

2021 年度

NGO スタディ・プログラム最終報告書

提出日	2022 年 3 月 23 日	
氏名	栗村友美	
所属団体(正式名称)	特定非営利活動法人 難民を助ける会	
派遣タイプ	実務研修型 / 研修受講型 （集合型・ オンライン型 ）	
研修国	日本	
受入機関名	特定非営利活動法人 日本評価学会	
研修期間	2022 年 2 月 19 日～3 年 13 日、3 月 20 日	
研修テーマ	自団体の事業評価ツール開発に資するプログラム評価の体系的学習	

1. 導入（研修前の問題意識、所属団体や NGO が持つ課題および課題解決方策の分析など）

NGOにとって、説明責任の遂行と事業の質の向上の二側面から、事業評価の必要性は近年ますます高まっている。欧米では評価の必要性が注目を集め、2012年にはUNICEF（国連児童基金）とIOCE（International Organization for Cooperation in Evaluation）が共同で、2012年にEvalPartners（国際評価パートナーシップ・イニシアチブ）を設立した。欧米の大手国際NGOの多くも評価実施能力を強化し、派生する形で途上国のCSOも事業評価への理解を深めている。一方で、日本のNGOにおいては、いまだ評価の実践は遅れている。当会においても、長年事業評価のニーズは感じながらも、助成金ドナーの報告に必要な範囲のみの評価を行うなど、限定的な実施に留まる現状があった。実施した事業を体系的に評価し、学びを抽出して次の事業実施に活かす、自主的な取り組みはいまだ不十分である。

当会が陥っているこうした現状の背景には、以下のような課題があった。1) 団体としての評価へのコミットメントの不足、2)（事業実施現場、マネジメント層に至るまでの）評価を事業運営の一部であり、必須のプロセスと捉えるマインドセットの不足、3) 評価に割ける人的・資金的・時間的リソースの不足、4) 職員間の評価実施に関する能力・知見の不足である。これらに対応するため、当会では2022-2024年度の行動計画の中で、事業評価を通じた事業の質向上を戦略の一つに掲げ、さらにこの実現のために、各国事業チーム単位で利用できる事業評価ツールの開発と導入を行うこととした。事業実施を担う駐在員、現地職員、東京本部担当職員、マネージャーを含む各チームが、時間的・労力的に無理なく使うことができ、かつ事業の効果を的確に測り、学びを抽出することのできるツールの開発を目指している。ツールの導入・定着過程では、各職員がツールの意図を的確に理解し、必要な知識をもってツールを利用し、事業評価を行っていけるよう、能力強化も行っていく必要がある。

しかし、これらツール開発及び職員への能力強化を行うにあたって、体系的に評価を行うための方法論や、具体的なデータ収集分析の方法等への理解が欠如していることが課題であった。

2. 研修テーマについて明らかになったこと、課題解決を前提とした研修実施内容の詳細報告

今回受講した、日本評価学会主催の評価士養成講座の講座内容は以下の通りである。

第一単元 評価の基礎	- 評価の基本的考え方（1） - 評価者倫理と評価者の社会的責任 - 評価の基本的考え方（2）	2/26 09:00-11:15
第二単元 プログラム評価の基礎と諸要素	- プログラム評価の基礎と概要 - プログラム評価の5階層（ニーズ評価、セオリー評価、プロセス評価、インパクト評価、効率性評価）	2/26 11:30-12:30 2/27 09:00-13:30 3/5 09:00-13:00
第三単元 業績測定	- 業績測定	3/5 14:00-15:00
第四単元 分析手法	- データ収集・分析（定量的手法） - データ収集・分析（定性的手法）	3/6 09:00-12:30
第五単元 評価報告と活用	- 評価報告と活用	3/6 13:30-14:30
第六単元 専門分野科目	- 日本の評価の現状レビュー - 大学評価の現状と課題 - ODA 評価の現状と課題 - 政府における政策評価の現状と課題 - 学校評価の現状と課題 - NPO 評価の現状と課題	3/12 09:00-12:45 3/13 09:00-12:00

以下、上記に記載した問題意識と課題の解決に特に有用であった学びを抜粋して記載する。

2-1. プログラム¹評価の定義と目的

日本評価学会はプログラム評価を、「評価対象である社会的介入行為の改善と説明責任の遂行に資するために、しかるべき情報に基づいて事実を確認し、明確な判断基準に基づいて何がしかの価値判断を下す行為」と定義している。つまり評価で行う作業は、体系的な情報収集を行って事実を特定し、価値判断をする、という3つのプロセスに集約される。プログラム評価の定義にはさまざまなものがあるが、共通する要素は①社会調査手法の活用、②体系的な調査、③何らかの基準と比較してプログラムの価値を判断する、④プログラムを取り巻く政策的・組織的な文脈を考慮して行う、⑤社会状況を改善するための活動に資する情報を提供する、と整理される。

評価には大きく2つの目的があり、それに応じた2つのアプローチがある。目的の一つは説明責任の確保であり、この場合は総括的評価（Summative evaluation）というアプローチが取られる。もう一方の目的は事業の開発および改善であり、この場合に取りられるのが形成的評価（Formative evaluation）と呼ばれるアプローチである。これら2つのアプローチは重視する評価項目等が異なる。2-2にて詳述する。

複数の講義担当者により繰り返し強調されたのが、評価を行う目的に応じて、評価で見べきポイントや評価項目、設問は変わる、という点であった。裏返せば、当会が開発しようとする評価ツールにおいても、内部評価かつ事業の質の改善を目指すという目的に沿うように、ある程度評価項目の絞り込みを行うことが可能であると理解できた。さらに、総括的評価か形成的評価かという視点に加え、誰がいつ、どのようにその評価を活用する想定であるのかにより、誰の問いを評価項目として含めるのか、どこまで問いを深掘りするのが変わる点も強調された。誰の問いか、という点では通常は評価の発注者（依頼者）が想定されるが、例えば受益者や事業関係者の問いが評価項目に反映される場合もある。当会が開発しようとする評価ツールは、基本的には事業実施者である当会が事業の改善を行うことに資するものを目指す、開発の段階では誰の問いを反映するか、を一度広く検討することも重要であることが理解できた。

2-2. プログラム評価の5階層とプログラム理論

プログラム評価を理解するうえで重要な理論的枠組みが、プログラム評価の5階層とプログラム理論である。プログラム評価の5階層は①ニーズ評価②セオリー評価③プロセス評価④アウトカム/インパクト評価⑤効率性評価で構成される。それぞれ以下の問いに答えるものである。

ニーズ評価： プログラムのニーズは何か、プログラム実施の必要はあるか

セオリー評価： プログラムはどのような仕立てか、その設計は目的を達成するために妥当か

プロセス評価： プログラムは意図どおりに実施されたか、実施過程で何が、なぜ起きているのか

アウトカム/インパクト評価： プログラムの成果は上がったか、それはプログラムのおかげか

効率性評価： プログラムにかかる費用等は妥当か

これら5つの階層は、1度の評価の中で5階層すべてが扱われることもあれば、評価の目的によってはいくつかの階層のみが選択されることもある。主に事業完了後に、説明責任の遂行を主目的として行われる総括的評価では、アウトカム/インパクト評価を中心とした評価が行われることが多い。一方、事業の開発段階で行われる形成的評価ではニーズ評価やセオリー評価の重要性が高く、また事業の実施中に事業改善目的で行われる形成的評価では、セオリー評価、プロセス評価、アウトカム評価に重点が置かれる。事業実施前、事業実施中、事業実施後に評価すべき項目が異なるため、当会が開発するツールにおいても、評価のタイミングに合わせた項目設定が必要となることを理解した。またこの点は、過去に当会有志で試行した事業評価において感じた困難を解消するヒントでもあった。過去事業の事業評価を行った際には、後述するDAC評価6項目に則った評価は行ったが、総花的で的が絞れず、評価結果のどの部分が重要であったのか、活かすべき学びはどれなのかが不明瞭に終わってしまったという反省があった。評価を計画

¹ プログラム評価の文脈においては、プログラムは何らかの社会問題や社会状況を改善するために設計された、組織的で計画された介入を指し、プロジェクト、施策と同義で用いられる。

するにあたり、目的を明確にし、目的に応じた評価階層および評価項目を選択することが肝要であることが確認できた。

プログラム理論（プログラムセオリー）は、プログラムの構造を手段と目的関係で整理し、プログラムは何を目指すのか、誰のために行うのか、どのような活動を行うのかを論理的に示すものである。インプット、活動、アウトプット、プログラムにより達成したいアウトカム（直接アウトカム、中間アウトカム、最終アウトカム）といった、プログラムを構成する要素を手段目的関係で並べ、プログラムが何をして、何を達成することを目指すのかの論理構成を明確にする。この概念をモデル化した、ロジックモデルでしばしば表現される。

図 1. ロジックモデル例 1

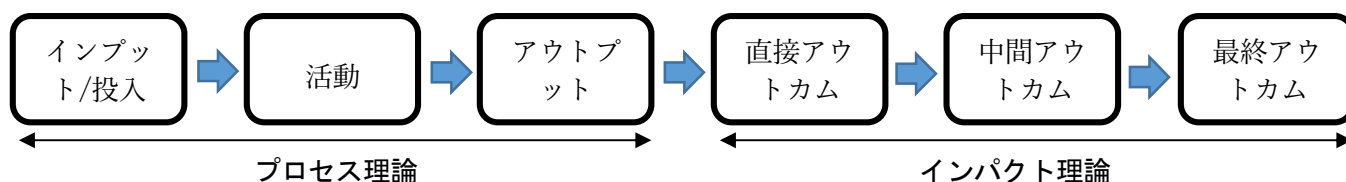
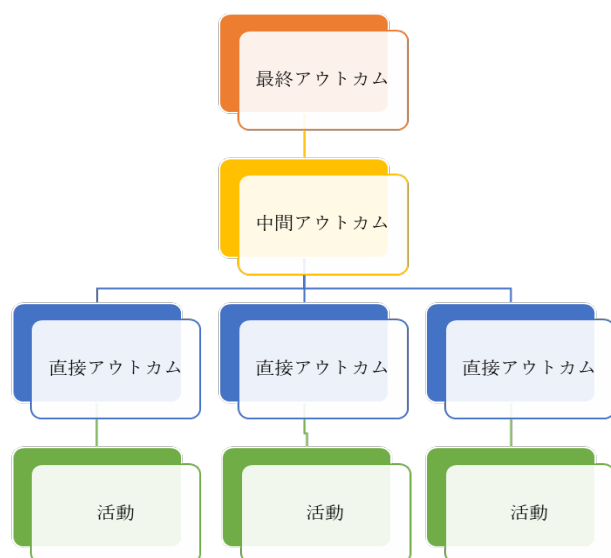


図 2. ロジックモデル例 2



プログラム理論は、図 1 に示す通り、プロセス理論とインパクト理論に分かれる。プロセス理論はプログラムの実施過程における活動計画や、活動実施のための運営体制等の投入が含まれ、活動の実施が受益者に届く道筋を示す。インパクト理論は、プログラムの目指すアウトカムの達成と、それをもたらすプログラムの活動との間の手段-目的関係を示すものである。この 2 つの理論は、プログラム評価にてアウトカムが十分達成されていないことが判明した際に、それがプログラムの持つ仮説（プログラムに定めるインプット・活動が直接・中間・最終アウトカムの達成を順次導く）の正しさの問題なのか、あるいは実施プロセスの問題なのかを理解するために有用である。プログラム理論およびロジックモデルの考え方そのものは、PCM 手法やログフレーム/PDM(Project Design Matrix)等の考え方と酷似しており、NGO にとっては非常に馴染み深いと言える。

2-3. DAC 評価 6 項目とプログラム評価の 5 階層

開発援助業界では、経済協力機構（The Organisation for Economic Co-operation and Development: OECD）の開発委員会（Development Assistance Committee: DAC）が開発した DAC 評価 6 項目（妥当性、有効性、効

率性、インパクト、持続発展性、整合性）に基づく評価が広く行われている。上述のプログラムの5階層は政策、大学、学校、NPO等幅広い分野での評価に活用できる基礎的な概念であるが、実際には各分野において少しずつ形を変えて適用されている。DAC6項目はその一例と理解できる。DAC6項目のうち、妥当性と有効性はプログラム評価の5階層のニーズ評価、アウトカム/インパクト評価とそれぞれ重なる部分が多い。一方で、DAC6項目の効率性、インパクト、持続発展性、整合性はプログラム評価の5階層とは単純には対応しない。以下に各項目の相違点を整理する。

DAC6項目の効率性は、各種資源の投入が計画通りに行われたか、アウトプットは予定通りに達成できたかという視点を用いる。一方で、プログラム評価の効率性評価では、プログラムに投入された費用に対し、発現された効果の大きさが見合っているかという視点で、費用効果分析や費用便益分析を行う。DAC6項目のインパクトと、プログラム評価のインパクトでは、インパクトという用語の定義から既に異なっている。DAC6項目では、インパクトは長期的な波及効果、二次的・副次的な効果、正負の予想しない影響を指す。一方でプログラム評価では、インパクトを（外部要因の影響を極力排除した）介入による純粋な変化の大きさ、と定義し、直接・中間・最終アウトカム指標において介入により発生した差異を測定、評価する。DAC6項目の持続発展性は、プログラム評価において該当する項目がない。整合性もプログラム評価5階層に項目がないが、一部の内容（同国内、セクターで活動する援助機関との連携、相乗効果）は、プログラム評価5階層のニーズ評価にてカバーされると理解できる。一方で、プログラム評価5階層のプロセス評価は、DAC6項目に該当するものがない。プロセス評価では、事業実施中に、活動が計画通り行われているか、意図した対象者にサービスが提供されているか、実施体制は適切か、関係機関との連携は十分か、プログラムに対する関係者の意識の変化はあったか、という項目を評価する。2-2で述べた通り、プログラム評価ではプログラムの仮説が正しかったのか、プログラムは予定通りに実施されたのか、という2つの視点が扱われ、後者はまさにプロセス評価にあたるが、DAC6項目には該当する視点がない。このためDAC6項目を機械的に適用して事業評価を行うと、この視点が抜け落ちる可能性がある。計画通りに事業を実施したか、は個々の活動について日々モニタリングがなされる部分ではあるが、プロセス評価の意識がなければ、モニタリングから得た個々の情報が集約されず、アウトカムの未達の原因となる事業計画遂行の失敗に気づくことは困難である。

2-4 セオリー評価

プログラム評価の5階層の中でも、報告者の課題意識上最も学びが深かったのがセオリー評価である。セオリー評価は、プログラムが発現させようとする効果に対して、プログラムの投入や活動が妥当であるかを評価するものである。プログラムの形成段階では、ロジックモデルを用いて手段—目的関係に整理したプロジェクトの論理の「もっともらしさ」を、文献や過去の報告書、過去の事業のアウトカム指標データ等のエビデンスを用いて、また、特に現場関係者との議論を通じて評価し、必要に応じてロジックモデルを改善する。事業の実施中には、プロセス評価とアウトカム評価を経て発見された学びをセオリーに反映させ、ロジックモデルを改善する。具体的には、プロセスが計画通りに行われているのに評価時点でのアウトカムの発現が想定以下であるといった場合には、プロジェクトの仮説に誤りがある可能性があるため、原因を究明の上、必要な変更をプロジェクトの仮説に加え、ロジックモデルを改変するということがある。NGOの活動領域では、事業実施前にプロジェクトを取り巻くすべての情報を収集することは不可能に近く、事業実施中にも多大な環境変化が起こりうる。このため、事業実施中のセオリー評価により、ロジックモデルを改変していくことは極めて重要である。複数の講師が、形成的評価によりロジックモデルを柔軟に変化させ、事業の効果を発現させていくことの重要性を再三にわたり強調していた。当会が開発する事業評価ツールでも、事業事前評価に必要な項目、事業実施中評価に必要な項目として、セオリー評価の視点は不可欠であると理解できた。

2-5 アウトカム/インパクト評価

前述の通り、援助業界で広く利用される DAC 評価 6 項目におけるインパクトと、プログラム評価において行われるインパクト評価は性質が異なる。後者のインパクト評価は、介入によって対象社会に生じる変化（アウトカム）の純粋な大きさを測り、その価値を判断することを指す。ただし、DAC 評価 6 項目との定義の差があるとはいえ、プログラム評価の指すインパクト評価が開発援助業界で軽視されているわけではない。むしろ近年は、信頼性の高いインパクト評価デザインの採用が各国ドナーからも推奨される傾向がある。採用により、効果の測定が適切になされるプロジェクトであるとの評価が得やすくなり、欧米ドナーの助成金獲得につながりやすくなるというケースも見られる。一方で、測定にかかる時間・労力・資金面でのコストや、手順の複雑さから、導入へのハードルが高いとの見方もある。インパクト評価は通例介入から一定期間が経過したのちに総括的評価で実施される場合が多く、当会の事業評価ツールは事業改善のための形成的評価の実施を主目的とするものであるため、ツールへのインパクト評価の組み込みについては慎重な検討が必要である。しかし依然として、介入効果を純粋に測定し評価することへの要請は助成金ドナーをはじめ高まっていくことが予想されるため、インパクト評価への理解は今回の研修を通じた有用な学びの一つであった。

インパクト評価では、介入対象群と、介入対象群にできるだけ性質の近似した非介入対象群のアウトカムの差を測る。インパクト評価の中心課題は、いかにして適切な（＝介入対象群に近似した）比較対象を用いた比較を実現するかにあり、これを実現するために複数のインパクト評価デザインが開発されている。研修の中で扱われたデザインは①ランダム化比較試験(Randomized Controlled Trial: RCT)、②クラスターランダム化比較試験(Cluster Randomized Controlled Trial: CRT)、③回帰分析デザイン(Regression Discontinuity Design: RDD)、④マッチング、⑤差の差であった。特に、インパクト評価デザインの中でも現在最も信頼性が高いとされる RCT について本報告では取り上げる。

RCT は、プログラムが介入対象者を決定するプロセスにおいて特定の層をターゲットにしたり、受益者自らが立候補して参加したりすることにより発生する介入対象者と非介入対象者の性質の乖離を避けるために、介入対象者を、サイコロを振って出た目やコインの裏表の結果などで選定する（＝ランダム割付）。サイコロの目やコインの裏表は偶然の結果なので、これにより二分された介入群と非介入群はほぼ同質の性質をもつ人々となる。このうち介入群にのみプログラムでの介入を行い、介入群と非介入群のアウトカムの発現程度を測定する。RCT は手順の煩雑さや、時間的・人的・資金的コストの大きさに加え、受益者の選定方法を偶然に頼る面、非介入群に対する倫理的説明の難しさといった面から、NGO にとっては特に実施ハードルが高いと考えられる。報告者自身も、RCT は導入にあたって最も各ステークホルダーの同意を取り付けることが困難であろうとの印象を持っていた。

しかし研修では、そうしたハードルを理解したうえで、RCT を採用するための工夫が紹介された。具体的には①応募超過法、②段階的導入法、③境界付近でのランダム割付、の 3 つである。応募超過法は、プログラムの予算でカバーしきれない数の対象者がいた際に、予算規模に見合う対象者数までへの絞り込みを、サイコロなどを用いたランダム割付で選定する方法である。段階的導入法は、プログラムが年次ごとに対象者数を拡大していく設計である際に、年次ごとの対象者数の選定をランダム割付で行う方法である。最後に境界付近でのランダム割付は、受益者の中でも特に困窮度が高いなどの理由で優先度が高い人々は、ランダム割付の対象とはせず介入対象者として設定し、受益者の中でも比較的優先度の低い層の人々においてランダム割付を行う方法である。優先度によって受益者を複数のグループに分け、最優先層は無条件で介入対象者に、続いて優先度が高いグループは 8 : 2 の割合で対象者に含まれるようにランダム割付し、優先度が下がることに 7 : 3、6 : 4 などのように対象者に含まれる割合を変えていくという方法も取られる。当会が実施する事業において、これらの工夫を取り入れれば RCT を試験的に実施できるものか、今後検討していきたい。

また、インパクト評価デザインの検討において重要なのは、事業形成の段階から評価デザインを想定する必要があるという点である。信頼度の高いインパクト評価デザインとして紹介された RCT と RDD は、受益者の選定過程に操作を加える手法である。つまり事業実施後に後から採用を検討することは不可能で

ある。受益者の選定過程が既に終了したのちに評価デザインを選択する際には、より信頼性が低く（＝アウトカムの発現程度から他の要因の影響を取り除くことが難しい）、かつ煩雑な手順を要するデザインを選択せざるを得ない。評価計画の策定時点のみならず、事業形成段階からの考慮が不可欠であることが確認できた。

3-1 結論：

本研修の受講により、目的としていた当会の評価ツール開発に向けて非常に有用な示唆を得ることができた。評価ツールの設計にあたり、評価の目的を、事業の改善という以上に絞り込んで設定する必要があること、当会として評価を通じてどのような視点から学びを抽出する必要があるのかを綿密に議論することが出発点であることが理解できた。また、DAC 評価 6 項目の評価を広く行うだけが方法ではなく、当会のニーズに合わせて評価項目を取捨選択できることも分かり、多忙な現地事務所であっても無理なく利用できる、的を絞ったツールの開発が可能であるとの示唆を得た。具体的な定量・定性データの収集や分析方法、指標設定方法の留意点等など、事後評価のみでなく、事業前、事業実施中の評価に必要な技術面の習得も進んだ。

3-2 本研修成果の自団体、NGO セクターの組織強化や活動の発展への活用方針・方法

本研修からのさまざまな学びは、前述通り当会の事業評価ツールの開発に活かしていく。2022 年度には評価ツールのプロトタイプを作成し、事業地での試験利用を通じて改善、導入を行っていく。あわせて、駐在員・現地職員・東京本部職員らへの評価に関する能力強化のため、内部研修を実施する。その際に、本研修での学びを盛り込んで研修コンテンツを作成していく。

報告者は当会の新入職員に対する事業評価についての内部研修を担当しており、その内容のブラッシュアップにも今回の研修の学びを反映する。これまでは、DAC6 項目を概観する内容の研修に留まっていたが、プログラム評価全体における DAC6 項目の位置づけや、事業の質向上において重視すべき評価項目と評価設問の立て方について新たに内容を加え、より実地的な内部研修に改変する。

当会は 2022 年度に引き続き 2023 年度も外務省より NGO 相談員を受託している。他の NGO から、事業地における行政手続きについて、また事業地のニーズ等についてなど、事業実施に関わる幅広い問い合わせを受けることがある。今後、NGO より評価の実践について問い合わせがあった際には、本研修の学びや、それを活かして当会が評価制度をどのように開発したかの知見を伝えることが可能である。リソースに限りのある NGO という共通の立場から、自分たちの持てる資源で行える、かつ効果性のある事業評価について学びを共有していきたい。

3-3 テーマに関する日本の国際協力分野への提言

講座受講中、行政や大学等を含む日本の評価の発展においては、ODA の評価が大きな役割を果たしてきたことが強調された。国際的なスタンダードに則った評価をいち早く取り入れ、かつ膨大な数の事業評価の結果報告書を公表してきたことで、日本の評価分野における知見の蓄積に貢献してきたためである。これは主に外務省および国際協力機構（Japan International Cooperation Agency: JICA）の取り組みによるものであり、講座の講師陣からも、評価の役割を明確に理解し、評価のためのリソースをしっかりと割いている JICA を高く評価する声が挙がっていた。一方で、NGO が実施する評価については、リソースを割くことへの理解が日本社会の中で浸透しているとは言い難い。助成金では、事業実施予算に評価のための予算を割くこと、事業効果の確認のためには最低限のプロセスともいえるベースライン・エンドライン調査に予算を付けることすら理解を得られないことがある。より質の高い事業の実施、事業効果についての説明責任の遂行には評価が欠かせないことは国際的にも広く理解されている。NGO 自身も評価のための能力強化に研鑽を積みつつ、評価が受益者や関係者にとってより良い事業の実践には不可欠であることを周知し続ける必要があり、国際協力に関わるステークホルダーの側にもそれを理解する努力が必要である。

4. 団体としての今後の取り組み方針

前述のとおり、当会は2022-2024年度の行動計画内にて、事業の質向上のための取り組みを行うことを定めている。このために、各国事業チームで利用できる事業評価ツールの策定を行っていく。さらに、行った評価の学びが各事業地の実施中の事業や次期事業の形成に活かされているかを、各事業チームのマネージャー、支援事業部長等の管理職レベルが把握できるよう体制を整えていく。

評価の実施には各種コストがかかる。これを組織として許容するには、組織の経営層が、より良き事業実施のために評価が果たす役割を十分に理解する必要がある。2022-2024年度の行動計画への承認は理事会で得ており、今後も定期的な進捗報告をもって、経営層への評価の重要性の浸透を図っていく。団体の年度予算においても、適正な範囲で評価に関する費用を計上していく。

また経営層のみならず、評価の仕組みの団体内への定着には、現地駐在員、現地職員をはじめ、常に事業実施や事務所運営で非常に多忙な面々が無理なく利用でき、かつその効果を実感できる精度の高い事業評価ツールの作成が不可欠である。当会の事業評価ツール開発担当者らは、元は団体内有志で立ち上がったモニタリング・評価タスクチームであり、数年の自主的な活動を経て会の行動計画の一部を担うようになった現在も、他の業務との兼任で評価関連業務を行っている。こうした状況でも事業評価の仕組みづくりが遅滞なく進行するよう、管理職レベルでタスクチームメンバーの業務量調整等を行い、また事務局長、支援事業部長等の責任者レベルで進捗をモニタリングしていくことで、目標年度には当会の各事業地に評価が根付き、評価を通じた事業の質の向上が行われていくよう取り組む。

5. その他

5-1 本プログラムや事務局側に対する提案・要望等

特になし。本来は2月中に終了する研修を対象としていたところ、3月にかかる講座受講を認めていただき、柔軟な対応が非常にありがたかった。