

「評価基準を熟慮しつつ適用するために」 (DAC評価基準に関する指針)



目次

要旨	1
1 ガイダンスの使用と目的：よりよい評価のためのよりよい基準	3
なぜガイダンスが必要なのか	4
ガイダンスの目的と使用方法	4
誰がこのガイダンスを活用できるか？	6
参考文献	7
注釈	7
2 評価基準：その目的と評価における役割	8
評価基準とは何か？	9
評価基準で、どのような種類の介入を評価できるか？	10
評価領域全体の中で、評価基準はどこに位置づけられるか？	11
参考文献	12
注釈	12
3 評価基準の実践への適用	16
熟慮して評価基準を適用	17
評価の目的に応じた基準の適用	18
時間軸の理解	21
よりよい評価の設問に向けた基準の活用	21
評価基準間の関係性	22
使用する評価基準の選択	23
組織レベルで評価基準の運用と適用	24
アジェンダ 2030 と持続可能な開発目標への対応	25
評価基準へのジェンダーレンズ（ジェンダーの視点）の適用	27
他の評価基準の使用	29
参考文献	30
注釈	31
4 6つの評価基準の理解：定義、分析要素、主な課題	32
妥当性	34
整合性	44
有効性	52
効率性	60
インパクト	67
持続性	76
参考文献	85
注釈	88

要旨

経済協力開発機構(OECD: Organisation for Economic Co-operation and Development)は、一貫性があり質の高い評価の実施を支援するため、6つの評価基準(妥当性、整合性、有効性、効率性、インパクト、持続性)の共通定義を設定しました。これらの基準は、介入(政策、戦略、プログラム、プロジェクト又は活動)の長所や価値を確認するための規範的な枠組みを提示するとともに、評価結果を判断する際の根拠としての役割を果たします。

評価基準は、介入とその結果の全体像を捉えるために相互補完的な見方を可能にする一揃いのレンズ(視点)と考えられます。この基準によって、評価者及び介入の計画策定や運営に従事する人々には、介入の特性、実施過程及びその結果についてより深く考察するよう勧めます。

OECD-DAC 開発評価ネットワーク(DAC Network on Development Evaluation)によって、6つの評価基準の定義とともに、以下の2つの使用原則が示されました：

使用原則1：評価基準は、質が高く有益な評価のために熟慮しつつ使用すべきである。

使用原則2：評価基準は、評価の目的に沿って活用すべきである。

評価基準は、熟慮しつつ適用し、介入及び評価の文脈に応じて使い分けることが肝要です。評価基準は、本来は国際協力で使用するために策定されましたが、現在では、どのセクターにも、そして公的機関又は民間団体による介入の評価にも適用できます。介入の様々な論題や形態(主題別や戦略的課題、政策、プロジェクトを含む)に適用できます。評価基準はあくまで評価の規範と標準全体の一部であることを念頭に置き、評価基準に沿って評価を行う際に様々な手法を使用できます。

評価者は、グループによって便益に差異が生まれてしまう事例について検討し、ジェンダーやその他の差別(例：年齢、人種・民族性、社会的地位、能力)から生まれる力関係が、介入の実施や結果にどのような影響を与えているかを熟考しなくてはなりません。

各評価基準の主要概念についての定義は、以下のとおりです。

・妥当性は、介入の目的やデザインが、どの程度受益者のニーズや優先事項に対応しているか、そして国家、地球的規模、パートナー／組織の政策や優先事項に対応しているかを確認します。妥当性の理解では、ジェンダーに基づく力関係に配慮し、「誰一人取り残さない」とするコミットメントについて考慮することが重要です。状況が変わっても介入は妥当であり続けているかどうかについても、評価では考察すべきです。

・整合性は、新しく追加した評価基準です。他の介入（特に政策）が、どの程度当該介入を促進又は阻害するか、及びその逆の場合を考察します。整合性には、（一つの組織又は政府内の）内部的整合性（国際規範や標準への適合性も含む）と、外部的整合性（同じ文脈における他者の介入との整合性）があります。整合性には、相互補完性、調和化や協調及び介入が重複を避けながらどの程度付加価値を生み出しているか、という概念が含まれます。アジェンダ 2030 に沿って、異なる政策領域間の相乗効果（又はトレードオフ）への視点を重ねつつ、整合性に対してより多くの注意を払うべきです。この新評価基準は、複雑な介入やその結果を理解する際に統合的アプローチが必要であることを示唆しています。

・有効性は、介入がその目的や結果をどの程度達成したか、又は達成することが見込まれるかを、その目的の相対的重要性に配慮しつつ確認します。新たな定義では、様々なグループ間に生じた差異ある結果や、介入がどの程度格差の縮小又は拡大をもたらしたかを分析することが求められます。有効性は最もよく使用される評価基準で、介入が全体的にどれだけ成功したかを測るためによく用いられます。

・効率性は、評価者に、介入がどの程度経済的かつ適時な方法で結果を生むか、又は生むことが見込まれるかを考えるよう促します。「経済的」とは、同じ条件下で実施可能な代替手段との比較で、インプット（資金、専門技術、天然資源、時間等）から結果への変換で費用対効果が最大限となる方法を指します。新しい定義では、「適時な実施」という側面を含めました。この評価基準は、実践的にも政治的にも重要性が高い「介入に投入された資源がその結果によって正当化できるか」について確認する機会を提供します。限られた資源を適切に使用して、より多くを達成することにつながることから、受益者を含めた多くのステークホルダーが効率性に注目します。

・インパクトは、介入により生じた、又は生じると見込まれる、重要な正又は負の、意図された、又は意図されない結果の連鎖で上位にある効果が生み出される度合いです。インパクトでは、介入の最終的な重要性と、潜在的に社会変容を伴う効果（社会組織や規範における全体的・永続的な変化）を確認します。インパクトは、有効性より上位の質問「結果として、介入で何が変わったのか（so what?）」への回答を用意します。この評価基準により、評価者は、介入が人々に本当に意義のある変化をもたらしたかどうかを確認します。

・持続性は、介入の純便益が継続する、又は継続する可能性の度合いです。評価の時点によって、純便益の実績フローの分析や、中長期にわたる純便益予測等を行います。便益の継続を確認するという基本的考えに変わりはありませんが、新しい定義では、分析のために財政・経済・社会・環境等各側面に関する要素を追加するとともに、これらの要素間の相互作用にも注目することが求められます。

各評価基準を適用する際には、様々な課題に直面するかもしれません。例えば、介入デザインやデータの入手可能性に関する問題点などが考えられます。本書では、評価者や評価管理者がどうすればこれらの課題を乗り越えられるかについて実践的な提言を示します。それは、評価可能性を評価工程の早い段階で確認する、意思決定を文書化する、又はベースラインを再設定するようステークホルダーと協力する、様々な期待に対処するために評価における限界を明らかにする、といった提言が挙げられます。このガイダンス内の表及び記述は、今後の経験から得られる教訓に基づいて適宜改訂されます。

1 ガイダンスの目的と使用： よりよい評価のためのよりよい 基準

本章では、ガイダンスの目的について説明し、読者が、改訂された評価基準の定義を理解し、業務における利用で、このガイダンスがどのように助力できるかを説明します。また、ガイダンスのデザインや内容づくりに貢献した全世界のステークホルダーが果たした役割を含め、ガイダンスの作成工程を説明した上で、評価の実施及びデザインを改善するためにこの評価基準をどのように熟慮して適用するべきか示します。

なぜガイダンスが必要なのか？

OECD-DAC は、1991 年に評価 5 項目を初めて示し (OECD, 1991^[1])、2002 年にそれらの定義を提示しました (OECD, 2002^[2])。評価 5 項目 (妥当性、有効性、効率性、インパクト、持続性) は、開発協力及び同分野を超えて評価者が参照する共通文書となっています。

持続可能な開発のための 2030 アジェンダに対応したいという熱意に促され、評価基準が最初に定義された以降に蓄積された長年の経験及び教訓を活用できる機会をとらえ、2017～2019 年間に広範な協議と改訂の過程を経て、2019 年 12 月には DAC により改訂版評価基準の定義が採択されました (OECD, 2019^[3])¹。この協議過程で評価基準の利用者たちからは、評価基準が多く面で有益であり、今後も介入の評価のために引き続き使用したいという指摘がありました。協議から得られた反応から、評価基準が広く使用、理解されて、一貫性と一定程度の比較可能性が確保されていることが分かりました²。評価基準により、平易かつ中立的な共通言語で、適応性があり、結果重視型で、説明責任を果たしつつ、学びに必要な主要分野を網羅する評価の規範的枠組みが設定されました。評価基準の改訂版は今後も継続の同意を得ており、結果的には従来の定義で上手く対応した内容を反映しています。

本ガイダンスは、「OECD-DAC 開発援助評価の基本原則」(OECD, 1991^[1])及び「OECD-DAC 開発評価の品質標準」(OECD, 2010^[4])と併せて読むことを奨励します。同基本原則は評価システムの管理と制度設計に焦点を当てており、同品質標準は評価プロセスや成果物に向けた情報提供をしています。このガイダンスは、事例を提示し、評価を実施する際に直面しがちな実践的課題を取り上げることで、より詳細な情報を提供し、読者にとって評価基準の定義が活用しやすくなることを目指しています。ボックス 1.1 には、評価基準の活用と理解を更に促進するための今後の出版物を記載しています。

過去 25 年間、DAC が策定した基準をどのように使用すべきかについては、多くの学びが得られました。2017-2019 年の協議過程で特定された主な課題は、その実践面での活用でした。協議では、単に基準をどのように定義するかだけではなく、どのように使用するか的重要性が強調されました。例えば、評価基準は文脈を考慮せずに機械的に使用されてしまう場合があります。このため、このガイダンスは、より高品質で役立つ評価につなげるために、様々な利用者のグループが実践において熟慮しつつ適用できるように支援することを目指しています。

同様に、本ガイダンスには、読者が評価基準の定義を理解して活用するための共通言語や指針が含まれています。ここには、評価基準の活用を導くために練られた 2 つの使用原則も含まれています (OECD, 2019^[3])。このガイダンスは、2 つの使用原則を膨らませ、定義について説明することにより、利用者が業務において評価基準を解釈し適用できるようにしています。

ガイダンスの目的と使用方法

この文書は、利用者がどのように評価基準を解釈して、業務において熟慮して適用できるように簡潔なガイドを提供するものです。評価基準は主に評価において適用されますが、モニタリング、結果重視型管理、介入のデザイン策定、戦略的計画づくりにおいても広く

適用可能です。利用者が直面しうる課題について思案し、解決方法を模索する際に手助けすることを目的とします。読者が定義を紐解き、評価基準がどのように相互に関連しているか、適用に際してどのような課題があるか、それらをどのように解決したらよいかを理解できるよう手助けします。現実の状況において、評価基準をどのように考察して適用できるかを示すために、実践的な具体例も選定されています。

このガイダンスは、評価の手引き書やテンプレート、或いは点検一覧表として使用されることを意図していません。評価計画を策定する際に評価者が依拠する所属組織の政策や基準、ガイダンスを補完する情報源の一つとして役立ちます。

ボックス 1.1. 更なる作業と今後の OECD の評価関連出版物

今後、新定義を利用した経験を踏まえてこのガイダンスを改訂する予定に加えて、経済協力開発機構開発援助委員会（OECD-DAC）による評価ネットワーク（EvalNet）とそのパートナーは、ジェンダー平等の視点での基準の使用や、人道的状況での基準の適用等、特定のトピックについて更に深く踏み込んだガイダンスを策定予定です。EvalNet と協力し、「人道支援活動における説明責任とパフォーマンスのための積極的学習ネットワーク（ALNAP）」では、人道的評価及び評価基準の使用に関して広く利用されているガイダンスの改訂も行っています（Beck, 2016^[5]）。

評価基準の定義及び使用原則の翻訳

OECD は、学びと評価能力向上を支援するという任務のもと、主要なパートナーと協力し、評価基準の定義を OECD の公式言語である英語及びフランス語に加え、様々な言語に翻訳しているところです。スペイン語及び中国語への翻訳は完了しており、2021 年にはアラビア語、ダリ語、パシュトー語、ポルトガル語、ロシア語、タイ語、ウルドゥ語の新版が作成される予定です。

「評価及び結果重視型管理における基本用語集」第2版

旧 DAC 評価 5 項目の定義は、OECD の「評価及び結果重視型管理における基本用語集」において初めて出版されました（OECD, 2002^[2]）。第2版は現在策定中で、評価基準及びその他の用語の改訂を反映させています。近々出版予定の第2版は、このガイダンス中で使用される用語（結果、アウトカム、インパクト、変化の理論、受益者等）の簡潔な定義を記載することから、このガイダンスと合わせて活用すると有益な資料となります。

同様に、このガイダンスは、OECD-DAC 又は他の機関が打ち出した既存の倫理ガイドラインや基準に取って代わるものではありません（OECD, 1991^[1]；OECD, 2010^[4]）。このガイダンスは、国連やその他の国際機関で使用される国際基準、例えば国連評価グループ（United Nations Evaluation Group: UNEG）の規範や基準、評価協力グループ（Evaluation Co-operation Group: ECG）のグッドプラクティス基準、「人道支援活動における説明責任とパフォーマンスのための積極的学習ネットワーク（ALNAP）」のガイダンス等を補完するものです。他の有益な情報源としては、二国間機関、国連機関、その他多国間機関の評価部門のウェブサイト、「よりよい評価（Better Evaluation）」のウェブサイト、EvalPartners 及び特定分野に

焦点を当てる関連機関、例えばインパクト評価のための国際イニシアティブ(3ie)等のウェブサイトが挙げられます。

また、このガイダンスは、評価基準そのものと同様、速やかな内省や批判的考察を促すことを目的としており、機械的に使用されるものではありません。そのようなアプローチは、アジェンダ 2030 のベストプラクティスや原則と合致していません（第 3 章で詳述予定）。評価基準が手法ではないのと同様に、ガイダンスは、特定の手法的アプローチや技法を規定するものではありません。そうした手法について詳しく解説する他の資料源の手掛かりとなる情報は、この文書に含まれています。

このガイダンスは、一貫性のある明確な定義と用語を提示する DAC 評価基準(OECD, 2019^[3])の定義やOECD 品質標準(OECD, 2010^[4])といった包括的な公式文書と、様々な文脈に応じて決断が必要となる組織内で日々使用されている評価基準との間を橋渡しするものと位置づけることができます。

評価基準が実際の評価においてどのように適用されてきたかを示す事例を提供することに加えて、このガイダンスは、それぞれの定義に含まれる概念を説明する分析要素も説明しています。これらの分析要素は、最終的なものではなく、別々に、或いはサブ基準として使用することを意図しておらず、むしろ各評価基準の定義をよりよく理解するよう手助けするものです。分析要素は、利用者が、それぞれの評価において各評価基準に対する文脈に応じた最善の解釈を見つける際に役立ちます。評価デザイン策定の段階では、評価者、評価委託者、評価結果の利用予定者の間で対話することにより、各評価基準の異なる要素をどのように捉え、全体的な結論を引き出すのにどの要素が最も重要になるのかを考察できます。なぜなら、介入、文脈、評価の目的に応じて、どの要素が重要になるのかは異なるからです。

第二章は、評価基準の役割を紹介し、より広い規範や標準の中でどのように位置づけられるのかを説明しています。第三章は、評価基準を理解して熟慮しつつ使用するための指針を示し、評価者がそれぞれの政策的文脈において各評価基準を使用することができるよう手助けします。第四章では、各評価基準の定義について更に詳細に説明し、共通する課題や事例を提示しています。アジェンダ 2030 の「誰一人取り残さない」という政策的優先事項を反映し、第四章には、それぞれの評価基準で考察できるように包摂性に関する主要課題が概説されています。

多くの事例は、旧定義に基づいて行われた評価から抽出されています。新しい定義と使用原則はごく最近打ち出されたばかりであることから、本ガイダンスは、作業用ドラフトとして位置づけられます。このガイダンスは、初期の試行的実施段階を経て、数年後には追加的な学習事例や最新の使用事例を盛り込んで改訂される予定です。特に、EvalNet 事務局は、国レベル及び地域レベルの評価における使用について、EvalNet 参加国・機関以外からも広く事例を収集したいと考えています。

誰がこのガイダンスを活用できるか？

このガイダンスは、地域、国家、国際的という様々な文脈において使用可能です。評価基準は、意図的に広範に、様々な介入やセクターに適用できるように設計されています。評価基準は、伝統的な政府開発援助（ODA）以外にも適用できます。熟慮しながら適用すれば、評価基準は、民間セクター、混合資金（ブレンディッドファイナンス）、又は貿易等の類似する政策領域に関連する介入にも等しく使用できます。

本ガイダンスは、主に評価者、評価管理者、評価委託者が使用することを想定しています。評価基準を初めて使用する学生や若い専門家だけでなく、複雑な評価を実施する経験豊富な実践者が使用するのにも適しています。介入デザイン策定、戦略的計画づくりや結果管理の指針として使用するというのも興味深い使用方法です。よりよい評価は、モニタリング、評価、学習制度をプログラムや政策のサイクルに効果的に取り入れられるかどうかにもかかっています。このガイダンスは、評価者や実践上のパートナーが、介入のデザイン策定段階等において、当該投入上の成功とはどのようなものなのかについて想いや記述を共有することで、介入の評価可能性を向上させ、全体的によりよい結果を生み出すことに役立ちます。

参考文献

Beck, T. (2016), Evaluating Humanitarian Action using the OECD-DAC Criteria, ALNAP, <https://www.alnap.org/help-library/evaluating-humanitarian-action-using-the-oecd-dac-criteria> (accessed on 11 January 2021). [5]

OECD (2019), Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use, DAC Network on Development Evaluation, OECD Publishing, Paris, <https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf> (accessed on 11 January 2021). [3]

OECD (2010), Quality Standards for Development Evaluation, DAC Guidelines and Reference Series, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264083905-en>. [4]

OECD (2002), Evaluation and Aid Effectiveness No. 6 - Glossary of Key Terms in Evaluation and Results Based Management (in English, French and Spanish), OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264034921-en-fr>. [2]

OECD (1991), Principles for Evaluation of Development Assistance, Development Assistance Committee, OECD Publishing, Paris, <https://www.oecd.org/development/evaluation/2755284.pdf> (accessed on 11 January 2021). [1]

注釈

¹ DAC評価基準「よりよい評価のためのよりよい基準」（OECD, 2019^[3]）は、従来の定義の変更点及びその理由について詳細な説明を記載。

² 協議結果の概要は EvalNet ウェブサイトに掲載：[oe.cd/criteria](https://www.oecd.org/criteria)。

2 評価基準：その目的と評価における役割

本章では 6 つの評価基準の目的を説明し、評価における役割を究明しています。評価基準を相互補完的なレンズとしてどのように捉えることができるか、各評価基準により、介入とその結果に対して異なる視点をどのように提供できるかを解説します。本章では、異なる種類の介入の価値や長所を判断する基盤として、評価基準を適用することで評価できることについても考察しています。本章の最後には、6 つの評価基準が他の評価規範や標準とどのように関連付けられるかを究明しています。

評価基準とは何か？

評価基準とは、評価において判断を行う際の基盤として使用される標準または原則です。

各評価基準の総体的意味は、以下の質問に要約されています。どの質問も、考察を行う上で重要な要素を提示しています。

- 妥当性：介入¹は正しいことをしているか？
- 整合性：介入はよく適合しているか？
- 有効性：介入は目的を達成しているか？
- 効率性：資源がいかにうまく使われているか？
- インパクト：介入はどのような違いをもたらすか？
- 持続性：便益は継続するか？

評価基準の目的は、共通の枠組みの中で一貫性のある質の高い評価を支援することにあります。評価基準は、ある特定の介入を評価するための規範的な枠組みを提供します。評価基準は、評価以外の工程、特に将来の介入改善を目的としたモニタリングと結果管理のための枠組み及び指標の定義、資金提供の承認、戦略的計画策定や介入デザイン作成などの工程にも活用できます。あらゆる介入に同じように適用できる共通の定義を持つことには意義があります。評価基準は、開発領域全体で一貫する共通言語を提供することによって、標準化を実現するとともに、全ての介入を比較し、さらに全ての介入から学びを得ることを可能にします。

評価基準は、介入を理解し、分析する際に用いるレンズとみなすべきです。評価基準は相互補完的な視点を提供し、介入の全体像を捉えるのに役立ちます。評価基準は、介入の特性、その実施、工程及び結果に関して、より深い考察を促します。評価基準は、総体として介入に望まれる属性を記述し、明示的予測を立て、規範を定めます。その規範とは、介入は文脈に照らして妥当で、他者の介入と整合性があり、効率的に成果を達成し、持続可能な開発に継続的で正のインパクトをもたらすべきだということです。評価基準は、評価へのアプローチ、包括的で体系的なアプローチ、そして評価工程の開始段階から使用される共通言語を開発するために広く受け入れられた枠組みです。

6つの評価基準は相互に関係性があり、各基準が介入の理解に対し異なる独自の視点をもたらす相互補完的な一纏まりの基準とみなすことができます。各評価基準の定義は、（評価基準毎に設けた上記の質問に示される）それぞれ独自の概念を示しますが、これらの概念は互いに関連性を持っています。開発介入は概して多面的であるため、相互に関連づけられた評価基準を使うことで、評価者は介入の全体像を考察することができます。

この評価基準は方法論ではなく、介入が達成しようと試みている目標でもありません。この評価基準は介入を評価する際に適切な設問を提起するためのきっかけとなるものです。全ての介入（評価の対象又は *evaluand*²）はそれぞれ異なるものであり、だからこそ評価工程では柔軟性が求められ、評価基準は熟慮しつつ、評価の目的と利用者に適合させて使用されるべきです。評価の品質基準に従って、工程と実施に関する結論と提言は、適切で信頼性の高い立証に基づいて提示すべきです。立証の論理、信頼性、解釈は

明確であり、かつ変化の理論（セオリーオブチェンジ）又は介入の論理に沿ったものでなければなりません。

評価基準で、どのような種類の介入を評価できるか？

このガイダンスにおいて「介入（intervention）」という用語は、評価の論題又は対象を意味するために使用されています。この用語には、基準を用いて評価できるすべての種類の取り組みが含まれます。こうした取り組みは、持続可能な開発や人道的な目標の達成を目的とする国際的又は国内的な取り組みです。

介入とは、プロジェクト、プログラム、政策、戦略、主題別分野、技術協力、政策助言、組織、拠出手法、手段、その他の活動を意味します。介入には、開発介入、人道的援助、平和構築支援、規範的活動及びその他の国際協力活動に加えて、国内の政策文脈における民間セクター及び中央政府の活動も含まれます。評価基準は、例えば、公立学校における給食プログラムの効率性から、紛争地域で様々な活動団体が行う国際支援の整合性まで、幅広い論題を評価するために使用されてきました。政策又は戦略の評価にも使用されています（例：ボックス 2.1 を参照）。一連の介入を評価する際にも、この評価基準を使用しています。例えば、教育セクターを支援する様々なプロジェクトや、長年にわたり複数のパートナーが関与する単一の大規模な「介入」（例：一般財政支援）の評価などです。

本文書では、それが最もわかりやすいため「介入」という用語を使用していますが、評価者は、評価工程の早い段階で、評価の論題——分析の対象となる介入——及び評価の範囲を明確に定義するよう心がけるべきです。プロジェクト又は政策など、その介入に関連するより具体的な用語を使用した方が、さらに有益でパートナーに理解されやすいはずです。

ボックス 2.1. 介入に合わせた評価基準の適用：ドイツ連邦経済協力開発省「障がい者インクルージョンに向けたアクションプラン」の評価

ドイツ連邦経済協力開発省（BMZ）の「障がい者包摂行動計画」の評価は、評価対象となる介入（この場合は戦略的行動計画）に合わせて基準をどのように適用できるか、また基準が評価に関する設問とどのように関係しているかを示す興味深い例を提供してくれます。この評価は、ドイツの開発協力において包摂の体系的主流化を推進する上での行動計画の成功度を調査するために企画されました。評価は、行動計画から導き出された評価に関する以下の基本設問を用いて実施されました。各基本設問の下に、評価基準に従ってさらに詳細な設問が設定されました。

1. 障がい者包摂に関して、BMZ は省内でどの程度模範例を示しているか？
 - 妥当性：選定された活動領域と施策は、国連障害者権利条約（CRPD）の規定にどの程度対応しているか？
 - 有効性：BMZ は、包摂的な体制や取り組みの確立にどの程度成功しているか？
2. ドイツの開発協力パートナー国において、行動計画は障がい者包摂推進をどの程度促進しているか？
 - 妥当性：選定された活動領域と施策は、国連障害者権利条約（CRPD）の規定にどの程度対応しているか？

- 有効性：ドイツの開発協力パートナー国の施策に、障がい者はどの程度効果的に統合されているか？

3. BMZ は、開発協力における障がい者の権利を擁護するパートナーとして、国内、地域的及び国際的にどの程度行動しているか？

- 妥当性：選定された活動領域と施策は、国連障害者権利条約（CRPD）の規定にどの程度対応しているか？
- 有効性：BMZ は、包摂という取り組みに、二国間及び多国間の活動団体を引き込むことにどの程度成功しているか？

4. 「包摂に向けた行動計画」の策定と実施は、どのように管理されたか？

- 有効性：どの管理系統や体制（管理工程における多様なステークホルダーグループの役割を含む）が、行動計画の策定と実施を決定したのか？
- 効率性：施策は、どの程度計画通りに実施されたか？

5. インパクトの範囲と政策手段としての触発効果を考えると、「包摂に向けた行動計画」の便益をどのように点数化すべきか？

- インパクト：行動計画は、どの程度広範なインパクトを生み出したか？
- インパクト：行動計画は、全体としてどの程度の触発効果を生み出したか？

Source: Schwedersky, Ahrens and Steckhan (2017^[1]), [Evaluation of the BMZ Action Plan for the Inclusion of Persons with Disabilities](#)

評価領域全体の中で、評価基準はどこに位置づけられるか？

すべての介入（評価される実体）は異なるため、評価工程には柔軟性が求められ、評価基準は熟慮して使用され、目的と利用者に応じて適用されるべきです。従って、評価基準の目的は、強制的な規則を提供することではありません。

評価基準は、単独で存在するものではありません。評価基準を最も有効に活用するには、他の規範や標準、手法、組織ごとの指針などとの関係から、評価基準をどのように位置づけるのが適しているかを理解することが重要です。

- マクロレベルには、開発評価の品質標準（Quality Standards for Development Evaluation）（OECD, 2010^[3]）に沿って公平性・独立性、信頼性、有用性及び参加性（OECD, 1991^[2]）などの評価の中核となる原則が位置付けられます。研究及びデータ収集の倫理基準もここに含まれ、すべての評価が全工程を通じて遵守すべきグッドプラクティスの包括的標準を提示します。
- 評価基準はメゾ（中層）レベルに位置し、介入を分析し理解するためのレンズを提供します。
- 組織レベルでは、各組織が独自の権限や優先度を反映させて、評価基準を自らのガイドラインや政策に合わせて解釈します（例：特定の評価基準を強調する、点数化を義務づける、追加基準を設ける）。
- ミクロレベル、すなわち個々の評価の文脈では、それぞれの評価で使用される評価設問及び手法の決定に加えて、その文脈や特定の評価対象に対して評価基準をどのように適用するかを検討します。

図 2.1. 他の規範及び標準を踏まえた評価基準の位置づけ



参考文献

OECD (2010), Quality Standards for Development Evaluation, DAC Guidelines and Reference Series, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264083905-en>. [3]

OECD (1991), Principles for Evaluation of Development Assistance, Development Assistance Committee, OECD Publishing, Paris, <https://www.oecd.org/development/evaluation/2755284.pdf> (accessed on 11 January 2021). [2]

Schwedersky, T., L. Ahrens and H. Steckhan (2017), Evaluation of the BMZ Action Plan for the Inclusion of Persons with Disabilities, DEval, <https://d-nb.info/1186644206/34> (accessed on 11 January 2021). [1]

注釈

¹ この箇所及び本文書では、「介入（intervention）」という用語を、評価の対象（評価されている事象）を意味するために使用しています。この点に関して詳しくは「3.2 評価の目的に応じた基準の適用」を参照のこと。

² 「evaluand」は、評価の対象——評価されている事象——を指します。

3 評価基準の実践への適用

本章では、評価のデザインと実施において評価基準をどのように使用すべきかという実践面について考えます。評価基準の適用の指針である 2 つの使用原則を振り返り、熟慮しつつ評価基準を活用するためのアプローチとはどのようなものかを考えた上で戦略的な優先事項、作業の仕方、文化が異なる多様な組織環境において、評価基準をどのように適用できるかを説明しています。評価基準により、ジェンダーの視点からみるためのレンズ（ジェンダーレンズ）を用いることで、ジェンダーによって異なった結果やインパクトがあることを評価者がどのように認識すべきかを説明します。最後に、評価基準を通じて、評価者及び評価管理者が持続可能な開発目標及びアジェンダ 2030 の達成に貢献する形で評価を行う方法を検討します。本章は、様々な評価への評価基準の適用方法を示す実用的事例も幅広く紹介します。

評価者及び介入のデザイン又は管理に関わる人々が、様々な文脈やステークホルダーに応じて適切で有益な評価を実施できるように、評価基準を使用する指針として、以下の2つの使用原則が策定されています。基準が機械的に適用されて、批判的思考や創造性及び参加者のオーナーシップが抑制されてしまうのを防ぐために、評価基準を使用する際には必ず、この使用原則に従わなければなりません（OECD、2019^[1]）。

- **使用原則 1**：評価基準は、質が高く有益な評価のために熟慮して使用すべきである。
- **使用原則 2**：評価基準は、評価の目的に沿って活用すべきである。

以下の節では、この2つの原則を詳しく説明し、個別の文脈に応じてどのように評価基準を適用するか、異なる時点でどのように評価基準を検討するか、或いは評価基準が相互にどのように関係しているかなど、評価基準を用いて評価を行うための主要な概念を説明しています。

熟慮して評価基準を適用

使用原則 1 は、熟慮して評価基準を使用すべきであると強調しています。これは実践においては、想定される利用者にとって価値ある質の高い有益な評価を実施するには、どの基準が最も有用かについて問題意識をもって考察することを意味します。ボックス 3.1 は、介入を評価する際にどうすれば熟慮しつつ評価基準を適用できるかの例を示しています。

以下の6つの側面とこれらに関連する設問を検討すれば、評価者は、熟慮しつつ評価基準を適用することができるでしょう。

- **評価の文脈**：介入の文脈はどのようなもので、個々の評価、介入、ステークホルダーの文脈において、評価基準はどのように理解されるか？
- **目的**：評価は何を目的としているのか、その目的を追求し達成するにはどのような設問が最も有益か？
- **役割と力関係**：ステークホルダーは誰か、彼らのニーズと関心は何か？彼らの間にどのような力関係があるか？どの評価基準を適用し、地域の文脈を踏まえて評価基準をどう理解すべきかを決定する際に誰が関与する必要があるか？オーナーシップに関する設問や、何を評価して、何を優先すべきかについては誰が決めるのかに関する設問を含むこともできる。
- **介入（評価対象）**：どんな種類の介入（プロジェクト、政策、戦略、セクター）について評価するのか？その範囲及び特性は？期待される結果はどの程度直接的又は間接的か？どれほど複合的・体系的思考が進行中か？
- **評価可能性**：情報へのアクセス、資源、データ（細かく分類されたデータを含む）について評価に影響する制約があるか、それが評価基準にどう影響するか？
- **評価時点**：介入のサイクルのどの段階で評価を実施するのか？介入実施の文脈は時間の経過とともに変化していくか、もしそうであれば、どのように変化してい

るか？評価中にそうした変化を考慮すべきか？タイミングは、評価基準の使用だけでなく評価根拠の情報源にも影響をもたらす。

ボックス 3.1.熟慮した評価基準の使用：ノルウェーの途上国向け市民社会助成金の評価

この評価は、ノルウェーの市民社会助成金を対象としたものです。助成金の目的は、途上国の市民社会を強化して、民主化、人権尊重、貧困削減を推進する能力と体力を備えた強い市民社会づくりに貢献することです。

評価では、評価基準を熟慮しつつ柔軟に適用しています。つまり、介入の特性（市民社会団体とのパートナーシップ及び能力構築）を反映し、ノルウェーの優先事項と文脈を考慮に入れた具体的な方法で評価基準を解釈しました。評価では、特定の観点に適応させた以下の評価基準を使用しています。

- インパクト – 民主主義、貧困削減につながる所得機会などの、主要なアウトカムと関連づけて定義した
- 妥当性 – 変更なし
- 有効性 – ここでは、パートナーシップの具体的な目的（すなわち、サービス提供、政策提案、能力強化）と関連づけて定義した。市民社会の活動の場の設置などの、いくつかの中期的アウトカムも含む。
- 持続性 – 変更なし

追加的な基準——「付加価値」——も使用しています。この基準は、専門的能力、組織的及び資金的な能力、ネットワーク形成能力として定義され、他の文脈であれば、有効性及び効率性の評価基準の下で扱われるであろう課題と関連しています。

Source: Helle et al. (2018[2]), [From Donors to Partners? Evaluation of Norwegian Support to Strengthen Civil Society in Developing Countries through Norwegian Civil Society Organisations](#)

評価の目的に応じた基準の適用

評価基準をどのように使うかを決定する際に最も重要な局面は、評価基準を評価の目的とその文脈に関連づけ、評価の目的を中心に評価基準及び設問を設定することです。ボックス 3.2 に、評価の目的を定義する方法について2つの例を示しています。

この評価基準は、すべての介入に標準的で固定的な方法を適用したり、チェックボックス形式で使用することは意図していません。評価対象となる介入と関連して、基準を慎重に解釈し理解する必要があります。これにより柔軟性が高まり、個々の評価に応じて評価基準を適用しやすくなります。評価では、評価基準のどの特定の概念を活用するか、またそれはなぜかを明確にしてください。

評価の目的は、慎重かつ明確に定義する必要があります。評価の目的と評価結果の活用方法を確実に理解してもらうため、評価に関わるステークホルダーをこの段階で参加させる必要があります。

評価の目的を決定する際に検討すべき主な設問には、以下が含まれます。

1. 評価に対してどのような要求があるか、誰のための評価か、その人たちは評価結果をどのように活用するのか？
2. 介入の特性と文脈を踏まえると、何が実行可能なのか？
3. 主な設問に答える際、どの程度の確実性が求められるのか？
4. いつ情報が必要とされるか？
5. 介入とその結果について、すでに分かっていることは何か？その情報を誰が保有し、どのように使用しているのか？

次に、品質基準及び倫理基準を参考にして、評価工程の方法論的アプローチ、デザイン、実施及び管理について考える必要があります。

個別の評価に合わせて基準を調整する際には、因果関係を徹底的に調べ、各効果についてその介入にどの程度起因しているかを特定できるかを判断することも重要です。こうした検討によってステークホルダーの期待値を調整し、どの評価基準を詳細に扱うか決定します。以下の 2 つの節で扱う有効性とインパクトにおいて、この検討が最も重要になります。例えば、効率性と持続性の基準には、有効性とインパクトのもとで評価したように、介入に起因する実際の（又は予想される）便益を使用できるでしょう。評価可能性の分析は、目的がよく理解されているか確認する上で有益なツールとなります¹。各評価基準の解釈の方法については、第 4 章で詳述しています。

ボックス 3.2. 評価の目的の定義：ウガンダ、ケニア、スウェーデンの例

デンマーク国際開発庁（Danida）が実施したウガンダにおける水・衛生・環境プログラムの評価

デンマーク外務省の評価部は、1990～2017 年にウガンダで実施された水・衛生・環境の改善のためのデンマークのイニシアティブについて外部評価を委託しました。

この評価の目的は、以下のとおりでした。

- 支援サブ・セクターの結果と実績を記録する
- Danida がこれらの支援サブ・セクターを支援したことにより生み出された「付加価値」を分析する
- 得られた教訓を導き出す

この目的に向けて、評価では①各セクター及びセクター横断的レベル、②サブ・セクターレベルの 2 段階における有効性の評価に重点を置きました。このイニシアティブの価値も評価し、結果の持続性を検討しました。

国連難民高等弁務官事務所（UNHCR）とデンマークによる合同評価

UNHCR とデンマークによる、カロベエイ（ケニア）周辺の難民対応に関する合同評価では、目的を以下のように定義しています。

「評価の主な目的は、カロベエイおよびその周辺の包括的定住モデルに関する学びに寄与することである。評価の意図は、人道的支援と長期的な開発支援をつなぐための具体的な取り組みから得た教訓を文書化することにより、包括的な解決モデルのデザイン及び実施の可能性と課題に関わる立証を示すことにある」

スウェーデン国際開発協力庁（Sida）による紛争中及び紛争後の平和構築支援の評価

この評価報告書は、評価に目的をいかに反映させるかに関して有益な例を示し、以下のように説明しています。

この評価の目的又は使用方法は、平和構築の実例から得た教訓を体系化することです。この評価は、スウェーデン国際開発協力庁（Sida）が使用する平和構築アプローチを概念化し確立する工程に役立つこととなります。さらには、そのアプローチが、紛争被害を被った地域における Sida の支援を今後戦略的に計画・立案する際に影響を与えるでしょう。加えて、この評価は概念及び実践としての平和構築について理解を高めるのに寄与することが期待されます。

評価の具体的な目的は、様々な文脈において Sida が戦略的なレベルで平和構築にどのように取り組んだかを評価することでした。そのために、以下の 4 つの評価基準に特に注目しました。

- 様々な文脈における受益者のニーズと Sida の政策的優先度に対する Sida の平和構築活動の妥当性
- 平和構築の全体的な目標への貢献に照らした Sida による平和構築活動の有効性と、それが受益国の平和構築活動に有益な枠組みを提示できる能力に係る有効性。
- Sida による平和構築活動のインパクト
- Sida による平和構築活動の持続性

4 つの評価基準すべてにおいて、ジェンダー平等、女性の権利とエンパワメントに加えて、被周縁化集団及び少数民族の包摂性に関連する結果を特に重視しました。ケーススタディで取り上げた 4 カ国すべてにおいて、文献調査及び面接調査で明らかになった平和構築と関連のある周縁化の側面に焦点を当てました。被周縁化集団は、主に民族の観点から、又は少数民族であることの影響として捉えられました。なぜなら、4 件の事例すべてにおいて、民族的な要素が紛争の根本的要因の大部分を占めていたからです。

Sources:

Danida (2019[3]), Evaluation of Water, Sanitation and Environment Programmes in Uganda (1990-2017), <http://www.oecd.org/derec/denmark/denmark-1990-2017-wash-environment-uganda.pdf>;

ADE (2019[4]), Joint Evaluation of the Integrated Solutions Model in and around Kalobeyei, Turkana, Kenya, https://um.dk/en/danidaen/results/eval/eval_reports/publicationdisplaypage/?publicationid=dd54bef1-1152-468c-974b-c171fbc2452d;

Bryld (2019[5]), Evaluation of Sida's Support to Peacebuilding in Conflict and Post-Conflict Contexts: Somalia Country Report, https://publikationer.sida.se/contentassets/1396a7eb4f934e6b88e491e665cf57c1/eva2019_5_62214en.pdf

時間軸の理解

評価基準は、様々な時点で適宜対応することができます。各基準は、介入の実施前、実施中、又は実施後に使用することができます。介入サイクルの様々な時点で使用することができます。ただし、時点によって評価基準の解釈や立証の情報源に違いが出るかもしれません。例えば、介入を実施する前の時点では、有効性と持続性は予測値である一方で、介入後には、より多くのデータが入手でき一層確かな結論を引き出せるでしょう。

そのため、評価基準及び関連する評価に関する設問には、評価時点に関わる次の 2 つの重要な要素を反映させてください。その要素とは、①介入のサイクルのどの時点で評価を実施するか、②評価では、介入のどの段階又は結果の連鎖におけるどの時点を重視するか、です。

介入の実施前、実施中、実施後のどの時点で評価を行っても、各評価基準の概念的な枠組みは変わりません。しかし、基準に沿った判断に使用されるデータや立証、そして使用方法は変わります。評価のステークホルダーと（予想される）結果を協議する際に、評価者はこうした違いに留意すべきです。この違いが、評価の有益性や信頼性に対する認識に影響を与える可能性があるからです。例えば、持続性の事前評価（介入開始前に実施）は、介入デザインや、予測される便益が持続するという推論の正当性に関する立証を分析することで、介入の便益が持続する可能性を検討できるでしょう。介入の完了後に持続性を評価する場合には、今度は、その介入によって実際に達成された便益に関するデータと立証に基づいて、便益が実際に持続しているかどうかを検討することになります。

評価を通して過去を振り返る場合、その時点の状況と入手可能なデータを考慮して、何ができたか、どうすべきだったかをめぐる合理的な推測に基づき判断を行う必要があります。当時は知りようがなかった現時点の情報に基づいて、過去のプログラム立案者の対応を判断するのは不公平というものです。とはいえ、多くの評価者は、入手可能な情報が（利用できたかもしれない場所で）十分に活用されていなかったと感じます。こうした情報を使えば、介入の妥当性がさらに高まっていたらと。例えば、現地の人に十分な相談がなされなかったり、彼らが介入のデザインに関与しなかったりした場合などが挙げられます。こうした状況はある程度回避できると想定されるため、評価において留意しておくべきです。

よりよい評価の設問に向けた基準の活用

よい評価設問を設定することは、評価工程の重要な部分であり、6つの評価基準の使用は、評価設問を決定する工程と相互に関連し、評価の促進に役立ちます²。設問の決定工程ではまず、評価の目的とその評価を誰がどのように使うかを考察します。適切な参加型過程を通してステークホルダーが効果的に関与することによって、評価者及び評価管理者は評価がどのように活用されるかを理解することができます。介入とその文脈、目的と「変化の理論」を深く理解することで、こうしたステークホルダーと協議する際の

助けとなります。初期段階においてこれらの設問について検討し、評価で対処する必要がある（数を絞った）重要な設問を設定することができます。

初期段階に続いて、評価基準によって何か漏れがないか様々な視点から確認するための技法が提供され、評価のデザインに欠かせない設問のさらなる精緻化が可能となります。こうして評価工程が体系化され、評価の包摂性が確保されます。これは「よりよい評価」の極めて重要な構成要素であり、この設問づくりの段階は、明確で首尾一貫した総合的アプローチによって取り組むことが強く推奨されます。

評価を委託する組織が、一定の基準に従って介入を評価するように特別に決めているかもしれません。その場合は、一貫性を保つために、熟慮して評価基準を適用すると共に、組織の指針にも従う必要があります。

例えば、Sida の評価手引書には、評価基準ごとに標準的な設問の例が示されています（Molund and Schill, 2004^[6]）。その他にも、組織がどのように基準を解釈して適用したかを示す例は、例えば世界食糧計画（WFP, 2016^[7]）で評価管理者及び評価者のために策定された指針やテクニカルノートなどで参照することができます。

評価基準間の関係性

評価基準は、介入とその結果を眺めるための複数のレンズから構成されます。評価基準は互いに関係しあっており、各基準の概念は、結果を生み出す過程における相互補完的部分を分析するのに役立ちます。例えば、有効性とインパクトは、介入の目標をどう定義するかによって、結果の連鎖の異なるレベルを確認するものです。従って、この 2 つの評価基準は因果の連鎖に沿って互いに関係し合っています。

評価基準は、しばしば相互に依存しています。例えば、受益者の優先度に対応していない介入が、意図されたインパクトをもたらすとは考えられません（通常とは異なる、インパクトを生み出す別の経路がある場合は別ですが）。適切に実施されなかった（有効性が低い）介入は、おそらく持続性も低いでしょう。一方、介入の妥当性は高いが有効性は低いといった可能性も考えられます。或いは、整合性が高い介入が、取引コストの増加のために非常に非効率なものになることも考えられます。評価者は、因果関係があるのか、どんな因果関係があるのかを検討することを含めて評価基準の相互関係性と相乗効果を考察する必要があります。

ほとんどの評価は、個々の評価基準の評価結果に加えて、全評価基準の総合的結論を踏まえて、時には実績を数値スコアによる点数化を行うなどして、結論を導き出します。アセスメントの目的にもよりますが、評価者は、介入に関する結論を出すに当たって、全体像を見てどうすれば適用した評価基準に適切に比重化ができるかを考える必要があります。組織によっては、最重視する（「ノックアウト」＝死活的）基準を明らかにして評価基準の比重化を行うこともあります。その場合は、ノックアウト基準の実績が不十分であれば、他の評価基準のスコアがどれほど良くても、その介入は失敗とみなされます（事前評価の場合には、資金が供与されないでしょう）。

評価基準の相互の関連性については、第4章で詳しく説明します。

使用する評価基準の選択

前述のとおり、介入と評価の目的、優先事項、範囲、文脈によって、どの評価基準により焦点を当てるかが決まります。評価者は、各基準がもたらす相対的な価値を考えなくてはなりません。ここでは2つの基本的な決定が必要です。一つ目は、この基準は今回の評価にとって重要な検討事項か。二つ目は、この基準に関する設問に答えることが実現可能かどうか、です。

文脈に関係なく、6つの評価基準すべてを単純に適用したくなる利用者がいるかもしれませんが、しかし、よりよいアプローチ（評価基準の本来の意図に合致し、最も質の高い評価につながるアプローチ）では、評価自体及びその評価で答えようとしている設問に相応しい形で、評価基準を慎重に選択し使用する必要があります。

これを実現するために、以下のような設問に答えてみてください。

- もし、この介入に関する設問をたったひとつしか設定できないとすれば、それは何か？
- 評価で取り組むのに最も適した設問はどれか、他の手段（研究事業、立証統合化、モニタリング演習、又は円滑な学習過程）を通して取り組むことができそうな設問はどれか？
- その設問に満足できる答えを出すために十分なデータがあるか？ない場合、今後、より質の高いデータ、またはより多くのデータを入手できるか？
- 誰に意見を聞いて設問リストを作成したか？重要な視点が欠けていないか？
- 関心のあるすべての評価基準に適切に取り組むための、十分な時間と資源があるか？あるいは、一部の基準の分析に集中した方が、価値ある情報が得られるか？

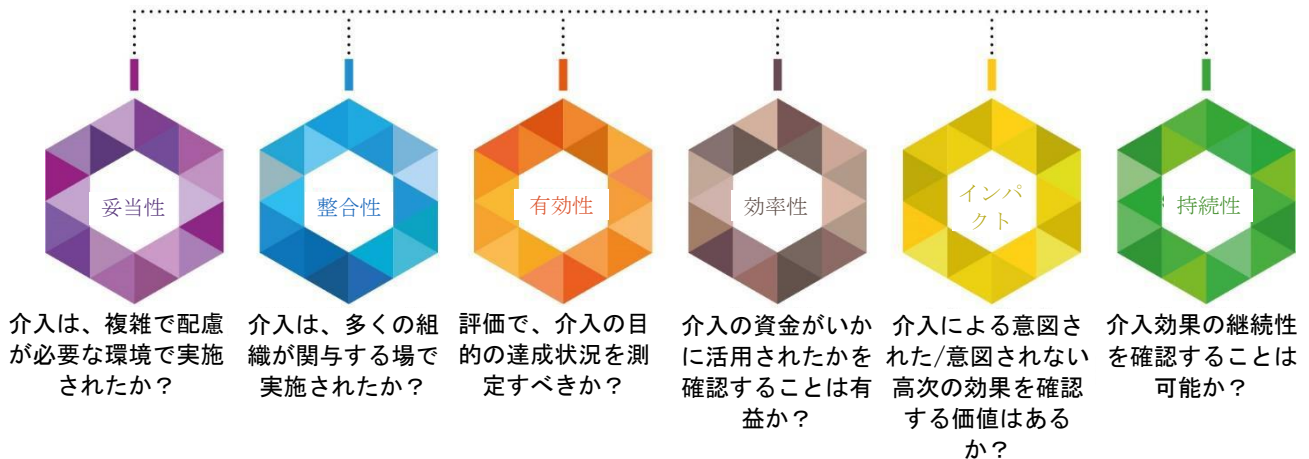
評価基準を使用するときは、柔軟性（すべての基準の機械的な適用を避ける）とご都合主義（最も簡単な基準のみ、又は望ましい結果が出やすい基準のみを選ぶ）のバランスをとることが重要です。特筆すべき点として、たとえその設問が時としてむずかしいものであっても、インパクトと整合性に関する重要な設問への回答を避けてはなりません。検討すべきポイントを図3.1に示します。ただし、この図は一例であり、包括的な点検表ではない点に注意してください。

図 3.1. 個別の評価における評価基準の取り上げ方



評価基準の検討プロセス

評価を実施する際に、まず6つの評価基準すべてを検討してください。
ただし、評価の目的や限界、介入の文脈を含むいくつかの要因に応じて、
評価でどの基準を重視するか慎重に選択する必要があります。



介入と評価の両方に関わるステークホルダーについて十分な知識があれば、評価のデザインを決めて実施する際に、彼らの異なる関心や優先事項のために生じ得る緊張関係を明らかにできます。受益者は有効性の把握（例：栄養失調治療プログラムへの参加を通じて子どもの健康が改善しているか）に最も大きな関心を抱く一方で、実施者は、さらに多くの世帯への治療の拡大を見据えて、効率性の把握により大きな関心を抱くかもしれません。ほとんどの場合、興味があるすべての設問に1回の評価で回答することはできず、選択が求められます。評価で選ばなかった設問を別の機会に対応する可能性を高めるため、ステークホルダーニーズの優先順位づけと評価設問の決定に関わる議論のプロセスと結果を文書に残しておくことが役立ちます。

組織レベルで評価基準の運用と適用

評価基準の定義とこのガイダンスは、共通のプラットフォームと合意された一連の定義を提供するものです。しかし、この定義を、組織の文脈に合わせてしつらえることは欠かせません。評価者と評価管理者は、多くの場合、開発機関、省庁、その他の組織を代表しており、それらの組織は固有の権限、政策的優先事項、評価政策、標準及び指針を有しており、これらすべては組織の管理組織によって決定されます。

評価基準を運用し適用する方法を検討するときには、評価者と委託者が、意思決定を行う背景として、組織の戦略的優先事項、文化、機会を注意深く検討することが重要です。例えば、インパクトなど特定の用語の使われ方は異なるため、評価基準とどの基準に焦

点を当てるかということに関して混乱や誤解が生じる可能性があることに、細心の注意を払う必要があります。そうすることで、評価基準を適用する際に評価者と委託者に役立ち、評価の結果を最大限に活用して意図していた利用者のニーズに対する妥当性を高めることができます。管理者は、評価基準をどのように適用し解釈すべきか検討するために、評価者、委託者及び評価を活用する予定の人々の間の議論を推進する必要があります。こうした過程を通じて、利用者のニーズに合致した、信頼性が高く適時な評価のデザインを作ることができます。

組織が定める方法論的な要件が、評価基準の適用方法に影響を及ぼす可能性もあります。評価者は、自身の組織又は委託者に特有の要件及び指針を参照しなくてはなりません。国連評価グループ（UNEG）の指針、ECG のグッドプラクティス基準、ALNAP の人道支援の評価ガイダンスなどの関連資料も非常に有益です。

アジェンダ 2030 と持続可能な開発目標への対応

使用する組織に応じて基準を適応させることに加えて、評価基準が理解され適用される方法にも、幅広い政策的な文脈が反映され、評価管理者、評価者及びステークホルダーによる評価基準の使用方法に影響を与えます。特に国際開発協力に携わる評価者にとって、今後 10 年間は、持続可能な開発目標（SDGs）とアジェンダ 2030 が唯一にして最も重要な包括的政策枠組みであり、グローバルな目標となります。

アジェンダ 2030 の重要な要素には、以下があります。

- 開発の便益への普遍的なアクセス
- 特に取り残されるリスクが高い人々に対する包摂性
- 人権、ジェンダー平等及び他の公平性への配慮
- 環境の持続可能性、気候変動、天然資源管理
- 文脈と開発介入の複雑さ
- 開発過程に関与するアクター間の相乗効果

この枠組みは、評価基準の解釈の方法や評価工程自体（評価基準の適用や優先すべき設問の特定に誰が関与するかを含む）をはじめとして、介入とその評価の両方に影響を及ぼします。ボックス 3.3 は、国の評価に関する検討課題にアジェンダ 2030 を関連付けるためのガイダンスを示し、ボックス 3.4 にドイツの開発評価制度における実施事例を示しています（BMZ, 2020^[8]）。コスタリカ、ナイジェリア、フィンランド等の政府及び非政府組織（NGO）が同様の取り組みを行っており、評価者と評価管理者に有益な教訓を与えています（D'Errico, Geoghe and Piergallini, 2020^[9]）。

改訂後の評価基準定義の文章にも、様々な形でこうした要素を反映しています。例えば、妥当性、有効性、インパクトを検討する際に、文脈や受益者の視点及び優先事項に特に配慮すること、有効性とインパクトでは結果の公平性を考慮に入れること、整合性を確

認する際には統合的アプローチを採用することなどです。また、使用原則にも、統合的な考え方を取り入れ、アジェンダ 2030 を反映しています。

ボックス 3.3. SDGs を国の評価に関する検討課題に関連付ける

「国家の優先事項を SDGs と関連づけるための評価、評価委託者及び管理者向けガイド」は、国が評価に関する検討課題を設定し、評価の価値を高めるために必要な 5 つの項目を示しています。評価者は「複合的システム思考」により、以下を行うことが推奨されています。

- 単一の政策、プログラム、プロジェクトにとらわれない発想をする
- 成否に影響を及ぼすマクロ要因を検討する
- 「成功」の微妙な意味合いを把握する
- 文化の重要性を認識する
- 評価的思考と適応型管理を採用する
-

Source: Ofir et al. (2016[10]), Briefing: [Five considerations for national evaluation agendas informed by the SDGs](#)

ボックス 3.4. アジェンダ 2030 及び SDGs に関する評価基準のマッピング

ドイツ連邦経済協力開発省（BMZ）は、評価者がアジェンダ 2030 及び SDGs への総合的貢献度を評価する際の助けになるよう、ドイツ開発評価研究所（DEval）の支援を得て以下の設問を設定しました。各設問は SDGs の原則に対応するものであり、評価基準にも関連しています。

普遍性、共有する責任と説明責任

- 介入は、SDGs の達成にどの程度貢献しているか？（インパクト評価基準を参照）
- 介入は、その活動を実施するために、どの程度（パートナー／他の支援国／国際機関の）既存の制度及び構造を使用する設計になっているか、また既存の制度及び構造が実際にどの程度使用されているか？（整合性評価基準を参照）
- 介入の実施に当たり、他の支援者や開発パートナーと連携しているか？している場合、どの程度か？（整合性評価基準を参照）
- モニタリング、学習及び説明責任遂行のために、どの程度共有制度を使用しているか？（整合性の評価基準を参照）

経済、環境、社会開発の相互作用

- 介入は、持続可能な開発に向けた全体アプローチ（社会、環境及び経済）にどの程度従っているか？（妥当性評価基準を参照）
- 社会、経済、環境面のアウトカムの間にはどの程度の意図された、或いは意図されない正又は負の相互作用があったか、また介入の全体としてのインパクトは何だったか？（有効性評価基準とインパクト評価基準を参照）
- 介入は、社会、経済、環境面のアウトカムの間、意図された、或いは意図されない正又は負の相互作用を生み出すことにどのような貢献をしたか、また介入の全体としてのインパクトは何だったか？（有効性評価基準とインパクトの評価基準を参照）

包摂性

- 介入は、最も不利益を被っている脆弱なグループの参加とその推進に関する国際的な規範及び標準に、どの程度合致しているか？（整合性評価基準を参照）
- 最も不利益を被っている脆弱なグループ（年齢、収入、ジェンダー、民族などによる差別を受ける可能性があるなど）に、開発によりどの程度の意図された、或いは意図されない正又は負の総合的な変化が見られたか？（インパクト評価基準を参照）
- 最も不利益を被っている脆弱なグループ（年齢、収入、ジェンダー、民族などによる差別を受ける可能性があるなど）における、開発を通じた意図された、或いは意図されない正又は負の総合的なインパクトに、介入はどのような貢献をしたか？（インパクト評価基準を参照）
- 介入は、最も不利益を被っている脆弱なグループ（年齢、収入、ジェンダー、民族などによる差別を受ける可能性があるなど）の強靱性の強化にどの程度貢献したか？（持続性評価基準を参照）

Source: BMZ (2020[8]), Evaluation Criteria for German Bilateral Development Co-operation

評価基準へのジェンダーレンズ（ジェンダーの視点）の適用

評価者は、ジェンダーの違いによって差異のある経験やインパクトと、それらが他の形の差別（例：年齢、人種・民族性、社会的地位）の特定の文脈における相互作用を熟慮する必要があります。どのような介入であるかにかかわらず、評価者は、ジェンダーに基づく力関係が、他の形の差別とどのように関係して相互作用するのかを確認の上、どのように介入の実施や結果に影響を与えるかを考慮することも重要です。こうすることにより、介入の政治経済的、社会文化的な文脈が、その目標の実現や達成にどのような影響をもたらすかについて究明することができます。

ジェンダーレンズに則して、ジェンダー平等の目標の達成を支援すると同時に、評価を通して学びを得て、説明責任を果たすために立証することが可能になります。評価基準にジェンダーレンズを適用する実践的手順は以下のとおりです。

- 評価者、管理者、委託者が、評価における意思決定、データ収集、分析及び結果の共有に適切に参加にできるよう、包摂的な形で評価を実施する。
- 介入において平等な機会提供を危うくする社会的障壁が、ジェンダーとどの程度の相互作用を起こしているかを考察する。
- 様々なステークホルダーの経験やインパクトをよりよく理解するために、介入が、法律、経済、政治、宗教、社会文化に関わる環境とどのように相互作用しているかを考察する。
- 男らしさ、女らしさ及びジェンダーの力学や役割の変化に関する社会的に構築された定義について考察する。
- 評価者を選ぶ際、ジェンダーに配慮した評価アプローチに関する評価者の技能や、様々な文脈で評価を実施した経験を確認する。

評価者が評価基準へのジェンダーレンズの適用方法を考える参考となるように、以下の表を作成しました。

表 3.1. 評価基準へのジェンダーレンズ（ジェンダーの視点）適用

評価基準	ジェンダー視点適用につながる質問
妥当性	介入は、すべてのジェンダーのニーズと優先事項に対応する形でデザインされたか？もしそうであれば、どのように？ 介入のデザインは、あらゆるジェンダーの人権をどの程度反映しているか、また被周縁化グループを含む地域の多彩なステークホルダーからの反応をどの程度含んでいるか？ 介入は、すべてのジェンダーの実践的で戦略的なニーズに対応しているか？
整合性	介入のデザイン、実施及び結果は、「女子に対するあらゆる形態の差別の撤廃に関する条約（CEDAW）、北京宣言及び行動綱領、国際人口開発会議の行動計画、アジェンダ 2030 等のジェンダー平等と権利に対する国際的な法及び合意にどの程度整合性があるか？ 介入は、ジェンダー平等と人権の向上を目指す国内の法律及びイニシアティブをどの程度支持しているか？どんな教訓を学べるか？
有効性	介入は、ジェンダー平等に貢献する形でその目標と期待される成果を達成したか？もしそうであれば、どのように？ 介入は異なる人々に対して差異のある結果をもたらしたか？もしそうであれば、なぜか？異なるジェンダーの人々まで及ぶようにするには、異なるアプローチが必要であったか？差異のある効果に対するモニタリングと分析は十分であったか？懸念点に対処し、効果を最大限に高めるために、介入を修正したか？ 「変化の理論」と結果に関する枠組みは、ジェンダー平等の分析、政治経済的分析及び人権を参考にしたものであったか？もしそうであれば、どの程度？ ジェンダーの違いによって有効性にどの程度の違いが生じているか、それはなぜか？
効率性	様々な投入資源は、ジェンダー平等を考慮した形で配分されたか？もしそうであれば、どのように分配されたか？差異のある資源配分は適切だったか？ 対象者 1 人当たりへの投資コストは、ジェンダーが異なる人々の異なるニーズに対応しているか？
インパクト	異なるジェンダーに対し平等なインパクトがあったか、あるいは介入への関与、経験やインパクトにジェンダーに関連する差異があったか？もしそうであれば、なぜこうした差異のあるインパクトが生じたのか？ ジェンダーに関連するインパクトが、人種・民族性、障がい、年齢及び性別指向等の社会的障壁とどの程度相互作用して、差異のある経験やアウトカムを生み出したか？ 政治、経済、宗教、法律及び社会文化に関わるさらに広範な環境においてジェンダーに関連する規範と障壁が、アウトカムにどのように影響したか？ インパクトは、ジェンダーの異なる人々同士の平等な力関係、及び社会的な規範と制度の変化にどの程度貢献したか？

持続性	介入は、より広範な法的、政治的、経済的及び社会的な制度におけるジェンダー平等の向上に貢献したか？もしそうであれば、どのようにどの程度？その結果、すべて又は一部のジェンダーの人々にとって障壁となる社会規範に持続性のある変化がもたらされたか？変化を実現できなかった場合、なぜか？ ジェンダー平等に関連する成果は、介入終了後も持続するか？その工程は便益の持続性向上に貢献したか？ジェンダー平等の達成を支持する仕組みが、長期的に策定されているか？
-----	---

他の評価基準の使用

この 6 つの評価基準は、評価で扱うべきすべての重要な概念を十分に反映した基準一式となることを目指すものです。熟慮して、かつ文脈を配慮して適用すれば、この評価基準は、持続可能な開発及び人道支援分野全体において評価者にとって適切なものとなるでしょう。

とは言え、ある文脈では他の評価基準が使用されます。例えば、多くの組織が、評価方針を通して特定の重点分野の分析を義務づけるでしょう。2020 年にイタリアがボリビアで実施した医療保健プログラムの評価では、妥当性、有効性、効率性、インパクト、持続性、整合性、イタリアの協力による付加価値、イタリアの協力の存在感、オーナーシップという 9 つの評価基準が使用されました（Eurecna Spa, 2020^[11]）。人道的な状況に評価基準を適用した別の例では、適切性、対象範囲、一貫性などの評価基準が高い妥当性を持つとされます³。追加性（アディショナリティ）という評価基準は、混合資金（ブレンディッドファイナンス）、民間向け融資、気候変動関連融資などの分野で適用されることがあります。様々な種類の財政的及び非財政的な追加性を含めて、追加性に対する多様な定義が使用されています⁴。使用する定義によっては、妥当性、有効性またはインパクトという評価基準の下で、追加性を検討することもありますし、他の評価基準とは全く別の横断的評価基準として扱うこともあります。

評価基準を追加すべきかどうか検討する際は、慎重になるべきです。追加が混乱を招き、評価の範囲があまりに広くなる（有益性の低い分析になる）可能性があるからです。評価基準を絞ることによって、十分に深い分析と概念的な明確性を確保することに役立つというのは、2017 年から 2019 年にかけて評価基準の定義を更新する際の協議過程で繰り返し指摘された点です。

他の評価基準を使用する場合には、その基準を定義することが重要です。その評価基準を追加する理由を説明することで、他の人が、追加基準と本文書に示す 6 つの評価基準との関係性を理解する助けとなります。評価全体からの学びを支援するためには、同じ評価基準の下で、同一の概念や要素を評価することが不可欠です。

どの評価基準を使用する場合にも、本文書に示した中核となる使用原則とガイダンスを適用してください。

参考文献

- ADE (2019), Joint Evaluation of the Integrated Solutions Model in and around Kalobeyei, Turkana, Kenya, UNHCR and Danida, https://um.dk/en/danidaen/results/eval/eval_reports/publicationdisplaypage/?publicationid=dd54bef1-1152-468c974b-c171fbc2452d (accessed on 11 January 2021). [4]
- Bamberger, M., J. Vaessen and E. Raimondo (2015), Dealing With Complexity in Development Evaluation - A Practical Approach, https://www.betterevaluation.org/en/resources/dealing_with_complexity_in_development_evaluation (accessed on 11 January 2021). [12]
- BMZ (2020), Evaluation Criteria for German Bilateral Development Co-operation. [8]
- Bryld, E. (2019), Evaluation of Sida's Support to Peacebuilding in Conflict and Post-Conflict Contexts: Somalia Country Report, Sida, https://publikationer.sida.se/contentassets/1396a7eb4f934e6b88e491e665cf57c1/eva2019_5_62214en.pdf (accessed on 11 January 2021). [5]
- D'Errico, S., T. Geoghe and I. Piergallini (2020), Evaluation to connect national priorities with the SDGs | Publications Library, IIED, <https://pubs.iied.org/17739IIED> (accessed on 22 February 2021). [9]
- Danida (2019), Evaluation of Water, Sanitation and Environment Programmes in Uganda (1990-2017), Evaluation Department, Ministry of Foreign Affairs of Denmark, <http://www.oecd.org/derec/denmark/denmark-1990-2017-wash-environment-uganda.pdf> (accessed on 11 January 2021). [3]
- Davis, R. (2013), "Planning Evaluability Assessments: A Synthesis of the Literature with Recommendations", No. 40, DIFD, <https://www.gov.uk/government/publications/planningevaluability-assessments> (accessed on 12 January 2021). [13] •
- 35 APPLYING EVALUATION CRITERIA THOUGHTFULLY © OECD 2021 Eurecna Spa (2020), Bolivia - Evaluation of Health Initiatives (2009-2020), Italian Ministry of Foreign Affairs and International Cooperation, http://www.oecd.org/derec/italy/evaluationreport-of-health-initiatives-in-Bolivia-2009_2020.pdf (accessed on 11 January 2021). [11]
- Helle, E. et al. (2018), From Donors to Partners? Evaluation of Norwegian Support to Strengthen Civil Society in Developing Countries through Norwegian Civil Society Organisations, Norad Norwegian Agency for Development Cooperation, https://www.norad.no/globalassets/filer2017/evaluering/1.18-from-donor-to-partners/1.18-from-donors-to-partners_main-report.pdf (accessed on 11 January 2021). [2]
- Molund, S. and G. Schill (2004), Looking Back, Moving Forward Sida Evaluation Manual, Sida, <https://www.oecd.org/derec/sweden/35141712.pdf> (accessed on 11 January 2021). [6]
- OECD (2019), Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use, DAC Network on Development Evaluation, OECD Publishing, Paris, <https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf> (accessed on 11 January 2021). [1]

OECD (2002), Evaluation and Aid Effectiveness No. 6 - Glossary of Key Terms in Evaluation and Results Based Management (in English, French and Spanish), OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264034921-en-fr>. [14]

Ofir, Z. et al. (2016), Briefing: Five considerations for national evaluation agendas informed by the SDGs, IIED, London, <http://dx.doi.org/10.3138/cjpe.30.3.02./11>. [10]

WFP (2016), Technical Note: Evaluation Methodology, DEQAS, World Food Programme, <https://docs.wfp.org/api/documents/704ec01f137d43378a445c7e52dcf324/download/> (accessed on 11 January 2021). [7]

注

¹ 評価可能性は、活動又はプログラムを信頼性の高い形でどの程度評価できるかを指します。評価可能性の判定は、活動の目標が適切に定義され、結果が検証可能であるかどうかを確認するために、活動案作成の早い段階で行う必要があります（OECD, 2002^[14]）。Davis（2013^[13]）も参照されたい。

² 評価設問の作成工程やステークホルダーの参加過程については、介入の複雑性も考慮に入れて、例えば、Bamberger、Vaessen 及び Raimondo（2015^[12]）が詳しく論じています。

³ 「人道支援活動における説明責任と実践のための積極的学習ネットワーク（ALNAP）」は、現在の ALNAP の包括的な人道支援活動の評価ガイドを補完するものとして、人道環境における評価基準の使用に関する 2006 年指針を更新しています。

⁴ EvalNet の混合資金評価に関する作業部会は、追加性と関連する概念の定義に関する研究を支援しました。その結果は、2021 年初めに発表予定です。

4 6つの評価基準の理解： 定義、分析要素、主な課題

本章では、6つの評価基準それぞれを詳しく検討し、それらの定義の実践面での意味、様々な分析要素、そして包摂性と公平性を厳密に反映した形で評価基準を適用する方法を探ります。評価基準毎に、その基準を適用する際の主な課題を、評価者と評価管理者がどのように課題に対処できるかについて実践的な提言を示しつつ、表にまとめています。評価基準の実践への適用例を幅広く紹介し、各評価基準の定義を構成する要素について批判的考察を行います。

本章では、6つの評価基準（図 4.1）について、最も論理的と考えられる順序で説明します。つまり、妥当性と整合性から始まり、有効性と効率性、最後にインパクトと持続性の紹介をします。それぞれの評価基準を定義し、その重要性を説明したのち、分析要素（定義に含まれる主な概念）の検討を通じて定義を更に詳しく解説します。これらの要素は二次的基準ではなく、文脈と目的に応じた評価基準の様々な適用方法を示すものです。他の評価基準との関係性についても考察します。

また、この章では、アジェンダ 2030 で義務づけられた、包摂性及び「誰一人取り残さない」という原則に関連する主な課題をめぐる初期的な考察も行います。ジェンダー平等、女性のエンパワメント、人権に関する具体的な指針など、この評価基準及び評価アプローチに公平性という視点を適用するために更なる作業が進められています。

各評価基準について共通する課題を表に示し、評価者や評価管理者がその課題を克服する上で役立つ考えをまとめています。今後新たな定義に基づく経験が共有されたら、この表を随時更新していきます。

最後に、各評価基準の解釈の仕方をするために、実際の例をあげています。

図 4.1. 6つの評価基準と関連する設問



妥当性の定義：

妥当性：介入は正しいことをしているか？

介入の目的やデザインが、受益者¹、世界、国、パートナー／機関のニーズ、政策、優先事項に対応し、状況が変わっても対応し続ける度合い

注釈：「対応」とは、介入の目的とデザインが、対象地域における経済、環境、公平性、社会、政治経済、および能力の状況に敏感に対応していることである。「パートナー／機関」とは、国・地域・地方の政府、市民社会、民間組織、および国際機関のことで、介入に対する資金提供や介入の実施・管理を行う組織を含む。妥当性の評価は、異なる優先事項やニーズ間の差異やトレードオフの考察を含む。妥当性評価では、介入が妥当であり続けられる（又はあり続けた）度合いを確認するために、当該文脈における全ての変化を分析する必要がある。

妥当性とは何か、なぜ重要なのか？

妥当性の評価は、介入が正しいことをしているかを理解するのに役に立ちます。妥当性の基準を通して、評価者は、介入の目標と実施が受益者及びステークホルダーのニーズや、介入の優先事項と明確に整合しているかどうか評価することができます。また、対象となるステークホルダーが、その介入を有益で価値あるものと見なしているかを確認します。

妥当性は、デザインから実施に至るプログラムや政策のサイクル全体に関係する検討事項であり、持続可能な開発目標（SDGs）などの国際目標と関連づけて考えることもできます。妥当性は、①受益者及びステークホルダーのニーズに対する妥当性、②文脈に対する妥当性、③品質とデザインの妥当性、④時間の経過を考慮した妥当性という、4つの潜在的な分析要素を通して分析できます。この点は、分析要素の節で詳しく扱います。評価の目的に応じてどの分析要素を含めるかを判断することが必要で、いつも全要素について分析するわけではありません。

妥当性の評価では、まず、介入の目的が現実的かつ実行可能な、適切に定義されたものであるか、そして、介入の結果は検証可能なもので開発介入の国際的標準に適合したものであるか、を判断することから始めます。これは、OECDの開発評価の品質基準（OECD, 2010^[1]）で詳述されている評価可能性の概念に沿うものです。結果や目的に関する記述が不十分で曖昧であったり、測定が困難、或いは活動やインプットに重点を置いてしまっている場合もあります。場合によっては、評価者が目的を明確にするために、

「変化の理論」を精緻化し再構築することが必要になります。評価者は、質が高く現実的な目的に対する評価を行うよう留意すべきです。目標の達成度を測定するための指標

は、例えば、SMART（Specific（具体的）、Measurable（測定可能）、Attainable（達成可能）、Relevant（妥当）、Timely（適時））指標（IDD and Associates, 2006^[2]）などの一般的に用いられている基準に沿って適正を確認する必要があります。妥当性を確認する評価では、適切な目的がグッドプラクティスを反映した形で設定されたか、その場合どのように設定されたかを考慮してください。実施パートナーの組織的な体力と能力、そして彼らが状況の変化にどう対応できるかも考慮する必要があります。

また、評価者は、妥当性の評価において、ニーズと優先事項を配慮すべきステークホルダーが誰なのかを、明確に特定する必要があります。これは受益者だけでなく、資金拠出、監督又は実施を担うパートナーや組織も含みます。特に受益者を重視すべきです。介入のオーナーシップは重要であり、優先事項とニーズを明確にする際には、受益者こそが最優先すべきステークホルダーです。介入によっては、国家的及び（該当する場合は）準国家的なニーズ、ニーズに対応するための地域戦略や、これらのニーズに介入がどの程度整合しているのかを考慮するのも適切かもしれません。組織のニーズには支援者のニーズも含められますが、これに限定されません。つまり、介入において明確な支援者ドナーはいない代わりに複数のパートナーが存在する場合（例：貿易政策の場合）など、政策的文脈全体に対して妥当性を評価することになります。

妥当性の理解：分析要素

妥当性の定義は、ニーズ、政策及び優先事項への対応、文脈への配慮と対応、デザインの質、時間の経過への対応という4つの主要要素から構成されます。

ニーズ、政策及び優先事項への対応

おそらく、妥当性の評価において最も重要な要素は、介入が受益者のニーズと優先事項にどの程度対応しているかを確認することです。これにより、介入が対応する課題は何か、それはなぜなのかについて洞察します。洞察では、受益者は介入において中心となるステークホルダーであり、介入の全期間を通じて考慮されるべきです。受益者が必ずしもサービスの直接的享受者とは限りません。介入の種類によっては、結果の連鎖の（はるか）上流に受益者がいる場合もあります。例えば、介入が政府監査当局の能力の向上を目指す場合、こうした能力強化は、公的財政管理を強化し、最終的には保健・教育分野の向上といった持続可能な開発目標の達成に寄与するでしょう。しかし、能力強化に対する評価の受益者であれば、監査当局の職員を一次受益者として着目することになります。（一次的及び二次的な）受益者を明確に設定することは、妥当性を評価するために必要な最初の一步です。

受益者のニーズとそのニーズが満たされたのか分析することは、ニーズへの対応だけではなく、介入のデザインと実施におけるオーナーシップや参加についても明らかにすることが可能です（これが他の評価基準に影響を及ぼすこともあります）。この分析を通じて、介入のデザインに誰が関与し、誰が関与していないか、それが介入のデザインと実施にどのような影響を与えるかを理解することができます。

妥当性の基準は、最大のニーズがある分野に、又はアジェンダ 2030 の表現を借りれば「最も取り残された人々を最優先に支援するために」取り組みを集中すべきことを示唆します。実際、誰が介入に関与し、誰のもとに便益が届けられているかを理解する上で、妥当性は有益です。妥当性は、周縁化された人々が、政策及び介入の優先事項に取り込まれているか、またどの程度取り込まれているかを検討する機会を評価者に提供します。たとえ介入が公的な政策と完全に整合的であったとしても、参加者の実生活上の優先事項とかけ離れているかもしれません。というのは、参加者が公的な優先事項と計画を設定する過程に関与しなかったかもしれないからです。

妥当性の評価では、介入が、関与する機関又はパートナーの優先事項にどう対処しているかも考慮に入れる必要があります。関与側には、介入の資金提供、実施、監督に関わる政府（国家、地域、地方）、市民社会団体、民間組織や国際機関が含まれます。妥当性は、これらの機関の戦略や方針との整合性を検討します。

グローバルなニーズ、政策及び優先事項に対する介入の妥当性を評価するためには、評価では、国際目標全体に対する介入の貢献度（すなわち、より広範な取り組みと比較してみた際の当該介入の相対的重要性）を検討する必要があります。こうした場合には、評価者には、特定の文脈／国家における（潜在的な）結果を、他の文脈／国家における代替手段と比較することを求められることがしばしばあります。こうした国際的な妥当性に関する設問は、介入デザイン策定中に必ずしも検討されません。評価者は、こうした設問に関する分析を行うことで、ステークホルダーが地域の文脈を超える介入の戦略的重要性を理解できるように役立てることができます。

妥当性の定義では、評価者に対して、介入を通して誰のニーズと優先事項を満たすかをめぐって生じ得る対立やトレードオフについても確認することを求めています。参加者や他のステークホルダーの様々な考え方が異なってしまうこともあるため、評価を通じてこうした違いを明らかにし、行なわれた選択が持つ意味を探る必要があるでしょう。一例をあげると、ポリオなどの疾患の世界的撲滅を目指す介入では、特定の国又は地域において国際的な優先事項の実施を決定するために、その疾患が相対的にどの程度の負荷となるのかを検討することになります。場合によっては、その介入の国際的優先事項（市中感染が残る最後の地域でのポリオワクチン接種）が、地域の優先事項（例えば水・衛生分野の課題を優先する受益者）と対立することもあるでしょう。評価者が、妥当性の入念な分析を通してこうした対立を明らかにしておくことが有益でしょう。

文脈への配慮と対応

受益者や他の主要なステークホルダーのニーズは、他から切り離して理解することは不可能で、文脈を通して理解する必要があります。文脈は多面的であり、経済、環境、公平性、社会、文化、政治経済、能力面に関する要因が含まれます。評価者は、どの文脈的な要素が介入に最も大きく関係するかを理解することが求められます。

介入のデザインと実施の両面から、文脈的な妥当性を分析できます。文脈の検討は、「事前」評価と「事後」評価のどちらであるかにも左右されます。例えば、評価者は、介入がデザインされた時点には、文脈がどのように理解され説明されていたかを問うことができます。事後評価の場合は、介入の開始時から終了時までに文脈の変化があったかを検討します。事後評価にはより多くの文脈が含まれ、これを分析に盛り込むことを目指さなければなりません。そうすれば、状況の変化に伴う介入の妥当性の変化を考慮することによって、妥当性の評価基準における時間の経過という分析要素を補強することができます。

過去の文脈も検討できます。例えば、以前に同じような介入が行われたか、ニーズの把握と目標の形成に影響を与え得る、過去の対立、法律、政治問題はあるか、過去の文脈には、介入の妥当性に関する過去の仮説も含み、現在の文脈でもその仮説が持続するのかを検証します。以前に評価が実施された場合、過去の文脈をたどり、以前の評価で得られた教訓が今回の介入に活用されているかどうかを把握する際に、その仮説は役立ちます。

デザインの質

「デザインの質」では、関係する優先事項とニーズに対処するため介入が上手く立案されているか、そして目標が明確に定められているかを確認します。さらに、関係者の優先事項とニーズが、介入の目的、その根底にある「変化の理論」、「行動の理論」、又は実施方法に明確に反映されているかを評価します。これにより、評価者は、介入の全体的な妥当性を損ないかねないプログラムのデザイン上の不備を把握できます。この分析要素は、最初の段階から介入デザインの質に重点を置くことで、介入全体の評価可能性にも影響をもたらします。また、介入が、実施機関にとってどの程度適切なものであるかも考察することができます。例えば、評価者は、介入が技術的、組織的、財政的な実現可能性を想定してデザインされているかどうか確認できます。

時間の経過への適応

評価者は、介入が時間の経過とともに変化する点を考慮に入れるべきです。例えば、紛争の勃発や政治経済的な文脈の変化は、実施に大きな影響を及ぼします。時間の経過に関して妥当性を評価する際には、順応的な管理を行ったかについても分析することが必要です。評価者は、プログラムの開始時と終了時の妥当性だけでなく、介入の期間全体を通して変化にどう対応したかも確認するべきです。これにより、実施環境の変化に伴う介入の妥当性の変化を検討できます。例えば、最も重要なニーズと優先順位に対応するために、どの程度適切な修正を行ったか、修正が長期的なデザインの質に影響を与えたかなどを分析できるでしょう。

ここでもやはり、外部の状況と内部の変更に継続的に対応したかを考慮に入れるべきです（例：資金調達の変更によりプログラムの変更を余儀なくされた場合など）。加えて、プログラムが介入の目的を損なうリスクをどの程度軽減できたか、プログラムが機会を捉えてニーズを一層満たすためにどの程度修正されたか、などリスクと機会についても

確認することができます。修正によって、誰のニーズを優先するかをめぐり対立が生じ、説明責任の問題が生じるかもしれません。この点については、修正によってプログラムの妥当性がどのように変化した／変化しなかった可能性があるかを把握するために十分に検討する必要があります。

他の基準との関係

評価基準は互いに関連しているため、評価において、妥当性を他の評価基準と結びつけることができます。妥当性は、多くの場合において、他の評価基準を達成するための前提要件とみなされます。

- 妥当性の評価は、有効性とインパクトの一環としてニーズが満たされているかどうか把握するための土台になります。有効性の査定を可能にするにはニーズと目標を明確に特定する必要があるため、評価基準としての妥当性は、有効性の前提要件です。
- 妥当性は整合性を補完します。どちらの基準も文脈の分析が必要です。すなわち、妥当性の場合には主な関係者の優先事項やニーズとの合致を把握するために、整合性の場合には他の介入とのつながりを理解するために、文脈の分析が必要になります。妥当性は、介入が文脈にどう対応するかに着目します。整合性は全体像を捉えるものであり、同じ文脈にある他の介入や、それらと評価対象となる介入との相互作用を検討します。妥当性と整合性の併用によって、実施される文脈に介入がどう影響しているか、また介入が文脈からどんな影響を受けているかをより明瞭に把握できるのです。
- 妥当性の分析は、インパクトの評価基準とも関係します。インパクトの評価基準は、影響を受ける人々への価値を含め、介入の最終的な意義を確認するものです。評価者は、介入の全体的な妥当性を把握できるよう、全ての関与者（関与者間に生じる潜在的対立を含む）のニーズ、優先事項及び政策について十分な時間を費やして考察し、その上でインパクトを確認する際に介入の重要性を分析する必要があります。
- 妥当性の要素の多くは、効率性と持続性にとって欠かせない要因です。つまり、妥当性の高い介入はステークホルダーからより大きな支持を得やすく、これにより適時な実施と資源の活用、さらには結果として得られる便益に対するオーナーシップ（及びその持続性）の程度に影響をもたらします。

包摂性の視点の包含

包摂性を理解するには、ジェンダーに基づく力関係を理解し、SDGsの「誰一人取り残さない」決意について考えることが非常に重要です。ジェンダーに基づく力関係及び人種／民族グループをはじめ特定のグループの周縁化は、特定の文脈で妥当性を把握する上で中心となる検討事項です。

力関係と周縁化されたグループに特別な注意を払い、介入デザインに誰がどんな形で関与したか理解することで、評価者は、デザインされた時点の介入の妥当性だけでなく、その介入が時間の経過とともに変化するニーズにどの程度対応したかを理解することができます。

上記の妥当性の定義では、介入により直接または間接に影響を受ける人々の公平性と力関係への一層の配慮を含めて、様々なニーズや優先事項間のトレードオフを考慮することの重要性を強調しています。

特に分野をまたがる視点を適用するとき、妥当性は人権や平等と深く関連します。その視点を通して、ジェンダー、障がい、民族性、性的特性、社会階級など社会的、政治的身分が重なり合って差別と不平等を生み出す仕組みを確認するからです。妥当性の分析のために優先事項を特定する際は、社会において主張が十分に代弁されていない周縁化されたグループ（サービス及び／又は権利へのアクセスが制限されている可能性があるグループ）に配慮し、彼らのニーズと優先事項が正式な文書と政策にどのように反映されているか、または反映されていないかを考える必要があります。加えて、特定のグループが直面する制約を踏まえて、介入に様々なレベルからのアクセスが組み込まれているかを考慮することが重要です。

妥当性の評価に関する課題と対処法

以下の表は、妥当性を評価する際に共通するいくつかの課題、つまり考慮すべきニーズと優先事項の範囲、目的の曖昧さ、文脈の変化—などを明らかにし、評価者と評価管理者によるこれらの課題への対処法を提案するものです。

表 4.1. 妥当性の評価に関する課題

課題	対処法：評価者	対処法：評価管理者
対立する可能性のある優先事項とニーズを抱えているかどうかを検討すべきステークホルダーが国内外に数多く存在する	妥当性を徹底的に分析するには、ニーズへの対応で想定される、または実際のトレードオフを考えるべきニーズと優先順位を検討する際、受益者の妥当性を最も重要な検討事項にする必要がある。この課題に対処し、対立する優先事項を説明し、可能であれば優先事項間のバランスを確保する上で、オーナーシップと参加も有益である。	介入の受益者をわかりやすく簡潔に定義する。「結果の連鎖」が非常に長い場合、妥当性の分析では直接的な受益者に重点を置き、最終的な受益者に関わる設問は、有効性とインパクトの基準に基づく分析に委ねる。 範囲を絞って、誰のニーズと優先順位が最も重要であるかを決定するには、評価の目的を考え、その目的達成を支援する上で妥当性のどの要素が最も有益かを考慮する。デザインの時点で、評価の有用性を慎重に検討すべき。
介入の目的が明瞭で適切であるか、その目的がどのように決定されたかを理解する	妥当性について綿密な査定を行うためには、介入の目的は重要な出発点になる。目的が明確に示されていない場合、これが課題になる場合もある。評価者は、目的が十分に挑戦しがいのあるものか批判的に検討し、意図された受益者ニーズとステークホルダーの能力を考慮し、その目的をどのように特定し検証できるか究明する必要がある。目的が、重要な社会変容を伴う変化に寄与するかどうか理解するために、インパクトの基準の下で目的の適切性を詳しく探ってもよい。 その場合、評価者は、プログラムのステークホルダーに密接に連絡して、プログラムの目的やそれが定義された方法を把握し、不測の場合の対応策が策定されているか、或いはいたかを確認すべきである。これらが、ニーズと文脈の健全な分析にどの程度基づくものかも評価する必要がある。これが妥当性を理解するための重要な出発点になる。	業務指示書を策定する作業の一環として、介入の目的を明確化しよう努めるべき。 業務指示書を通じて、評価者が、ニーズや外部的な文脈に照らして目的の適切性を批判的に検討し、関係者からの反応を考察できるようにすべき。

介入に、変化の理論、ロジックモデル、又は他のデザイン根拠が明確に示されていない	妥当性の査定は、変化の理論、ロジックモデル、又は他のデザイン根拠に支えられている。これらの根拠は、デザインの妥当性を評価するために重要な洞察を可能とする。評価者は、この根拠が存在するか確認し、必要に応じて、評価の出発点として関係者が変化の理論を再構成したり明確に説明したりできるよう、関係者と協力すべきである。	評価管理者は、評価を開始する時点で、プログラム／介入の変化の理論、ロジックモデル、又は他のデザイン根拠を記した文書を、評価者に提供しよう努めるべきである。
文脈は次第に変化したが、介入が修正されたか—、どう修正されたかが不明である	残念ながら、修正が必ず文書にきちんと残されているとは限らない。妥当性を評価するために、まずは、どんな変化が起きて、介入がどのように修正されたか（又はされなかったか）を記述する必要がある。 妥当性に影響しかねない形で、文脈が大きく変化したことが明らかである場合、介入に関する文書（年次又は四半期報告書、モニタリングデータに加えて、可能であれば会議議事録やスタッフ・参加者への通知などの連絡文書）を参照して実際に起きた変化を「再構成」し、重要な決定事項とその決定を促した要因を明らかにすべきである。	評価管理者は、介入の中で行われた修正の筋書きの「再構成」に協力できる主な意思決定者を探し出す手助けができる。決定がきちんと文書に残されていない場合、何が起きたか理解するため面接方法を活用できる。

妥当性の評価の例

この節では、一般財政支援、農家生計向上及び保健セクタープログラムの妥当性の評価から得られた、代表的な事例を紹介します。

ボックス 4.1. 一般財政支援の妥当性の評価

EvalNet は、財政支援評価のための方法的指針において、財政支援デザインの妥当性と適切性、及びパートナー国の政治、経済、社会的な文脈、政府の政策枠組み、開発パートナーの開発援助戦略に関連する様々な投入財に焦点を当てることを提案しています。

妥当性に関わる重要な設問の一つの例は、その国の政治、経済、社会的な文脈、政府の政策枠組み、及び外部の開発パートナーの開発援助戦略を踏まえると、財政支援プログラムのデザインはどの程度適切で妥当なものであったか、です。

例えば、（この指針に引用された）チュニジアにおける財政支援の評価では、財政支援において、技術援助、政策対話及び資金が、チュニジア政府の組織的文脈とその優先事項に相応しい適切な組み合わせで使用されたかを検討しました。第一段階として、この評価では（受益者としての）政府にとってのアウトカムを検討しました。最終受益者であるサービス利用者（高等教育を受ける若者など）にとってのアウトカムは、有効性とインパクトを分析する第二段階に含まれます。

Source: OECD (2011[3]), [Evaluating Budget Support: Methodological Approach](#)

ボックス 4.2. ミャンマー乾燥地帯における「農家生計向上」プロジェクトの評価

この介入と評価は、ミャンマーの中央乾燥地帯における貧困弱者階層社会の生計と食糧安全保障に重点を置いたものであり、妥当性の評価基準を個々に応じて適切な方法で適用した好例を示しています。

妥当性の査定として、ステークホルダーのニーズと優先事項に対する介入の対応性及びデザインから実施の全期間を通じたステークホルダーグループの関与の方法を検討しました。

この評価では、実施中の環境の変化に合わせたプログラムの修正を含む、長期的な妥当性を重視しました。プロジェクト期間全体での妥当性に加えて、中間評価に基づく妥当性査定と評価時点における現在の妥当性との比較を行い、複数の重要な時点における妥当性も検討しました。

評価の結果、ニーズと文脈に対して介入の妥当性が高く、特に環境及び政策面の変化に合わせ巧みに修正されていることが分かりました。この評価は、デザイン段階と実施中の双方の文脈を用いてニーズを検討することで、妥当性を適切な形で適用し、介入とその進化に対する包括的な理解を提供しています。

Source: FAO (2020[4]), [Evaluation of "Improving farmer livelihoods in the dry zone through improved livestock health, productivity and marketing"](#)

ボックス 4.3. ボリビアにおける保健イニシアティブの妥当性の評価

この評価は、2009～2020 年のボリビアにおける保健セクター支援を検討しました。妥当性の評価として、ボリビア政府、特に保健省の正式な承認を得た要請を分析しただけでなく、受益者を含むステークホルダーのより具体的なニーズへの対応状況も確認しました。評価では、次のように結論づけています。

妥当性：このプロジェクトは、保健省から承認を受けた、各種の組織・当局によるイニシアティブ及び要請に対応している。これは、イタリアの協力が、ボリビア政府が必要と表明した内容に効果的に対応していることを意味する。しかし、プロジェクト全体のニーズを特定する適切な工程が存在しないため、結局は妥当性と一貫性の問題が生じている。評価対象のうち最も成功度が高いプロジェクトでは、満たすべき具体的ニーズに関して少なくとも適切な情報を示すことができた。

Source: Eurecna Spa (2020[5]), Bolivia - [Evaluation of health initiatives \(2009-2020\)](#)

整合性

整合性の定義：

整合性：介入が良く適合しているか？

国、セクター、組織に対する当該介入と他の介入との適合性

注釈：他の介入（特に政策）による当該介入に対する支援又は阻害の度合い、あるいはその逆の関係。内部的整合性と外部的整合性がある。**内部的整合性**では、同じ組織／政府が行う当該介入と他の介入との相乗効果と相互連関、及び組織／政府が従う国際規範や標準との一貫性を扱う。**外部的整合性**では、同じ文脈のもとで他の関与者による介入との整合性を考慮する。そこでは、他の介入との補完関係や調和・調整の状況、及び取組の重複を避けたことによって生まれた付加価値の度合いを検討する。

整合性とは何か、なぜ重要なのか？

今日の世界では、異なる政策領域どうしの相乗効果（又はトレードオフ）に一層大きな焦点が当てられ、政府内の横断的な調整への注目も高まっており、整合性により多くの注意を払わなければなりません。紛争や人道支援に関わる場面や、気候変動に関わる危機に対処する場合はなおさらです。

アジェンダ 2030 と SDGs に沿って、この新たな基準は統合的なアプローチをとることを推奨し、相乗効果、政府横断的な協調、及び国際的規範や基準との調和を含めて、整合性を評価するための重要な視点を提供します。整合性の基準は、様々なトレードオフや対立を検討し、政府全体又は異なる機関を横断して政策を実施する際に、取組の重複やアプローチの一貫性の欠如によって全体的な進捗が妨げられかねない状況を特定するためのものです。

この評価基準は、評価者が、排他的介入主義又は組織中心主義の視点に偏ることなく、特定の制度（組織、セクター、主題領域、国）における介入の役割を理解することも促します。外部的整合性は、異なる関与者の政策目的と国際的な開発目標が密接に調和しているか、どのように調和しているかを理解しようとするものですが、他の外部のアクターの関心、影響、権力を考慮しなければ完全な理解は得られません。そのため、介入の整合性を理解するには、幅広い政治経済的な視点が重要になります。

加えて、持続可能な開発の財源が（国際的にも国内的にも）ますます多様化しています。定義にある「国際的規範と標準」の言及は、汚職防止法や人権条約といった、国際法や国際的合意に基づく当該アクター自身のコミットメントと、介入との一貫性の分析を推奨するものです。これは、当該主体が既に約束している合意に該当するものであり、従って内部的整合性の下で扱われます。以前は、このような整合性は十分に分析されませんでした。国際的規範と標準との関連は、地球規模的な優先事項への対応という視点から妥当性の基準でも確認することができ、妥当性は補完的な視点を提供します。

整合性の理解：分析要素

整合性には、内部的整合性と外部的整合性という 2 つの面があります。

内部的整合性

内部的政策の整合性では、次の 2 つの要素を考えます。1 つ目は組織全体のより広範な政策枠組みとの調和、そして 2 つ目は開発介入の実施を担当する他部署が実施する介入や同一の運営体制に影響を与える可能性がある介入等、当該組織が実施する他の介入との調和です²。これらの活動が調和しているか、取組や活動に重複がないか及び介入どうしが相互補完的かどうかを確認する必要があります。

中央政府内（場合によっては、他の行政レベル）では、異なる種類の公共政策間、異なる行政レベル間、また、異なるステークホルダー（国家か非国家か、営利か非営利かを問わず）間において整合性に課題が生じます。整合性を評価する際は、介入がこうした構図のなかでどのような位置づけにあるか、そして、より広範な文脈に適用される政策とどの程度調和しているかについて理解するために、こうした課題を慎重に考慮する必要があります。

例えば、日本外務省の「ODA 評価ガイドライン」（外務省、2019^[6]）は、評価の委託者と実施者が、日本政府全体の外交・開発に関わる政策と活動の整合性を評価することを推奨しています。このガイドラインは、評価者が外交戦略と政府開発援助（ODA）戦略の相互関連性、補完性及び整合性を検討する上で役立つ枠組みと助言を提供し、異なるセクターと国々に対する日本の関与と支援について包括的な分析を行うことを支援します。

政策の整合性は水平的視点から理解することができます。例えば、人道支援－開発－平和の連携では、ある関与者が開発、軍事、安全保障政策に関連する介入を実施する場合もあるため、整合性が強く求められます。環境分野では、水－エネルギー－食糧の連携、又はジェンダー平等－気候変動の連携において整合性確認の必要性があるでしょう。その他の文脈では、貿易などの開発以外の政策領域が介入にどう影響するかも検討できます。

垂直的な視点に立つと、ある組織内の異なるレベルや、一国の政府内で開発資金協力を担う異なる組織（二国間援助機関、開発金融機関 [DFI] 及び多国間援助など）について政策の整合性を把握することができます。介入が地理的範囲を超えて政策目標をどのように支援するか（又は損なうか）を、検討することもできるでしょう。例えば、ある地域の開発介入が、国家の開発目標や国家レベルの介入とどのくらい調和しているか、又はその逆はどうかを検討できます。

外部的整合性

外部的整合性には、次の 2 つの主要検討事項があります。つまり、外部的な合意政策との調和及びある特定の文脈で他の関与者が実施する介入との整合性です。

外部的整合性では、政策的な視点に基づき、介入と SDGs などの外部的な合意政策との調和や、これらの合意政策が介入のデザインや実施にどのように考慮されているかを検討します。この時点で組織の SDGs に対する決意について考えることが重要です。なぜなら、SDG の目標 17 では、二次目標 17.14 として「持続可能な開発のための政策の一貫性を強化する」ことを目指しているからです。これらは、SDGs に対する政策の調和と説明責任の両方が実務においてどのように主流化され実施されているかを確認する上で重要な検討事項となります。

具体的な文脈において介入実施を見る際に、評価者は、他の関与者が実施する介入との整合性を考える必要があります。例えば、広範な関与者によってどのようにサービスを提供しているか、関与者の間に重複や相違はあるかなどの確認をします。整合性では、当該介入が他の介入と関連してどのような形で付加価値を生みだしているか、或いはどのように重複を避けているかを考察します。

外部的整合性の評価は、対象となる特定の介入又は組織に焦点を当てながらも、その介入や組織を、人道支援と持続可能な開発に関わる様々な関与者を含むより広範な文脈の中に位置づける必要があります。これには、国又はセクターレベルの調整システムなどの既存の制度や組織体制の枠内で、かつそれらを用いて、介入がデザインされているかどうかの考察も含まれます。

他の評価基準との関係

整合性は、特に妥当性、有効性及びインパクトと関連しています。

- 妥当性が、直接関与するステークホルダーと受益者のニーズや優先事項という次元で介入を評価するのに対して、整合性は一段高い次元に立ち、評価体制全体における介入の調和を検討します。妥当性と整合性はどちらも、介入が文脈にどう調和しているか検討するものですが、視点が異なります。
- 整合性は、有効性とインパクトの基準で検討される意図されない効果の検討を始める上で、有益な切り口になります。介入がその目的（有効性）を達成できても、同じ文脈にある他の（整合性のない）介入によって、得られた便益がマイナスの影響を受けるかもしれません。
- 同様に、効率性ともつながりがあります。整合性のない介入は重複を生み、資源の無駄遣いを生みます。

包摂性の視点の包含

内部的整合性は、特に人権に関わるコミットメント、規範及び標準に関連することから、包摂性を検討するための有益なレンズ（視点）となります。このレンズによって、評価者は、実施機関の国家レベル又は組織レベルの包摂性と平等に関する規範や標準、また、草の根レベルの先住民団体や障がい者団体といった地域組織の考え方と、当該介入との適合性を考察できます。整合性の確認を通じて、排除をなくし、周縁化されたグループや脆弱なグループに手をさしのべ、ジェンダー不平等を解消することを目指す活動の価値や整合性に関して、有益な洞察を行うことができます。

妥当性、整合性、インパクトの評価には相乗効果があるため、評価者が、インパクトがどの程度包摂的で介入が妥当であったかを考察する際には、整合性の視点からの包摂性について分析することを検討する必要があります。

整合性の評価に関する課題と対処法

以下の表に、対象範囲、権限及びデータの入手可能性など整合性を評価する際の主な課題を挙げ、評価者と評価管理者による各課題への対処法を提案します。

表 4.2. 整合性の評価に関する課題

課題	対処法：評価者	対処法：評価管理者
厳密な査定を行うには、大量の政策文書、データ及び情報提供者を得る必要があるため、多くの文脈や国においては政策の整合性評価が困難である、	利用可能なデータ／情報源に基づく定性的査定適用によってのみ、この課題を解決できる。評価開始時に評価可能性を検討し、限界を率直に認めることも重要である。情報源にアクセスできない場合は注意を促すべきである。 法的環境と組織の方針を理解し、知識を活用して適切な一次データの収集・分析計画を立案する。	評価管理者は、評価を委託する際に、評価可能性、データの制限やアクセス可能性を検討すべき。これらをできる限り予測し、開始段階で評価チームと協議して問題を軽減する必要がある。
データ保護法、又は組織の方針のために、組織がデータ入手を制限している		評価デザインの策定と委託を行う際に、データ制限とデータへのアクセスを確認する。データへのアクセスを支援し、適切なリスク軽減戦略を策定するために、評価チームと継続的に協議すべきである。

関連するすべての政策領域の整合性の評価が、非常に広い範囲に及ばなくてはならない可能性がある	評価管理者と連携して、評価開始時に優先事項と対象範囲を明確に定義する。	検討の範囲と重点政策領域を定義することにより、整合性にどのように優先順位をつけるか決定する。評価の目的や利用法に基づいて、これを決定すべきである。 ある特定のセクターにおける政策の整合性を評価する際は、国家機関がそのセクター内の整合性評価をできる限り調整すべきである。
一部の評価部門の権限が、人道支援・開発援助のみに重点を置いていて、整合性の問題を政府横断的に評価することが制限される	評価デザインの段階で、評価範囲の定義と見直しを行い、評価部門の権限に沿って整合性を探る方法を実現するか、又はその過程での限界を明らかにすることに、特別な注意を払うべきである。	整合性では、開発以外の政策領域が評価に影響を与える可能性があるという認識が求められる。整合性の評価によって、評価者は理論上、介入の直接的な影響を超えた領域を検討する大きな自由を手にする。とはいえ、当該組織の権限を超えた他の政策領域を検討する場合、その実現に数々の課題が伴う。これは、できれば評価前に対処すべき評価可能性に関する検討事項である。組織は、限られた領域の整合性を評価するかどうかを判断すべき。それが優先事項であれば、評価者又は評価管理者は他の部門と協議して、合同評価を実施することも可能であろう。

整合性の評価の例

整合性は OECD-DAC の新しい評価基準ですが、長年にわたり多くの評価で整合性が取り上げられてきました。整合性の基準は、人道支援の評価でも恒常的に使用されています。この節では、ノルウェーにおける開発に対する政策整合性の戦略的評価、フィリピンの自然災害からの復興、モンテネグロにおける国別開発協力全体系の評価で整合性がどのように評価されてきたかを表す代表的な事例を紹介します。

ボックス 4.4. ノルウェーにおける開発のための政策の整合性確保に向けた取組の評価

ノルウェー政府は、介入における開発のための政策の整合性確保にかねてから取り組んできました。政策の整合性は、「開発援助に加えて、貿易、移民、投資、気候変動、安全保障などの開発のさらに広範な局面に取り組むこと」と理解されます。この評価では、異なる関与者の政策や政策領域の間に明らかな矛盾があるという政策的「板挟み」を検討しています。報告書は、「真の板挟みや行うべき選択があるため、実際の矛盾や優先事項に関してより開かれた透明性のある議論」を求めています。

この評価の目的は、開発のための政策の整合性の達成に関わる政策上の板挟みと課題を理解することであり、国別の文脈及び地球的規模での整合性を検討しています。内部的整合性及び外部的整合性に関して、以下の要素を探っています。

- 外務省が主導する介入と、ノルウェーの他の関与者が主導する介入の内部的政策の整合性
- 関与者の政策目標が国際的な開発目標とどう調和しているかを理解するための、外部的整合性

この評価は、ミャンマーの事例を事例研究として、開発政策と他の政策領域の政策上の矛盾を検討しています。ミャンマーが選ばれたのは、ノルウェーの様々な伝統的及び非伝統的な開発主体がこの国に関わっており、各関与者の現地での政策上の優先事項の間に矛盾が生じる可能性があるからです。政策上の矛盾を理解するため、この評価は興味深い方法論的アプローチを採用し、ノルウェーの優先事項と介入全体にわたる政策の不整合を示すサインとして、非政府機関（NGO）、政治団体又は国際機関が課題を提起した分野を検討しています。例えば、2つの矛盾が指摘されました。それは、ミャンマーにおけるノルウェーの産業界の関与と平和構築への合意及び天然資源開発への関与と平和への合意です。この分析は、整合性を評価する際の課題と整合に基づいて活動する際の課題も明らかにしています。実際、外務省は他省庁の政策決定に介入する権限を持たないため、不整合の把握と解決が難しくなっています。地球的規模の視点では、ノルウェーの国際的な合意との外部的不整合も検討しました。その結果、国内政策と国際的な開発の優先事項を調和させるという課題も発見しました。

この評価は、運用上の文脈と組織の優先事項の両方に配慮した、政策的整合性のある介入の実施に伴う様々な課題を示すことで、整合性の基準を適用して入念な考察を行った例を示すものです。

Source: Norad (2018[7]), [Evaluation of Norwegian Efforts to Ensure Policy Coherence for Development](#)

ボックス 4.5. フィリピンにおける自然災害からの復興支援の整合性の評価

この評価は、自然災害に関するカナダの国際援助プログラムの策定と支援国としての役割を、検討しています。介入の整合性を、外部、内部双方の視点から検討しています。この評価では、20.5 百万カナダドルの台風 30 号（ハイエン）被災者支援プログラムと、2013～14 年から 2018～19 年までのカナダとフィリピンの二国間関係を検討しています。

整合性の評価

この評価は、OECD-DAC 評価基準に整合性が追加される前に実施されたにもかかわらず、明確な記載はないものの、整合性にはっきり焦点を当てています。分析の外部的要素と内部的要素の両方に関連づけて、整合性を検討しているのです。

- 外部的整合性 – ドナーとしてのカナダの役割、運営と政治の文脈におけるその行動及び介入のこれらの文脈との調和性を検討しています。フィリピン政府の優先事項や国家的な復興プログラムとの調和性に注目し、この地域で運用されている他のプログラムとの相互補完性を検討しました。
- 内部的整合性 – 評価では、カナダの国際援助プログラム作成を検討し、より広範な国際援助プログラム作成の事例を踏まえてプログラム作成の文脈を明らかにしました。このプログラムとカナダの総合的な政策との調和に特に重点を置き、当該介入が、災害復興対応におけるカナダのアプローチとしてどの程度典型的なものであるかを評価しました。

整合性と妥当性がどれほど密接に交差しているかも示されます。これは、この介入が妥当性の分析要素であるステークホルダーと文脈のニーズや優先事項——と継続的な関連性をもっているからです。とはいえ、この評価はさらに一歩踏み込んで、介入の対応性だけでなく、フィリピンにおけるカナダの全体的な関心に合致し、同国の他の災害復興プログラムとも一貫性のある明確で政策的な整合性のある介入をデザインし実施する上での調和性や検討事項も理解しようとしています。

Source: Global Affairs Canada (2019[8]), [Evaluation of Natural Disaster Reconstruction Assistance in the Philippines, 2013-14 to 2018-19](#)

ボックス 4.6. スロベニアのモンテネグロとの開発協力の整合性の評価

この評価は、スロベニア共和国とモンテネグロの 2013～2016 年の開発協力の実施を対象としています。評価では、「実績の証拠を示し（開発協力の目標をどの程度達成したか）、実績の成否を左右した理由を分析し、今後の政策、プログラム、プロジェクトの策定に向けた提言を示す」ことを目指しました。

この評価は、8 つの評価基準（プログラム管理、実行、スロベニアにとっての付加価値も含まれたもの）のひとつに整合性を含め、次の設問を投げかけています。

- 他の政策との矛盾により、開発目標の実行と達成が妨げられたか、それとも政策は互いを補強しあっているか？

評価では、EU の交渉過程や環境インフラへの投資など、開発協力に関連する他の政策との外部的整合性を検討し、プログラムは他の政策と調和していると結論づけられました。

他のドナーの活動との相互補完性も検討し、モンテネグロに十分な支援者調整システムがないために、相互補完性の実現や重複の防止が妨げられていることが判明しました。

Source: Slovenia's Ministry of Foreign Affairs (2017[9]), [Evaluation of Slovenia's Development Cooperation with Montenegro 2013-2016 period: Final Report](#)

有効性

有効性の定義：

有効性：介入は目的を達成しているか？

介入の目的と結果の達成度、又は達成見込みの度合い（グループによって差異のある結果を含む）

注釈：有効性の分析は、短期的目的や直接的結果の相対的な重要性の考慮を含む。有効性という用語は、介入が妥当で持続性のあるインパクトを効率的、整合的にどの程度達成したか、又は達成することが見込まれるかを示す集合的尺度としても使用される。

有効性とは何か、なぜ重要なのか？

有効性は、介入がその目的をどの程度達成しているか、又は達成したかを理解するための基準です。その介入が、予定された結果を達成したか、それが達成された過程、その過程において決定的であった要因、意図されない効果の有無について考察を行います。有効性は、介入に最も直接的に起因する結果に関する基準であり、より上位の結果やより広範な変化を確認するインパクトと区別することが重要です。

結果の連鎖や因果経路に沿って目的を達成したかを検討するには、介入のねらいや目的の明確な理解が必要です。そこで、有効性の視点を使えば、評価者、プログラム管理者又は担当者達は、明確な目的を策定（又は評価）することができます。同様に、有効性は、評価者が、結果の達成（又は未達成）が介入の実施面又はデザイン面の欠陥に由来するものかどうかを確認する上で役立ちます。

評価者は、有効性の評価基準の下で、意図されない効果も明らかにします。プロジェクト管理者が、デザイン段階でリスクを特定していることが理想的であり、そうすれば評価者は評価開始時にこの分析を活用できます。負の結果（例：対立の力関係の悪化）と正の結果（例：有効性を高める革新）両方について、意図されない効果を探ることが重要です。人権侵害などの深刻な意図されない結果を伴う場合は、評価委託機関は、意図されない効果を確認するための最低基準のガイダンスを評価者に提示することもよいでしょう。

有効性の定義は、評価者と管理者は、意図されるかされないかにかかわらず、異なるグループにおいてどのような結果がそれぞれ現れるのかという重要な問いを投げかけることを求めます。このことは、「誰一人取り残さない」ための SDGs の政策優先事項に沿って、公平性を掘り下げて検討することを明確に目指しています。評価者は、公平性が介入の明白な目的であるかどうかにかかわらず、公平性の問題を検討するよう求められ

ます。こうした分析にはデータが必要で、資源投入が必要な場合もありますが、評価から貴重な分析結果を得るために、資源投入が正当化されることが多いです³。

有効性に関して結論を引き出す際、評価者は、介入の文脈及び評価結果のユーザーにとって最も重要な結果に集中すべきです。「相対的重要性」という用語が示すように、有効性に関して結論を出す際は、意図されない結果も含め、達成された、或いは達成されなかった目的や結果の重要性を比較検討し、評価的判断をすることが必要です。

有効性の理解：分析要素

有効性の定義には、目的の達成、目的と結果の変化する重要性、受益グループによって差異のある結果、アウトカムに影響を与える要因の理解、という重要な概念が含まれます。

目的の達成

有効性を評価する際の最大の焦点は、介入が結果の連鎖の様々な次元において意図された結果（通常は産出（アウトプット）と成果（アウトカム）ですが、場合によってはインパクト）を達成したかどうかを確認することにあります。結果の連鎖は、介入のデザインの一部として認識されますが、管理、モニタリング及び評価においても重要な参照ポイントとなります。

設定された目的や予定された結果が曖昧で不明瞭であったり、介入過程において更新や再構成を経ずに変更されたりした場合は、活動の有効性を評価するのは非常に困難です。介入管理者は、最低限、目標を変更した理由と新たな目標が何であるかを説明しなくてはなりません。この説明がなければ、評価者は、介入に関する文書を参照するかステークホルダーに面接して、時を経て変化した介入の論理を再構成しなくてはなりません。その後、評価者は、再構成した論理に基づいて、新たな目的がどの程度妥当で効果的に達成されたかを判断できます。

ステークホルダーからの反応や生み出される結果、文脈の変化などに対応して何度か変更が行われる順応性の高いプログラムでも、有効性の評価は重要です。順応型のプログラムでは、介入のデザインと実施段階においてどんどん様々な変化が起こります。こうした場合、評価者は、介入の背景となる評価体系全体に照らして変化の理論について考察と見直しを行い、どのような変更が、どうして行われたかを示すあらゆる記録を考慮に入れることが重要です。有効性の評価では、現在の目的を反映します。介入の実施や目的を変更する論理とニーズ（改訂された変化の理論や結果の連鎖の枠組みの中で見出されることが多い）を見直すことは、有効性の評価において有益です。

有効性の評価では、介入の実施において対象グループや環境の目に見える変化を確認した上で、その変化が異なる次元でどのような因果関係にあるのかを確認することが必要です。すなわち、その変化は当該介入によって引き起こされた、或いは他の環境要因や他の介入ではなくまさに当該介入がその変化に貢献したことを示すことが求められます。結果がどのように生み出されたか、結果が達成された、達成が不十分だった、又は未達

成だった場合の理由は何かを評価者が導き出せるような手法をデザインする必要があります。

有効性の評価では、介入の結果の検討も行います。OECD-DAC は、結果を、介入のアウトプット、アウトカム、インパクト（意図された／意図されない、正又は負）として定義しています（OECD, 2002_[10]）。したがって、有効性の評価は、介入の結果として生じた正と負両方の意図されない効果の分析も含みます。介入の実施は、常に社会、経済又は環境面において意図されない効果を引き起こす可能性があり、意図されないが予測可能であった効果をもたらすこともあります。したがって、評価では、介入の目的として示されたもの以外の効果を慎重に検討すべきです。これが意図されない（予測可能か否かを問わず）効果から生じうる便益やリスクの検討に発展する可能性もあります。介入が、同じ文脈にある政府の開発目標や目的、又は他の関連する開発目標や目的の実現にどの程度寄与したかも有効性に含まれます。ただし、これらが寄与する潜在性そのものは、妥当性の基準で検討されます。

達成されたものの相対的重要性の検討

有効性を評価する際、評価者は、介入の様々な目的や結果の達成（又は未達成）について検討します。目的のすべてではないが一部は達成された場合、その目的の相対的重要性を確認して有効性に関する結論を出す必要があります。これには妥当性の分析を活用することができます。妥当性の分析では、様々なステークホルダーの優先事項やニーズの間に想定される違いを扱うからです。相対的重要性を考慮することで、評価者は、介入がある意味では効果的だったが他の意味ではそうではなかった、又は一部のステークホルダーにとっては効果的だったが、他のステークホルダーにはそうではなかったと結論づけることができます。

差異のある結果

評価者は、受益者が個人であるか組織であるかに関わらず、受益者グループ間における結果の包摂性と公平性を考える必要があります。差異のある結果について理解するためには、介入のデザインと実施に包摂的なアプローチがどの程度担保されたかを検討します。例えば、評価では、介入の目的が設定された工程を確認できます。ニーズ分析やステークホルダー（主な対象グループを含む）との協議過程に基づいて目的が策定されたかどうかを確認します。この評価工程を通じて、不平等の削減などのより長期的な変化への貢献を含めて、介入が対象グループや受益者のために結果を生み出す機会を見逃していないかについて考察できます。評価者は、変化や変容の中には微妙なものであったり、或いは長期にわたるために数値化が困難な場合もあるという事実を考慮に入れて、意図された結果だけでなく、意図されない、又は想定外の結果も確認することができます。

影響を及ぼす要因

評価者にとって、介入の結果に影響を及ぼす要因を検討することは重要です。なぜなら、一つの介入によって目標を達成できた理由や達成できなかった理由そのものを理解するのに役立つからです。理由を知ることによって、受益者は改善すべき領域を特定しやすくなります。要因は、介入によって内部的なものもあれば外部的なものもあります。これらの要因は、管理面、人材面、財政面、規制面、実施における修正、或いは計画からの逸脱に関連するものかもしれません。実施の質（及び実施要綱の遵守）は有効性を高める要因になることが多く、有効性、効率性、インパクト及び持続性の評価を行う前に説明されるべきです。

外部要因として評価者は、介入の文脈から生じる正と負の効果を検討する必要があります。こうした効果は結果の達成又は未達成に関与します。文脈の変化に応じて介入が適応できたかの確認も含まれ、結果が適時であったかも検討できます（例：結果の実現につなげるための対象グループ又は環境支援の段階的实施）。

他の評価基準との関係

有効性は、他の評価基準、特に妥当性やインパクトと関連性があります。

- 妥当性の基準では、介入の目的を明らかにします。その目的に向けた進捗を、有効性の基準で見極めます。もちろん、妥当性はないが効果的に実施されるということもあり得ます。こうした乖離がある場合、総合的な結論を出すに当たって評価者は判断を行う必要があります。ある評価基準で得られた評価結果を、別の基準を優先したいがために無視することはできないからです。
- 有効性とインパクトは、結果の連鎖の異なる次元に焦点を当てた相互補完的な評価基準です。有効性は、介入の目的に対する結果の達成度、すなわちアウトプットやアウトカムの次元を確認するのに対して、インパクトは上位の結果を対象に、結果の連鎖に設定された上位の結果とは何なのか、その結果に何が寄与しているかを検討します。一般的に、介入管理者と評価者は、異なる結果の次元（インプット、アウトプット、アウトカム、インパクト）を明確に区別し、それぞれの評価基準の下でどの側面を評価するかをはっきりさせることが重要です。

包摂性の視点の包含

有効性の定義は、異なるグループ間の公平性を詳しく検討することを求めます。評価者は、介入において異なる受益者グループに対する包摂性がどのように配慮されたか、公平性、差別禁止、説明責任などの重要な原則が、介入のデザインから結果までのすべての段階にどのように取り入れられているかを評価する必要があります。「誰一人取り残さない」という理念に基づき、介入が最も周縁化された人々のニーズにどの程度合致しているかに特に注意を払うべきです。幅広い受益者グループと関連づけて結果の達成度と配分を確認し、そこに差異があれば説明することが重要です。

さらに、評価者は、結果が、不平等の解決にどのように、そしてなぜ貢献しているかを考える必要があります。この評価基準に基づき、評価者には、特定の活動が特定のグループの福祉にどう影響しているか、それらの活動が参加者に能力向上の機会を与えているかどうかを検討することが求められます。

有効性の評価に関する課題と対処法

以下の表では、データ、目的の設定、及び結果の要因に関連する課題を明らかにし、評価者と評価管理者両方のこれらの課題への対処法が提案されています。

表 4.3. 有効性の評価に関する課題

課題	対処法：評価者	対処法：評価管理者
参考になる適切なデータ、又は細分類されたデータがない	データの入手可能性（例：ベースライン評価や中間評価の欠如）によって、目的の達成度の測定が妨げられることがある。データがないと因果関係の特定がさらに困難になり、誰が何に貢献しているか判断しにくいためである。ベースラインデータは再構築できるが、適切に収集されたベースラインデータに比べると概して精度と信頼性に劣る。代替アプローチとして、評価者は、1) 入手可能なデータの三次元測定（トライアングレーション）と検証、2) 代替的な手法の活用、3) 寄与度をめぐる筋書きの構築、を採用できる。	介入期間を通して適切で参考になる細分類されたデータを収集できるよう、介入デザイン策定時に、できればデータの入手可能性について計画を立てるべきである。評価管理者は、入手可能なデータと考えられる代替データ源を明確に把握し、データがない場合は、評価者と一緒に代替手段を考えるべきである。 同一領域の他の評価や調査報告書の分析結果を集める。SDG の 2 次目標を含め、関連する指標の開発に関する過去の文献から学び、これを適用する方法を審査し検討する。
表現が不十分、曖昧、または測定が難しい目的／目標の記述に基づいて作業している	評価開始段階で、介入の目的が検証可能で現実的で、開発介入のための現在の国際標準と調和するものであることを確認する。目的の達成度を測る指標も、SMART（Specific（具体的）、Measurable（測定可能）、Attainable（達成可能）、Relevant（妥当）、Timely（適時）などの一般的に使われている基準に沿って検証すべきである。できればこの評価工程は、評価者と評価管理者双方が参加する形で実施されるべき。	評価者と連携して、一般的に使われている基準（例：SMART 指標）に従って目的の達成度を測るための明確な基準を定める。ジェンダーへの配慮が必要な評価では、SMART 及び RACER 指標から、CREAM（Clear（明確）、Relevant（妥当）、Economic（経済的）、Adequate（適切）、Monitorable（モニタリング可能））及び SPICED（Subjective（主観的）、Participatory（参加型）、interpreted（解釈）、Cross-checked and Compared（照合・比較）、Empowering（能力向上につながる）、Diverse and Disaggregated（多様化・細分類化））指標への移行を検討する。 その時点で定義されたプログラム理論／変化の理論の枠内で、目標と結果の定義及び具体性を改善する。変化の理論がない場合、介入の内部的論理を文脈を踏まえて検討するために、妥当な理論の再構築を試みる。

		関係者の関与を得て、参加型の取り組みに基づきプロジェクトの位置づけを見直すことで、目的や目標の記述の整合性を事後的に高める。
結果の帰属	因果関係の帰属や貢献度の判断に関する問題は、特に長期的に亘る有効性を評価する際の課題である。貢献度をめぐる筋書きの作成と、データの流れの適切な三次元測定によって、結果を評価対象となる介入に帰属させられるかどうか、評価者が判断するのを支えられる。	介入／組織について詳しい知識を持ち、介入の実施や開発結果の達成においてパートナーになる可能性があるステークホルダーの意見、評価、判断を考慮に入れる。

有効性の評価の例

この節では、選挙支援及びカーボベルデにおける国別プログラムの評価で有効性がどのように評価されてきたかを示す代表的事例を紹介します。

ボックス 4.7. オーストラリアの選挙支援の有効性と包摂性の評価

オーストラリアは 1970 年代以降、正式で深く根付き、かつ機能する民主主義への移行とその定着を支援するために、様々な途上国に選挙支援を行ってきました。この評価は、オーストラリアが 2006～16 年にアジア太平洋地域の 8 カ国に行った国政選挙支援を対象とします。経時的な目的の変化に特に配慮して有効性の評価基準を適用するとともに、包摂性、ジェンダー、脆弱なグループと関連づけた有効性の検討も行った優れた例となっています。

この評価は、オーストラリアの選挙支援の全体的な目的を検討したのち、選挙管理制度の強化、参加の推進、選挙組織の支援という 3 つの重要な領域における有効性と包摂性に注目しました。

評価では、選挙支援に関するオーストラリアの目的が、実施期間中に变化したことに触れ、分析のなかで目的の変化について説明しています。加えて、有効で包摂的な選挙支援の提供に伴うリスクも入念に検討し、現在及び今後の介入の有効性を損ないかねない外的要因を説明しています。

目的の検討においては、特に結果の包摂性を重視しました。これは、有効性を評価する際に、様々なグループにおける差異のある結果をどのように分析できるかを示す上で有益です。ジェンダーレンズを通して結果の達成度を確認し、定性的な調査結果を用いた評価に基づき、社会的規範と法的・制度的な障壁のせいで女性の有権者投票率が過度に低いことを明らかにしました。

この評価は、弱者集団の包摂性、具体的には障がい者の包摂性についても説明し、物理的なアクセス改善に向けた活動にもかかわらず、文化的障壁が原因で異なるグループ間の結果の公平性が限定的であることを確認しました。これらの調査結果に基づき、この評価は、結果の公平性の向上に重点を置いて今後の介入に向けた提言を行うことができました。そしてこのことが、介入の結果と開発目標のさらなる広範な達成を支えています。

Source: Arghiros et al. (2017^[11]), [Making it Count: Lessons from Australian Electoral Assistance 2006-16](#)

ボックス 4.8. カーボベルデにおけるアフリカ開発銀行の国別戦略・プログラムの有効性の評価

アフリカ開発銀行（AfDB）によるカーボベルデの電力セクター支援を対象としたこの評価は、ジェンダーと社会的包摂性に関する目的に特に注目して、デザインと実施の妥当性を含めて、有効性の評価基準を実際に適用した例を示すものです。AfDB は、援助に対し広範な目的をいくつか掲げていました。例えば、インフラ開発と良い統治改革を通じたカーボベルデの包摂的な成長の推進や、「グリーン成長」すなわち温暖化ガス低排出で気候変動に強靱な開発への移行支援という戦略的な目的が含まれていました。

介入の有効性とその目的の達成の有無を評価するために、この評価では、介入によって次の予定された開発アウトカムが達成されたかどうかを確認しました。

- 国民への電力供給増大
- 再生可能エネルギーの割合拡大
- 耕地と水供給の増大
- 良い統治の向上

その過程で、デザインの堅牢性や実施の進捗といった内的要因だけでなく、介入に影響を及ぼす外的要因も分析しました。結果の公平性を評価する際は、ジェンダー平等と社会的な包摂性にも特別な注意が払われました。

インフラ開発と良い統治改革を通じた包摂的な成長支援に関しては、確実な進歩が認められたものの、「グリーン成長」すなわち温暖化ガス低排出で気候変動に強靱な開発への移行支援という戦略的な目的は、利用できる資源と手段にそぐわないものでした。この評価から、有効性の検討によって、計画された介入の結果を達成できたかどうかや、結果の実現に影響を及ぼす要因を明らかにすることが分かりました。

Source: IDEV (2018^[12]), [Cabo Verde: Evaluation of the Bank's Country Strategy and Program 2008–2017: Summary Report](#)

効率性の定義：

効率性：資源がいかにうまく使われたか？

経済的かつタイムリーな方法で結果を生む、又は生むような介入実施の度合い

注：「経済的」とは、同じ文脈で実現可能な代替手段との比較で、インプット（資金、専門技術、天然資源、時間など）からアウトプット、アウトカム、インパクトへの変換で費用対効果が最大限となる方法を指す。「適時」とは、計画期間内、又は変化する文脈に合わせるべく合理的に調整した期間内での実施を指す。これは、運営の効率性の評価を含むこともある（介入がいかに上手く管理されたか）。

効率性とは何か、なぜ重要なのか？

この評価基準は、介入のための資源が結果によって正当化できるかを確認する機会であり、これは実務的及び政治的に非常に重要です。政府、市民社会、受益者をはじめ多くのステークホルダーグループにとって、効率性は重要な基準です。限られた資源をよりよく使用するとは、例えば莫大な資源ニーズがある SDGs 達成への進捗において、開発協力によってさらに多くを達成できる可能性を意味します。効率性は、納税者への説明責任を負う政府にとって特に関心が高い基準です。納税者は様々な政策やプログラムの投入額に注目し、特に国際開発協力に関する決定に疑問を抱きます。

効率性は、運営面でも重要です。多くの介入では実施可能性と実施に関わる問題が発生し、特に資源の使用方法是重要な問題となります。効率性の評価は、マネジャーが決定の経緯やリスク管理の方法について説明責任を負いながら、プログラムの適切な実施を確保するための管理改善意欲の向上に役立ちます。

次のとおり、いくつかの重要な前提と留意点があります。

- 資源は最も広義に理解され、すべての経済的費用（人材、環境、財務、時間）を含めるべきです。資源は、プログラム予算や支出額と同一ではありません。
- 結果についても、アウトプット、アウトカム及びインパクトという結果の連鎖全体を対象として、包括的な意味で理解すべきです。評価の様式にもよりますが、一部の組織は効率性をアウトプットのみと関連づけています。しかし、本書では、困難な場合も多いもののインパクトなどの（結果の連鎖における）上位の効果とも関連づけた効率性の評価も促すために、この基準を定義して概念化しています。
- 評価可能性：有効性、インパクト、整合性及び持続性を評価できるかどうか、効率性についてどう評価できるかに影響します。
- 効率性とは、所定の資源の範囲内で同じような結果を実現できる、実施可能な選択肢の中から選択を行うことです。費用対効果の比較を行う前に、質と結果の観点から、純粋に実施可能で比較可能な選択肢を明らかにすることが必要です。

こうした理由から、効率性の分析は文脈の分析にしっかりと基づいて行う必要があります。例えるなら、意図された受益者に便益をもたらすには費用面では割高だけれども、開発のインパクトという意味では、より重要で正当化できることもあるかもしれません。

効率性の理解：分析要素

効率性の評価では、経済的な効率性、運営面の効率性、適時的であるか、という主要な項目について確認します。

経済的な効率性

これは、効率性を分析する上で最も重要な要素です。ここでは経済的な効率性とは、費用効率が最も高い方法でインプットが結果に変換され無駄がないことを指して使用されます。経済的な効率性では、アウトプット、アウトカム、インパクトという結果の連鎖のすべての次元において、結果の効率性を評価します。これには、デザイン段階と実施中にどの程度適切な選択がなされ、トレードオフへの対処がなされたかを評価することも必要です。この選択には、市場の状況に応じてインプットを購入できるという選択肢に加えて、複数の対象グループや複数の期間に対する資源の配分方法も対象とします。

運営面の効率性

運営面の効率性も、確認すべき重要な要素です。実施中に資源をいかに上手く使用しているかを点検します。運営面の効率性を考えるのに有益な設問としては、人材と財源は計画どおりに適切かつ十分に活用されたか（又は資源の配分間違い、予算の未消化や超過はあったか）、ニーズの変化に応じて資源は転用されたか、リスクは管理されたか、新たな情報に応じて効率性の向上を促す決定が下されたか、物流と調達に関する判断は最適であったか、などが挙げられます。

タイムリーであるか

タイムリーであることについては、経済面と運営面両方の効率性と密接に関連し、まず意図された期間内に結果が達成されたか、どの程度達成されたかを確認します。そもそもスケジュール自体が現実的又は適切なものだったかを確認する機会でもあります。加えて、多くの介入では外的要因やプログラムの変更が起こりがちである点を踏まえて、介入の期間中にスケジュールは適切に調整されたかを確認します。評価者は、状況の変化に伴って介入をどのように管理して障壁を乗り越え遅延を防ぐための努力が行われたかを確認することが必要です。

他の評価基準との関係

すでに指摘したように異なる評価基準の間にはつながりがあり、基準間には厳格な境界線はなく、介入を検討する際にどちらかを選ぶ視点とみなされるべきです。他の評価基準との間には、相互に次のような関係があります。

- 妥当性と効率性：運営面の妥当性の重要な要素は、介入デザインが文脈によく見合ったもので、実施可能性と能力の検討を可能としているかどうかです。つまり、デザインは実施可能であるか、実際に実施できたかも、効率性に直接的な影響を与えます。従って、こうした特定の側面では、評価者が妥当性と効率性両方を合わせて検討することになる可能性もあります。
- 効率性と結果：効率性には、使われた資源のどの程度が結果に変換されたかの評価が含まれるため、結果のあらゆる側面（すなわち、有効性、インパクト、持続性を評価する際の設問）を検討すべきです。運営面の効率性は、有効性やインパクトと密接に関係しています。多くの場合、介入がどの程度上手く機能しているか知るには、有効性と効率性を同時に検討します。例えば、隘路とその対処法を明らかにする、又は必要とされる場所に資源が配分されているかを確認するといった場合には、これが特に当てはまります。

包摂性の視点の包含

効率性の評価基準という視点を通して、評価者は、包摂性が介入の管理にどのように取り入れられ、どのように理解されているか、資源の使用に、異なる人々の間で差異のある経験や結果がどの程度反映されているかを把握することができます。成果の達成にかかる費用は、受益者によって異なることが多く、「最も取り残されている」人々は手を差し伸べるのが最も困難で、最も費用がかかります。したがって、効率性の分析では、当該文脈における不平等と力関係を明確に理解するだけでなく、根底にある不平等の解決に向けた社会変容を伴う変化へのニーズにその介入がどの程度適合しているかを把握する必要があります。効率性の分析は、「誰一人乗り残さない」というアジェンダ（及び周縁化されたグループに対して社会変容を伴う変化の達成というアジェンダ 2030 の目標）への合意が、管理、意思決定及び資源配分の全体において有意義かつ効果的に運用されているかどうかを確認するために重要な領域です。

ここでは、介入の対象となる様々なグループに資源をどのように分配しているか、どうしてそのように配分しているのか、及び資源配分が周縁化されたグループのニーズと彼らへの関与にどの程度基づいているかを分析します。評価者は、合理的な費用で包摂性に留意した公平な結果を達成しているか、その「合理的な費用」がどう定義され決定されているか、異なる受益者グループ間でその費用にどのような違いがあるかを検討します。例えば、介入が特定グループへの働きかけに取り組んでいる場合、それを成功させるために十分な資源が配分され正当化されているかも確認します。

政策立案の方法を決定する際に誰の意見に耳を傾けているか、資金をどのように支出し、誰が工程管理と統括を行っているかを理解することが重要な検討事項です。介入の論理

と計画に、不平等な構造や力関係を変えることが含まれる場合は、評価を通して、その介入がどの程度成功を収めたか、介入によって既存の不平等な構造や力関係が、意図しないうちに強化されていないかを確認する必要があります。

実施者が、包摂性に配慮した開発の活動と目的や資源について妥当な決定を行うために、介入において適切に細かく分類されたモニタリングデータが収集されているかどうかを確認することも重要です。

効率性の評価に関する課題と対処法

効率性の評価基準の適切な適用法は、介入自体の特性に完全に左右されるものであり、プロジェクト、プログラム、国家計画や政策によって異なります。以下に紹介するオランダ外務省の例は、水分野における政策次元での効率性の基準の適用を示すものです。この例では、効率性はもっぱら、一連の複雑な関係性における計画策定とパートナーシップの調整と実務的な局面で理解されています（ボックス 4.11）。

表 4.4 は、効率性を評価する際の主な課題を明らかにし、評価者と評価管理者両方のこれらの課題への対処法を提案しています。

基本的に決定すべきは、効率性を評価するために、費用便益分析、収益率、費用対効果分析、ベンチマーク比較などの従来型の経済学的方法や関連技法を使用するかどうか、そしてどのように使用するかです⁴。この決定は、評価の目的、介入と意図される結果、実施可能性、入手可能なデータ／資源、評価結果の利用者に左右されます。

様々な分析技法の有益性は、当該組織においてデザイン／承認段階でどんなアプローチが使用されたかにも左右されます。このアプローチが、各種の分析の実施に必要とされる情報の入手可能性に大きく影響するからです。多国間開発銀行の内部、一部の公共セクター資本投資プログラムでは、経済、社会、環境面の事前評価に対して非常に明確なガイダンスが提供されています。さらに、こうしたプロジェクトやプログラムにおけるジェンダー分析についても明確なガイダンスの提供が拡充されています。プロジェクトの審査又は承認時（一般的には介入実施前の事前評価）と評価の時点とで、効率性の評価に同じ技法を使用するのは理にかなったことです。組織が採用する政策面の規則とガイダンスも、評価者が利用できるデータと指標を決定する一因です（例：経済性審査を実施している場合、その審査で収益率を試算したか、別の選択肢を検討したか）。

表 4.4. 効率性の評価に関する課題

課題	対処法：評価者	対処法：評価管理者
適切な比較対象を見つける	評価者は通常、完璧な比較対象を見つけるのに苦心するため、入手できるものを使用する。とはいえ、比較の際は慎重さが求められ、感度分析を行うべきである。例えば、紛争の被害を受	評価管理者が、評価者に有益な背景情報を提供できそうなデータや分析を入手できる場合、その情報源を、評価者に提供する文書に含めるようにする。

	けた脆弱な環境における効率性は、同程度に困難な環境とのみ比較すべきである。	
効率性に関する様々な概念や、関連する手法と技法のどれを利用するか	効率性に関する経済的概念を扱った文献は大量にあり、目下の評価に対してどの手法と技法が最も適切かを決める上で、評価者はこれらの文献を参照できる。BMZ の作業文書に、優れた要約が示されている (Palenberg, 2011 ^[13])。	
効率性の分析に関する十分な専門技術が不足している	効率性を分析するには何が必要か、それを評価工程にどう取り入れるかを、確実に理解する (アプローチと専門技術いずれの点でも)。	評価の業務指示書に、効率性の分析に関して必要とされる専門技術を指定しておく。
データと測定	便益、結果、費用に関して利用可能なデータの入手が難しいこともある。例えば、インパクトの評価は、介入によるアウトカムの変化の概算値を定量的に示せるが、総じて費用の概算は省かれている。ましてや費用総額は推計されていない。	評価者が、効率性の分析に必要なデータに確実にアクセスできるようにする (例: 予算、人的資源、時間などの観点から見た、特定の介入又は行動の費用/便益に関するデータ)。
時間への配慮	経済性審査と効率性の評価のいずれでも、資源が使用される時期と成果が実現される時期の選択が、重要な要素である。結果と費用総額 (環境費用など) は長年ののちに生じることもあるため、プログラム期間又は評価時点における実際の効率性が、全体像を反映していない場合もある。	効率性は、管理者が優先事項の変化にどう対応したか、リスクをどう想定又は対応したか、費用超過にどう対処したかなど、プログラム又は介入実施中の一連の選択によって影響される。

効率性の評価の例

この節では、農村電化、農業投入財補助金及び水管理に関する重要開発協力の事例を対象に、効率性がどのように評価されてきたかを示す代表的な事例を紹介します。

ボックス 4.9. 世界銀行による農村電化支援の効率性の評価

この研究は、120 件のプロジェクトの総体事業報告、10 件の国別事例研究、既存の調査データ（12 カ国のエネルギー及び世帯調査）の分析を含む、1980～2006 年に幅広い国で実施された世界銀行の農村電化支援事業に関するものです。

長年にわたってこのセクターに携わった大規模多国間開発銀行が持つ豊富なデータセットを活かして、費用便益の比較を含めて、多様な角度から効率性を検討しました。インフラ案件の検証済の事業完了報告書には通常、収益率が含まれているため、この総体事業報告は、世銀の全開発投融资事業に含まれるプロジェクトの収益率推計値を利用しています。さらに、消費者余剰の推計値を用いて福祉へのインパクトも見積もっています。これにより評価者は、介入の便益と関連費用を評価できます。この評価は、インパクトや公平性の要素（貧困層への到達度）を分析し、また、持続性、経済効率と分配効率の間、総合的なインパクトと対象グループ別のインパクトの間のトレードオフを理解するための一つのアプローチを提示しています。

Source: IEG (2008[14]), [The Welfare Impact of Rural Electrification: A Reassessment of the Costs and Benefits](#)

ボックス 4.10. サハラ以南アフリカにおける農業投入財補助金の効率性の評価

この研究は、最低限の費用で最大限の効果を目指すいわゆる「スマート補助金」をはじめ、ガーナ、マラウイ、タンザニア、ザンビアにおける農業投入財補助金を評価しました。

効率性の評価

この研究では、主題の性格から、経済的な効率性（ここでは、補助金が様々な点で費用に見合うものかという意味で使用）を、中心的な課題としています。

農業生産性の変化に関する推計に基づき、もし収益が相応の水準であれば、こうした生産性の変化は担った費用に見合うものか、また事業目標の選定を含めて、どんな要因が一定の役割を果たしたかを検討しています。例えばザンビアに関連して、次のように指摘しています。

農業投入財の拡大は、トウモロコシ生産量に大きな効果を与えているようだが、産出量増大に非常に高いコストを伴っている。最も正確な推計によると、現在の形でZFSP（ザンビア食糧安全保障プログラム）を通してザンビア国内の生産量を増やすより、都心部で消費するトウモロコシを輸入する方が安上がりと考えられる。公平上の配慮や投入財セクターの長期的な持続可能な開発が、外見上の非効率性を上回るのであれば、それでも投入財補助金を正当化できるだろう。だが、そのような効果は疑わしい。

以上のように、この研究は効率性を検討しながら、同時に持続性に関する一部の要素（例：どんな出口戦略を使用しているか）や補助金のインパクト／有効性を含む他の基準だけでなく、全体的なインパクトについても検討しています。

Source: Baltzer and Hansen (2011[15]), [Agricultural input subsidies in Sub-Saharan Africa: Evaluation Study](#)

ボックス 4.11. 政策次元での効率性の評価：水管理の改善に向けたオランダの開発援助政策

この報告は、2006～2015 年のオランダ外務省による総合的な水資源管理の改善に向けた開発援助政策に関するものです。これはオランダにとって重要性が次第に増している政策優先事項であり、多国間組織、各国政府、NGO 及び民間セクターとの幅広いパートナーシップを対象としました。評価で用いられた 2 つの主な評価基準は、政策の有効性と効率性でした。

効率性の評価

効率性に関して使用された評価に関する具体的な設問は、次のように設定されました。

- 水の利用者グループ、技術的な質、及び物的インフラの保守に関して、意図された主な成果の達成を支援するために、政策は適切に組織され運営されたか？

評価者は、この設問に答えるに当たり、主な成果の達成状況と計画の間に（費用と期限という面で）どの程度の乖離があったか、それはなぜかを確認しました。品質、アウトカム、持続性を含めて、結果の連鎖の各次元で結果を検討しました。例えば、評価では、地域固有の課題の社会的・技術的な複雑性から、政治面と制度面の要因がさらに悪化し、実施の遅れと費用超過を招いたため、参加型の灌漑管理計画に時に遅れが生じていることが明らかになりました。

Source: IOB (2017[16]), [Tackling major water challenges: Policy review of Dutch development aid policy for improved water management](#)

インパクトの定義：

インパクト：介入はどのような違いをもたらすか？

介入により生じた又は生じたと予期される、重要な正又は負の、意図された又は意図されない（結果の連鎖における）上位の効果の度合い

注釈：インパクトは、介入の根本的な意義と潜在的に変容をもたらす効果を扱う。インパクトは、有効性の基準で既に捕捉されているものよりも、長期的又は広範囲で、介入の社会的、環境的、経済的な効果の特定を究明する。これに伴う評価基準は、直接の結果のみならず、介入による間接的、二次的、潜在的な結果の把握を試みる。そのためには、社会組織や規範の全体的かつ持続的な変化及び人々の幸福、人権、ジェンダー平等、環境への潜在的な影響の調査を必要とする。

インパクトとは何か、なぜ重要なのか？

インパクトの評価基準は、「それで、介入で何が変わったのか？（so what?）」という広範な疑問について思索することを求めます。この基準をもとに、介入による最終的な開発の効果が検討され、評価者は、介入が人々に本当に意義のある変化をもたらしたかどうかを判断します。これは、さらに広い視点を取り込み、全体像を捉える機会となります。現実的には、介入の日常的な側面は着目し易く、介入に携わる人々の視点のみを追求しがちです。インパクトの評価基準は、評価者がさらに踏み込んだ考察を行い、どんな変化が誰のために達成されたかを理解するよう促します。こうした重要性は、スウェーデン国際開発協力庁（Sida）の評価手引き書（Molund and Schill, 2004^[17]）において強調されています。

「インパクトの評価基準は、開発介入者の意図に基づいた過度に偏狭な先入観や、それに付随して起こる、対象グループや他の主要なステークホルダーの視点を軽視したり、放置したりする、などの対応を修正する重要な手段を提供する」

インパクトという用語はよく使われますが、この用語ははっきりと理解されていないことが往々にしてあり、これが、ステークホルダーによる評価の理解に影響を与える可能性があることに留意が必要です。まず、インパクトは、政治的な文脈において、最も広義で漠然とした「結果」を意味し、本文書で定義する有効性とインパクトの両方が含まれるほか、実績に関するその他の要素も含まれます。第二に、近年では、介入と観察された効果の間の統計的に重要な因果関係を立証する特定の手法である「インパクト評価」という用語に含まれるインパクトとしばしば混同されます⁵。「インパクト評価」

において使われるインパクトは、アウトプットを含めた結果の連鎖上のあらゆる次元の結果を指すことがあり、多くの場合「期待される効果」を意味します。こうした理由から、初期の段階でステークホルダーはこのインパクトという用語をどう理解しているかを確認し、インパクトが、評価の文脈において（結果の連鎖において）上位にある効果を考察するための基準として、どのように使用されるかを説明することが重要です。

インパクトの評価基準が取り上げるであろう設問には、以下が含まれます。

- 介入が、意図された受益者の生活に著しい変化をもたらしたか？
- どのようにして介入が、上位の効果を生んだか（規範や社会組織の変化など）？
- 最も恵まれない弱者層を含む、意図されたすべての対象グループが、介入から等しく便益を受けたか？
- 介入は社会変容的で、ジェンダーに関する規範を含む規範や社会組織に、意図されたか否かを問わず、永続的な変化をもたらしたか？
- 介入は、「拡張性」や「再現可能性」がある結果を含めて、その他の変化につながっているか？
- 介入は、社会のより良い変化にどのように貢献すると予想されるか？

インパクトの理解：分析要素

インパクトの定義には、（結果の連鎖における）上位の効果、意義、（グループによって）差異のあるインパクト、意図されない効果、社会変容を伴う変化という重要な概念が含まれます。

意義

インパクトの評価基準は、評価において「それで、介入で何が変わったのか？」という疑問に対して答えを探るものです。介入の意義と上位の結果、つまり関わった人にとってどれほどの重要性があったかを検討します。

この定義は、状況に応じて様々な視点を考慮するよう評価者に促すことを目的としています。評価者は、文脈に加えて、介入の意図された受益者のニーズと優先事項、関係組織間で合意された政策目標、介入の特性について慎重に考える必要があります。介入による意図されない結果を考察する際にも、この分析要素を適用することができます。

「意義⁶」の評価において、評価者は、異なる視点に配慮すること、そしてステークホルダーのニーズを参考にした体系的アプローチを用いることの重要性を認識すべきです。また介入の意義に対する評価が、（暗黙の）偏見や価値判断に影響されないよう対策を講じる必要があります。

（グループによって）差異のあるインパクト

ジェンダー平等を含めて、「誰一人取り残さず」人権を守るという SDGs の理念を踏まえて、グループによって差異があるインパクトの違いを評価することが重要です。結果的に、正の効果によって重大な負の分配効果が隠れてしまうことがあります。対象グループ毎のインパクトを確実にモニタリングし、また、確実に評価するために、評価デザインの段階、或いは、むしろ介入をデザインする段階において、グループによって差異があるインパクトについて検討することは不可欠です。そのためには、デザインと評価の初期段階において、可能な範囲で細かく分類されたデータを入手する必要があります。また、排除や包摂性に関する幅広い測定基準の確認も伴うかもしれません。細かく分類されたデータが入手できれば、その詳細な分析が必要になるでしょう。

意図されない効果

評価者は、介入が意図されない効果を生んでいないかを検討する必要があります。こうした分析では、様々なインパクトが、介入をデザインする段階でどの程度意図された、又は予測されたものであったかを含めるべきです。意図されない効果には、正負いずれもあると考えられます。正の効果である場合、その全体的な意義を考察して、革新につながったり、拡大したり、もしくは、他の介入において正のインパクトを再現する余地がないかを検討すべきです。評価者は、負のインパクトに対して十二分に注意することが必要です。特に、環境へのインパクトや弱者層への意図されないインパクトなどの非常に重大な影響を与える可能性が高い負のインパクトに対して、特別な注意を払うべきです。

社会変容を伴う変化

ここでは、社会変容を伴う変化を「社会組織や規範における全体的・永続的な変化」と定義します。社会変容を伴う変化は、根本的な原因や貧困、不平等、排除、環境被害の体系的な要因への対処として考えることができ、アジェンダ 2030 では、持続可能な開発目標の達成に必要なものとして認識されています。介入が社会変容を伴う変化への貢献を目指すことは次第に一般的になりつつあり、インパクトの基準を評価する際に、評価者には、規範や社会組織（社会、経済又は政治制度）に及ぼす効果についての設問に応じることがますます求められています。例えば、評価において、性別に課せられた伝統的な役割が、何らかの方法によって、どの程度変化したかを調査することもあり得ます（ボックス 4.12 を参照）。

他の評価基準との関係

他の評価基準と同様に、インパクトの評価基準も他の評価基準と概念上関連しています。

- インパクトと有効性：インパクトと有効性はどちらも、介入によって達成される結果、受益者への効果、意図されない結果を検討します。インパクトと有効性の基準の違いは主に、介入と目的がどのようにデザインされたかによって決まります。有効性は一般的に（結果の連鎖上のあらゆる次元に）設定された目的の達成

度に注目する一方で、インパクトは（結果の連鎖の）上位にある効果に着目します（この上位の効果は、介入の目的としては認識されないため、インパクトの基準で確認しなければ自動的に分析されません）。両者を区別するもうひとつの考え方は、多くの介入は時の経過とともに有効であると判断されるかもしれませんが、それでも望まれる上位次元の変化や社会変容をもたらすような変化まで至らないかもしれない、というものです。もしインパクトを評価しなければ、こうした傾向は見落とされてしまうでしょう。結果の連鎖を定義する要件は組織によって異なるため、介入による意図された結果の解釈も組織ごとに異なってきます。すべての介入を上位の目標と関連づけるよう求める組織がある一方で、直接的・短期的な効果しか検討しない組織もあります。これら2つの評価基準を適用する場合は、組織が介入デザインの要件を明瞭化し一貫性を持たせると役立つでしょう。プロジェクトなどの比較的小規模な介入の場合、上位の目標にどのようにつながっているのかを恒常的に意識しないので、評価方針においてインパクトの分析を義務づけることが重要です。そうでなければ、上位のインパクトが見落とされてしまいます（例としてボックス 4.12 を参照）。

- インパクトと整合性：インパクトには全体的な視点が採用されることから、整合性の検討事項にも当然適応します。介入によって達成される効果は、他の介入、政策目標、トレードオフ、介入が実施される場所の社会組織に多くの場合左右されるからです。その一例として、途上国の公衆衛生組織強化を支援する開発協力プログラムが挙げられます。こうしたプログラムでは、インパクトは、当該プログラムだけでなく、医薬品の価格設定と規制や医療従事者の採用に関する国内又は国際的な政策に影響されます。
- インパクトと効率性：効率性の基準で指摘したように、最も広義な意味での効率性を検討するためには、評価者は達成された結果の全体像（例：インパクト、持続性）を考慮しつつ、結果と資源を比較する必要があります。
- インパクトと持続性：インパクトと持続性は、長期間ある程度結果が持続するかどうかを検討します。インパクトは、社会変容を伴う変化（本質的に永続的なもの）の確認という視点から時間的な要素（将来を見据えた時間の流れ）に注目します。持続性は便益の継続性に着目します。評価基準としては、持続性はインパクトに比べてより幅広い意味を持っています。これは持続性の基準が、それを達成するための条件や、介入の経済面、社会面、及び環境面における持続性の間のつながりを検討するからです。

包摂性の視点の包含

社会変容を伴う変化、差異のあるインパクト、意義は、どれも本質的に**包摂性**と結びついています。ここでは、誰にどんなインパクトが起きたかを理解することが重要となります。抑圧的なシステムに変容をもたらす有意義な貢献がなされたか？それが周縁化された弱者層の持続的な変化につながる可能性はあるか？評価者は、意図されないインパク

トが、社会組織全体の構造的不平等にどの程度根差すものであるか、介入がその社会組織に及ぼす影響を理解することを目指すべきです。

改訂後のインパクトの定義では、長期的な社会経済的効果を含めて、介入による上位の結果を重視します。インパクトは、定義に記された「人々の福祉、人権、ジェンダー平等、環境に与え得る効果」をはじめとする、社会組織や規範に係る社会変容を伴う変化を含みます。インパクトの評価を通じて、評価者は、介入が社会変容を伴う変化、公平性、人権、能力向上にどのように寄与しこれらを向上させるかに関する全体像を把握できます。

インパクトの評価に関する課題と対処法

6つの評価基準のうち、インパクトは、評価し、理解するのがしばしば最も困難になり得る基準です。インパクトの評価に伴う主な課題のうち4つを、対処法に関する提案とともに以下にまとめています。

原則として、インパクトを評価するには、概して他の基準以上に多くの資源と大量の基本データが必要になるため、こうした資源を利用又は入手できそうな場合に限って、評価デザインにインパクトを組み込むべきです。その一方で、インパクトは、介入が効果を生んだか、どのような効果を生んだかに着目するため、利用者の注意を最も集めがちな評価領域です。したがって、インパクトを評価に組み込む場合に必要となる時間と労力は、正当とみなされることが多いです。

インパクトを評価する場合のもうひとつの課題は、評価を委託する組織が定める期限に関連します。この期限は、多くの場合、介入の終了日と同一とされます。より長期的にインパクトを検討し、介入の終了後に初めて明らかになる可能性があるインパクトへの理解を深めるためには、この期限に関する条件をもっと柔軟に設定できれば有益と思われます。

次の表は、インパクトを評価する際の主な課題をいくつか明らかにし、これらの課題への対処法を評価者と評価管理者両方に提案しています。

表 4.5. インパクトの評価に関する課題

課題	対処法：評価者	対処法：評価管理者
評価の関係者やユーザーにとって、インパクトが概念として何を意味するかを明確にする	OECD DAC の用語か別の概念かなど、どの定義を使うかに最初の段階で合意し、一貫してその定義に従う。用語の解釈が変化していないか、協議中に定期的に関係者に確認する。	

その介入にはどんなインパクトが意図されているのか、それをどうやって達成するかを明確にする	これは、介入当初のデザインや、変化の理論に基づく、明確な結果の枠組みの有無に左右される。枠組みがない場合、理論に基づくインパクトの評価（これを行うことが非常に望ましい）は困難になるため、具体的な対象グループへの言及も含めて、枠組みを再構築する必要がある。介入によっては、目標があまりに広範囲に及び、その介入が達成できる範囲を超えている場合もある。	
介入がどの程度のインパクトをもたらすかを確認する	介入の結果に対する貢献、又は帰属の経路を、上位の次元までたどれない／明確ではない場合、この経路を作成する。ここでは、貢献分析などの手法を使うことが有益かもしれない ⁷ 。 インパクトの測定法に関する具体的なガイダンス／方法論に基づいて選択肢を探り、評価の目的、文脈、答えるべき設問、データ入手の可能性、評価に利用できるリソースとその実現可能性を考慮して、手法について慎重に現実的な選択をする。重要な設問は、その手法が単にインパクトの規模を評価するものか、それともインパクトが生じた理由と経緯も明らかにするものか、である。	評価に利用できる資源と、関係者の（手法の適用に関わる）技能が、インパクトの評価を行う上で十分なものかどうか確認する。
基準値や指標を含む、データの入手の可能性	上位の効果を評価する上で、基準値や指標の欠如を含めて、特にデータが問題になることがある。これはある意味、評価開始の時点で考えるべき評価能力の問題であるが、様々な種類の定量・定性データを可能な限り最適な形で活用する上で、適切な手法を選択する参考にもなる。	評価業務指示書の草案を作成する工程のなかで、データの入手可能性を確認する。
介入が、想定外の、又は意図されない重大な効果をもたらす	意図しない、もしくは想定外の結果は、モニタリングの枠組みやデータ収集から漏れがちであるため、把握しにくいことが多い。受益者や介入によって（正又は負の）影響を受ける人々にとって何が最も重要かを知するには、評価者は、正式なプロジェクト文書や想定されたデータ以外にも目を向け続けるべきである。効果の相対的な重要性は、関係者によって異なる可能性が高いからである。	評価者に、介入によって（正又は負の）影響を受ける人々の声を聞く場を提供し、こうした人々を支援して、彼らにとって重要な、想定外の効果に評価者の注意を向けさせる。

インパクトの評価の例

この節では、家族の能力向上（エンパワメント）、平和構築、研究支援、土地利用計画の作成、女性・少女への暴力防止に関する介入を対象に、インパクトがどのように評価されてきたかを示す代表的な事例を紹介します。

ボックス 4.12. コロンビアの「More Families in Action（家族の行動推進）」プログラムの有効性とインパクトの評価

コロンビアにおける「More families in action」プログラムの運営面と福祉に関するインパクトの評価は、有効性とインパクトの基準の下でどう評価されるかを含めて、上位レベルの社会変容を伴う変化について、少し異なる角度から理解し明らかにする方法について興味深い事例を提供しています。この評価は、プログラムの実施を支援するための訓練を受け、地域共同体の擁護者として活動する、*Madres Lideres* と呼ばれる女性たちの役割を検討しました。

- 運営：プログラムの有効性を支える *Madres Lideres* の役割を確認し、実施と提供における女性の重要な役割を明らかにしました。
- 上位レベル：加えて、こうした女性のプログラムへの関与が、個々の女性の能力向上（エンパワメント）に正の効果をもたらすだけでなく、社会における女性の役割に対する認識を変えることで、地域社会全体を変容させ得る効果につながったことを確認しました。こうした規範における上位レベルの変化は、インパクトの評価基準の下でも確認されています。

Source: Economía Urbana and Ipsos (2019[18]), [Evaluación de operaciones del programa "más familias en acción" y de resultados del componente de bienestar comunitario](#) [Evaluation of the "More Families in Action" programme]

ボックス 4.13. 紛争下・紛争後に影響を受けた国における Sida による平和構築支援のインパクトの評価

この評価は、不安定な環境での長期間に及ぶ複雑な介入に対するインパクトの評価を検討した興味深い例となっています。Sida の平和構築活動の全体的な効果について、直接的／間接的、意図された／意図されない、正／負の結果の観点からインパクトを評価しました。

この評価は、1990 年代初めから 2010 年代末のソマリア、ルワンダ、ボスニア、グアテマラにおける平和構築活動を対象として、非常に長期的な視点を取ることで、長期的なインパクトを捉えました。全体的なインパクトに加えて、特に女性や周縁化された集団へのインパクトに注目しました。評価では、介入の貢献度を評価するために、定性データと参加者による記述に注目した「最も重要な変容」手法を使用しました。

Source: Bryld (2019[19]), [Evaluation of Sida's Support to Peacebuilding in Conflict and Post-Conflict Contexts: Somalia Country Report](#)

ボックス 4.14. 戦略的な政策評価におけるインパクトの評価

18 の受益国で 40 件のプロジェクトに携わった、デンマークの 21 の当局が実施した援助・研究の評価においては、インパクトの評価基準に基づく主な調査結果として、受益国機関にまたがる上位の能力向上効果の重要性が明らかになりました。この評価は、介入の主要な成果のひとつが、プロジェクトの正式に設定された目的を超えて受益国機関の能力向上に貢献したことにあると、結論づけました。これは、デンマークが主要な公的・私的双方の関係者や市民社会と交流・連携して、幅広い政策成果やプログラム成果に影響を与えたことを指します。こうした成果は両当事国機関にとって互恵的なものですが、通常、介入モニタリングでは「捉えることはできません」。しかし、これによって付加された価値はすべて、プログラムの重要な成果となります。

Source: PEM Consult (2020[20]), [Evaluation of the Danish Strategic Sector Cooperation](#)

ボックス 4.15. フィリピンにおける包括的な土地利用計画のインパクト、普及及び拡大の評価

土地利用への介入のインパクトを調べるに当たり、この評価は、環境指標と社会経済的指標に基づいた、介入の中期から長期的な効果を検討しました。

この評価では、多次元調査のパネルデータ、定性的面接、特定化グループ討論、文献調査、土地利用計画作成に関する文書の分析、地理的データ・情報を活用した、理論に基づくアプローチと混合研究法デザインを手法として採用しました。評価者は、介入の変化の理論を包括的に再構築しました。

この評価の中心は、管理構造と計画管理状況の改善、天然資源の処理、災害リスク管理に関する対策・活動、地方統治の機能、影響を受ける集団の福祉向上という 5 つのインパクト領域の効果を測定し数値化した綿密なインパクト査定です。

パネルデータは、2 つの時期（2012 年と 2016 年）に測定された、ヴィサヤ地域の 11 州 100 自治体の 300 の集落に暮らす 3000 世帯から構成されます。これには、ドイツの開発援助機関 GIZ

（*Deutsche Gesellschaft für Internationale Zusammenarbeit*）から過去に援助を受けたことがある世帯と、援助を受けたことがない世帯が含まれました。インパクト査定では、自治体、集落、世帯に関する数十項目の特徴に基づいて「統計上の類似的なペア」を見つけだすための傾向スコアマッチング（傾向点数合わせ方式）を採用した準実験デザインの手法も採用しました。

Source: Leppert et al. (2018[21]), [Impact, Diffusion and Scaling-Up of a Comprehensive Land-Use Planning Approach in the Philippines: From Development Cooperation to National Policies](#)

ボックス 4.16. リベリアにおける「ガールエンパワー」プログラムのインパクトの評価

国際救済委員会によるこの評価は、女性と少女に対する暴力を防止する多面的介入「ガールエンパワー：リベリアにおける思春期の福祉向上のためのメンタリングと現金給付による介入」のインパクトを評価した具体例を示すものです。この介入は、2016年にリンバ郡で実施されました。有効性とインパクト双方の基準に基づき評価された具体的な結果に、介入の目標をどのように定義するかに応じて違いが生じることを示す、有益な事例となっています。

インパクトの評価

この研究の目的は、リベリアの女性・少女に対する暴力防止を目的とする特定の種類の介入（すなわち、思春期初期の少女を対象とする経済強化とジェンダー変容を組み合わせた介入に、現金給付の要素を加えたもの）のインパクトを具体的に評価することです。

評価では、介入実施中と1年後の両時点において、若い女性・少女に対する暴力の程度と介入の具体的な効果に関する、基準値と最終値の詳細なデータを集めました。結果は、村落次元でのクラスターランダム（集合無作為抽出）化対照試験を使って分析されました。この研究デザインは、①反事実群（対照群）と比べたプログラムの総合的なインパクト、及び②参加誘因となる給付を追加する効果（「GE+プログラム」）の検証を目的としました。すなわち、少女たちのプログラム参加率を高めることを通じて、現金という意欲付与が少女の保護と能力向上（エンパワメント）という便益をもたらし、性的暴行を受ける危険を減らせるかどうかを測定することを目指しました。

この事例によると、能力向上（エンパワメント）効果、及び社会における少女の位置づけや暴力に関する社会的な規範と態度の変化が、変容的なインパクトの例になる可能性があります。

Source: Hallman et al. (2018[22]), [Girl Empower Impact Evaluation: Mentoring and Cash Transfer Intervention to Promote Adolescent Wellbeing in Liberia](#)

持続性

持続性の定義：

持続性： 便益は継続するか？

介入の純便益が継続する、又は継続する可能性の度合い

注釈： 長期的に純便益を維持するために必要な社会組織の財政的、経済的、社会的、環境的、制度的能力の検討を含む。強靱性、リスク、潜在的なトレードオフの分析を含む。評価の時点に応じ、純便益の実際のフロー分析や、中長期にわたる純便益の継続見込みの分析を行う。

持続性とは何か、なぜ重要なのか？

持続性の評価では、評価者は、介入の便益が財政面、経済面、社会面、環境面から継続するかどうかを判断します。便益の継続性という基本的な概念に変わりありませんが、以前の定義と比べて、改訂後の基準はより簡潔で範囲の広いものになっています⁸。持続性は、分析要素として財政、経済、社会、環境を含み、これらの要素間の相互作用にも注意を払う必要があります。

結果の継続性という意味での持続性と、環境の持続可能性や未来の世代に配慮した資源の利用との間に混同が生じることがあります。環境の持続可能性はひとつの関心事で、妥当性、整合性、インパクト、持続性を含む複数の評価基準の下でも検討できますが、持続性の評価基準の最大の意味は、そうした環境の持続可能性ではありません。持続性を記述する場合、評価者はその評価基準を自分がどう解釈しているかを明確にする必要があります。

結果の連鎖や介入のプロジェクトサイクルの各時点で、持続性を確認します。評価者は、動的で複雑な環境に対する強靱性と適応に関連する持続性についても考察すべきです。これには、介入終了後の投入財の（財政的又はその他の）持続性及び介入の全体的な文脈におけるインパクトの持続性を含みます。例えば、受益者の能力を考慮して、実施期間の初めにオーナーシップを確立したかどうかに加えて、介入終了時点で資金調達を続ける意欲と能力があるかどうかも評価できるでしょう。一般的に、評価者は、介入デザインにおいて介入の活動ごとに定められたもの、又は定められなかった持続性の条件や、必要に応じて修正が行われたかどうかを確認することができます。

評価者は、支援者や外部資金のフローの視点のみに基づいて、持続性を検討してはなりません。委託者も、介入の開始前か、又は資金提供や活動が継続している間に、持続性の評価を検討すべきです。持続性を評価する際、評価者は、① 純便益（これは、継続的な費用を含む、介入の継続的な便益が持つ全体的な価値を指します。）を考慮に入れ、② 想定されるトレードオフや、便益の継続性の土台となる能力／組織の強靱性を

分析します。例えば、便益の財政的な持続性と政治的な持続性（政治的支持の維持）の間に、トレードオフがあるかもしれません。

持続性の評価によって、中長期にわたる介入の純便益の継続又は継続見込みに関する貴重な洞察が得られます。様々なメタ分析によって、こうした洞察を得るのは実務では非常に困難であることが示されています。例えば、水・衛生などのセクターや、地域共同体主導型開発などの介入においては、しばらくすると便益が徐々に失われがちです。こうした場合の評価の役割は、持続性の達成方法に関する変化の理論の仮説を精査することかもしれません（Mansuri and Rao, 2013^[23]; White, Menon and Waddington, 2018^[24]）。

評価によって持続性の様々な側面を慎重に検討すれば、介入を通じて将来的に持続可能な開発を保証するような変化を計画・実施する方法について、重要な洞察が得られる可能性があります。この教訓から、現在の文脈において介入の持続性の尺度を拡張できる可能性や、別の文脈でも再現できる可能性が明らかになるかもしれません。

持続性の重要な要素は、出口計画の策定です。評価では、適切な出口戦略が考案され適用されているか判定すべきです。これによって、財政面・能力面の検討事項を含む（ただし、これらに限定されない）正の効果の継続性が確保されるでしょう。事後的に評価を行う場合、評価者は、以下の事例のように背景状況の変化を認めつつも、意図された正の効果の継続性を保証するために、出口戦略が適切に実施されたかどうかを検討することができます。

持続性への理解を深め、共通する課題に対処する上で有益な資源として、ドイツ開発評価研究所（Deval）が近年実施したメタ評価（Noltze, Euler and Verspohl, 2018^[25]）と評価の統合（Noltze, Euler and Verspohl, 2018^[26]）があります。この2件の研究は、評価において検討できる持続性の様々な分析要素を明らかにしています。メタ評価は、SDGsの原則を取り入れた持続性の評価を強く訴え、そうすることで評価に付加価値を与えられる領域を明確にしています。このメタ評価では、プロジェクトの意図されない効果、持続性の異なる要素間の相互作用やトレードオフを見つけだして評価することによって、継続性の基準を適用する際の知識と説明責任をどのように支えられるかを分析しています。

持続性の理解：分析要素

持続性の定義を理解するには、実現可能にする環境、正の効果の継続性、リスクとトレードオフという要素の理解が必要です。

持続可能な開発を可能にする環境の構築

評価は、介入が開発を可能にする環境の向上に多くの面でどのように貢献したかを、考えることができます。これには、今後の開発活動や人道的な活動を支援するために、介入が社会組織、団体又は能力をどのように強化したかを含みます。そのため、評価は介入の結果として構築又は強化された開発受益者の能力だけでなく、外的な変化や衝撃を

吸収するために確立された強靱性を確認することを求めます。そうすることで、前述のように純便益を将来に亘って継続させられるでしょう。

開発を可能にする環境の改善方法について例をあげると、介入による貢献には、能力の強化（個人、地域共同体、又は組織次元の）、オーナーシップ又は政治的意思の強化、国家的な（場合に応じて準国家的な）財政面・予算面の合意の強化、政策又は戦略の変更、法改正、組織改革、統治改革、公的支出に対する説明責任強化、開発計画策定における市民との協議過程の改善などが考えられます。

正の効果の継続性：実際の持続性と将来の持続性

様々な時間軸で持続性を評価することができます。評価者は、実際の持続性（介入がもたらした既に明らかな純便益の継続性）と将来の持続性（今後も継続する見込みの主な関係者にとっての純便益）の両方を評価できます。評価の時点と意図された便益の期間に応じて、実際の／将来の持続性を確認するための適切な評価アプローチを慎重に考えるべきです。多くの上位の変化は、完全に実現するために何年も、場合によっては何十年もかかります。

実際の持続性の評価において、評価者は、介入による正の効果が、意図された受益者をはじめ主なステークホルダーに対して、介入終了後も、どの程度明確に継続したかを確認できます。介入による正の効果の継続性を促す機会が特定、予測、計画されたかどうかや、その方法、正の効果の継続性を阻みうる障壁があるかどうか、確認できます。これにより、必要に応じた介入の適応能力を証明する評価結果を裏づけることができます。

将来の持続性の確認には、少し異なるアプローチが含まれます。今後の持続可能性を確認する評価は、基本的には現在の状況が続くと想定して、介入による現在の正の効果、又は予定される正の効果が継続する見込みを評価することになります。実現された正の効果の安定性と相対的な永続性、及び組織の持続性、経済的・財政的な持続性、環境面の持続性、政治的な持続性、社会的な持続性、文化的な持続性などの継続性の持続条件を評価する必要があるでしょう。

リスクと想定されるトレードオフ

持続性の評価は、正の効果の継続見込みだけでなく、介入に伴い想定されるリスクや継続的な費用も確認します。したがって、評価管理者は、長期的な純便益の持続性を高め得る要因に加えて、持続性を妨げ得る要因も考慮すべきです。介入の持続性に関連するリスクの確認には、介入の結果の持続性に影響を与え得る、特定可能又は予測可能な、正又は負の背景要因がどの程度存在するかに関する評価が含まれます。

ここから、改訂後の評価基準の重要な要素である、トレードオフの問題も提起されます。介入に関連するトレードオフの評価は、短期的なインパクトと想定される長期的な効果、又は費用とのトレードオフに加えて、財政、経済、社会、環境面の要因間のトレードオ

フを確認することを求めます。例えば、評価によって、ある介入は経済成長を支えたが、長期的な経済成長に負のインパクトを与え得る多額の環境費用が原因で、介入による成長に持続性がないことが判明するかもしれません。この要素は、持続可能な開発に関する SDGs の定義に沿うものであり、介入による正の効果の継続見込みにとどまらず、持続性を検討する評価の対象範囲を広げます。

他の評価基準との関係

持続性は、他の評価基準と密接につながっています。

- 持続性は妥当性と関連していて、主な関係者にとっての妥当性の次元は、最終的な便益への賛同とオーナーシップに影響を与える重要な要因になります。ひいては、それが持続性を高めることにつながります。
- 同様に、整合性は、持続性に関して有益な洞察を提供します。整合性では、一定の文脈において、介入の長期的な便益を高める可能性もあれば、損なう可能性もある他の介入を検討するからです。
- 有効性とインパクト：結果の継続性の評価は、第一に達成された結果（有効性）、第二は証明された上位の効果（インパクト）に依存します。したがって、有効性とインパクトは、持続性にとって最も重要な評価基準とみなすことができます。有効性とインパクトの分析によって、介入がアウトプット、アウトカム又はインパクトを達成したことを示せなければ、持続すべき明確な便益自体が存在しないからです。ボックス 4.19 は、インパクトと持続性を併せて検討した例を示しています。介入が終了した後も結果が継続するために必要かつ十分な条件の評価を通じて、インパクト、有効性、持続性の相乗効果を確認することによって、評価者は、より長期的な有効性とインパクトを探ることができます。
- 効率性への懸念が、便益の持続性を低下させることもあります。例えば、短期的な費用のせいで、長期的な効果を損なう意思決定が促された場合、持続性は低下するかもしれません。

包摂性の視点の包含

持続性の改訂後の定義と注釈は、持続性の「財政的、経済的、環境面、社会的な」側面と、こうした側面が、介入の結果がもたらす継続的な長期的便益をどのように支えるかに注意を向けています。評価者は、異なる受益者グループに対する便益の継続性がどのように計画されているか、また事後評価の場合には、便益の継続性が異なるグループにどのように表れているかを確認すべきです。ここでは、「誰一人取り残さない」という原則、周縁化されたグループが継続的な正の便益をどのように享受しているか、及び異なるグループ間に起こり得るトレードオフに注目する必要があります。

インパクトと持続性両方の基準に基づいて、介入が包摂性に配慮して公正な開発を可能にする環境をどの程度構築し、根底にある社会組織上の課題に取り組んだか（「症状だ

けでなく病気を治療する」）を確認することも有益です。オーナーシップとジェンダー能力向上に関する設問が重要です。社会組織の持続性には能力強化が必要となるため、評価者は、誰の能力が構築されたか、その能力が既存の不平等な社会組織や構造とどう関係しているかを理解する必要があります。ジェンダー平等と女性の能力向上を中長期的に実現する環境を作り、支持するための能力と、これに対する異なるステークホルダーの合意の両方が存在するか。ない場合は何が障壁になっているか。その障壁は介入の対象範囲となっているか。

持続性の評価に関する課題と対処法

次の表は、介入時点、正の効果の欠如、持続性に影響する他の要因に関連して、持続性を評価する際の課題を明らかにし、評価者と評価管理者両方のこれらの課題への対処法を提案しています。

表 4.6. 持続性の評価に関する課題

課題	対処法：評価者	対処法：評価管理者
介入が、（有効性の評価で）意図された結果や、意図されない正の便益を達成しなかった場合、又はインパクトに貢献しなかった場合	介入に全く便益がない場合、持続性の分析は不要になる。評価者は、持続性の評価の限界を率直に明確にして、他の評価基準に目を向けることを考えるべきである。	評価の範囲を修正して、便益が達成されなかった理由の理解を含め、他の評価基準に資源を振り向けるべきである。
評価の時期	介入の全期間にわたり評価が行われることが多いが、持続性が顕在化していない時点で評価が実施されることもある。従って、評価者は、介入がまだ実施中の場合や最近終了したばかりの場合、持続性の条件に注目することを考えるべきである。持続性を促し得る運営環境の種々の要因を明らかにすることによって、将来的な持続性を厳正に評価する必要がある。 その意味では、ジェンダー分析が鍵であり、達成される構造的変化の一貫性や、それが不平等解消に及ぼす効果について有益で妥当な情報が得られるだろう。	まだ介入を実施している最中に評価を行う場合、持続性の検討事項とこれを確認する目的をどの程度優先するか、評価管理者が明確にすべきである。例えば、評価結果は、出口計画や移行計画の作成に使用されるのか？
持続性に影響する見込みが高い要因の評価	将来的な持続性を支える（又は損なう）背景要因を、定量的・定性的に評価することができる。これには、ステークホルダーのオーナーシップと従事の状況、吸収力、政治的意思、国内資源の入手可能性などが含まれるが、これらに限定されない。	面接や調査を行うために、ステークホルダーを特定する手助けを行える。

持続性の評価の例

この節では、持続性を評価する方法を示すために、アフガニスタンにおける農村開発、一般財政支援、妊産婦の健康に関する事例を紹介します。

ボックス 4.17. アフガニスタンにおける日本の「地方開発支援プロジェクト」の事後評価での持続性の確認

地方開発支援プロジェクトは、国際協力機構（JICA）の緒方イニシアティブの一環として実施され、アフガニスタンのバルフ州、バーミヤン州、カンダハル州において、復興と農村開発に寄与する住民主導型地域開発モデルを構築することを目標としました。この事後評価は、アフガニスタンにおける内戦後の復興支援に関連して持続性を検討しました。

介入の実施が成功を収め、その目的は達成されたと判断したため、評価では続いて便益の持続性を確認しました。持続性の評価において、正の効果の継続性を確保するための政策制度、体制、技術、財務状況に問題がないかを分析しました。この分析では、プロジェクト終了後もアフガニスタン政府が実施を継続するための能力と意思も対象としました。

評価の結果、他の支援者の協力を受けたモニタリング体制と支援国の資金を利用した財政基盤が確立されており、紛争の被害を受けたアフガニスタンの国内体制の能力を補完し構築していることがわかりました。アフガニスタン政府には、支援国の支援を受けて地域共同体開発委員会を集合化して、住民主導型地域開発を継続する決意があることも確認されました。このプロジェクトは、国家連帯プログラムにも組み込まれました。従って、正の効果継続させる政策、体制、技術、財政状況が整っており、大きな問題は見られないと結論づけられました。

Source: Watanabe (2016[27]), [Ex-Post Evaluation of Technical Co-operation Project “Inter-Communal Rural Development Project”](#)

ボックス 4.18. マラウイ、ルワンダ、ウガンダ、ザンビアにおける一般財政支援撤回の評価から得られた持続性に関する教訓

評価において熟慮して持続性の基準を使用した興味深い例として、この研究は、一般財政支援の終了後も正の効果を継続させるために、介入がどの程度適切な準備をしたか、開発を可能にする環境の構築を支援したかについて確認しました。この研究は被援助国の一連のスキャンダルと支援国の説明責任への圧力が原因となって、2010 年頃から各国で起き始めた一般財政支援の撤回を受けて実施されました。このような展開は、2005 年の援助効果にかかるパリ宣言に対応して、財政支援に軸足が移されたことに伴うものです。財政支援は、開発結果を支援する有効で持続性のある手段と広くみなされていました。多くの評価や統合研究が、財政支援は貧困層のための支出増大、就学率の向上、公共財政管理の改善などの重要な開発結果に正の影響を及ぼしたことを示す証拠を提示しました。

この評価では、マラウイ、ルワンダ、ウガンダ、ザンビアにおいて比較事例研究デザインを用いて、撤回後も一般財政支援に伴う正の効果を継続させるために、適切な戦略が実施されたかどうかを評価しました。一般財政支援の撤回が決定された後もほとんどの国の援助総額は一定のままでしたが、構築された体制はほぼ消失していることが明らかになりました。正の結果の大部分が負の影響を受けるか、悪い方向に転じていました。共同融資はいくつかのセクターやプログラムに限定され、二国間援助方式の新たな形として独立プロジェクト型の融資が広がっていることがわかりました。これが負の外部性と援助の高度な細分化につながり、国家のオーナーシップを生み出したり、体制面の幅広い統治的課題に対処したりすることは困難になっていました。

Source: Watanabe (2016[27]), [Ex-Post Evaluation of Technical Co-operation Project “Inter-Communal Rural Development Project”](#)

ボックス 4.19. DFID の妊産婦の健康プログラムのインパクトと持続性の評価

イギリス国際開発省（DFID）の援助インパクト独立委員会（ICAI）が実施した、様々な国における妊産婦の健康向上に関する成果の報告は、インパクトと持続性を併せて確認する方法の例を示しています。ICAI の報告は特有のアプローチを採用しており、この例では 3 つの主要領域、すなわちインパクト、公平性、持続性を重視しています。報告は、以下を含む設問を取り上げています。

- インパクト：妊産婦の健康プログラムのインパクトを、DFID はどの程度最大化したか。全世界的、国別、プログラム次元の妊産婦の健康面の成果に関する DFID の主張は、どれくらい強固なものか。投資には、妊産婦の健康向上には何が効果的かをめぐる国際的な証拠が、どの程度反映されているか、また DFID のプログラムは、妊産婦の健康向上のため求められる成果をどの程度達成したか。
- 公平性：プログラムは、援助が届きにくい周縁化された女性をどの程度ターゲットにしたか。DFID は、医療保健制度の強化とその制度への女性のアクセス向上にどれくらい寄与しているか。
- 持続性：DFID の妊産婦の健康面の結果は、持続性がある見込みが高いか。結果の持続性を確保するため、DFID は、国家的な政策や制度の策定と実行をどれくらい支援したか。

ここでの公平性は、介入の性格に沿った、貧困層と若い女性への具体的効果の直接的な確認を含みます。差異のある結果や特定の集団に関するこうした設問を、妥当性、有効性、インパクトの新たな定義に基づき様々な角度から扱うことができます。実際の持続性（事後）は長期的（将来的）にしか観察できない可能性があることを踏まえて、持続性は、成果が維持される「見込みがどれくらいあるか」という視点で検討されています。

この事例の手法には、DFID によるモデル化をはじめとする、妊産婦の健康向上支援に関し DFID が主張している成果に対する（ICAI の役割と権限に沿った）独立評価を含みます。独立評価は、文書、面接、現地視察に基づいています。

Source: ICAI (2018[29]), [Assessing DFID's results in improving Maternal Health: An impact review](#)

参考文献

- Arghiros, D. et al. (2017), Making it Count: Lessons from Australian Electoral Assistance 2006- 16, Australian Government Department of Foreign Affairs and Trade, <http://www.oecd.org/derec/australia/australia-electoral-assistance-2006-2016.pdf> (accessed on 11 January 2021). [11]
- Baltzer, K. and H. Hansen (2011), Agricultural Input Subsidies in Sub-Saharan Africa: Evaluation Study, DANIDA, International Development Cooperation, Ministry of Foreign Affairs of Denmark,, <http://www.oecd.org/derec/49231998.pdf> (accessed on 11 January 2021). [15]
- Befani, B. and J. Mayne (2014), "Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation", IDS Bulletin, Vol. 45/6, pp. 17-36, <http://dx.doi.org/10.1111/1759-5436.12110>. [31]
- Belcher, B. and M. Palenberg (2018), "Outcomes and Impacts of Development Interventions", American Journal of Evaluation, Vol. 39/4, pp. 478-495, <http://dx.doi.org/10.1177/1098214018765698>. [34]
- Bryld, E. (2019), Evaluation of Sida's Support to Peacebuilding in Conflict and Post-Conflict Contexts: Somalia Country Report, Sida, https://publikationer.sida.se/contentassets/1396a7eb4f934e6b88e491e665cf57c1/eva2019_5_62214en.pdf (accessed on 11 January 2021). [19]
- Chambers, R. et al. (2009), "Designing impact evaluations: different perspectives", No. 4, 3ie, <https://www.3ieimpact.org/evidence-hub/publications/working-papers/designing-impactevaluations-different-perspectives> (accessed on 11 January 2021). [35]
- Economía Urbana and IPSOS (2019), Evaluación de operaciones del programa "más familias en acción" y de resultados del componente de bienestar comunitario, [Evaluation of the "More Families in Action" programme], Departamento Nacional de Planeación, Bogotá D.C., https://colaboracion.dnp.gov.co/CDT/Sinergia/Documentos/Evaluacion_MFA_Informe_Resultados.pdf (accessed on 15 January 2021). [18]
- Eurecna Spa (2020), Bolivia - Evaluation of Health Initiatives (2009-2020), Italian Ministry of Foreign Affairs and International Cooperation, http://www.oecd.org/derec/italy/evaluationreport-of-health-initiatives-in-Bolivia-2009_2020.pdf (accessed on 11 January 2021). [5]
- FAO (2020), Evaluation of "Improving farmer livelihoods in the dry zone through improved livestock health, productivity and marketing", Food and Agriculture Organization of the United Nations, Rome, <http://www.fao.org/3/ca8463en/ca8463en.pdf> (accessed on 11 January 2021). [4]
- Gertler, P. et al. (2016), Impact Evaluation in Practice: Second Edition, World Bank Group, Washington D.C., <https://publications.iadb.org/en/impact-evaluation-practice-second-edition> (accessed on 12 January 2021). [30] Global Affairs Canada (2019), Evaluation of Natural Disaster Reconstruction Assistance in the Philippines, 2013-14 to 2018-19, Global Affairs Canada, <https://www.international.gc.ca/gacamc/publications/evaluation/2019/endra-earcn-philippines.aspx?lang=eng> (accessed on 12 January 2021). [8]
- 78 • APPLYING EVALUATION CRITERIA THOUGHTFULLY © OECD 2021 Hallman, K. et al. (2018), Girl Empower Impact Evaluation: Mentoring and Cash Transfer Intervention to Promote Adolescent Wellbeing in Liberia, International Rescue Committee, <https://www.rescue.org/sites/default/files/document/4346/girlempowerimpactevaluationfinalreport.pdf> (accessed on 12 January 2021). [22]

ICAI (2018), Assessing DFID's Results in Improving Maternal Health: An Impact Review, The Independent Commission for Aid Impact, <https://icai.independent.gov.uk/wpcontent/uploads/ICAI-review-Assessing-DFIDs-results-in-improving-Maternal-Health-.pdf> (accessed on 11 January 2021). [29]

IDD and Associates (2006), A Joint Evaluation of General Budget Support Evaluation of General Budget Support: Synthesis Report, <https://www.oecd.org/development/evaluation/dcdndep/37426676.pdf> (accessed on 11 January 2021). [2]

IDEV (2018), Cabo Verde: Evaluation of the Bank's Country Strategy and Program 2008–2017: Summary Report, Independent Development Evaluation, African Development Bank, <http://www.oecd.org/derec/afdb/AfDB-2008-2017-cabo-verde-bank-strategy.pdf> (accessed on 11 January 2021). [12]

IEG (2008), The Welfare Impact of Rural Electrification: A Reassessment of the Costs and Benefits, World Bank, <http://dx.doi.org/10.1596/978-0-8213-7367-5>. [14]

IOB (2017), Tackling major water challenges: Policy review of Dutch development aid policy for improved water management, Policy and Operations Evaluation Department, Ministry of Foreign Affairs, Netherlands, <https://english.iob-evaluatie.nl/publications/policyreview/2017/12/01/418-%E2%80%93-iob-%E2%80%93-policy-review-of-dutch-developmentaid-policy-for-improved-water-management-2006-2016-%E2%80%93-tackling-major-waterchallenges> (accessed on 12 January 2021). [16]

Leeuw, F. and J. Vaessen (2009), Impact Evaluations and Development: NoNIE Guidance on Impact Evaluation, The Networks on Impact Evaluation, Washington D.C., <http://search.oecd.org/dac/evaluation/dcdndep/47466906.pdf> (accessed on 12 January 2021). [33]

Leppert, G. et al. (2018), Impact, Diffusion and Scaling-Up of a Comprehensive Land-Use Planning Approach in the Philippines: From Development Cooperation to National Policies, German Institute for Development Evaluation (DEval), Bonn, https://www.deval.org/files/content/Dateien/Evaluierung/Berichte/2018/Zusammenfassung_D_deutsch_%20DEval-2018_Philippinen_final_web-2.pdf (accessed on 12 January 2021). [21]

Mansuri, G. and V. Rao (2013), Localizing Development, The World Bank, <http://dx.doi.org/10.1596/978-0-8213-8256-1>. [23]

Ministry of Foreign Affairs Japan (2019), Japan ODA Evaluation Guidelines, https://www.mofa.go.jp/policy/oda/evaluation/basic_documents/pdfs/guidelines11th.pdf (accessed on 18 February 2021). [6]

Molund, S. and G. Schill (2004), Looking Back, Moving Forward Sida Evaluation Manual, Sida, <https://www.oecd.org/derec/sweden/35141712.pdf> (accessed on 11 January 2021). [17] •

79 APPLYING EVALUATION CRITERIA THOUGHTFULLY © OECD 2021 Noltze, M., M. Euler and I. Verspohl (2018), Evaluation Synthesis of Sustainability in German Development Cooperation, German Institute for Development Evaluation (DEval), Bonn, http://www.deval.org/files/content/Dateien/Evaluierung/Berichte/2018/DEval_Evaluierungssynthese_EN_web.pdf (accessed on 12 January 2021). [26]

Noltze, M., M. Euler and I. Verspohl (2018), Meta-Evaluation of Sustainability in German Development Cooperation, German Institute for Development Evaluation (DEval), Bonn,

http://www.deval.org/files/content/Dateien/Evaluierung/Berichte/2018/DEval_NH_MetaEvaluierung_EN_web.pdf (accessed on 12 January 2021). [25]

Norad (2018), Evaluation of Norwegian Efforts to Ensure Policy Coherence for Development, Norad, <https://www.norad.no/contentassets/4ac3de36fbdd4229811a423f4b00acf7/8.18-evaluation-of-norwegian-efforts-to-ensure-policy-coherence-for-development.pdf> (accessed on 11 January 2021). [7]

OECD (2011), Evaluating Budget Support: Methodological Approach, DAC Network on Development Evaluation, OECD Publishing, Paris, https://www.oecd.org/dac/evaluation/dcdndep/Methodological%20approach%20BS%20evaluations%20Sept%202012%20_with%20cover%20Thi.pdf (accessed on 12 January 2021). [3]

OECD (2010), Quality Standards for Development Evaluation, DAC Guidelines and Reference Series, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264083905-en>. [1]

OECD (2002), Evaluation and Aid Effectiveness No. 6 - Glossary of Key Terms in Evaluation and Results Based Management (in English, French and Spanish), OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264034921-en-fr>. [10]

Orth, M., M. Birsan and G. Gotz (2018), The Future of Integrated Policy-Based Development Cooperation: Lessons from the Exit from General Budget Support in Malawi, Rwanda, Uganda and Zambia, German Institute for Development Evaluation (DEval) , Bonn, http://www.deval.org/files/content/Dateien/Evaluierung/Berichte/2018/DEval_EN_The%20Future%20of%20Integrated%20Policy-Based%20Development%20Cooperation..pdf (accessed on 12 January 2021). [28]

Palenberg, M. (2011), BMZ: Tools and Methods for Evaluating the Efficiency of Development Interventions | Managing for Sustainable Development Impact, BMZ Evaluation Division – German Federal Ministry for Economic Cooperation and Development, <http://www.managingforimpact.org/resource/bmz-tools-and-methods-evaluating-efficiencydevelopment-interventions> (accessed on 12 January 2021). [13]

PEM Consult (2020), Evaluation of the Danish Strategic Sector Cooperation, Evaluation, Learning and Quality Department, Ministry of Foreign Affairs/Danida, Denmark, https://um.dk/en/danidaen/results/eval/eval_reports/publicationdisplaypage/?publicationID=CBE77158-1D4D-46E3-A81F-9B918218FAFF (accessed on 12 January 2021). [20]

Slovenia's Ministry of Foreign Affairs (2017), Evaluation of Slovenia's Development Cooperation with Montenegro 2013-2016: Final Report, Slovenia's Development Cooperation, <https://www.gov.si/assets/ministrstva/MZZ/Dokumenti/multilateral/razvojnosedelovanje/Development-cooperation-with-Montenegro-evaluation-final-report.pdf> (accessed on 12 January 2021). [9]

80 • APPLYING EVALUATION CRITERIA THOUGHTFULLY © OECD 2021 UNEG (2013), Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management, <http://www.uneval.org/document/detail/1433> (accessed on 12 January 2021). [32]

Watanabe, K. (2016), Ex-Post Evaluation of Technical Cooperation Project "Inter-Communal Rural Development Project", JICA, https://www2.jica.go.jp/en/evaluation/pdf/2015_0603847_4.pdf (accessed on 12 January 2021). [27]

White, H., R. Menon and H. Waddington (2018), "Community-driven development: does it build social cohesion or infrastructure? A mixed-method evidence synthesis", No. 30, 3ie, <https://www.3ieimpact.org/evidence-hub/publications/working-papers/community-drivendevelopment-does-it-build-social-cohesion> (accessed on 12 January 2021). [24]

注

¹ 「受益者」という言葉は、「目標対象であるかどうかを問わず、開発介入によって直接又は間接に便益を受ける個人、集団又は組織」と定義されています。それ以外に、「権利の保有者」「影響を受ける人々」などの言葉が使用されることもあります。

² 一方は実施担当者として、他方は資金提供者として、2つの組織が介入に参加する場合もあるかもしれません。このような構成においては、資金提供者と実施者双方の視点から内部的整合性を評価することができます。

³ これに関する重要な注意点として、インパクトの評価は、純粋な説明責任の行使である場合よりも、インパクトの達成方法や、インパクトを高めるために何ができるかの分析を伴う場合の方が有益である傾向が見られます。

⁴ 効率性を定義し評価するための方法論的な選択肢をめぐり、包括的で詳細な考察は、Palenberg による BMZ のワーキングペーパー（2011^[24]）に示されています。この文書では、次の3つの異なる次元を明らかにしています。次元0は、援助介入の効率性に関するいくつかの要素について意見を記述し提供します。次元1は、援助介入における効率性向上の可能性を特定します。これは、実施過程、インプットのアウトプットへの転換、又はアウトプットの成果（アウトカム）への転換に関する部分的な情報を提供するものです。次元2は、代替案やベンチマークと比較できる形で援助介入の効率性を評価します。これは、すべての主要な便益と費用の信頼できる計算を含む包括的なアプローチです。

実務面において、次元2の査定が適用されることは滅多にありません。世界銀行のように大きな組織でも、2010年まで、費用便益分析が最も適切とされるセクターにおいても審査段階でのこの種の分析の適用が大幅に減少していることを認めました。この報告書は、こうした詳細な分析を使用することにより、厳密さとその後のプロジェクト実績との関連性の観点からの肯定的な側面だけでなく、関連する課題も浮き彫りにしました（IEG, 2010^[25]）。農村電化の例（IEG, 2008^[26]）は完全な次元2分析が実施された比較的稀な事例です。評価者が利用できる他の選択肢には、多基準意思決定／モデリングが含まれます。この種の効率性の分析は、例えばインフラ、保健、農業などの特定のセクターに適用される可能性が高いことも指摘しておく必要があります。幅広い経済的な推計と調査データに基づく国別事例研究アプローチを用いた、農業投入財補助金に関する興味深い例を、ボックス4.10に引用しています。

⁵ この題材を扱った文献は豊富にありますが、例えば UNEG（2013^[32]）、Chambers et al（2009^[35]）、Leeuw and Vaessen（2009^[33]）、Belcher and Palenberg（2018^[34]）、Gertler et al（2016^[30]）を参照。

⁶ 特定の種類のインパクト評価でしばしば取り上げられる、統計学的な有意性と混同しないこと。定性的なインパクト評価における検出力の計算と、関連する技術的概念を論じた Gertler et al.（2016^[30]）を参照。

⁷ Befani and Mayne（2014^[31]）を参照。

⁸ 2002年の用語集（OECD, 2002^[10]）における持続性の定義は、「開発介入の終了時における、開発介入による便益の持続性。長期的便益が継続する見込み。時間の経過に伴う純益の流出という危険に対する回復力」となっています。