一般来说都很不错，问题出在报告的方式，不在实质内容。

25. 无论如何，相当多的评价都属于“拙劣”类。⁵ 在这里不妨从以前的元评价质量的角度评论最近的全球评价报告监督系统的评估结果。从这些最需要质疑的评价可以看到两个趋势。

26. 第一，2011 年儿基会的评价有 23% 被评级为“拙劣”，这比 2000-2002 年的 33% 的基准数已经有了改进。⁶ 可是，基准数虽然在 2004-2006 年大举加强能力行动后下降了，自那次以后就很少改变。在当前安排下要减少占 20-25% 的拙劣绩效类别似乎很困难。因此，第五和第六节概述正在进行的新举措或加强举措，并提出供今后几年审议的其他措施。

27. 第二，在进行评级时，一些评价做法得到推广。根据严格的标准，将人权、性别和公正纳入的评估从 2010 年的 18% (评价这些问题的第一年) 增加到 2011 年的 33%。同样，利益攸关方令人满意的参与率从 2010 年的 40% 增加到 2011 年的 52%。

指标 5：评价的使用情况，包括管理部门的答复

28. 2008 年评价政策规定，每一个评价都必须完成管理部门的答复。宣传、培训和监督使得在遵守规定方面取得显著进展：一是在准备管理部门答复方面，一是在把完成的答复上载到跟踪系统方面。把管理部门的答复提交全球跟踪系统已经达到 93%，而 2009 年的基线是 10%。这样高比例的答复是过去三年来的巨大成就。⁷

29. 不过，评价的使用不只是依靠拟订答复，管理部门答复的承诺得到实施才是关键。遵守规定的情况从 2011 年的 62% (实施管理部门答复跟踪系统的第一年) 增加到 2012 年的 82%。这些数字都是由办事处自己填报，因此需要独立的核查工具，确认答复得到实施，并符合声称的完成程度。

指标 6：机构一级评价

30. 机构一级评价由评价办公室管理，列入综合监测和评价框架的名单。2010-2011 年综合监测和评价框架提出承诺，要进行 15 个机构一级的评价，其中 13 个 (86%) 已经完成，2 个 (13%) 取消 (因为数据和资金不足)。现在正在执行的 2012-2013 年全球综合监测和评价框架是在广泛协商后拟定的，以确保评价的主题相关并符合要求。核准的 14 个评价主题中，7 个正在进行，5 个订于 2013 年开始。为了让评价办公室能够充分监督，已提议把余下的 2 个改为 2014 年开始 (见附件)。

⁵ 2009 年是 15%，2010 年 30%；2009-2011 年的平均数是 23%。

⁶ 这个类别现在的名称是“没有信心采取行动”，原来的名称是“拙劣”。

⁷ 许多证据显示，过去根本没有提交过管理部门答复：2000-2002 年的元评价显示，75% 的办事处没有对评价提出管理部门答复的纪录。

31. 全球评价议程的相关性已有改进。不过，它没有完全涵盖主要的方案和战略。2014–2017 年战略计划提供了机会，就证据的产生、统一评价、国家统计监测和研究采取全面的办法，并得到公平结果监测系统、包括人道主义绩效监测组成部分的支持。在这个框架内，并与战略计划密切整合，提议把全球评价计划提交 2014 年执行局。

其他绩效指标

32. 现在使用的绩效指标推动在儿基会各级采取行动，表明有助于管理评价职能的信息不用太多费用就可产生。不过，评价范围仍然有一些重要缺口。关于专门用于评价的费用的数据，是计算效率所必需的。除了跟踪管理部门的正式答复，还需要关于评价的使用和附加价值的信息，以确保该职能产生最大利益。也需要关于证据的辅助来源的信息，以拟订整体的证据产生战略。最后，必须衡量绩效以管理和加强举措，提高内部、伙伴和国家能力。为此，儿基会认为，今后几年有必要试行一套简单而翔实的补充指标。

四. 儿基会的评价职能：人力和财政资源

33. 几年来，儿基会每两年向执行局报告用于评价职能的人力和财政资源的最新具体情况。这样做可以报告真正的趋势，而不是一年的变动起伏情况。继去年的两年期最新情况之后，⁸ 本报告从比较大的角度介绍人力资源和供资情况。

人力资源基础

34. 与几乎所有其他发展机构相比，儿基会对于评价人员的投资最大。⁹ 大约 75% 的国家办事处设有职称中含“评价”二字的专业员额，这是很突出的配置。国家工作人员的百分比相对较低(32%)，但也提供了职业发展机会。女性工作人员的百分比相对较高(44%)。

⁸ E/ICEF/2012/13。

⁹ 儿基会在各发展机构中确是如此，但明显落后于联合国负责人道主义工作的其他机构；以专门负责紧急人道主义援助的工作人员与评价产出的比率比较，儿基会是 1: 12，世卫组织是 1: 4.7，粮食署是 1: 3.7。

表 3
评价的人力资源

指标	2012 (%)
员额的职称中含“评价”二字的办事处百分比	75
女性工作人员百分比	44
监测和评价干事用于评价时间的百分比	14
评价职责捆绑部门职责员额的百分比	32
儿基会监测和评价员额任职者的在职时间	<2 年: 20% 2-5 年: 49% >5 年: 31%
区域或国家办事处评价员额的百分比	89

35. 然而，必须为以下问题找出解决办法：

(a) 大约 25%的办事处没有这样的员额；

(b) 还有 25%这样的员额职等是 1 或 2 等，不适合管理高度技术、复杂和敏感的任务；

(c) 这些员额明显集中在比较大的办事处。比较小的办事处，特别是在中等收入国家，必须采取创新的方式和解决办法才能对儿基会类型、规模和环境各异的办事处支持高质量的评价；

(d) 监测和评价干事具有多重角色和职能，他们平均用于评价的时间只占全部工作时间的 14%；

(e) 监测和评价员额往往与其他任务捆绑在一起，包括可能需要评价的方案职务，因此可能产生利益冲突；

(f) 监测和评价工作人员从事监测和评价工作的时间平均只有 4.5 年，这是因为鼓励和惩罚的因素参半，包括评价人员的晋升途径不明显，而监测和评价人员竞争管理职位的成功率很高。需要一批长期从事专业工作的干部，才能担负产生证据这样越来越困难的工作。

36. 这些员额的工作地点明显反映出儿基会是一个权力分散的组织。儿基会把 89%的监测和评价人员分派到外地办事处，让他们与方案工作人员密切接触。这个分散运作模式与大多数其他机构的模式明显不同，其他机构大都依靠它们的中央评价干事。不过，这也造成区域同事要求给予支持和保证质量，而由于其他职责太多，难以满足这些要求。

37. 儿基会的评价干部要从事各种机构间活动。这些活动要付出机会成本，包括无法满足儿基会方案的需要。许多其他机构没有像儿基会那样在外地具有众多监

测和评价人员，对国家一级的监测和评价人员配置投资有限。随着联合国系统的演变，必须采取协调一致的办法来满足设置方案的监测和评价人员的需要，包括人道主义情况。

加强内部评价人员能力

38. 具有适当技能的工作人员是获得高质量评价的重要推动力量，其他因素还包括坚定的领导和充足的供资。相当数额的投资继续用于改进内部各级评价能力。这方面的工作包括知识管理活动，例如网上实践社区、网上研讨会、服务台功能、电子公报和时事通讯、以及常规网络会议、考察访问和培训研讨会等。

39. 2012 年进行了重大的调整和改进，包括：

(a) 使用在线培训课程的自学活动。这包括一个针对方案工作人员以及监测和评价专家的新课程，讲授如何管理高质量的评价和有效利用评价；

(b) 关于重视公正的评价的新的工具和培训。这些在机构评价领域的工作，既有一般性说明，也有详细的方法指导。一份关于设计和管理关注公正的评价手册已翻译成阿拉伯文、法文、俄文和西班牙文分发。手册分发之后，在儿基会内外进行培训以及办理非常成功的新电子学习方案，约有来自 168 个国家的 5 000 人参加。

40. 新的活动很重要，但是否能够把儿基会的评价提升到新的高度并改进评价的范围和质量，还很难说。新的评价政策还必须要有全面的战略配合，从而侧重建设工作人员能力和增加使用外部人力资源。

财政资源

41. 根据 2012 年的支出数据，儿基会方案预算大约 3%用于评价、分析、研究、学习、数据获取、调查和统计。这符合 2008 年评价政策规定的把支出总额的 3%到 5%用于各种产生证据的途径。不过，鉴于现在的趋势是更多上游方案编制、支持国家政策和宣传、以及全球宣传以支持儿童权利，应该大幅度增加对证据产生的投资。

42. 2010 年关于评价的具体数据显示，外地办事处的方案预算有 0.33%用于评价活动(不包括人员费用)。全球来说，如果把评价办公室的资源算进去，支出增加到 0.4%。考虑到本报告别处提到的范围和质量问题，该职能显然资源不足。这个比率低于联合国全系统评价最近审查提到的 1%的名义目标，¹⁰ 更远低于几个参照组织(包括妇女署和美国国际开发署(美援署))规定的 3%的支出准则。¹¹

¹⁰ Angela Bester and Charles Lusthaus, 同上, p. 39: “虽然没有正式的预算比例标准，但一项广泛使用的指南是评价支出占总预算的 1%。”

¹¹ 《联合国促进性别平等和增强妇女权能署的评价政策》，妇女署，2012 年；《美援署评价政策》，美援署，2011 年。

43. 而且，过去几年来评价办公室的资金来自核心资源和其他资源，由进行某一评价的方案单位直接捐给该办公室或指派给该办公室。过去三年来，预算的 38% 来自其他资源。2012 年，全球专题资金第一次提供拨款，因为认识到供资与要评价的部门脱钩有助于提高独立性。不过，专题资金只能断续地提供，也不能提供给所有评价主题领域。因此，必须增加提供核心资源的资金来支持全球评价议程。这将保障全球评价的独立性，并且由于核心资源可以预测，将加快资金的提供速度。

表 4

评价的财政资源

指标	2010 (%)
外地办事处方案预算用于评价的百分比	0.33
儿基会整个方案预算用于评价的百分比(外地加纽约总部)	0.4
儿基会整个方案预算用于评价、监测、研究、学习、调查的百分比	2.95
	38
来自非核心资源的评价办公室预算：实际支出的百分比	(2010–2012)

44. 儿基会最近改变了管理信息系统，使得分配给评价的资源更难衡量。评价支出的数据编码不如以前准确，使得监测和报告更为困难。这个问题将在更新机构系统时纠正。

45. 必须增加对评价的投资，但这只不过是问题的一部分。即使资金充足的评价也可能得出拙劣的结果。在整个评价过程都必须明智地管理资金，并具备良好评价所需的先决条件，包括设计详尽的方案以及能够提供充分监测数据的完备数据库。

五. 加强评价系统

审查评价的职能和政策

46. 2012 年，评价办公室委托外部咨询人收集和分析关于权力分散的评价职能的证据。关于儿基会评价职能和政策的综合报告可在评价办公室网站上查阅。¹² 以下各节引述该报告的数据，并加上来自其他来源的证据。

47. 综合报告认为，评价政策仍然相关，并密切配合联合国评价小组的规范和标准。综合报告也认为，虽然评价政策是各办事处了解机构目标及其问责制的基础指南，但政策的一些部分必须更新。修订的评价政策提出了需要更新的问题，在此无需详述。审查还发现一些妨碍政策执行的因素。本节阐明一些需要注意的问题。

¹² 《关于儿基会评价职能和政策的综合报告》，Universalia，2013 年。

治理和领导

48. 综合报告认为，权力下放的评价职能相当适合儿基会权力下放的方案环境的业务。因此，区域和国家各级高级管理人员的领导作用就很重要。为了支持领导人员，经常通过各级仪表板和计分卡提供关于主要绩效指标的全球评价报告监督系统的结果和其他数据。综合报告认为这项工作效果很好，有助于提高认识和持续行动。中东和北非已经就计分卡开始区域协作。

49. 鉴于需要加强评价职能，导致各区域提供有效的监督，变革大都限于评价的经常方案。在中欧和东欧，独立国家联合体、中东和北非以及美洲和加勒比区域，区域主任在 2012 年向所有代表发出正式备忘录，促请更加注意评价职能。中欧和东欧、独立国家联合体以及美洲和加勒比区域确认监测和评价是核心职能，在具有特殊需要和能力的中等收入国家的区域内属于优先事项。这些区域办事处已开始探讨这对于加强用于评价职能的人力和财政资源会有什么影响。它们的战略构想倾向于采取不同于以前模式的监测和评价结构(即多个国家共同使用监测和评价干事)，必须修改职务说明，高度重视伙伴关系，并支持迅速改变政策层次的评价。

50. 一般来说，现在重新重视赋予更多工作人员进行评价的职权，并实行问责制。2008 年评价政策没有充分说明不同评价利益攸关方的作用和责任。修订的评价政策、修订的内部方案手册以及关于管理评价的新的线上工作人员培训课程都比较清楚地规定了评价问责制。

51. 虽然取得上述进展，儿基会从来没有就区域和国家各级的评价职能规定明确的绩效标准。现在有必要定出标准，同时保持包容不同政策、方案和业务环境的灵活性。采用明确的绩效标准后，将清楚说明期望达到的成果，又让办事处能够灵活决定如何克服评价的挑战。

评估评价的需求

52. 鉴于评价范围出现缺口，评价的数目相对较低，必须了解需求是怎样产生的。一个关于评价需求和使用的评估调查了儿基会代表的意见，提出了一些答案。¹³ 值得注意的是，许多代表认为关于评价的内部和外部需求都在增加(内部 44%，外部 46%)，少数代表认为需求在下降(内部 16%，外部 5%)。其余代表认为没有改变。需求的来源包括：工作人员和国家伙伴的专业判断(占需求的 32%)，捐助方(29%)，强制性规定(26%)。

53. 需求虽然增加，投资或评价数目却没有相应地增加。财政限制可能是部分的原因，但我们看到许多资源相对充足的办事处也出现评价缺口。因此，财政限制不能完全说明问题。

¹³ 《内部审查儿基会评价政策和职能的调查报告》，Universalia，2012 年。

54. 另一个矛盾现象是关于评价的使用情况。儿基会的代表们表示，评价最重要的目标是支持基于证据的政策宣传(占答复的 50%)，其次是方案决策需要(28%)和问责制(17%)。这些目标完全符合儿基会评价政策的目的。然而，如果政策宣传如此重要，就应该看到更多的影响评价和累积评价，而不是更多的构成评价。

55. 展望今后的需要，代表们最感兴趣的是重视公正的评价(60%)、影响评价(52%)和政策评价(42%)。创新是另一个日益增加的需要。儿基会在加强用于重视公正评价的能力方面已经做了很多，但在其他方面还需要更多支持，而如何针对儿基会多样的个案环境制订新的方法将是很大的挑战。

加强评价的战略规划和范围

56. 面对要求提供有用的分析的压力，促使在战略规划和评价方面进行创新和协作。区域办事处和区域管理小组再度成为推动力量。

57. 在中东和北非，面对叙利亚危机，国家办事处在区域支持下制订了救助难民的人道主义方案的次区域决策和评价系统，中西部非洲区域办事处(中西非办事处)目前正就马里日益扩大的危机启动实时评价。在中东欧/独联体，区域管理评价小组核准了区域知识和领导议程，内容包括有关减少侵害儿童权利和缩小公正差距的 10 个重要成果领域。重点是正在相当多国家执行的方案，许多材料都随时可以纪录和评价。面对如此丰富的经验，中东欧/独联体开始就五个主题进行多国评价，评估减少侵害儿童权利和公正差距的程度，并确认儿基会如何帮助进行改变。

58. 所有区域都有政策规定，在国家一级综合监测和评价计划执行前加以审查；这项工作的范围和内容都正在扩大。好几个国家已经在审查国家方案文件和国家方案行动计划的“可评价性”，帮助办事处调整其方案的设计以及监测和评价战略，以确保必要时在方案周期进行彻底评价。

促进和支持质量评价

59. 高质量评价不仅需要必要的人力和财政资源，还需要采取适当的程序和技术。拟订职权范围是评价设计的关键时刻。2000–2002 年的基线显示，被评为“拙劣”的评价有 75%从一开始就错了，因为没有制订好职权范围。最近的全球评价报告监督系统数据显示，职权范围制订得好，就会对最后报告的质量有好的影响，这样的案例大约占 50%。可以由区域办事处或外聘公司订约来审查和评论职权范围，对制订职权范围提供支持。

60. 中东欧/独联体的分析指出，另一个系统的特色——对最后报告草稿进行正式审查——有系统地改进了评价报告的质量，不是提出建议，就是要求收集更多数据和创新分析。未经审查的报告草稿评级总是最低。现在已经证明在关键时刻的独立审查很有价值。

61. 因此，今后的一个清楚目标是在内部或由外部专家协助，扩大质量支持和质量保证的范围。必须对职权范围和报告草稿进行审查，成为高级管理人员监测和使用的一个绩效标准；这种审查可以委托别人进行，花费不会太多。

62. 有些国家办事处已经在这方面建立创新安排和伙伴关系。例如，布隆迪国家方案与加拿大卡尔顿大学就质量保证建立了伙伴关系，津巴布韦方案设立了外部英才中心，审查评价并支持其他研究和知识管理任务。

加强评价的使用和管理部门的答复

63. 由于有效的监测和后续行动，在准备管理部门答复和提交全球跟踪系统方面都取得了进展。儿基会除了管理部门正式的答复机制，也采用妥当的传播和使用办法。上文提到的委托别人进行的调查，现在第一次提出了关于需求和使用的全系统数据。

64. 代表们的回答表明，评价的使用率很高。评价结果被纳入基于证据的政策宣传(53%回答“大量使用”)和方案决策(42%)。不过，调查显示还需要改进以下方面：75%的代表表示，评价结果“有时候”系统地向外传播，只有18%回答“总是”(还有7%回答“不确定/不知道”)。儿基会的办事处必须改进传播工作，才能实现评价投资的重大利益。还有一些良好做法的最新例子：制作概况介绍(中东欧/独联体、乌干达和津巴布韦的方案)；通过联合国国家工作队传播评价结果(伊朗)；通过国家评价协会向议会说明评价结果(摩洛哥)；就评价的传播和使用进行系统记录和分享创新(东亚和太平洋区域办事处)。

65. 除了传播问题，评价证据是否有效使用的问题也还没有得到具体审查。2013年将有两个全球评价处理这个问题：全面评价儿基会的政策宣传绩效，和部门评价儿基会在上游教育宣传的工作。这两个评价既审查成功的故事，也注意宣传失败的例子。

六. 国家评价能力建设

加强评价能力的全球伙伴关系

66. 强大的国家评价能力能够切实发挥促进发展的作用。例如，我们已经看到，拉丁美洲的几个主要社会方案有效使用评价来指导决策。鉴于强大的国家评价能力可能产生的影响，使得儿基会和其他捐助者提供投资差不多20年，并得到方案国家的评价专业人员和 supporters 的支持和鼓励。这项工作迄今主要由国家和区域评价协会担任。然而，这里需要的是使得全球工作更加相互关联和得到更多资金。过去一年取得了重大进展，形成了协调一致的全球伙伴关系，以此促进各级的评价能力建设：

(a) 创建了EvalPartners，这是加强民间社会评价能力的举措(与加强公共部门的能力相辅相成)。31 个成员国包括所有区域评价协会、联合国评价小组、开发署、妇女署、非洲开发银行、西班牙政府、美国政府、洛克菲勒基金会和其他主要利益攸关方；¹⁴

(b) 建立了基于公正的特派团，为社会的处境不利者服务；

(c) 商定概念框架，着重加强国家和区域各级的评价协会(与加强个人技能和知识的举措相辅相成)；

(d) 启动 EvalPartners 网站(与社交媒体挂钩)来方便信息传播；

(e) 调查区域各国家评价协会或“专业评价支援组织”，查明 110 个国家的 150 个协会，以及关于它们的切实绩效的个案研究；¹⁵

(f) 召开关于民间社会评价能力的国际论坛。¹⁶ 这个论坛在 2012 年 12 月成立，有助于分享评价协会和其他重要利益攸关方的良好做法和教训，确定 2013 年的EvalPartners优先事项。所有伙伴承诺实施《清迈宣言》的保证，与民间社会伙伴合作，支持评价工作；¹⁷

(g) 拟订宣传战略，促进执行 2013-2015 年评价的有利环境：特别是把有力的评价承诺纳入实现预期的可持续发展目标过程。

67. 儿基会是 EvalPartners 的主要推动者和积极参与者，并得到好几个捐助机构和基金会的慷慨支持。随着 EvalPartners 趋于成熟，这种支持将继续下去。要实现伙伴关系的宏伟目标，必须明确地设定目标，仔细监测，并在适当时候进行独立评估。

全球学习

68. EvalPartners得到全球学习平台MyM&E的帮助，这是一个方便用户的知识管理系统，旨在加强个人和国家的评价能力。MyM&E的许多加强能力方法都可以在网站上检索。¹⁸ 该网站自 2011 年以来访问者增加了五倍，2012 年几乎达到 250 000 访问者(来自 168 个国家)，下载增加了八倍，2012 年几乎达到 100 万页，足以说明其对用户的价值。

¹⁴ 见 <http://www.mymande.org/evalpartners>。

¹⁵ 见 <http://www.mymande.org/evalpartners/international-mapping-of-evaluation>。

¹⁶ 见 <http://mymande.org/evalpartners/forum>。

¹⁷ 见 http://www.mymande.org/sites/default/files/Declaration_evalpartners_English.pdf。

¹⁸ 见 www.mymande.org。

69. MyM&E经常更新它的产品。例如，2012 年启动了关于发展评价的新的电子学习方案，由 33 位世界级的主旨发言人讲课。¹⁹ 到 2012 年 12 月，已有 168 个国家约 8 000 人登记。用户的满意率极高，72%的参与者表示满意，88%表示希望参加未来的电子学习课程。发展费用经过匀支后，效率极高，参与者的平均费用大约是 2 美元。

加强国家主导的监测和评价系统

70. 儿基会像过去几年一样，继续加强国家评价系统，特别是与其他联合国机构建立伙伴关系。执行局一直鼓励这方面的工作，并强调联合国大会关于四年度全面政策审查的决议。

71. 2012 年的一个调查活动确认大约有 120 个儿基会国家办事处在从事加强国家评价机构和帮助提高国家需求数量的工作。东亚和太平洋区域办事处以及南亚区域办事处通过联合国亚洲和太平洋评价发展小组尤其积极推动该工作。同时，联合国评价小组与联合国亚洲和太平洋评价发展小组商定，在 EvalPartners 举措内制订针对区域的电子学习方案。

72. 现在已经累积了以大学为基地的国家英才中心的必要数量的经验。以前的报告提到拉丁美洲和加勒比以及南亚的工作。中东和北非的新活动在埃及(Assyout 和 Helwan 大学颁发关于评价的专业证书)、黎巴嫩(贝鲁特美国大学)和苏丹(喀土穆/的 Ahfad 女子大学)。在东部和南部非洲地区，布隆迪方案与国立大学缔结了谅解备忘录，以加强国家评价能力。

73. 在自发性南南协作的范围内，儿基会帮助摩洛哥人类发展国家办公室与墨西哥评价社会发展政策国家委员会联系，制订评价公共政策的系统。儿基会还找到南方伙伴来支持柬埔寨、北朝鲜和缅甸的需求。乌克兰和津巴布韦执行以本国为重点的工作，评估其新的国家评价系统，拟订国家评价能力建设战略。这些国家都可以从上文提到的联合国评价小组工作队的文件《如何加强国家评价系统的切实可行的建议》得到帮助，该文件已经译成多种语文广泛分发。

74. 鉴于该领域的活动范围日益扩大，有必要就评估国家评价系统的优点和缺点的工具和方法，以及就国家一级工作的日常监测和分享，达成协议。这将有助于加强评价工作的这一重要领域。

¹⁹ 见 <http://mymande.org/elearning>。

七. 结论：儿基会评价的挑战和展望

75. 本报告扼要介绍了儿基会评价职能如何在儿基会的工作中继续发挥重要作用，以及如何通过联合评价活动在联合国系统的工作中发挥作用。报告叙述评价职能的绩效如何通过加强管理和支持机制得到改进，但必须更加注意改进评价范围和质量。在全球一级，儿基会及其众多伙伴已经为评价能力建设做出广受欢迎的贡献；不过，许多国家还要多多努力，使有效的国家评价系统制度化。

76. 评价职能十分适合于支持儿基会内新的战略计划的执行，衡量和评估新战略的进展情况，并通报新的政策和方案。随着儿基会面对不断变动的世界继续扩展维护儿童的工作，一个更新的评价政策、一个基于战略计划的评价以及一个加强评价职能的执行战略将是整个儿基会评价活动的基础。

八. 决定草案

77. 执行局，

1. 注意到关于儿基会评价职能和主要评价工作的年度报告(E/ICEF/2013/13)；
2. 重申儿基会评价职能发挥的中心作用及儿基会评价政策(E/ICEF/2008/4)所列各项原则的重要性；
3. 欣见报告中所述的继续在权力下放层次加强评价职能的证据，并鼓励儿基会继续采取行动；
4. 注意到跟踪评价制度有效性的关键绩效指标以及关于人力资源和财政资源的行政数据；
5. 注意到为增强国家伙伴的技能、能力和制度以及加强南南学习而做出的积极努力；
6. 请儿基会：
 - (a) 报告为确保在拟订关键政策、战略和方案时系统地考虑和利用相关评价成果而采取的步骤。

附件

全球评价：2012-2013 年综合监测和评价框架的执行进展情况

(截至 2013 年 3 月 31 日的现况)

全球评价标题	现况	评论
1. 多指标类集调查	执行	预计 4/13 完成
2. 重新重视公正的可评价性和评价程序	执行	预计 9/13 完成
3. 群组牵头机构在人道主义行动的作用	执行	预计 6/13 完成
4. 紧急情况下的儿童保护	执行	预计 6/13 完成
5. 社区牵头的全面卫生办法	执行	预计 7/13 完成
6. 应急准备系统	执行	预计 9/13 完成
7. 教育部门的上游工作	执行	征求建议于 1/13 发出。预计 9/13 完成
8. 暴力侵害儿童	尚待执行	儿童保护部门供资。预计 6/13 调动
9. 防止发育迟缓	延迟	性质复杂。需要在 2013 年界定范围。2014 年开始
10. 社会保护	尚待执行	预计 5/13 调动
11. 宣传和政策改变	尚待执行	预计 5/13 调动
12. 使用基于成果的管理	替代；尚待执行	战略计划的可评价性似乎更为关键。5/13 开始
13. 信息流通促进发展	尚待执行	预计 7/13 调动
14. 国家能力建设	延迟	性质复杂。需要在 2013 年界定范围。2014 年开始
调动：制定职权范围或拟订中；调集资源		
执行：工作正在进行或完成；分析在进行中		
完成：提交最后报告；正在分发		
尚待执行：时间表依旧；未采取行动		
替代：原来的重点改变到更加优先事项		
延迟：提议到 2014 年开始		