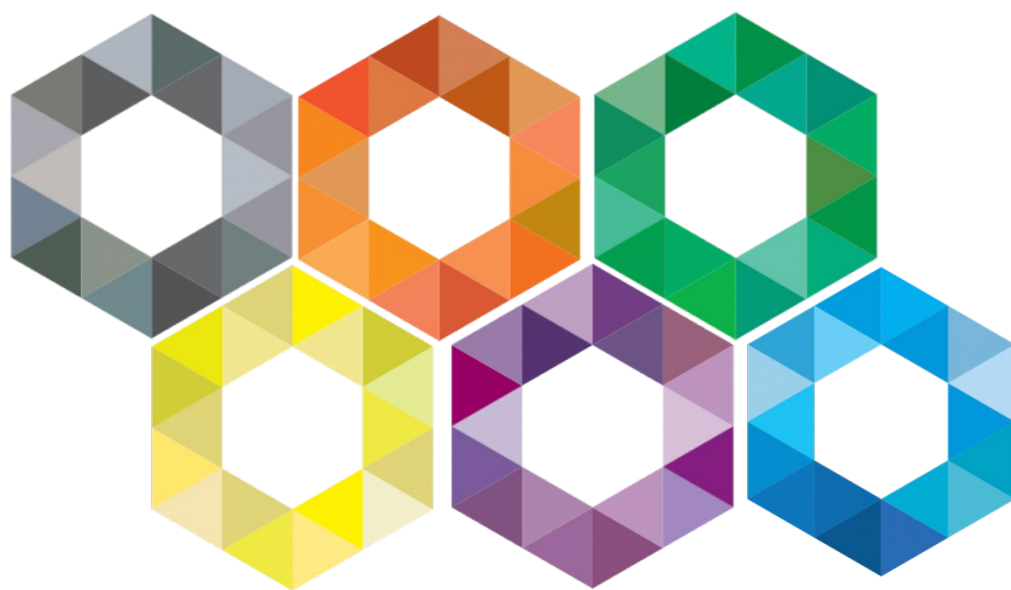


采用更好的准则实现更优质的评估

评估准则的概念界定和应用原则（修订版）



经济合作与发展组织发展援助委员会（OECD/DAC）发展评估网络（EvalNet）

致谢

《采用更好的准则实现更优质的评估》(Better Criteria for Better Evaluation) 中文版由联合国儿童基金会驻华办事处 (UNICEF China)、亚太财经与发展学院 (AFDI) 和经济合作与发展组织 (经合组织, OECD) 共同编写。

联合国儿童基金会驻华办事处, 亚太财经与发展学院的刘晓强、王蕾和刘曜桀, 以及北京师范大学做出了重要贡献并提出了深思熟虑的意见, 使本报告的编写工作受益匪浅。

在 Megan Kennedy-Chouane 的指导和 Ola Kasneci 和 Evelyn Garland 和 Stephanie Coic 的支持下, Anita King 代表经合组织负责管理本报告的编写工作。

Acknowledgements

This Chinese language adaptation of Better Criteria for Better Evaluation has been produced through a partnership between UNICEF China, the Asian-Pacific Finance and Development Institute (AFDI), and the Organisation for Economic Co-operation and Development (OECD).

The work greatly benefited from the contributions and thoughtful engagement of UNICEF China, Scott Liu, Wang Lei and George Liu of AFDI and Beijing Normal University.

The process has been managed on behalf of the OECD by Anita King with support from Ola Kasneci, Evelyn Garland and Stephanie Coic, under the supervision of Megan Kennedy-Chouane.

1. 背景

1.1. 修订原因

1. 经济合作与发展组织（OECD）发展援助委员会（DAC）在 1991 年 OECD DAC《发展援助评估¹原则》（*Principles for Evaluation of Development Assistance*）中首次提出了评估准则：相关性（relevance）、有效性（effectiveness）、效率（efficiency）、影响（impact）和可持续性（sustainability），其后在 2002 年版《评估和基于结果的管理关键术语表》（*2002 Glossary of Key Terms in Evaluation and Results Based Management*）中对这些准则做了定义。这五条准则已经逐步成为评估国际发展与人道主义项目、方案和政策的核心参考。除了发展合作外，评估者和评估委托方也将这些准则应用于公共政策的其他领域。

2. 基于 25 年间应用这些准则所取得的经验，并遵循 2015 年达成的《2030 年可持续发展议程》，包括《可持续发展目标》《2030 年议程》和《联合国气候变化框架公约》下的《巴黎协定》，全球评估界开始讨论如何重新审视这些评估准则。在 2017 年 10 月 31 日的高级别公报中，OECD DAC 决定，“依照《2030 年议程》，探索如何修订五条关键评估准则并应用到项目评估中”。作为 DAC 附属的负责支持评估工作，并为政策制定和学习提供循证依据的机构，EvalNet 于 2018–2019 年期间开展了此次准则修订工作。

3. 修订工作的目标是在总结经验教训的基础上对评估准则进行完善，以期用更好的准则来支持更优质的评估。更优质的评估也将有助于制定更好的政策，推进《2030 年议程》，协助各国达成《巴黎协定》和其他目标。

1.2. 针对准则的咨商：优势与提升空间

4. 修订过程中，我们在 2018 年 5 月至 11 月期间开展了广泛的咨商，包括公开问卷调查、访谈、在各种国际会议上的讨论以及文献回顾等。在最初草案的基础上，咨商活动一直持续到 2019 年底。咨商意见小结可从网上查询²。

5. 咨商结果显示，评估准则得到了广泛的支持和应用。受访者明确表示，评估准则已被普遍接受而且实用性强，因此倾向于保留原有准则的主体部分。许多受访者强调了该准则在促进评估专业和评估实践标准化和一致性方面的价值。咨商还表明，有必要继续保持准则的简洁性，这可以通过只保留数量有限的一套准则、并保持定义的一致性来达成。

1. 对于 evaluation 一词的中文翻译，在当前中国公共财政支出和国际金融组织贷款项目绩效管理等领域通常译为“评价”。

2. 参见 DAC 评估网络（EvalNet）（2018 年），OECD/DAC 评估准则：咨商意见小结（2018 年 11 月），参见 oe.cd/criteria

6. 同时，有些人提出了希望澄清某些概念的要求。许多人指出，这些准则在实践中的应用方式面临挑战，尤其突出的问题是在实际应用中倾向于选用过多的评估准则和评估问题。尽管《发展评估质量标准》已经明确指出，不强制评估者使用所有的准则，评估者也可以使用其他准则，然而在实践中，可能结果是在没有充分考虑具体评估的背景和意图的情况下，机械地应用这些准则。也有人担心原来的一套准则没有充分涵盖《2030年议程》的相关表述和当前的政策重点。还有一些人则认为，这些准则过于聚焦在项目层面，没有充分关注复杂性和权衡取舍（trade-offs）、公平以及人权和性别平等等问题。许多人要求加强对评估准则执行情况的指导，进而改善准则使用情况，提高评估质量。

7. 除评估准则及其定义方面的问题之外，咨商也反映了在评估实践和评估体系中的一些问题。EvalNet 致力于与全球评估领域的合作伙伴一起解决这些问题，推动评估事业更好地发展，并且正在探索开展更多的相关工作。

1.3. 修订版主要特点

8. 为了回应《经合组织发展援助委员会高级别公报》的要求以及咨商提出的意见和建议，EvalNet 修订了评估准则的定义和使用原则。EvalNet 的成员和合作伙伴对两轮草案发表了意见，随后举行了一系列在线研讨活动，就这些定义进行了深入的交流和互动。我们也邀请了国际评估专家对草案提出反馈意见。

9. 汲取所有的意见和建议之后，修订后的准则具备以下特点：

- 完善并更新了之前的五条准则：为了保持和加强概念的清晰性这一关键优势，我们对定义进行了细化，并使用注解来解释相关概念，同时使文本尽可能保持简洁明了。
- 增加了一条新的重要准则，即“一致性”，以更好地体现关联性、系统思考、伙伴关系动态和复杂性。
- 为了更好地支持准则应用并澄清相关疑惑，增加了以下内容：关于准则意图的介绍；应用指导原则；以及一份随附的指南来进一步解释每个准则的维度，和如何在不同的评估中应用这些准则（即将发布）。
- 确保准则适用于不同的干预措施，而不仅仅限于项目层面：考虑到发展、人道主义活动和机制等评估对象的多样性，以及在发展合作之外的领域对该准则的应用，修订版中使用了“干预”一词（而不是像以往一样，使用外部资金支持、方案或项目等词语进行表述）。修订版中“捐资方”的提法也已删除。

- 更好地响应公平（equity）、性别平等（gender equality）、“不让任何人掉队”（leave no-one behind）议程等当前政策重点：尤其是相关性和有效性的定义鼓励对公平问题进行更深入的分析。这些准则有助于评估国家、地方或国际社会为实现《2030年议程》所设定的可持续发展目标和《巴黎协定》所做出的努力。同时，修订后的准则具有足够的高度，以保证其能被广泛应用，并随着政策重点的转移和目标的变化仍能具有参考价值。
- 定义和应用指南反映了可持续发展的整体性特征，也提升了各个准则之间的内在联系，包括对协同增效和权衡取舍的考量。



2. 修订后的评估准则

2.1. 评估准则的目的

10. 评估准则的目的与评估的目的是相互联系的，即能够确定干预措施的优点、价值或者重要性³。本文通篇采用“干预”一词来指代评估对象（见专栏 1）。每一条准则都从不同角度来审视评估对象，从而共同形成对干预措施、实施过程和结果的全面认识。

11. 这些准则起着规范作用。整套准则描述了干预措施应具备的特征：所有干预措施都应该与具体的背景相关，与其他干预措施协调一致，达成目标，有效率地实现结果，并产生持久的积极影响。

12. 在评估和评价中应用这些准则，以实现：

- 支持问责，包括向公众提供信息；
- 通过形成和反馈评估发现和经验教训来促进学习。

13. 这些准则也可以应用于评估之外的领域，包括监测和结果管理、战略规划以及干预措施的设计。

14. 这些准则既可以用于审视过程（变化是如何发生的），也可以用于审视结果（发生了哪些变化）。所有准则都适用于干预措施实施前、实施中和实施后的评估⁴。

专栏 1. 评估干预措施

本文通篇采用“干预”一词来指代评估对象。干预包括可以应用这些准则进行评估的各类发展和人道主义活动，例如项目、方案、政策、战略、专题领域、技术援助、政策建议、制度、筹资机制、举措或其他活动。它包括发展干预、人道主义援助、维和、减缓和适应气候变化、规范性工作和非主权业务。

这些准则既可用于国际合作活动和国际金融组织贷款项目，也可用于私营部门、非政府组织、国家或地方政府在国内政策背景下开展的干预措施，以及公共财政支出。

3. 参见经合组织发展援助委员会《发展援助评估原则》（1991年）中关于价值和重要性的表述。

4. 但是，本文中对准则的定义方式反映了当前事中评估、终期评估和事后评估的主流做法。

2.2. 应用原则

15. 准则的应用需遵循以下原则。随附的指南(即将发布)中提供了更多建议和案例。此外,经合组织发展援助委员会在《发展评估质量标准》中也描述了规划和实施一项评估活动应该达到的质量标准。重要的是要在一个更广泛的背景下来理解这些准则,并且要与其他关于如何高质量地开展有用的评估的原则和指南结合起来解读。

原则一

审慎应用这些准则来支持高质量和有用的评估。应用评估准则应考虑不同情境,结合个案的评估背景、被评估的干预措施,以及涉及的利益相关者的具体情况来理解。应该根据评估问题(你试图发现什么)和你想要怎么使用问题的答案来对这些准则进行解读和分析。

原则二

准则的应用取决于评估的目的。不应该机械照搬准则,而应该根据利益相关者的需求和评估背景来决定是否选用准则。针对每条准则进行评估分析时投入的时间和资源是多还是少,是由评估的目的决定的。数据可得性、资源限制、时间和方法方面的考虑因素也可能会对如何(以及是否)选用某条特定准则产生影响⁵。

2.3. 定义和修订说明

16. 以下部分对各准则进行了逐条定义,并使用了注解来阐明定义中的相关概念。《评估和基于结果的管理关键术语表》(OECD, 2002)中载有原始定义,解释性的专栏中描述了修订版中的改动。《术语表》第二版正在制作中,它定义了很多本文中使用的术语,包括干预(intervention)、结果(results)、产出(output)、成果(outcome)和目标(objective),可以为读者提供有用的参考。

相关性: 干预措施是否在做正确的事?

在多大程度上干预目标和干预设计回应了受益人⁶、全球和国家层面以及合作伙伴/机构的需求、政策和重点领域,以及在情况发生变化时持续回应的程度。

注解:“回应”指干预措施的目标和设计对所处的经济、环境、公平、社会、政治经济以及能力等背景要素具有敏感性。“合作伙伴/机构”包括为干预提供资助、参与实施或进行监督的(国家、地区、当地)政府、民间社会组织、私营实体和国际机构。相关性评价还涉

5. (在评估开始前开展的)对可评估性的评价有助于为评估设定一个合理预期,包括评估可以提供哪些信息、可以收集到哪些证据,以及评估如何回答相关问题。

6. 受益人被定义为“从发展干预中直接或间接受益的个人、群体或组织,而无论其是否是干预所针对的”。替代用词包括权利持有人、受影响的人群等。

及审视不同重点领域或需求之间的差异和权衡取舍。它需要对干预背景的变化进行分析，以评价在何种程度上可以（或已经）通过调整干预措施以保持其相关性。

专栏 2. “相关性”定义修订说明

我们在定义中增加了“设计”二字，以涵盖与相关性有关的其他质量因素，即干预措施是否经过精心设计，以回应相关的重点领域或需求（即其目标、内含的变化理论（theory of change）、行动理论、操作方式、风险分析、背景分析等）。这旨在深化相关性分析，同时用有效性来评价目标的实现程度，用效率来评估干预措施的实施状况。

注解中描述了干预措施应对哪些背景要素具有敏感性，以促进对背景条件的深入分析。

我们增加了与受益人有关的“重点领域”，以阐明由受影响的人（不仅仅是捐资方和政府）来决定最重要/紧急事项的重要性。聚焦受益人也反映了当前的政策重点：“不让任何人掉队”。我们使用了“需求”（needs）一词，比“要求”（requirements）更容易理解（“要求”有时会与法律要求混淆）。

在咨商过程中，有人提出“受益人”（beneficiaries）的表述欠妥，因为“受益人”可能意味着一种被动的接受者角色，或者是因为它假定人们正在受益而真实情况也许并非如此。我们在此保留该表述是因为它在评估相关性上有具体含义：干预措施应针对特定人群，且对于这些人群应该是相关的。为避免混淆，我们在脚注中给出了该术语的定义并提供了替代用词。

我们保留了“全球重点领域”（global priorities）的表述，如今该表述已在《2030 年议程》和《巴黎协定》中得到体现。我们并没有在定义中明确提及这些细节，因为政策重点领域是会改变的，并且相关性准则可以被应用于不同的评估对象。

我们在定义中明确指出了时间维度（之前版本在注解中提及了时间维度），因为相关性在干预设计之时和之后都可以被评估。包含时间维度很有用，因为适应并应对复杂或变化中的环境对于有效性也很重要。而之前版本的相关性通常用来对政策重点领域做一次性的评价。

为了更好地反映准则在发展合作领域之外的应用，我们删除了“捐资方”（donors）一词，并增加了注解，以解释合作伙伴/机构（partner/institution）包括资助、实施、监督等方面的合作方（可能是地方、多边、私营等）。

定义中没有包含参与（participation）和所有权（ownership）的概念，因为这些是影响相关性（以及有效性和可持续性）的因素，而不是准则本身的维度。指南中会有对这些概念的讨论。

《术语表》中对相关性（relevance）的定义：“发展干预的目标与受益人的要求、国家需求、全球重点领域以及合作伙伴和捐资方的政策相符的程度。注解：从以往的经验来看，相关性通常会变成干预的目标或者其设计是否仍适应变化之后的环境的问题。”

一致性：是否与其他干预措施协调相容？

与国家、部门或机构中其他干预措施的兼容性。

注解：其他干预措施（尤指政策）对该干预措施支持或削弱的程度，反之亦然。包括内部一致性和外部一致性：**内部一致性**指该干预措施与同一机构/政府实施的其他干预措施之间的协同增效和相互联系，以及与该机构/政府所遵循的相关国际规范和标准的一致性。**外部一致性**指该干预措施与其他行动方在同一背景下所实施的干预措施的一致性，包括互补性、协调性和合作性，以及该干预措施在多大程度上带来增加值并同时避免重复劳动。

专栏 3. 对增补“一致性”及其定义的说明

这条准则体现了一个之前没有涉及的视角。缺乏一致性会导致重复劳动而且破坏整体进展。增加这条准则将有助于加强对这些重要问题的分析。增加一致性准则也有助于评估者理解某项干预措施在特定系统（组织、部门、专题领域、国家）中的作用，而不是仅仅局限于以干预措施或机构为中心的视角。

在当前的背景下，随着不同政策领域间的协同增效（或权衡取舍）以及跨政府的协调越来越受到关注，尤其是在冲突和人道主义回应以及应对气候紧急状况的情况下，一致性需要引起更多的重视。此外，可持续发展的资金来源（国际和国内）也正在变得多样化。

“国际规范和标准”的提法有助于对干预措施是否与其行为主体本身基于国际法或者国际协定（例如反腐败法规或者人权公约）所做出的承诺相一致进行分析。这适用于该实体已经承诺的协议，因此属于内部一致性。在此之前，这种一致性通常得不到足够的分析。从回应全球重点领域的角度，也可以对干预措施与国际规范和标准的相关性进行评价，与一致性分析互补。

有效性：干预措施是否能实现其目标？

干预措施的目标和结果在多大程度上已经实现或预计会得以实现（包括不同群体之间的差异化结果）。

注解：有效性分析涉及考虑目标或结果的相对重要性。

专栏 4. “有效性”定义修订说明

需要明确的是，有效性应该分析结果链/因果路径中实现目标的进展。影响（**impact**）关注的是干预可能带来的更高层级的效果和更广泛的变化，而有效性（**effectiveness**）不同，它更关注可归因于干预的结果。

我们增加了“结果”（**results**）和“差异化结果”（**differential results**）的表述，以便提出结果在不同群体间分布的重要问题，并且有助于关注预期之外的目标。这符合“不让任何人掉队”的政策重点。它鼓励评估者审视边缘群体的公平问题和结果，而不是直接假定公平是干预的目标。

这样的措辞赋予评估以灵活性，将重点放在最值得关注的目标和/或结果上。

我们在注解中保留了“相对重要性”的提法，因为在得出总体有效性结论时，评估者应该权衡已实现/未实现/预计实现的目标和结果的重要性。另外，我们用相关性准则来评价干预措施对主要利益相关者的重要程度。而干预效果的整体意义是通过下述的影响这一准则来评估的。

《术语表》中对有效性（**effectiveness**）的定义：“在考虑到不同目标的相对重要性的情况下，发展干预措施的各项目标在多大程度上已经实现或预计得以实现。注解：也可用于综合衡量（或判断）一项活动的优点或价值，即干预措施在多大程度上高效、持续地实现了或预计将实现其主要相关目标，并对制度建设产生积极的影响。”

效率：资源是否得到充分利用？

干预在多大程度上能够或可能以经济且及时的方式实现结果。

注解：“经济”是指与该背景下可行的替代方案相比，以尽可能最具成本效果的方式将投入（资金、专业知识、自然资源、时间等）转化为产出、成果和影响。“及时”是指在预期时间范围内，或者根据不断变化的环境的要求在合理调整的时间范围内交付结果。这可能包括对运行效率（干预的管理水平）进行评价。

专栏 5. “效率”定义修订说明

好的评估实践表明，我们在对效率进行分析时可以关注与整个结果链（产出、成果和影响）相关的投入。人们认为，分析整个结果链，尤其是分析投入产生影响的效率，在方法上困难重重。基准数据通常缺失。但这是一个值得关注的领域，所以我们保留了与影响有关的表述。这将有助于使用者可以灵活地将效率分析集中在与他们的评估最相关的那部分结果链上。

增加“与该背景下可行的替代方案相比”的表述旨在强化这样一个观点，即有效的成本效益分析需要将干预的价值与相关的反事实进行比较：如果无法产生足够的效益，成本最低的干预方案不一定是最好的选择。

我们增加了及时性维度，因为衡量干预价值的这一重要因素有时会被忽略。

我们明确提到运行效率，因为评估的利益相关者高度重视实施过程和管理问题。例如，这有助于检验协调机制在提升了一致性的同时是否也增加了交易成本。

《术语表》中对效率（**efficiency**）的定义：“对资源/投入（资金、专业知识、时间等）如何经济地转化为结果的衡量。”

影响：干预带来哪些改变？

干预在多大程度上已经产生或预计将产生显著的正面或负面、预期或非预期的更高层级的效果。

注解：影响表述的是干预措施的最终意义和潜在的变革性效果。它力图明确干预措施在社会、环境和经济方面产生的比有效性范畴更长期或者范围更广的效果。除了直接结果之外，这一准则还力图体现干预措施带来的间接、次要和潜在的结果。它通过分析制度或规范整体而持久的变化，以及人民福祉、人权、性别平等和环境受到的潜在影响，来做到这一点。

专栏 6. “影响”定义修订说明

我们引入“更高层级”是为了体现效果的重要性、范围和变革性特征。注解解释了其含义。在《2030年议程》和《巴黎协定》呼吁变革的时代背景下，这一含义具有政策相关性。

现在的定义更符合“影响”一词的通常用法，指有意义或重要的变化。

我们认识到，影响一词有时会和被某些使用者采用的有效归因的概念混淆（在影响评估（*impact evaluation*）中经常使用），但是我们认为有效归因是内嵌在所有评估准则中的。

我们删除了“直接和间接”以及“主要和次要”，因为直接和主要的效果已涵盖在有效性中。该注解有助于进一步澄清有效性和影响之间的区别。

《术语表》中对影响（*impact*）的定义：“由发展干预直接或间接、预期或非预期地产生的正面和负面、主要和次要的长期效果。”

可持续性：是否能持久获益？

干预的净效益在多大程度上能够持续或可能持续下去。

注解：包括对相关体系为了长期维持净效益所需财务、经济、社会、环境和制度能力的考查。涉及对于韧性、风险和对潜在的正面或者负面结果的权衡取舍的分析。根据评估的时间点，这可能涉及对实际净效益流的分析，或对净效益在中长期持续的可能性的估计。

专栏 7. “可持续性”定义修订说明

该定义简明扼要，且注解也明确了可持续性具备多个维度（财务、经济、社会和环境等）。

原始定义过于以捐资方为中心，且仅关注外部资助（主要的发展援助），没有充分反映如今发展评估的背景。此外，即使在资助或活动仍在进行的情况下，评估可持续性也可能是有用的。取消外部资助的提法将有助于在各种情况下对可持续性进行评估。

该注解还鼓励对潜在的正面或者负面结果的权衡取舍，以及支撑持续受益的能力/体系方面的韧性进行分析。相比之下，之前的定义包含了重复的措辞（“持续长期受益”和“抗风险能力”）。

考虑到所有与干预相关的持续成本，我们保留了“净”效益一词，以关注干预总体价值的延续。

《术语表》中对可持续性（sustainability）的定义：“在主要的发展援助结束后，干预措施的效益的持续性。包括持续长期受益的可能性，以及净效益的抗风险能力能否经受时间考验。”

摘要

本文介绍了经济合作与发展组织发展援助委员会发展评估网络（OECD DAC Network on Development Evaluation，下文简称为 EvalNet）于 2018–2019 年间是如何对 OECD DAC 评估准则的定义和应用进行重新审视和修订的。文中对原有五条评估准则即相关性（relevance）、有效性（effectiveness）、效率（efficiency）、影响（impact）和可持续性（sustainability）的定义进行了修订，增加并定义了一条新准则：一致性（coherence）。本文还描述了应如何审慎地应用这些准则，以及如何结合具体的干预背景和使用者需求进行调整。

此次评估准则的定义和应用原则的修订是在对原版准则进行了全球范围的咨商，并对这些准则在评估以及评估以外的领域的应用情况进行了总结回顾的结果。咨商后 EvalNet 的成员和外部评估专家深入讨论了这些概念并审阅了数轮草案。修订版定义更加清晰，有助于对公平议题与协同增效（synergies）等符合当前政策重点的问题进行更严谨和细致的分析。为了厘清相关问题，修订版增加了对各条准则预期目的和应用指导原则的介绍。关于准则应用的详细指南将在准则被采纳后另行提供。

本文件于 2019 年 11 月 20 日获得 EvalNet 的批准，并在 2019 年 12 月 10 日的发展援助委员会（DAC）会议上被采纳并公开。

Translation of the document, “Better Criteria for Better Evaluation – revised definitions and principles for use” (OECD/DAC Network on Development Evaluation, 2019).

www.oecd.org/dac/evaluation

中文版由联合国儿童基金会驻华办事处（UNICEF China）于 2020 年组织翻译，亚太与财经发展学院（AFDI）校译。



携手为儿童