

No.101

普及プログラムの評価方法
How to Conduct Evaluation of
Extension Programs

Murari Suvedi
Kirk Heinze
Diane Ruonavaara

共著

ミシガン州立大学普及部

評価研究センター

(1999 年 12 月)

平成 17 年 3 月

(財) 農林水産奨励会

農林水産政策情報センター

本書の翻訳に当たって

わが国でも、これまでのパフォーマンス・メジャーメント（業績測定，又は実績評価）に加えて，プログラム評価を実施する動きが中央省庁を中心に始まっている。

プログラム評価を実施するためには，人的・資金的な資源の投入が必要であり，また，評価結果は，納税者や当該プログラムの関係者に対して説得力のあるものであり，かつ，プログラムの実施改善や次期プログラムの立案に活用されるものでなければならない。

わが国では，プログラム評価を実施した経験が少なく，また評価の専門家も育っていないと言いがたい。この「普及プログラムの評価方法」は，1の「導入」で評価を実施するための基本的な事項を述べ，2の「評価の実施に至るステップ」では，ステップごとに手法，特徴などを解説している。本書は，普及プログラムの評価方法を解説したものであるが，他のプログラムにおいても適用可能である。

本書の翻訳は，ミシガン州立大学普及部の Dr. Murari Suvedi, Dr. Kirk Heinze, 及び Dr. Diane Ruonavaara から翻訳許可を得て行ったものである。ここに，翻訳の許可を与えて頂いたことに対して感謝の意を表したい。

本書がプログラム評価を計画し，実施する場において利用されることを願っている。

平成 17 年 3 月

（財）農林水産奨励会
農林水産政策情報センター

Preface and Acknowledgments

Japanese government organizations, mainly which in the central government have began to conduct program evaluation as well as performance measurement.

In conducting program evaluation it is necessary to invest human and capital resources. The result of evaluation should be convincing not only to decision makers and finance authorities, but to all other stakeholders such as taxpayers, consumers and farmers. It should be utilized in program improvement and program planning for the next programs.

In Japan examples of program evaluation are not enough yet. Also there is really a call for specialists with expertise in the area.

This document consists of two sections. Section 1.*Introduction* indicates basic points of program evaluation. Section 2.*Steps to Evaluation* describes important features of each evaluation step. It focuses on evaluating in extension, but also includes full of suggestions applicable to other evaluations.

This document is the Japanese translation of "How to Conduct Evaluation of Extension Programs", the text published from Michigan State University Extension, translated in our institute by permission of all coauthors. We would like to express our sincere thanks to Dr. Murari Suvedi, Dr. Kirk Heinze and Dr. Diane Ruonavaara for letting us translate such a wonderful text. The publication will be also available in our website, intended for the use of evaluators in various organizations.

Agriculture, Forestry and Fisheries Policy Research Institute

目 次

本書の翻訳に当たって

1. 導入	1
1-1 評価とは何か	1
1-2 なぜ、評価を行うことが必要か	1
1-3 いつ評価を行うべきか	2
1-4 評価者の役割	2
2. 評価の実施ステップ	4
2-1 評価対象となるプログラムを特定し、その内容を説明する	5
2-2 プログラムの進捗状況と必要な評価作業のタイプを特定する	5
2-3 評価作業実施の可能性を検討する	7
2-4 主要なステークホルダーを特定し、その意見を聴取する	8
2-5 データ収集の方法を特定する	11
2-6 データ収集方法を選択する	14
2-7 母集団を特定し、サンプリングする	47
2-8 データを収集し、分析し、解釈する	51
2-9 調査結果を報告する	59
2-10 調査成果を実地に適用し、活用する	61
2-11 評価を評価する	61
参考文献	64
普及プログラムの評価実施のための資料	65

1. 導入

普及活動の評価は、これまで、主として、プログラムのメリット又は価値を判定することに焦点を当てていた。更に、これまで行われてきた評価の方法は、基本的に、定量的に表現される活動を中心に行われてきた。しかし、ますます複雑さの度合いを高め、要求が厳しくなっている今日の世界にあっては、評価は、アカウンタビリティ（説明責任）が十分果たされているか、優れたマネジメントが行われているか、知識の創造と共有が行われているか、組織的な学習と能力開発が行われているか、問題の特定と方針の形成は適切かといった問題を扱わなければならないとなっている。評価の範囲が拡大するのに従って、定性的アプローチや、複合的手法がますます必要となってきた。同時に、今日、普及活動の評価に携わる者は、自分自身が、複数の役割を果たし、数多くの方法論を熟知している必要があることに気づいている。このマニュアルは、普及活動に適用される評価の範囲の拡大に対応できるように設計されており、評価者に、幅広いさまざまな方法論を備えたツールボックスを提供し、それらを適切に使いこなすためのアドバイスを与えるものである。

1－1 評価とは何か

プログラム評価は、普及活動の価値もしくは潜在的な価値を評価する継続的かつ体系的プロセスであり、プログラムの将来に関する決定に指針を与えるものである。

評価を行う場合には、

- 既存のプログラム又は提案されているプログラムの基礎となっている前提条件を検討する。
- 当該プログラムの目的と目標を検討する。
- 当該プログラムに必要なインプットとアウトカムに関する情報を収集する。
- その情報を、あらかじめ設定された基準と比較する。
- 当該プログラムの価値判定を行う。
- 当該プログラムにおいて利用できるような方法で発見した事柄を報告する。

ことが求められる。

1－2 なぜ、評価を行うことが必要か

プログラムの効率、プログラムの効果、公に対するアカウンタビリティに関する普及活動へのニーズが、ますます増大してきている。評価は、こうした要求をさまざまな方法で満たすうえで、大いに助けとなる。

- プラニング
 - ニーズを見極めること
 - 優先順位を設定すること
 - 資源の割当てを指示すること

- 方針を指導すること
- プログラムの効果又は質の分析
 - プロジェクトの目標の達成状況を判定すること
 - プログラムの長所、短所を明らかにすること
 - 受益者のニーズに合致しているか否か判定すること
 - プログラムの費用便益を判定すること
 - 成功又は失敗の原因を明らかにすること
- 方針決定の指示
 - プログラムの運営と有効性を改善すること
 - 必要な変更事項を明らかにし、促進すること
 - プログラムを引き続き拡大すること、又は、プログラムを終了すること
- アカウンタビリティの維持
 - ステークホルダー（関係者）に対して
 - 資金提供者に対して
 - 一般市民に対して
- プログラムの影響評価
 - 個人又はコミュニティに対するプログラムの影響を明らかにすること
- 意見の主張
 - 政策立案者や諮問機関からの支持を獲得すること
 - 特定のステークホルダー・グループのニーズに注意を払うこと

1－3 いつ評価を行うべきか

いつ評価を実施すべきか否かを決定するか、という基本的な質問が必ず出てくる。もし以下の質問に対する答えが「いいえ」であれば、まだ評価を行う時期ではないことになる。

- 当該プログラムは評価を行うに足る十分な重要度を持つに至っているか。
- 評価を行うべき法的要件があるか。
- 評価の結果が当該プログラムに関する方針の決定に影響を及ぼす可能性があるか。当該評価が、ステークホルダー、又は当該評価に関心持つ者によって提起された質問に答えるものとなっているか。
- 当該評価を実施するに必要で、十分な資金が利用可能か。
- 当該評価を完了するために必要な時間が十分にあるか。

1－4 評価者の役割

評価に携わる者の役割は、絶えず拡大を続けている。従来、評価者に与えられていた役割は、明確な因果関係を発見する専門家、科学者、及び研究者を兼ねるということであった。現代の評価者は、しばしば、教育専門家であり、ファシリテーター（まとめ役）であり、コンサルタントであり、通訳者であり、仲介人であり、更には、変革推進者でもある。

評価者の信頼性 (Evaluator's Credibility)

評価者は、自分自身のコンピテンス（職務遂行能力）や、個々人の仕事のやり方を判定される。コンピテンスは、訓練や経験を積むことによって作り上げられていく。一方、評価者一人ひとりの仕事のやり方は、訓練、経験、及び個人的な性格が組み合わされたもので、時間の経過に伴って進歩して行くものである。

コンピテンス (Competence)

- 評価の対象となっているプログラムの分野に関する経歴
- 当該プログラムの内容、目的、目標を理解する能力
- 評価を組み立てていく思考能力
- 評価データの収集を行うのに必要な定性的アプローチと定量的アプローチに熟達していること
- 基礎的な定性的データ分析と定量的データ分析のスキルを有すること
- 報告書の作成とプレゼンテーションを行えるスキルを有すること

個人的資質 (Personal Style)

- コミュニケーション・スキルを有すること
- 自信に満ちていること
- 対人関係能力に優れていること
- 他人との間に信頼関係と友好関係を育む能力があること
- センシティブティのある報告が行えること

2. 評価の実施ステップ

プログラムの評価というのは、決して生半可にできるものではない。プログラムの評価を、よりプレッシャーが少なく、管理可能なものとするために、それを複数の管理可能なステップに分解することができる。それぞれのステップの細目は、プログラムの性格、範囲、複雑さ、及び、評価を行うに当たって使用可能な資源がどれほどあるかによって変化する。それぞれのステップの詳細については、後に論じることとする。

評価実施の 10 ステップ（フローチャート）

		ステップ 1 評価対象となるプログラムを 特定しその内容を説明する。		
		→		
→	→	ステップ 2 プログラムの進捗状況と必要 な評価作業のタイプの特定す る。	→	→
→				→
→				→
ステップ 10 調査成果を実地に適 用し、活用する。				ステップ 3 評価作業実施の可能性を 検討する。
→				→
ステップ 9 調査結果を報告する。				ステップ 4 主要なステークホルダー を特定し、その意見を聴 取する。
→				→
ステップ 8 データを収集し、分析 し、解釈する。				ステップ 5 データ収集の方法を特定 する。
→				→

ステップ 7 母集団を特定し、サンプリングする。	β	β β β β β β β	β	ステップ 6 データ収集の方法を選択する。
-----------------------------	---	---------------	---	--------------------------

2-1 評価対象となるプログラムを特定し、その内容を説明する (ステップ1)

- 評価の対象とするプログラムを特定し、その内容を説明する。

プログラムの内容を説明するに当たっては、以下の項目を含むものとする。

- プログラムの目的と目標
- プログラムが対象とする地理的範囲
- 対象となる顧客
- プログラムの資金提供者
- プログラムの運営に当たるスタッフ

- 情報収集の対象となるオーディエンスを特定する。

2-2 プログラムの進捗状況と必要な評価作業のタイプを特定する (ステップ2)

ニーズ評価、基準の検討、形成的評価、総括的評価、及び、フォローアップ作業など、評価作業にはさまざまなタイプのものがある。実際にどのタイプの評価作業を行うかの選択は、プログラムの進捗状況、プログラムの要求項目、及び、ステークホルダーの関心に基づいて行う。

プログラムの進捗状況と必要な評価作業のタイプの特定

質問		プログラムの進捗状況を特定する。		評価作業のタイプを選択する。
プログラムは企画段階にあるか。	○	プログラム企画段階	○	ニーズ評価
プログラムは開始されたばかりか。	○	プログラム開始段階	○	基準の検討
プログラムは実施中か。	○	プログラム実施段階	○	形成的評価
プログラムは終了段階	○	プログラム取りまとめ段階	○	総括的評価

階か。		階		
プログラムは終了したか。	0	プログラムのフォローアップ段階	0	フォローアップ作業

評価作業のタイプ

ニーズ評価 (needs assessment)

ニーズ評価では、対象となるオーディエンスが持つニーズの特定、プログラムの理由付けの明確化、必要とされるインプットの特定、プログラム内容の決定、プログラム目的の設定に焦点が合わされることになる。ニーズ評価では、現に何が存在し、何が求められているかについての質問が行われることになる。

- ・何をする必要があり、その理由は何か。
- ・オーディエンスが期待しているものは何か。
- ・プログラムを実施するにはどのような資源が必要か。

基準の検討 (baseline study)

基準の検討では、将来実施するプログラムやプロジェクトの影響を判定するための基準を策定することになる。基準の検討では、以下のように、現に存在するものについて質問が行われる。

- ・プログラムの現状はどうなっているか。
- ・オーディエンスの、現状における知識、スキル、態度、及び、信念のレベルはどの程度か。
- ・優先的に介入すべき分野は何か。
- ・手元にある利用可能な資源は何か。

形成的評価, プロセス評価, 又は発展的評価 (formative, process, or developmental evaluation)

形成的評価, プロセス評価, 又は発展的評価と称される評価では、プログラムの改善, 修正, 及び, マネジメントにかかわる情報が提供される。形成的評価では、以下のような記述的な質問が問われることになる。

- ・何を行おうとしているのか。
- ・現に何をしているか。
- ・どのようにすれば改善が図られるか。

総括的評価, 影響評価, 又は判定評価 (summative, impact, or judgmental evaluation)

総括的評価, 影響評価又は判定評価と称される評価では、プログラムの、全体的に見た成功度, 効果, 及び, アカウンタビリティを判定することに重点が置かれる。この評価は、当該プログラムを継続すべきか, 拡大すべきか, 縮小すべきか, 終了すべきかについて重要な決定を行う上で役に立つ。総括的評価では、以下のように、すでに行われたことについて質問が行われる。

- ・プログラムのアウトカムはどうであったか。
- ・プログラムに誰が参加し、その役割は何であったか。
- ・費用はどのくらいかかったか。

フォローアップ作業 (follow-up study)

フォローアップ作業では、長期的な観点からプログラムの効果を検討する。フォローアップ作業では、以下のような、長期的な影響に関する質問が行われる。

- ・プログラムの影響はどのようであったか。
- ・プログラム参加者にとってもっとも有益であったものは何か。
- ・長期的な影響は何か。

2－3 評価作業実施の可能性を検討する (ステップ3)

プログラム評価を実施する可能性を検討することによって、当該プログラムが有意義に評価されるようにし、更には、当該評価が、プログラムの設計、又は業績を改善することに貢献できるようにする。以下に示す質問事項を注意深く検討し、今がプログラム評価を行うべき適当な時期か否かを決定することが必要である。これらの質問の多くに対する答えが「いいえ」であれば、今はまだ評価作業を行う適当な時期ではないということになる。

- ・当該評価に基づいてなされるべき重要な決定事項はあるか。
- ・当該評価に基づく調査結果を活用するという約束があるか。
- ・当該評価に基づく調査結果に関わりなく、重要なプログラム上の決定が行われることになっているか。
- ・評価を行うべき法的責務が存在するか。
- ・当該プログラムが、公式な評価を行うことを正当化するだけの十分な影響、又は、重要性を備えているか。
 - ・当該プログラムは一時的なものか。
 - ・当該プログラムは継続的なものか。
 - ・当該プログラムの実施費用は、評価を必要としないほど安価であるか。
- ・当該評価を行うことによって、有用で、信頼性の高い情報を得る可能性はあるか。
- ・当該評価は受け入れ可能な正当性基準を満足できる可能性があるか。
 - ・当該評価は専門的職業に関わる原則に違反する恐れはないか。
 - ・当該評価が利害関係の衝突に脅かされることはないか。
 - ・当該評価がプログラム参加者の安らぎを脅かすようなことはないか。
- ・当該プログラムは評価を実施し得る状態にあるか。
 - ・総括的評価を行うことが適当と思われる場合、当該プログラムは、明確に定義されたアウトカムを提供し得る程度に十分長期にわたって実施されてきているか。

- 評価を実施するのに十分な人的・財政的資源が用意されているか。
- 評価を完了するのに十分な時間的余裕があるか。

2-4 主要なステークホルダーを特定し、その意見を聴取する (ステップ4)

ステークホルダー (stakeholder) とは、評価結果に利害 (stake)、あるいは既得権 (vested interest) を有する者を言う。ステークホルダーは、当該プログラムに対する出資者であったり、スタッフ、管理者、顧客、あるいは、何らかの形で当該プログラムに参加している者であったりする。評価を開始するに当たっては、主要なステークホルダーとの間で、評価の目的、手順について明確にし、確認しておくことが重要である。このプロセスを入れることによって、必要とされる評価のタイプを確定し、当初考えられていたよりも生産的な評価の理由を付け加えることも可能になる。

ステークホルダーと以下の点について合意に達すること

- 評価対象のプログラムはどれか、また、評価対象の中に何を入れ、何を除くか。
- 評価の目的。
- プログラムの目的及び目標。プログラムの目的及び目標は、当該プログラムが達成すべきこと、及び当該プログラムの目標が達成されたか否かを判定するために使用される基準をどうするかを示したメッセージの形で記載することが可能である。

個々の目標は、以下の要件を備えるものとする。

- 一つのアウトカムを含むこと
- 対象となるオーディエンスを特定していること
- 当該プログラムへ参加することによる成果として何を変えることを期待しているのかを具体的に記載すること
- 達成度が測定可能な程度に具体的なものであること

例：インガム郡に居住する全世帯の成員は、ミシガン州立大学が実施する調査に参加することによって、水質に関する意識を高める。

- プログラムの価値を判定するために使用する指標及び基準。プログラムの目的が明確に記述されている場合は、プログラムの価値を判定するための指標及び基準は明確に記述されているはずである。
- 当該評価が取り扱う質問と問題
- 当該評価に誰が参加するのか。
- 当該評価に割くことができる予算と時間
- 評価者の役割
- 当該評価の成果を受け取る者は誰か。

評価に関わる質問、課題、指標、及び基準を明確にすること

評価は、具体的な質問に答え、プログラムに関わる課題を取り扱い、将来のプログラムを計画し、あるいは、既存のプログラムの価値を判定する基準を適用することを目的に実施される。評価に使用される質問や課題が明確にされていなかったり、プログラムの価値を判定するために使用される指標、基準が十分な検討を経ていないものである場合には、当該評価は、焦点がぼやけたものとなり、見当違いの評価をしたり、重要な分野の項目を見落としていたり、あるいは、根拠のない結論を導いたりすることになる。

質問、課題、指標、及び基準を選択するに当たっての基本的ステップ

協議を行ったすべての当事者から提起された質問、課題、基準をリストアップする。

得られた材料を管理可能な数のカテゴリーにまとめる。プログラムのレベルを指標に合わせる。一つの評価で、関心のあるすべての分野に対応することは不可能であることを忘れてはならない。

与えられた金銭的、時間的な制約の中で、どの程度の不完全性であれば許容できるかについて、ステークホルダーとの間で合意しておく。

評価の範囲を、欠かすことのないもので、かつ実施可能な事項に限定する。

ステークホルダーと話し合うことに加えて、評価の目的を明示し、質問、課題、指標、基準を作成するに当たっては、以下の点も考慮に入れるものとする。

- 各種の評価モデル、入手可能な文書を検討すること
- プログラムの対象分野に関連する専門的な基準、指針を調査すること
- 当該分野の専門家に相談すること
- 自分の専門的判断を活用すること

指標に関する合意に到達する

指標は、状況によって変化する変数である。変数は、システムの属性(品質、特性、特徴)を運用可能な形で表現したものである。指標は、さまざまな状況、プログラム、及びアウトカムに関する意図された、あるいは、現実には生起している諸条件を指し示す目に見える現象であり、自然のさまざまなシステムや、人間の努力による業績を判定することを容易にするものである。

指標とは、何かが変化したことを示す観測可能な目印である。指標は、プログラムの影響が現れ始めた初期の段階において、そうした変化を人びとに気づかせるものである。

指標が備えるべき特性

1. 評価対象のプログラムの目的に関連したものであること
2. 理解可能なものであること。すなわち、シンプルで、あいまいさを排除した明瞭なものであること
3. 支援体制、時間、技術などの各種の制約条件が満足されれば、実現可能なものであること
4. 概念として、十分な根拠を有しているものであること
5. 数が限定されており、一定の間隔で更新することができるものであること

指標を選択する基準

1. 測定可能であるか。
2. 関連性がある、利用しやすいものであるか。
3. 標準的な状況を提示できるようなものであるか。
4. 解釈しやすいものであるか、また、長期にわたる傾向を示すことが出来るものであるか。
5. 変化に対応できるものであるか。
6. 比較するための参考資料がある、ユーザーがその価値の重要性を評価することが可能なものであるか。
7. 妥当なレベルの費用で計測することが可能であるか、また、更新することができるか。

ベネットの証拠に関する階層

ベネット (Bennett) の証拠に関する階層 (Hierarchy) は、プログラムの異なるレベルにおける目標とアウトカムの関係を概念化する方法を提供するものである。この階層は、ある目的が達成されたかどうかを判定する手段として、どんな情報が適切であるかということを示してくれる。この階層は、あなたが収集した情報が、あなたが評価しようとしているプログラムのレベルに対応した適切なものであるかどうかを確認する上で役に立つ。

プログラムのレベル	指標
最終成果 (End results)	プログラムへの参加による成果として参加者の個人的生活や職業生活に生じる変化
習慣、及び行動の変化 (Practice and behavior changes)	プログラムへの参加による成果として生じた参加者の習慣の変化
知識、態度、スキル、及び願望における変化 (Knowledge , Attitude, Skill and Aspirational Changes, KASA)	プログラムへの参加による成果として生じた参加者の知識、態度、スキル、及び願望の変化
反応 (Reactions)	参加者や顧客が当該プログラムに対して示した反応
参加 (Participation)	誰が、何人参加したか
活動 (Activities)	プログラムを通じて参加者が行った活動。プログラムへの参加者が交流するために使用した情報や方法の種類
インプット (Inputs)	プログラムの実施に使用した人的資源及びその他の資源

2-5 データ収集の方法を特定する (ステップ5)

データ収集の基本的な方法として、定量的方法と定性的方法の2種類がある。定量的データは、数量データに焦点を当てる傾向があり、一方、定性的データは、文章で表現される。

定量的方法 (Quantitative Methods)

定量的方法では、あらかじめ定められ、限られた複数のアウトカムを測定するもので、効果を判定し、原因を探り、成果を比較あるいはランク付けし、分類し、かつ、一般化するのに適している。定量的方法は、以下点で優れている。

- 大規模なプロジェクトに適していること。
- 因果関係を判定するのに有用であること。
- 信頼性が高いという評価を得ていること。
- 大規模な母集団に適用可能で、一般化できること。

定量的方法は、一般に、普及プログラムの評価に使用される。しかし、既存の情報には限らない。	情報及び知識をテストする
調査	費用便益分析
グループを対象とした質問書	個人ベースのインタビュー

定性的方法 (Qualitative Methods)

定性的方法は、対象者、場所、会話、及び行動に関する豊富な説明を含め、数多くの形式を取る。定性的方法では制約条件がないため、インタビュー対象となった個人は、自分の考えに基づいて質問に答えることができる。定性的方法は、以下の点で優れている。

- プログラムが実施されている背景を理解することができる。
- 複雑な問題やプロセスに関わる問題に対応することができる。
- プログラムの目標と実施の間の関係を明らかにすることができる。
- プログラムが意図していない結果を特定することができる。
- 記述的情報を収集することができる。
- プログラムの運用と効果を理解することができる。
- プログラムの影響を詳細に分析することができる。

定性的方法は、通常、必ずしもこれらに限定されないが、以下に例示したような各種の普及プログラムの評価に使用されている。	
既存情報	個人インタビュー
フォーカスグループ	簡易農村調査
参加者観察	ケーススタディ
グループインタビュー	

多角的方法 (Multiple Methods)

多角的方法とは、一つの評価作業において定性的方法と定量的方法を組み合わせたものである。二つの方法を組み合わせることによって、定性的方法、定量的方法を単独で実施した場合に発生する偏りを排除し、それぞれの方法が備えるメリットによって相互に補完することが可能となる。多角的方法を使用する場合には、選択した方法が評価のための質問に適したものであり、資源があまりに拡散されることがないように配慮しなければならない。多角的方法は、以下の点で優れている。

- 複雑な社会現象を理解することができる。
- 非常に多くの観点と関心事項を許容することができる。
- 標準的なケースと例外的なケースのいずれについても理解を深めることができる。
- より深く、広い考察を行うことができる。

多角的方法の一例：果樹園プロジェクトの評価	
文化的、政治的に複雑な状況にあつては、多角的方法が特に適している。以下に示す方法は、グアテマラのペテンにおける先住民グループと移住民グループが行った果樹園プロジェクトの評価について、定性的方法と定量的方法が組み合わされたものである。	
コミュニティへの紹介	果樹園訪問と果樹調査
○ 簡易農村調査	● 参加者観察
● コミュニティ・マップ	● 写真
	● 地図作成
標本抽出方針の策定	● 果樹調査ツアー
● 連鎖インタビュー，便乗インタビュー	● 果樹の特定
● 重要な情報提供者を特定する	
○ 重要な情報提供者への接触	フォーカスグループ・インタビュー
焦点は定まっているが，体系化されていないインタビュー	データ分析
	● 定性的分析
果樹園プロジェクト訪問	● 果樹データの定量的分析
● コミュニティ 6 か所を訪問	
● 普及員とのインタビュー	

証拠の品質（Quality of Evidence）

「データ収集用質問」（date collection instrument）の有効性と信頼性が、定量的方法における証拠の品質を決定する。

有効性（Validity）

データ収集用質問は、測定しようとするものを測定するためのものである。収集されたデータは、特定の状況又は特定のオーディエンスに関連したものである。

- 測定しようとするものをはっきりと定義する。
- データ収集用質問に含めるアイテムを明確化し、あるいは、展開する。
- 大まかなドラフトを作成する。
- 内容、フォーマット、及び、オーディエンスに対する妥当性を判別するため、データ収集用質問をレビューする「専門家パネル」を選定する。
- 専門家のアドバイスをもとに、データ収集用質問を改訂する。
- 明確さ、内容、言い回し、及び文章の長さの可否を測るためにフィールド・テストを実施する。
- 必要に応じて、データ収集用質問を改訂する。

信頼性 (Reliability)

データ収集用質問は、同じ条件の下では、同じ対象グループについて、同じ結果を導き出すように、首尾一貫した測定結果を出せるものでなければならない。

- 対象としているオーディエンスが持つ特性と類似した特性を有する小グループについて、パイロット・テストを実施する。
- 同一グループを対象として、1週間後に同一のデータ収集用質問を再度使用して回答を収集し、結果を比較する。信頼性テストを実施するためにコンピュータ・ソフトウェアを利用することができる。
- 結果が食い違う場合、質問又はアイテムを変更する。場合によっては、質問の文言を変更したり、信頼性を高めるために、若干のアイテムを追加したり、あるいは、質問自体を削除することも必要になる。

2-6 データ収集方法を選択する (ステップ6)

プログラム評価に当たってデータを収集する場合、ベストの方法はこれしかない、ということとは決していない。ある一つの方法を選ぶか、あるいは、複数の方法を選ぶかは、必要とされる情報の種類、作業に使うことができる時間、及び費用によって大きく影響される。大事なことを一つ言い残したが、収集された情報が、あなたが所属する組織によって、信頼し得る、正確で、利用価値の高いものと見なされるようなものであるか否かを考えなければならない、ということをつけ加えておく必要がある。これは、最後に付言したことだからといって重要ではないということではない。

評価に使用することができる方法は数多くある。ここでは、次の9つの方法と費用効果分析について述べる。

定量的方法	定性的方法
(1) 既存の情報	(5) フォーカスグループ
(2) 情報と知識のテスト	(6) 簡易農村調査
(3) 調査 (電話、郵送、個人インタビュー)	(7) ケーススタディ
(4) グループを対象とした質問	(8) 準構成的インタビュー ¹
	(9) 参加者観察

¹ 訳注：回答の選択肢が予め用意されているような質問が周到に構築されたインタビューのこと。ここでは、これに準じる形のインタビューとあるので、指示的インタビューと自由回答式インタビューとが折衷された形のインタビューを指す。

(1) 既存の情報 (Existing Information)

データ収集を開始する前に、どのような情報が既に存在しているかを確認する必要がある。既存情報は、各種の文書、報告書、プログラム記録、過去に作成された報告、会議の議事録、書簡、写真、国勢調査資料、及び各種の調査記録などで見出すことができる。

既存の情報は、以下のことに対して有用である。

- プログラムのニーズを確定する。国勢調査資料、各種メディアの特集記事、地図、あるいは、各種のサービスや事業に関する統計などを使用する。
- プログラムの実施方法や対象者の特定方法を明らかにする。プログラム文書、詳細な会議録、登録者記録、各種メディアの発表文書などを使用する。
- 成果をアセスメントする。各種の公的記録、地方の雇用統計、当局のデータ、及び類似のプログラムの評価結果などを使用する。

既存の情報を使用する利点

- ほとんどの場合、入手しやすい。
- 少ない費用負担と努力で入手できる。
- 広範な種類の特性を備えたデータが入手できる。
- 単発的でなく、長期にわたって、連続的に入手することができる。
- 信頼性が高い。

既存情報をデータ・ソースとする場合の不利な点

- データが記述的なものである場合が多く、利用する場合に、分類、選別、及び相互の関連付けが必要になることがある。
- 数字の中には、実績値ではなく、推定値や見積値を表しているものがある。
- 現在生じているトレンドのベースとなっている価値、理由、あるいは、信念を明らかにしてくれない可能性がある。
- 地域のコミュニティのデータには、しばしば、限界があり、いつも最新のものであるとは限らない。
- 実態について偏った見方を示していることがある。

(2) 情報と知識のテスト (Testing Information and Knowledge)

特定のプログラムに関して個人が保有している知識、理解、及び能力のレベルを計測するツールとして、テスト結果を利用することができる。

情報と知識のテスト結果を使用する利点

- 特定のプログラムに関わる知識レベル、その他の変化を示すことができる。
- 比較的実行しやすい。

- グループ設定をして実行することができる。
- 安い費用で実行できる傾向がある。

情報と知識のテスト結果を利用する場合の不利な点

- 成人は、しばしば、自分の知識をテストされることに抵抗を示す。
- 知識の獲得と行動の変化が無関係である場合がある。
- テスト結果を有効で信頼性の高いものとするには、特別なスキルが必要で、そうしたスキルを開発するには時間がかかる。
- 識字能力に劣るオーディエンスにはあまり適していない。

知識と情報の変化をテストする場合の基本的ステップ

1. プログラム実施中に提起された特定の問題に焦点を当てた質問からなるテストを構成する。
2. 批判的ではないトーンで質問の文言を作成する。
3. 口頭又は文章でプログラム参加者に対してテストを行う。
4. グループ・ディスカッションで回答を得た場合には、あらゆる種類の回答を引き出すようにする。
5. 個々人の回答は、定型化された文書に記載してもらうか、非公式な論議によって入手する。
6. 回答を格付けするか、又は、正しいとされている回答と比較する。
7. プログラムを通じて達成された理解の程度を総括し、分析する。

(3) 電話調査, 郵送調査, 面談調査 (Telephone Survey, Mail Survey, Personal Survey)

調査は、評価データを収集する上で、ごく一般的に使用されている方法であるが、郵送、電話又は個人面談によるインタビューの場合には、慎重にデザインされた質問書が必要である。調査は、参加者の知識、態度、スキルと願望、いつものやり方、及びプログラムの利益と影響に関するデータを収集するために使用することができる。倫理基準が常に維持されるようにすることは、評価者の責務である。つまり、調査に参加するか否かは自発的意思によるもので、調査結果は、秘密を保持できるような方法で公表されることを意味する。

調査を利用する利点

- かなり複雑な内容の質問も設定できる。
- 回答者の匿名性を確保できる。
- 費用があまり多くかからない。
- 多数の人々を対象者とするのが容易である。
- 母集団が広範囲に分散している場合に有用である。

調査を利用する場合の不利な点

- 調査では、因果関係が容易に明らかにならない。
- 調査は、異文化環境では利用し難い。
- 調査を実施するには、かなり識字能力のある母集団が必要になる。
- 正確な、最新の回答候補者リストを見つけ出すことが困難な場合がある。

調査を実施する前に答えを見出しておくべき重要な質問事項

- 調査を有効なものとするには、何人の対象者が必要か。
- 質問書に回答すべき者は誰か。少数の人間か、あるいは、特定のグループか。
- 標本はどのようにして選定するか。
- 回答率はどの程度を期待するか。
- 調査結果の正確度はどの程度になるか。

調査方法を選択するに当たっては、どんな資源が利用可能かを検討すること

- 調査を実施する調査担当者に報酬を支払うか、否か。
- あなたに与えられた期間と予算。
- 調査作業に経験を持つ者がアシスタントとなってくれるか。
- 電話によるアクセスなど、どんな施設が利用可能か。

電話調査

電話調査は、選択されたグループを対象に、電話で、記載された質問書を読み上げて調査を行うものである。調査標本は、電話帳又はその他の人名録から選び出される。調査リストに記載された調査対象者に対しては、一人ずつ、電話でインタビューが行われる。

電話調査の利点

- 回答者が地理的に広範囲に散在している場合に利用可能である。
- 傾向的に高い回答率が期待できる。
- 郵送質問書による場合よりも、複雑な質問設定が行える。
- 回収が迅速で、かつ、効率的なデータ・ソースとなる。
- 郵送質問書による場合よりも、特定の回答者を選定することが容易である。
- 質問者は回答者に対して質問を詳細に説明することができる。

電話調査の不利な点

- 質問は、明瞭で、簡略であることが求められる。
- 調査に時間がかかることがある。
- 回答 1 件当たりの費用は、郵送調査と同程度である。
- インタビュー担当者のインタビュー・スキルと訓練を受けた監督者が必要である。
- 電話を備えていない所帯や電話帳等に記載されていない所帯が調査対象から除外されるため、調査結果に偏りが出る可能性がある。

- 電話をするタイミングが非常に重要であり、それによって、偏りが出る可能性がある。
- インタビュー担当者の口調やマンネリ化した質問の仕方によって偏りが出る可能性がある。
- 信頼性の高い電話システムとインタビュー調査を行うのに適した十分な場所が必要である。

電話調査の実施

1. 調査を実施するのに適した施設と必要な設備を確保する。
2. 標本単位の中でどの回答者を選択するかの方法を含め、標本設計について決定する。標本抽出に使用する電話番号プールを作成する方法を選択する。
3. 調査マテリアルを準備する。氏名と住所が判明している場合には、電話調査を行う旨の事前通知、質問書、識別番号を記録するカバーシート、コールシート、インタビュー担当者用のヘルプ・シート
4. インタビュー担当者を訓練する。調査に関する背景情報、電話調査の基本点、装置の使用方法、質問書、コールシートの記載方法
5. インタビュー予定を作成する。仕事中か、仕事をしていないときか、回答者にコンタクトする時間帯としてどちらにするか検討する。ほとんどの調査の場合、インタビューを完了するには約 50 分もあれば十分である。回答を拒否された場合にどう対応するかを決めておく。
6. 電話をする。電話番号 1 件当たりの呼出し回数を決めておく。ローカル調査では、呼出し回数を 6 回ないし 7 回とするのが通例である。駄目な場合は、かけ直す。

電話インタビュー用コールシート（サンプル）

抽出枠から選び出した電話番号ごとにコールシートを作成する。インタビュー担当者は、対応した電話番号ごとにどう処理すべきか監督者が決定できるような情報を記録する。電話インタビューが完了したら、コールシートは質問書に添付する。

電話インタビュー・コールシート				
調査タイトル				
質問書 ID 番号 _____				
エリア・コード, 電話番号 () _____ - _____				
コンタクト回数	日付	時間	結果コード, コメント	インタビュー担当者 ID
1				
2				
3				
4				
5				

6				
<u>追加コメント</u>				
結果コード				
コード	説 明			
01	7 回呼出し後でも応答なし			
02	1 回リダイヤルした後，話中となる			
03	留守番電話(自宅)			
04	自宅であるが，話を通じない			
05	非居住者が応答した			
06	自宅であるが，質問を拒絶した			
07	接続しない，その他電話番号が使用できない			
08	一時的に接続しない			
09	オフィス又はその他の業務用電話番号である			
10	回答資格要件を満たす者がいない			
11	該当電話番号にコンタクトできただけである			
12	選定された回答者が一時的に捕まらない			
13	選定された回答者が調査対象期間中不在である			
14	選定された回答者が身体的，精神的障害のために対応できない			
15	選定された回答者に言語的障壁が存在する			
16	選定された回答者が回答を拒否した			
17	一部の質問項目についての回答にとどまった			
18	回答者にコンタクトした。インタビューを完了した			
19	その他			

インタビュー担当者用ヘルプ・シート(サンプル)

1. 後援団体の名称
・ 調査研究の目的
・ 調査担当者
・ 調査規模

・ インタビュー担当者の身元
・ 回答者の氏名の入手方法
・ 秘密保持に関する事項
・ 調査結果写しの入手方法
・ 調査結果の利用方法

郵送調査

郵送調査は、普及プログラムの評価を行うときに、もっとも多く利用されている調査方法で、必要な資源も最も少なくてすむものである。

郵送調査を利用する利点

- 大規模な標本規模の調査に利用することができる。
- 母集団が広範囲に分散している場合や電話や個人インタビューではアクセスできない場合に利用することができる。
- 質問を目に見える形で提示する。
- インタビュー担当者による偏りが発生する恐れがない。
- 回答者は十分考慮した上で回答することができ、回答に要する時間や順序も自分で決めることができる。
- 他の方法に比べて費用がかからない。

郵送調査を使用する場合の不利な点

- 質問事項は簡明で、十分注意して作成する必要がある。
- 回答率は、回答者との接触回数や郵送時期に左右されることが多い。
- 回答が完全な形で行われているかどうかをコントロールできない。
- 回答者が対象としている母集団に属する人でない場合がある。
- 費用に響くミスを避けるために、質問書を事前にテストする必要がある。
- 郵送調査では、データ収集に 4 週間ないし 6 週を要する。
- 識字能力が高い母集団が存在していること、信頼性の高い郵便システムが存在していることが要求される。

郵送調査を実施するに当たっての基本的なステップ

1. 調査マテリアルを準備する。書面の質問書を作成し、回収の有無を追跡するために、質問書の表紙に ID 番号を付ける。
2. データ収集用質問の事前テストを行い、有効性、信頼性を確認する。
3. 標本母集団を選択する。
4. 郵送スケジュールを策定する。a) 2 週間前に質問書を郵送、事前予告レターを発送する。
b) 添え状、切手を貼付した返信用封筒とともに、質問書を郵送する。c) 1 週間後程度にはがきを郵送し、既回答者には謝意を伝え、未回答者には調査に協力してくれるよう依頼する。d) 最初に質問書を郵送してから 3 週間後に、同じ質問書、切手を貼付した返信用封筒を同封し、回答を受領していない旨を記載したフォローアップ・レターを郵送する。
郵送スケジュールを検討するに当たっては、休暇期間、特に 12 月を避ける。ほとんどの場合、回答率 60%ないし 90%で十分であると考えられている。

面談調査

面談調査は、回答者に個々にインタビューし、質問ごとに回答者の答えを系統的に記録していく方法である。

面談調査を行う利点

- 広範囲に分散している母集団について利用することができる。
- 代表標本が取り出せない母集団に適している。
- 識字能力が低い集団に対しても利用することができる。
- 調査回答者を誰にするかコントロールできる比率が高い。
- インタビュー担当者が、回答者の質問に対する回答意欲を高めることができる。
- 調査の質問意図を分かりやすくする視覚資料を使うことができる。
- かなり複雑な質問をすることができる。

面談調査を行う場合の不利な点

- 費用も時間もかかる可能性がある。
- インタビュー担当者を十分注意して選び、適切な訓練を与える必要がある。
- 能力に優れた監督者が必要である。
- 電話調査、郵送調査に比べて、より多くのマテリアルが必要になる。

面談調査を実施するに当たっての基本的なステップ

1. 調査マテリアルを作成する。a) 回答依頼先の氏名、住所が判明している場合に郵送する事前予告レター、b) インタビュー担当者のネーム・プレート、c) 回答者の手元に残しておく調査目的を説明した紹介状、d) インタビュー担当者用のマニュアル、e) インタビュー担当者用の標本抽出情報、f) 質問書
2. インタビュー担当者を指定し、訓練する。
3. インタビュー調査の手順を説明し、訪問先を明記したレターをインタビュー担当者から

郵送する。

4. 公的機関に調査の実施を通知する。
5. インタビューを実施する。インタビューに当たって発生するあらゆる問題进行处理する監督者を、常時待機させ、電話で連絡できるようにする。
6. 監督者は、インタビュー担当者と定期的に会合を持ち、質問書の編集、変更を行い、インタビュー担当者からの質問に適宜答えるものとする。インタビュー担当者が犯す、費用を伴うミス、誤解、不正行為は、この際に発見することができる。
7. インタビュー調査が完了すると、質問書は調査監督者のもとに回収される。

調査インタビューの一般的な手順

インタビュー担当者による偏りを最小限にするには、

- 小奇麗な容貌を保つ。
- 標本抽出プランに従って回答者を特定する。
- 回答者に対して、丁寧に、正直に対応する。
- 調査の目的を十分に理解してもらう。
- 質問は、書面どおりに正確に行う。
- 回答を正確に記録する。
- データ収集用質問書に十分習熟する。
- 標本抽出指示事項に従う。
- データ収集を完全に行ったかチェックする。

回答者に接触を開始するには

- 自己紹介し、資格認定書を提示する。
- インタビュー担当者訪問の数日前に、回答者が事前通知レターを受領したか否かを確認する。
- 調査の目的を説明する。
- 回答者の回答の機密が保持される旨を伝える。
- なぜ回答者として選定されたかを説明する。

インタビュー調査実施に当たってのガイドライン

- 標本抽出基準に従って回答者を選定する。
- インタビューを実施するか、又は、都合の良い時間を確認する。
- 回答者の注意が散漫になることを避けるために、同席者がいない状態でインタビューを行うよう努める。
- 回答者に、インタビューは任意のものであって、回答内容の機密が保持される旨を喚起する。
- 回答者が回答してくれたこと、積極的に参加してくれたことを感謝することを伝え、意思の疎通を図る。
- 質問書に記載されているとおりに質問を読み上げ、回答を正確に記録する。

- インタビュー担当者の意見は一切表示しない。
- 自由回答式の質問に対する回答が完全なものでなかったり、関連がないと見られる場合には、明確な回答を得るよう確認を行う。
- 回答者が質問への回答を拒否した場合は、回答するよう無理強いしない。回答を無理強いすることで、インタビュー全体が台無しになる恐れがある。

エラーを避け、品質の高い調査を行うには

調査にも、精度の高いもの、低いもの、いろいろなものがある。調査の精度とは、調査結果が、標本を抽出した母集団を密に代表しているか否か、ということである。調査の誤りは、標本範囲の誤り、標本抽出の誤り、標本選定の誤り、構造上の誤り、非回答による誤り、判定の誤りなどによって引き起こされる。

誤りの種類	誤りの原因	誤りの防止法
標本範囲の誤り	抽出枠に母集団のすべての要素が含まれていない。	母集団のすべての要素が含まれるように、標本を抽出するリストを再度作成する。
標本抽出の誤り	全数調査を実施するのではなく、母集団に属するすべての人の中の小集団又は標本が調査対象となっている。	標本の規模を大きくする、無作為抽出法を使用する、重複リスト分を削除する。
標本選定の誤り	抽出単位の中に、他の単位に比べて選定される機会が大きいものがある。	無作為抽出法を使用する。
構造上の誤り	標本リストが正確でなかったり、特定の抽出単位が除外されている。	最新の正確な標本リストを使用する。
非回答による誤り	対象者の所在が特定できない、又は、回答してこない。	期限内に回答してきた回答者と期限に遅れて回答してきた回答者の回答結果を比較する。明確な差がない場合には、回答者を合算して、大きな母集団のものとして調査結果を一般化することができる。 非回答者の約 10%と接触し、データを収集する。当該データを回答者データと比較する。明確な差がない場合には、回答者を合算して、大きな母集団のものとして調査結果を一般化することができる。

		判明している特性について回答者と非回答者とを比較する。明確な差がない場合には、回答者を合算して、大きな母集団のものとして調査結果を一般化することができる。
判定の誤り	一部の回答者の回答が正確でなかったり、他の回答者の回答との比較がうまくできない。この原因は、質問が明確でない、回答についての指示が明確でない、回答者が正しい情報を知らない、意図的に虚偽を言っているというのではなく、社会的に見て正しい回答をしている、などが考えられる。	適正なデータ収集方法を選択し、評価を行う。 回答者が正確に回答し、回答意欲が沸く、明確な、あいまいなところがない質問を作成する。 インタビュー担当者の訓練を十分に行う。 有効で、信頼の置けるデータ収集用質問を使用する。

(4) グループを対象とした質問

グループを対象とした質問書は、ワークショップ、セミナー、プログラムなどの終了時に、参加者一人ひとりに直接、手渡される。回答者は、一人ひとり質問に答え、評価者に質問書を返す。

グループを対象とした質問の利点

- 回答者がよく理解している目的又は目標に直接関係する質問を行うことができる。
- 費用が安く、管理がしやすい。
- 回答者にとって質問に回答しやすい。
- 直ちに結果を出すことができる。
- 全母集団を標本とすることができる。
- 論議を行うベースとして使用することができる。

グループを対象とした質問の不利な点

- プログラム参加者が大きな母集団の代表ではない可能性があり、調査結果を一般化できない場合がある。
- 質問書は、単一のトピック又は課題しか取り上げられないことが多い。
- プログラムが終了した後、時間がかかる。

グループを対象とした質問書により質問を実施するに当たっての基本的なステップ

1. 調査用質問書作成ガイドラインに従って質問書を作成する。ただし、質問書に回答する目的、回答方法については、インストラクター、監督者又は実務作業者が説明する。説明に当たる者は、参加者に対して、匿名性を保持する旨を保証するものとする。
2. 質問書は参加者一人ひとりに配布し、参加者それぞれが記入する。
3. 質問書は、収集の上、完全に記載されているかチェックする。

質問書の設計

質問書の設計の全体としての目的は、参加者に質の高い回答を求めることにある。回答の質は、調査、調査対象となるトピックス、インタビュー担当者、質問の文言、質問の仕方について回答者が信頼を置くか否かにかかっている。質問書が郵送され、直接、回答者に手渡されるのか、電話調査の手法を使用するのか、あるいは、個人インタビューの手法を使用するのか、検討が必要である。調査を開始するには、調査結果を評価するに当たってどのような証拠を知りたいのか、情報をどのように利用するのかを事前に確認しておくことが重要である。

調査を開始する前に...

- 知りたいこと、情報をどのように利用するかを記載したリストを作成する。
- 情報が他の機関、組織ですでに存在していないことを確認する。
- 絶対に必要な質問以外はすべて削除する。
- 質問を作成するに当たっては、回答者の目で見してみる。

質問書の作成

1. 質問書のタイトルは、回答者にアピールするものとする。
2. 使用する字型は、大きく、読みやすいものとする。
3. 質問書は、専門知識に基づく形式とし、回答しやすいものとする。
4. 序文では調査対象者や調査目的を明示し、質問書の回答方法を指示する。
5. 質問は、内容が重なり合った印象を与えない。
6. 質問は、項目ごとに番号を付し、項目の中を分ける場合には、それが分かるように枝番号を付ける。
7. 質問は、論理的な順序に従って行うものとし、一般的事項を先にし、具体的な事項は後にする。回答しやすい質問を先に行い、複雑で、考えることが必要な質問や、デリケートな質問は後で行う。個人的な質問や督促するような質問は、最後に置く。
8. それぞれの質問で回答が求められている事項をはっきりと示す。
9. 自由回答式の質問については、十分な記入スペースを確保する。
10. 質問の枝分かれがどこで行われ、どこで一般的な質問に戻るかを明確に示す。
11. 読み間違えることがないように、キーワードは太字、又は、大文字を使用するものとする。
12. 性別、年齢等の情報を記載するよう求める文言は質問書の末尾に記載する。
13. 郵送調査については、回答者が質問書を返送すること、送り先が記載された、郵送料

支払い済みの封筒を投函するように注意喚起する。

14. 質問書の最後には、「ご協力ありがとうございました」と必ず記載する。

質問書作成に当たって考慮すべき特別な事項

電話調査に使用する質問書

電話調査に使用する質問書は、口頭によるコミュニケーションの仕方で受ける側の印象が変わる。そのため、質問書の策定に当たっては、特に注意が必要で、回答者の注意を可能な限り引くよう、質問担当者の努力を支援するものであることが求められる。質問書の設計、構造は、美的感覚よりも、使いやすさに重点を置くべきである。

導入

回答拒否が発生するのは序文の書き方、内容が原因となることが多いので、序文の作成には特に注意が必要である。序文には、以下の項目を記載するものとする。

- インタビュー担当者の氏名
- インタビュー担当者が電話している機関、市町村の名称
- 回答者の電話番号の入手方法
- インタビューの見込み所要時間(控えめなものとする)
- 調査の説明、調査結果の使用目的、男性又は女性を選定した理由、調査の正当性を確認するためのコンタクト先、電話番号など、回答者の質問にインタビュー担当者が回答するためのコールバック・シート

郵送調査の質問書

郵送調査の質問書の外見は、非常に重要である。郵送された質問書は、必ず回答がなされ、返送されるように、回答者にとって「魅力ある」ものでなければならない。このため、質問書の設計に当たっては、相当な注意が必要である。

- 簡単な冊子は、縦横 8.5 インチ × 11 インチの数のものを半分に折りたたむことができるものとする。
- 質問に答えるのに次のページを開かなければならないことがないように、質問、回答は 1 ページ内に収まるようにする。
- 質問の回答についての指示事項は、対応しやすいものとする。
- 質問と回答が上から下のフローになるようにする。回答者がページを左右に見るのではなく、上から下に直線的に見て行けるように、選択する回答は質問の右にではなく、すぐ下に置く。

質問表のカバー・レターをデザインする

第1パラグラフ：

- 調査目的を説明する。
- 質問書に回答を求められているのはどんな人々かを説明する。
- 回答については秘密が保持されることを保証する。

第2パラグラフ：

- 回答者に対して、この調査が有意義なものであることを保証する。
- この調査の成功にとって、回答者の協力が重要であることを理解してもらう。

第3パラグラフ：

- 何時、どのような方法で質問書を返送するか、指示する。
- 質問書に付された ID 番号はフォローアップを容易にするためのものであることを説明する。

第4パラグラフ：

- この調査の社会的意義を再度、強調する。
- もし希望されるなら、調査結果の報告書の写しを提供することを約束する。
- もし質問があれば、どんな質問にも喜んで答える用意があることを伝える。
- ご協力に感謝する旨の言葉を添えて、結びとし、発送者の名前と肩書きを明記する。

質問の作成

質問書に盛り込まれる質問は、あなたが行おうとしている調査の有効性を左右する基本的な構成要素である。良い質問を作成することは、それほど簡単なことではなく、また、通常は、何度も書き直しが必要となるものである。質問書に、どのような情報を含めるべきか、質問をどのように組み立てるか、また、回答者はこれらの質問に正しく答えることが出来るかといった要素を熟考する必要がある。優れた調査を可能とする質問は、焦点を絞った、明確な、的を射た質問である。

どんな質問も一つの課題や特定のトピックスに焦点を当てたものでなければならない。

下手な質問：あなたはどのブランドが一番好きですか。

上手な質問：これらのブランドの中であなたが一番買いたいものはどれですか。

これらの質問の目的は消費者の好みを判定するためのものである。最初の質問は焦点がぼやけている。消費者は特定のブランドを好むかも知れないが、それは価格が高いため購入しないかも知れない。

すべての回答者に質問の意味が明確に分かるものでなければならない。すべての回答者がその質問を全く同じように解釈できるようにするためには、質問は明確でなければならない。

下手な質問：あなたが自分自身の健康診断のために最後に医者に行ったのはいつですか。

上手な質問：あなたが最後に健康診断に行ったのは何か月前ですか。

最初の質問の場合、回答者は、週、月、年、あるいは、日付で答える可能性がある。

質問は出来るだけ短くする。短い質問には答えやすい上、インタビュー担当者にも回答者にも解釈の相違による誤解が生じる余地は少ない。長い質問は、得てして、焦点がぼやけ、明瞭性を欠いたものとなり易い。

下手な質問：あなたはお子さんを何人お持ちですか。その子たちは男の子ですか、女の子ですか。また、何才ですか。

上手な質問：あなたのお子さんたちの年令と性別を教えてください。

回答者は最初の質問にはあいまいに答える可能性がある。例えば、「私には男の子が二人と女の子が一人います。その子たちの年令は、5才、7才、10才です。」しかし、この回答では、どの子が何才かを特定することは出来ない。

質問は偏見を排除するような形で作成されなければならない。

下手な質問：我々の職員はいつも長時間働いているというのは本当ですか。

上手な質問：普及員は、平均して、何時間働いていますか。

情報のタイプ

質問は、すべて次の4つのタイプの情報を引き出すような形で構成される。すなわち、①知識、②信念、態度、及び意見、③行動、④属性である。一つの質問には、これらの情報タイプのどれか一つ、もしくは、複数を組み合わせたものを含めることが出来る。

①**知識**に関する質問とは、人々が何を知っており、何かについてどれほど良く理解しているかを問う質問である。

・家庭における子供たちの事故死の主な原因は何ですか。

②**信念、態度、及び意見**に関する質問とは、人々の理解や、思想、感情、判断、あるいは、考え方を問う質問である。

・マイナー郡のクリアーウオーター地域教育センターは、大学レベルの、もしくは、成人教育のコースやプログラムを今後とも提供し続ける必要がありますか。

③**行動**に関する質問は、人々に、自分が過去にやったこと、現在やっていること、もしくは、将来やろうと計画していることを問う質問である。

- ・あなたや、あなたの家族は、これまで、マイナー郡のクリアーウオーター地域教育センターのクラスを受講したことがありますか。

④**属性**に関する質問とは、人々の年齢、学歴、職業、あるいは、収入といった人間的特性である。属性に関する質問とは、回答者に、あなたは今、何をしているのですかと問うのではなく、あなたはどんな人ですかと問う質問である。

- ・あなたは今、どちらにお住まいですか。
- ・あなたは現在、何人の子どもをお持ちですか。
- ・あなたの家庭の収入の何パーセントが農業以外の仕事からの収入ですか。

質問のタイプ

調査において使用される質問には基本的に 2 つの明瞭に区別されるタイプの質問がある。すなわち、回答選択式質問（Closed-ended question）、自由回答式質問（Open-ended question）である。

回答選択式質問

回答選択式質問は、幾つかの回答の選択肢をあらかじめ提示しておき、回答者に、その中から選択してもらう質問である。この回答選択式質問を行う場合、選択肢の中に必ずすべての回答が含まれているように配慮する必要がある。

回答選択式質問の長所

- 回答がしやすい。
- 回答にあまり時間を必要としない。
- インタビュー担当者の訓練をあまり必要としない。
- インタビュー担当者の偏見を排除できる。
- 個々人の回答を比較し、定量化しやすい。
- データ入力と分析がしやすい。

回答選択式質問の短所

- 質問を作成する前に回答の選択肢が分かっている必要がある。
- ありそうな回答のすべてを選択肢に含まれない可能性がある。
- 回答の選択肢に中身の無いものや偏見が入り込む可能性がある。
- 回答の選択肢が、回答者によって、異なる解釈をされる可能性がある。
- 選択肢が多過ぎると回答が困難になる。
- 回答者に選択を強制することになる。

回答選択式質問の参考例

1. あなたや、あなたの家族は、これまで、この郡の地域教育センターのクラスを受講したことがありますか。 YES _____ NO _____
2. あなたは、新しい土地利用区分規制に、どの程度、賛成ですか、反対ですか。
 - 1) 全面的に反対である
 - 2) どちらかといえば反対である
 - 3) どちらでもない
 - 4) 全面的に賛成である

自由回答式質問

自由回答式質問は、回答者に選択肢の中から回答を選ばせるのではなく、自分の言葉で回答させるものである。

自由回答式質問の利点

- 自由な発想を促す。
- 回答者に自分の感情や意見を表明する機会を与える。
- 回答者に自分の言葉で自分を表現することを可能にする。
- 説明のための引用に飾りや補足説明を加えることができる。
- 予備調査に適している。
- 回答選択式質問を作成するための材料を提供してくれる。

自由回答式質問の不利な点

- 分析が難しい。
- 回答に時間を必要とする。
- 回答者の記憶に頼るところがある。
- かなり熟練したインタビュー・スキルを必要とする。
- 質問の意図を明確にするための回答の選択肢を提示できない。
- 手書きの回答はしばしば判読が困難である。

自由回答式質問の参考例

1. あなたは、今回のトレーニングで得た情報をどのように使おうとお考えですか。
2. あなたは、この郡が実施している 4-H プログラムを改善するにはどうすれば良いとお考えですか。

評価のための質問書の事前テスト

定性的方法や個人参加型方法では、若干異なるフォーマットを用いて事前テストを行うこともあるが、通常、事前テストは、定量的方法に関連して実施される。事前テストのためには、評価プロセスを開始するに先立って、また、高くつくエラーや無駄骨をなくすために、評価手法と評価のための質問書を試験的に実施することが必要となる。事前テストは、可能な限り、実際の評価が行われる期間中に予想される環境条件と類似の環境条件下で実施される必要がある。もし可能であれば、事前テストのためのミニ標本を選ぶ際、実際の評価に当たってあなたが採用しようと考えている標本と同じサンプリング・プランを使用するのが良い。

事前テストにおいて、我々は、下記の質問をする。すなわち、

- 議論の対象となる課題、なされるべき質問、及び、使用される言語は明確で、あいまいさを排除したものになっているか。
- 使用される手法と質問書は、インタビューや観察の対象となっている人々に適しているか。
- インタビュー担当者、もしくは、観察担当者に対する指示は、遵守しやすいものとなっているか。
- 入手した情報を記録する手法とフォームは明確で使いやすいものであるか。
- すべての手順は標準化されているか。
- その手法と質問書は、必要な情報を提供してくれるものか。
- その手法と質問書は、採用されたデータ収集方法の指標を用いれば、信頼できて、かつ、妥当な情報を提供してくれるか。

実地テストを実施したあと、あなたは、採用した手法と質問書を修正する必要があることを認識する可能性がある。もし、大掛かりな改定がなされる場合、再度、実地テストを実施することが必要となる。

(5) フォーカスグループ (Focus Group)

フォーカスグループとは少人数のグループを指し、典型的には、相対的に均質な 8 人～12 人の人で構成される。これは、脅迫的でない雰囲気の中で、特定のトピックスを議論してもらうために選定されるグループである。このフォーカスグループは、熟練したインタビュー担当者が司会 (moderator) と記録係を勤める。フォーカスグループは、特定のトピックスに対する市民の態度、理解度、及び意見、更には、特定のプログラムが個人やコミュニティに及ぼす影響といったコミュニティのニーズや課題を判定するために利用される。

フォーカスグループの利点

- 編成が容易である。
- 迅速，かつ，比較的費用を掛けずに実施できる。
- プロジェクト担当者と当該プロジェクトが対象とする受益者との間の距離を縮めることができる。
- 対話を促す。
- 別の調査方法に含めることの出来る「評価のための質問」のアイデアを提供してくれる。

フォーカスグループの不利な点

- 誤用されやすい
- 司会者に特別なスキルが要求される。
- データの解釈が単調で冗長となる。
- 偏見を排除しにくい。
- 新たに出てきた大きな問題は捕捉するのが難しい。
- フォーカスグループの結果をターゲットとする母集団に一般化することが出来ない可能性がある。

フォーカスグループに対するインタビューのステップ

æ 10	簡単な報告書を作成し、調査結果をステークホルダーと共有する。
æ 9	録音された議論の結果を分析し、発言されていることを要約する。 意味を解釈し、提言書を作成し、インタビューを要約する。
æ 8	インタビュー終了後、司会者と助手は体験したこととそれに対する自分の見解を議論し、次のフォーカスグループが実施される前に、協力して録音をレビューする。
æ 7	フォーカスグループへのインタビューを実施する。司会者は目的を説明し、会議は録音されるが、匿名性は守られることを保証する。
æ 6	会議室を用意する。座席とテーブルのアレンジを確認する。
æ 5	フォーカスグループのインタビューを実施できる司会者と助手を指名する。 司会者は穏やかで友好的な雰囲気作りをし、スムーズな会話の流れを維持し、メモを取る。
æ 4	参加候補者を選定して、個人宛の招待状を送り、コンタクトを取る。 彼らに会議の目的を説明し、各人の参加が会議にとっていかに役立つかを示す。 彼らが参加できることを再確認する。
æ 3	中立的で、圧迫的でなく、便利で、見つけやすい場所に会議場所を設定する。 会議を記録する手段を決定する。（テープでの録音、メモの筆記など）
æ 2	インタビューでなされるべき質問を特定する。 それぞれの質問の内容を確定する。 質問の論理的な順序付けをする。
1	フォーカスグループのインタビューを実施する目的を自分自身、よく考える。 フォーカスグループの議論で出された情報のユーザーは誰かを特定する。 所要時間と必要な資源などについて、暫定的なプランを作成する。

フォーカスグループの議論の始め方

フォーカスグループの議論の最初の数秒間は極めて重要である。この短い時間の間に、司会者は、思慮深く、自由に発言できる雰囲気を醸成し、基本的なルールを定め、議論の基調を誘導しなければならない。グループインタビューの成否の大部分は、こうしたオープンな環境作りの成否に懸かっている。グループ・ディスカッションを開始する際のオープニングの挨拶の

パターンとして、次の事項を含めることを推薦する。すなわち、歓迎の挨拶、全体的な概観とトピックスの紹介、議論の基本的ルール、そして、最初の質問である。

典型的な開会挨拶の参考事例

こんばんは、本日は皆様方のこの会議へのご参加を心より歓迎いたします。わが郡の教育サービスに関する討議に皆様方の貴重なお時間を割いて頂いたことに、心から御礼申し上げます。私の名前は_____と言い、_____を代表して、この会議に出席しております。本日、私の助手を務めて頂くのは_____の_____です。私どもは、この会議を通じて、我々のコミュニティにおける教育の機会に関する情報を収集したいと考えております。今日は、それぞれのご意見を頂くために、この郡の異なる幾つかの地区にお住まいの方々にお集まり頂いております。

皆様方は、私どもに特別な関心のある事項について、共通して、一定の条件を備えているということで、本日の討議に参加するメンバーとして選ばれました。皆様方は、全員、外に出て働いておられます。また、この郡の都市郊外地域に住んでおられます。だから、私どもは、皆様方がこの郡の住民を代表する人々であるということに格別の関心を抱いています。

本日は、我々は、このコミュニティにおける非公式教育の課題について、皆様方に議論して頂く予定です。この課題には、あなた方が、ご自分に関心のある分野に関する新しい情報を入手されているすべての方法が含まれます。この課題には、正しい答えも、間違った答えもあります。異なった見解があるだけです。あなたの見解が、別の人の見解と異なったものであっても、それを皆さんと共有して頂くために、自由にご発言下さい。

議論を始める前に、幾つかの基本的ルールをご説明させていただきます。まず、自由にご発言下さい。ただし、一度に何人もが同時に発言するようなことは避け、常に、誰か一人が発言するという形を守って頂きます。もし同時に何人もの方が発言されますと、テープがすべての音を拾いますから、聞き取り辛くなり、大切な発言を聞き逃します。今日はお互いファースト・ネームで呼び合うことにします。今日の議論は後日報告書にまとめられますが、そこでは、どのコメントをどなたが発言されたか分からないようにします。従って、皆様方の発言の匿名性は完全に守られます。私どもは、ネガティブな意見もポジティブな意見も同じように扱います。時には、ネガティブなコメントも極めて有益です。このことをまず念頭に置いてください。

本日の討議の予定時間は約1時間です。途中で休憩時間をとるようなことは致しません。

お互いをもう少し良く知るために、全員に順番に自己紹介をして頂きます。まず最初に、あなたのお名前と、どこにお住まいかを教えて頂けますか。

フォーカスグループでの質問の仕方

1. 議論を活発化するために、自由な回答を求める質問をする。
 - ・このプログラムについて、あなたはどのように思われますか。
 - ・あなたは、どのような方法で新しい情報を手に入れておられますか。
 - ・提案されているプログラムについて、あなたは何が一番気に入っていますか。
2. YES か NO で答えられるような二分法的質問は避ける。
3. あまり「なぜですか」という質問はしない。
 - ・「なぜですか」という質問は相手を防御的にさせ、何かを答えなければという気分させる。
 - ・「なぜですか」という質問をすると、人々は、常に、属性や影響について答えようとする。
 - ・むしろ「何があなたをそうさせたのですか」とか「あなたは、どんなところが好きですか」といった質問をするほうが良い。
4. 相手に、将来のことを予想させるような質問ではなく、過去のことを思い出させるような「思い出してみてください」といった質問をする。
5. 焦点となる質問を注意深く用意する。
 - ・質問したほうが良いと思われる質問を特定する。
 - ・5つのタイプの質問は、
 - ① 最初にする質問（round-robin とも言う）
 - ② イントロの質問
 - ③ 話題転換のための質問
 - ④ キーとなる質問
 - ⑤ 締めの質問
6. まず「キューのない質問」からはじめ、次に「キューのある質問」をする。

（ここで「キュー（cue）」というのは、参加者に、話題に関する特定の事柄や詳細を思い出させるようなヒントや助言のことを言う）
7. 標準化された質問をするように心がける。（これを説明してください。これはどんな意味ですか。）
8. 一般的な質問から特定の問題に焦点を合わせた質問へと、順序立てた質問をするように配慮する。

フォーカスグループでの質問（サンプル）

畑作物生産諮問委員会

注：以下の質問は、フォーカスグループミーティングが開催されている期間中、すべての諮問委員会メンバーに送付される。

1. お互いをもう少し良く知るために、全員、順番に自己紹介しましょう。まず、あなたのお名前とどこにお住まいかを教えてください。それから「ミシガン大学」という言葉を聞いた時、あなたは、まず、最初に何を思い浮かべるかを教えてください。

（ファシリテーターのメモ：重要な単語を記録しなさい。そのうちの幾つかは、あとの議論の際に、再度、言及しなさい。）

2. あなたは、あなたのコミュニティにおける農業資源と自然資源に関するプログラムの普及活動について、人々がどんなことを言っているか、聞いたことがありますか。それはどんなことでしたか、教えてください。最近、普及活動の内容がどのように変わりましたか。

（ファシリテーターのメモ：ポジティブなコメントとネガティブなコメントを細大漏らさず収集するようにしなさい。）

3. 現在実施中のミシガン州立大学普及部の活動に関してあなたが過去に経験したことを思い出してください。そして、それを説明してください。
4. あなたが失望したミシガン州立大学普及部の活動で、あなたが過去に経験したことを思い出し、それを説明してください。その普及活動のプログラミングをどのように変えれば良かったと思われますか。
5. ミシガン州立大学普及部は、普及活動への取組みのために専門家チーム（AOE）を編成していますが、あなたは、この専門家チームの恩恵を受けたことがありますか。昨年 1 年間のこの専門家チームの活動業績について、あなたはどのような印象をお持ちですか。
6. ミシガン州立大学普及部の畑作物専門家チームは、どうすれば将来のプログラムの提案の内容を改善できると思いますか。畑作物専門家チームのプログラムの改善策について、あなたのご意見をお聞かせ願えますか。

（ファシリテーターのメモ：参加者全員が答えを出してくれるようにする。各人が答えてくれるまで我慢強く粘るようにしなさい。）

7. ミシガン州立大学普及部や、同部が推進している農業・自然資源プログラムに関して、

まだ言い忘れたコメントや、助言、あるいは、考え方はありませんか。

8. 我々は、何か見落としとしてはいませんか。

(6) 簡易農村調査 (Rapid Rural Appraisal)

簡易農村調査とは、迅速、かつ、フレキシブルで、適応性があり、しかも、調査対象と高い関連性のある多数のデータ収集手法を使用する調査方法を言う。この簡易農村調査、我々が、地域の人々の状況や、経験、更には、地域の観点から見た問題点を知る上で役に立つ。

簡易農村調査の利点

- 安い費用でできる。
- 少ない時間でできる。
- 地域の人々の参画を促進できる。
- 外部の者が持ちやすい偏見を減らすことができる。
- しばしば見逃されてきたグループの人々の参画を促すことができる。
- 手法の選択がフレキシブルにできる。

簡易農村調査の不利な点

- 結果に季節的な偏りが出やすい。
- 母集団となるメンバーへのアクセスに偏りが出やすい。
- (選ばれた人々が) 自分はエリートであると認識することにより偏りが出やすい。
- 推測に基づく偏りが出やすい。自分は選ばれたと自認することから来る過剰な思いが原因。
- 一貫性の偏り。データの首尾一貫性が欠けることに起因する。
- 一般化しにくい。

簡易農村調査に使用される各種のツール

既存の情報	ビジュアル化の手法
個人インタビュー	・活動マッピング
・重要な情報提供者	・時系列マップ (例：穀物カレンダー)
・口伝の歴史	・資源のマッピング
グループインタビュー	・社会的組織のマッピング
・フォーカスグループ	ランク付けゲーム

	・所得によるランキング
	・好みによるランキング
	マトリックス

簡易農村調査の基本的ステップ

1. 簡易農村調査の目的と問うべき質問を明確にする。
2. 簡易農村調査に使用できる資源。時間、スキル、スタッフ、顧客、公的組織、非公式なグループ、企業、交通手段、電話・郵便、メディアなどを明確にする。
3. 既存の文書類をレビューする。
4. 必要なデータや、必要とされる分析のタイプを明確にする。
5. 可能性のある情報ソースを明確にする。
6. データ収集方法を明確にし、適切なものを採用するか、もしくは、新たに開発する。
7. 実地テストの方法を明確にする。
8. 質問や、情報ソース、あるいは、アプローチの方法を修正する。
9. 訪問先と訪問時期に関する計画を立てる。
10. 状況にフレキシブルに対応しながら、データ収集を開始する。適切な方法を採用し、質問や活動が正しいものとなるよう修正する。
11. 収集されたデータを体系的な方法で記録する。
12. 集められた回答を検証し、それらに対する理解を深め、それぞれの回答間の違いや関連性を明らかにすることによって、データを継続的に分析する。

(7) ケーススタディ (Case Study)

ケーススタディは、特定のケース、すなわち、ある一つのプログラム、ある特定の参加者グループ、ある特定の個人、あるいは、ある特定の場所などについてなされる、掘り下げた分析である。ケーススタディには、説明的な(explanatory)もの、記述的な(descriptive)ものがある。説明的なケーススタディは、因果関係を明らかにすることが出来る。記述的なケーススタディは、ある特定のプログラムが実施される背景や、当該プログラムそのものを記述するために使われる。また、説明的なケーススタディは、将来の評価のための指標を決定し、あるいは、仮説を提示する上で役に立つ。すべてのケーススタディは、特定のケースについて、出来るだけ完璧な絵を描けるようにするためには、多数の情報ソースと手法に依存する。

ケーススタディの利点

- 質問を「なぜ」、「どのように」するのかという問題に取り組む上で優れている。
- 問題や解決策に対して具体性を与えてくれる。

- 進化的プロセス²や方針の決定プロセスを調査するために使用できる。
- ある一つの設定や、グループ、あるいは、組織について掘り下げた情報を提供できる。
- 特定の状況に合わせて調整することが可能である。
- 更に突っ込んだ調査を行う際の指針となるバックグラウンド情報を提供できる。
- さまざまな人間関係や人間の感情に対する洞察を深める上で役に立つ。
- 基本となっている前提条件や一般的知識の深化を助けてくれる。
- その他の手法の補助的手段として利用できる。

ケーススタディの不利な点

- 時間が掛かり、大量のデータを必要とする。
- より大きな母集団に対して一般化できない可能性がある。
- ある問題の一つか二つの側面に関するデータしか提供できない可能性がある。
- 優れた観察能力と、優れた記録・報告スキルを必要とする。
- 調査を担当する人間の偏見のゆえに、集められた情報が主観的なものとなる可能性がある。

ケーススタディの事例

ある特定の生産者がある総合財務管理プログラムに参加する前と、参加した後で、どのように変わったかを詳細かつ体系的に記録すれば、そのプログラムが、より大人数の生産者を対象に実施するのは有用かも知れないといった、当該プログラムの影響度に対する貴重な識見を提供してくれる可能性がある。

ケーススタディの立案と実施のステップ

1. そのケーススタディで何を期待するかをレビューする。
2. 基本的な質問と仮定を明らかにする。ここで言う「仮定」とは、変数間の関係、例えば、農業者は、そのプログラムに参加することで何らかの利益が得られるから、そのプログラムに参加する、といった関係を示すものを指す。
3. 対象とするケースの境界を明確に定義する。
4. 当該ケーススタディを実施する評価者の能力、すなわち、質問能力、多くの新しい情報を偏見なく吸収できる能力、適応性、及び弾力性などアセスメントする。
5. あらかじめ、調査で直面しそうな重要な問題、出来事、属性、問題人物の想定をしておく。
6. 初期計画を作成する。これには、現地オブザーバーの役割も明記しておく。
7. アクセスの手段と折衝プランをアレンジする。
8. 秘密保持のための措置を議論する。
9. 対象とする活動について予備的観察を行う。

² 訳注：諸要素を少しずつ変えて行きながら、最適解を求める手法で EVOP と言う。

10. 情報提供者と情報ソースを特定する。
11. 記録システム、ファイル、テープ起こし、コーディング・システムなどを確立する。また、保護されたデータ蓄積手段をアレンジする。
12. 新たに出現する属性や、問題、出来事、オーディエンスなどに基づいて、再調査する際の優先順位を定める。
13. データ収集の指針に何か課題はないか、あるいは、何らかの論理的基礎はないか、良く検討する。
14. 観察し、インタビューを実施し、記録を集め、各種の調査結果を活用する。
15. 象徴的エピソード、特別な証言、事例を選別して収集する。
16. 生のデータを分類し、データの解釈を始める。
17. 対象とする課題とケースの範囲を再定義し、必要であれば、ホストと一連のアレンジについて再折衝する。
18. 重要な観察結果を検証するために三角測量法を用いて追加的データを収集し、ある種の生のデータについて考え得る多くの解釈がある場合、そうした生のデータをレビューする。
19. データのパターンを探す。
20. 当該プログラムのアレンジと、活動、及びアウトカムに関連性を探し出す。
21. 仮説的に幾つかの結論を出し、それらを系統的にまとめ、最終報告書をまとめる。
22. データをレビューし、新しいデータを収集する。意図的に反証となるような証拠を探す。
23. 当該ケーススタディが実施された状況設定を記述する。
24. 報告書を作成し、オーディエンスのために材料を作り直す。報告書を一つの物語と考え、物語に不完全な部分がないか、良く読み直し、欠けている情報を補充する。

(8) 準構成的インタビュー (Semi-structured Interviews)

プロジェクトへの参加者やその他の重要な情報提供者に対する準構成的インタビュー³は、当該インタビューでカバーされるトピックスと、質問すべき自由回答式の質問をリスト化したインタビュー指針（又は、インタビュー・ガイド）を用いて始める。インタビュー対象の人々の理解や、信念、態度、更には、心配事などを反映した回答を得、かつ、それらを反映した新しいトピックを提起するために、さまざまな調査手法が使用される。

準構成的インタビューの利点

- 質問に対する回答が予め想定できないような複雑な状況に有用である。
- 評価の指針となる仮説を立てるために使用される。
- 質問に答える際に、回答者は予め設定されている選択肢に縛られない。
- インタビューの構成は、事前に決定されたものではないが、インタビューが進むうちに構成が出来あがってくる。
- 課題を明確化したり、新しい課題領域を探し出したりするために、追加的質問を行うことが出来る。

³ 訳注：「構成的インタビュー」は、回答の選択肢が予め用意されているような質問が周到に構築されたインタビューのこと。ここでは、これに準じる形のインタビューとあるので、構成的インタビューと自由回答式インタビューとが折衷された形のインタビューを指す。

準構成的インタビューの不利な点

- 定量的方法に精通した人々からは有効なものとなみなされない可能性がある。
- より完全な回答をいつ、どのようにして求めるのが良いかを知っており、新たに出現してきた重要なトピックを認識できる能力を持ったインタビュー担当者を必要とする。

準構成的インタビューのためのガイドライン

- トピックスを明らかにし、自由回答式の質問を作成する。
- あなたの外見や身振りの癖がインタビューに影響を与えるということを覚えておく。相応しい服装をし、相手に威嚇的な印象を与えない口調で話をし、分かりやすい単語を使用するようにする。
- サンプリングの母集団を選択したあと、回答者を選定する。
- インタビューを実施し、お互いに都合の良い帰りの時間を決める。
- 不審感を抱かれないようにする。インタビューは第三者がいない場所で行うようにする。
- 回答者にインタビューの目的を説明し、このインタビューは、回答を強制するものではないこと、また、回答については必ず秘密が保持されることを理解させる。
- インタビューの冒頭では、回答者が興味を示しそうな中立的な話題に関する一般的な会話で始め、お互いの個人的なバックグラウンドを共有し、相手の回答に対する感謝の意を表する。
- まず簡単な質問から始め、長い答えを要する質問や、熟慮しなければ答えられないような質問は避ける。そうしながら、徐々に複雑で、微妙な質問に移る。
- 回答はすべて逐語的に記録する。
- あなた自身の個人的な意見は表明してはならない。
- もし回答が不十分であったり、質問のポイントから外れていたりする場合には、明瞭な回答を得るように努力する。
- もし回答者が答えを拒否する場合には、出来るだけ答えを得るように努力するが、インタビューを台無しにしてしまうようなことは絶対に避ける。

(9) 参加者観察 (Participant Observation)

参加者観察とは、重要な情報提供者に対する直接的な観察やインタビューを通じて、また、観察の対象となっている活動への参加を通じて、参加者の行動や反応に関する情報を集めることである。評価に使用する場合、参加者観察による評価者は、ステークホルダーの目で世界を理解しようとする意図を持って、調査対象となっている環境に自らを没入させる必要がある。参加者観察は、コミュニティ内の利害の衝突や誤解を明らかにし、当該コミュニティのニーズや問題を評価し、あるいは、ローカルの住民を問題解決に参画させる方法を明らかにする上で有用である。

参加者観察の利点

- 自然な環境の中で観察ができる。
- 参加者観察は、事前に構成されておらず、フレキシブルである。
- その他の方法と容易に併用することが出来る。
- 隣人、クラスルーム、あるいは、ある特定のグループといった小規模な単位に対して有用である。
- 各種のプログラムや実践が地域の住民に与える長期的な影響を評価する上で役に立つ。
- 参加者自信が気づいていない行動上のパターンや、社会的プロセス、あるいは、問題の多い課題を明らかにしてくれる。

参加者観察の不利な点

- 評価に役立つ十分な情報を提供してくれることが少ない。通常、インタビューなどの他の方法と併用して使用しなければならない。
- 評価者には、よく訓練された観察スキルが求められる。
- 評価者が状況をコントロールするようなことはあまり出来ない。
- 評価者の存在が、観察の対象となっているグループの行動を変える可能性がある。
- 観察の結果を一般化は出来ない。
- 観察者も一人の参加者となることから、観察が客観性を失う可能性がある。
- しばしば、時間が制約要因となる。
- 大きなグループ、あるいは、異質な分子を多数含むグループには適さない。

参加者観察の実施に対する一般的指示事項

1. 評価に使用される参加者観察は、理論を構築するためではなく、実践的な問題を解決しようとするニーズが動機である。このため、この手法を用いる評価者は、最初に一定の概念的なフレームワークを持って、現場に臨むことが必要である。このフレームワークには、予備的な諸課題や、それらの諸課題の間に関係がある可能性があることがある。
2. 当該フレームワークにおける主たるコンセプトは何かを定義づける。例えば、学習スタイルとか、リーダーシップとか。
3. 情報ソースを特定する。
4. 参加者観察が実施されるサイトを選択する。2 つ、もしくはそれ以上のサイトを選択すれば、データの比較分析が可能となる。この参加者観察では、非公式なサンプリング手法が使用される。選択されるサイトは、観察の対象となっているタイプのプログラムや組織を代表するような場所でなければならない。また、当該組織は、参加者観察の評価者を喜んで受け入れる必要があり、参加者観察の評価者は、観察の対象となっている活動の内部に入り込むことができないとなければならない。一回限りの活動、季節的なイベント、あるいは、日常的にルーティン化された活動の場合には、タイミングが決定的に重要となる。
5. アクセス手段を用意し、秘密を保持できるアレンジを整える。

6. 観察に必要なツールを揃える。ノートブック、ペン、テープレコーダーなどである。データをどのような手段で記録するかは、状況次第である。あなたは現場でノートを取る方法を選ぶかも知れず、観察を終えた後でノートを取りたいと思うかもしれない。写真やテープレコーダーは記録を取る助けとなる。しかし、ある種の状況下では、それは邪魔になるかも知れず、また、観察の対象となっている状況に影響を与えるかも知れない。
7. 観察を開始する。あなたは、進行中のすべてを観察する必要はない。しかし、評価に係る活動については、あらゆる側面に焦点を当てて観察することが必要である。下記の質問が、あなたの観察の指針となるかも知れない。
 - a.あなたが観察している場面のセッティングはどうなっているか。
 - b.何が現在進行しているか。
 - c.観察の対象となっている場面との関係で、貴方はどこにいるのか。
 - d.観察されている場面の内容はどうなっているか。例えば、一日のうちの何時頃か。どんな天気か。参加者は何名くらいいるか。参加者の年令は。性別は。人種は。また、階層は。参加者間には何か相互の関係があるか。各人のお互いの関係は。同じ場面で別の活動が進行していないか。
 - e.あなた自身は、その活動の参加者か。もしそうなら、あなた自身が参加していることが、あなたの観察にどのような影響を与えているか。
 - f.あなたが観察している人々の年令、性別、人種、及び、階層はどうなっているか。
 - g.複数以上の人間に同じ質問をしたら、同じ回答が得られたか。
 - h.あなたは、あなたが観察している場面を見て何を考えさせられたか。あなたを困惑させるようなものが何かあったか。あなた自身は何を理解したと思うか。
 - i.その活動は、あなたが同じようなセッティング（環境）で観察したことのある別のタイプの活動と似ているか、それとも、全く異なっているか。
8. あなたが観察している活動は、何らかの理論的フレームワークに関連付けることが出来るか。

(10) 費用便益分析 (Benefit/Cost Analysis)

費用便益分析は、プログラム評価に対する一つの典型的な代替手段とみなされている。しかしながら、それは同時に、評価プロセスの延長と見ることもできる。そのような訳で、費用便益分析は、プログラムへの金銭的なインプットとプログラムのアウトカムを定量的に評価し比較する手段を提供する。費用と便益を金銭的に評価すれば、当該プログラムの正味の影響を判定し、代替的なプログラムやプロジェクトの比較を行い、プログラムの計画策定を助け、当該組織のアカウンタビリティを増進し、あるいは、当該プログラムに対するサポートをするために、費用と便益を直接的に比較することが可能となる。

費用便益分析の利点

- 情報ソースの一つとして高い信頼性がある。
- 予算を正当化し、当該プログラムの価値をデモンストレートし、あるいは、当該

プログラムへのインプットから最大のアウトカムを引き出すことに役立つ上で有用である。

- それは、寄付してくれた人や資金を提供してくれた人に有益な情報を手供する。

費用便益分析の不利な点

- 費用や便益を金銭的に定量化することが困難な場合がある。
- 機会費用、埋没費用、あるいは、仮定費用を算定することが困難な場合がある。
- プログラムの間接的便益を算定することが困難なことがある。
- 各費用や便益に金銭的な評価額を割り当てる際に偏見を生じる余地がある。
- 基本的な前提条件や立証されていない前提条件に偏見が入り込む余地がある。

費用便益分析のステップ

1. 当該プログラムのパラメーターを定義づける。
2. 多くの情報ソースを利用して、費用項目と利益項目の一覧表を作成する。プログラム・費用には、直接費用、暗示的費用もしくは間接費用、潜在的費用もしくは推定費用が含まれる。プログラム内容の記述、専門書籍、あなた自身の知識、更には、分析の初期段階を通じて集められた情報などを利用する必要がある。プログラムの便益というのは、当該プログラム、あるいは、プロジェクトから生じるプラスのアウトカムあるいは成果を指す。これらには、間接的な便益や時間の経過に伴って発生する便益が含まれる。費用と便益を決定する場合、それらの費用と便益が同じレベルで測定されるようにすることが大切である。

A. 費用等式： 費用=L+K+I-i

L=労務費。労働の時間当たり費用で、給与と労務副費を含む。労務副費はさまざまであるが、通常は、給与全額の 22%～35%の範囲内にある。時間当たり総労務費（L）の等式は、給与=S，労務副費=S.35，年間労働日数=260，一日当たり労働時間=8 として， $(S+S.35)/260/8$ で表される。

K=直接費用。当該プログラムに割り当てられた、あるいは、予算付けされた直接的なプログラム費用で、支給品、通信費、旅費・日当、設備及びオーディオ機器などである。もし幾つかの費用項目が複数のプロジェクトで共有されている場合には、総額を費用配分公式に基づいて各プログラム分を算定する。機会費用というのは、各参加者が当該プログラムに参加することによって失った機会のことである。機会費用には、分析のレベルにもよるが、各参加者、プレゼンター、あるいは、ステークホルダーにとっての直接的費用が含まれる。

I=間接費用。各参加者には間接的に関係するが、当該プログラムには直接的に関係する費用を指し、例えば、施設の賃借料などの管理的費用、コピー代、報告書作成費用、電話代、設備費と支給品の比例割当額を含む。

i=割賦償却費。時間の経過に伴って発生する計測可能な利益又は費用（＋と－の両方を含む）時間の経過に伴って追跡できないものは、この項目には含まれない。

B. 利益等式： $B=Cr+DB+IB$

Cr=当該プログラムの活動によって享受できた費用の削減額

DB=直接便益。当該プログラムへの参加者、及び、その他の直接的な関係者が享受する主たるアウトカム。これらのアウトカムは、当該プログラムの目標から直接的に得られる典型的な便益である。

IB=間接便益。当該プログラム又はプロジェクトへの参加者、非参加者、あるいは、社会一般が、当該プログラム、プロジェクトから二次的に享受する、あるいは、無形のアウトカムを指す。これらのアウトカム、又は、結果には、プラスもあれば、マイナスもある。

3. 費用をこれらの便益と、便益から費用を直接的に控除することによって、あるいは、便益の費用に対する比率を計算することによって、比較する。ここで、第1の等式は、ある特定のプログラムにおいて、費用を便益と比較する方法を提供し、第2の等式は、各プログラムの相互の比較を可能とする。

費用便益分析シート（サンプル）
便益推計と費用推計のワークシート

受益者数							単位数	単位価格	総費用
直接便益						直接費用			
1.						労務費	時間		
2.						1.			
3.						2.			
4.						直接費用			
5.						1. 賃借料			
						2. 電気代			
間接便益						設備＋材料費			
1.						1. 印刷費	枚数		
2.						2. 家具備品			
3.						3. 教材費等			
4.						4. 旅費	距離		
5.						機会費用			
6.						1. 子供養育費			
						2. 食料費			
						3. 旅費			
						間接費用			
便益合計						費用合計			
便益/費用 比率									

2-7 母集団を特定し、サンプリングする (ステップ7)

サンプルは、調査目的の対象である大きな母集団から選別された一定の数の回答者である。サンプリングが適切に行われれば、サンプルは当該母集団の特徴を正確に代表する。サンプリングは、正確性と精度を犠牲にすることなく、時間と費用、材料、及び努力を節約できる手法である。

サンプリングの5つのステップ

5 æ		結論を推測して全母集団に適用する。
4 æ		サンプル情報に基づいて結論を引き出す。
3 æ		サンプルを抽出し、サンプリング計画を実行する。 サンプル数を決定する。
2 æ		サンプリング・エラーをどの程度許容できるかを決定する。
1 æ	母集団を決定する。母集団の数はどのくらいか。 当該母集団にはどの程度の多様性が認められるか。	

サンプル数

サンプルの数は、どの程度の大きさにすべきであろうか。大きな母集団に対しては、回答者数を100名とするサンプル数がしばしば最小の数とされている。一般的には、30名以下の回答者では、有用な分析を提供できるだけの十分な確実性はないとされている。しかしながら、実際のサンプル数を決定するに当たっては、サンプル数以外の要素をいくつか考慮する必要がある。

母集団の特徴

母集団の多様性を代表できるサンプルを選ぶ。母集団が比較的均質である場合、サンプル数は比較的小さくても良い。逆に、異質性が高い母集団の場合、もっと大規模なサンプル数が必要になる可能性がある。

サンプリング誤差

サンプリング誤差は、同じデータ収集方法が使用された場合、母集団から引き出される推測値と、サンプルから引き出される推測値に生じる誤差を指す。このサンプリング誤差は、サンプル数が小さい場合、大きくなる。このため、時間、費用、及び材料が許す範囲で、出来るだ

け大きなサンプルを使用する方が良い。

精度

精度は、同じデータ収集方法が使用されたと仮定して、ある推定結果が、母集団から得られると推定される結果に近似している度合いを言う。サンプリングをデザインする場合、評価者は、要求される精度を明らかにすることから始めることが必要となるかも知れない。

誤差の許容範囲

誤差の許容範囲は、調査の目的によって異なり、選択の問題である。得られる結論を比較的安全なものとしたい場合、誤差 5%というのが許容範囲である（添付参照）。一般的には、0.1 アルファ・テストの場合、0.05%アルファ・テストの場合よりも多くの調査対象を必要とする。ツー・テイルド・テストの場合、ワン・テイルド・テストよりも大きなサンプル・数を必要とする。

信頼レベル

信頼レベルは、母集団に見られるある価値を、サンプルから得られたそれに対応する価値と比較した場合、一定の定量的範囲内に収まっている確率を言う。一般的に、95%の信頼性レベルがあれば、サンプルから得た結論を大きな母集団に当てはめても安全であるとされている。

費用

サンプル数が小さければ、費用は余り掛からない。

一定の母集団からサンプル数を決定するための試算表

n *	s *		n	s		n	s
10	10		220	139		1200	291
15	14		230	143		1300	296
20	19		240	147		1400	301
25	24		250	151		1500	305
30	28		260	155		1600	309
35	32		270	158		1700	313
40	36		280	161		1800	316
45	49		290	165		1900	319

50	44		300	168		2000	322
55	48		320	174		2200	327
60	51		340	180		2400	331
65	55		360	185		2600	334
70	59		380	191		2800	337
75	62		400	195		3000	340
80	66		420	200		3500	346
85	69		440	205		4000	350
90	72		460	209		4500	353
95	76		480	213		5000	356
100	79		500	217		6000	361
110	79		550	226		7000	364
120	91		600	234		8000	366
130	97		650	241		9000	368
140	102		700	248		10000	369
150	107		750	254		15000	375
160	112		800	259		20000	377
170	117		850	264		30000	379
180	123		900	273		40000	380
190	127		950	277		50000	381
200	131		1000	284		75000	382
210	135		1100			1000000	384

* n = 母集団の数

s = サンプルの数

サンプリングの手法

ランダム法、又は確率法

ランダム法、又は確率法によるサンプリングは、対象となる母集団からランダムにサンプルを選別する方法である。ランダム・サンプリング法には下記が含まれる。

単純ランダム・サンプリング法：母集団に属するすべての個人がサンプルに選ばれる機会を均等に、かつ、独立して持っている。ランダムに選ばれたランダム・サンプルを抽出するスタート・ポイントが決定されたあと、しばしば、ランダム・ナンバー表が使用される。(添付参照)

体系的サンプリング法：母集団のすべてのメンバーが、ランダム選別表にリストアップされ、ランダム抽出のスターティング・ポイントが選定されたあと、第 n 番番目の人間がすべてサンプルに抽出される。

層化サンプリング法⁴は、母集団の中の一定のサブ・グループが、母集団に占めるそれぞれの割合に応じてサンプルを選ぶために使用される方法である。それぞれのサブ・グループに対して別個にサンプル数が決定され、それぞれのサブ・グループごとにサンプルのランダム抽出が行われる。各サブ・グループを選定するための明確な理屈付けがなければならない。

クラスター・サンプリング法⁵は、地域で選べば調査者の交通費が節減できるのでマーケット・リサーチでよく用いられる。サンプリングの単位が個人ではなく、クラス・ルームとか、組織とか、コミュニティといった、自然に成立している個人のグループをサンプルの単位とする方法。

マトリックス・サンプリング法は、サンプルとなった一つの集団の人々がサンプリングのための一定の質問を受ける一方、別の集団の人々は、別の質問を受けるという方法。

意図的サンプリング法

意図的サンプリング法とは、特別に選択された重要な特性に基づいて、多様な人々をサンプルに含めるために使用される方法である。この意図的サンプリング法は、サンプルの抽出にランダム抽出法を使用しない。

偶然サンプル法は、その時にたまたま利用可能であった個人の中からサンプルを抽出する方法。このサンプルは、サンプルとしては一番弱いサンプルである。推計結果を大きな母集団に一般化して適用することはできない。

⁴ 訳注：年齢別、性別など階層ごとに単純無作為抽出する方法。サンプリング以前に、階層の情報がわかっていなければならない。

⁵ 訳注：多段サンプリング法の1つで、例えば、複数の村を含む県の調査などで、村をランダムに選んで、選ばれた村については全数調べる。集落抽出法とも呼ばれる。

評価サンプル法は、誰がその母集団を代表する「典型的な」人間であり、誰がそうでない人間かという判断を基に、サンプルとなる人間を選んで調査やインタビューの回答者にする方法。

2-8 データを収集し、分析し、解釈する (ステップ 8)

定量的データと定性的データの双方に対して、さまざまな種類のデータ分析方法がある。あなたは、その分析法が、当該評価によって求められる質問に答え得る情報を本当に提供できるものであるのか、また、評価者はその分析方法に関するスキルを本当に持っているのか、良く考える必要がある。

(1) 定性的データ分析

定性的データの分析、解釈は、定量的データのように単純な手法ではない。定性的データの分析は、データを積み上げる過程であり、またパターン化、カテゴリー化、基礎的な記述を積み上げることである。定性的データの解釈は、分析し、パターンを説明し、記述された特質の間における関係を探ることである。その結果、評価者やステークホルダーは、分析し、解釈したことに価値があるかどうかを判断する。

定性的データ分析の特徴

- データ収集が開始されると同時に開始される。
- データ収集の期間中、反復的に実施されるプロセスである。
- 妥当性と信頼性が、明瞭性、検証可能性、及び反復可能性によって表現される。

定性的分析が実施される場合

- 参加者によって使用される用語とそれらの用語の意味
- コンテンツ。各コメントをコンテンツに即して解釈する。
- 内容的な首尾一貫性と非一貫性。非一貫性がある場合、その理由を特定する。
- 各コメントの頻度と範囲
- 各コメントの強さ
- 回答の特異性
- 支配的なテーマ

定性的データ分析の各種のタイプ

コンテンツ分析 (Content Analysis)

これは、情報のパターンとデータの意味を調査し、一定の概念的フレームワークの中で、それらをコード化し、分類する手法を指す。

コンテンツ分析は、一連の手順によって、ニュースレター、会議議事録、通信文書、インタビューの記録などのテキストから得られた推論の妥当性を検証する調査手法である。推論は、当該メッセージの送り手、メッセージそれ自体、もしくは、当該メッセージの聞き手に関するものである。コンテンツ分析は、通信内容を目的に照らして監査する、調査における自由回答式の質問をコード化する、コミュニケーションに対する個人的見解や個人的対応を記述する、何かに対する個人的な、グループや組織としての、あるいは、社会的な関心を記述するなど、さまざまな目的に使用される。コンテンツ分析の中核をなすものは、テキストに含まれている多くの言葉を少数の内容別カテゴリーに分類することである。それぞれのカテゴリーは、一つの単語で成り立っているかも知れず、数語、あるいは、多くの単語から成り立っている可能性もある。同じカテゴリーに分類された単語、フレーズ、あるいは、その他の言語単位は、同じような意味を持つものと推定される。コンテンツ分析の手順は、テーマや、カテゴリー、あるいは、課題などの文化的単位で表現される関心の度合いや心配事の度合いを定量的指標として表わすことを可能にする。その上で、調査担当者は、適切な理論を援用して、分析結果を解釈し、説明する。この作業には、下記の三つのステップがある。

1. 測定 (Measurement) : 単語、フレーズ、コンテンツカテゴリー、あるいは、テーマなどの意味を持つ単位が使用される頻度を数える。こうした計算によって、それぞれのテキストをより正確に比較することが可能となる。我々はまた、こうした比較によって、どの課題が、他の課題よりもより多くの（あるいは、より少ない）関心を持たれているかを知ることができる。定性的分析手法は、もし、その他の方法が採用されていたなら、不可能とは言えないまでも、困難であったであろう各テキスト間の類似性や相違性を明らかにしてくれる。
2. 具象化 (Representation) : これは、コンテンツ分析において、当該文章やテキストで基本的な構文 (syntactic) 又は意味 (semantic) の特徴が省略されていて、文章の豊かさをコード化したり、具体化したりすることが困難な場合に、これを取り扱う手順である。単語、フレーズ、あるいは、その他のテキスト単位の意味を描出する一つの方法は、一連のカテゴリーに分類することである。ある意味をあるカテゴリーに分類するに当たっては、必ずしも、それらの単位に含蓄されているすべての意味を取り上げる必要はない。例えば、“kind” や “state” という単語は無数の意味を有しているが、カテゴリーに分類するに当たって、そのすべての意味を考慮する必要はない。

3. 解釈 (Interpretation)

信頼性

テキストから得られた推論を妥当なものとするために、分類の手順が首尾一貫性を有しているという意味で信頼できるものであることが重要である。別の人間が同じテキストを同じ方法でコード化しても、同じ結果になるということが大切である。コンテンツ分析において、言葉の意味が不明確な場合、カテゴリーの定義があいまいな場合、あるいは、コーディングのルールが不明確な場合に、信頼性の問題が発生する。コンテンツ分析においては、次の三つのタイプの信頼性が重要である。すなわち、安定性、復元可能性、及び正確性である。

安定性 これは、コンテンツ分類の結果が、どの程度、時間の経過に対して不変であるかという側面に関わるものである。すなわち、そのコンテンツが、同じコード分類担当者によって、同じ方法で、2回以上コード化された場合、同じ結果が得られるということを意味する。

復元可能性 これは、同じテキストが二人以上のコード分類担当者によってコード化された場合、どの程度まで同じ内容分類結果が出るかという側面に関わる。

正確性 これは、テキストの分類が、どの程度まで、標準あるいは規範に対応しているかという側面に関わるものである。これは信頼性の中でも最も強い信頼性を表すが、通常は、これを適用することは難しい。しばしば、これはコード分類担当者を訓練するために利用される。

妥当性

分類の手順は、妥当な変数を設定できるようなものであることが必要である。すなわち、それは、調査担当者が測定しようとしているものを適切に測定、あるいは、表すことができるものでなければならない。信頼性に関して生じたと同様に、言葉の意味、カテゴリー、あるいは、変数の定義にあいまいさがあると、妥当性の問題が生じてくる。

額面の妥当性 (Face validity) (最も弱い妥当性)。これは、調査担当者の各概念に対する定義づけと、それらを測定するカテゴリーの定義付けの間の対応性に関わるものである。あるカテゴリーが測定される構成概念をどの程度まで測定できるかというのが額面の妥当性である。

ある測定は、同じ構成概念を別の測定結果とどの程度関係付けることが出来るかというのが構成概念の妥当性 (Construct validity)である。従って、構成概念の妥当性は、当該構成概念を、多くの測定や方法にわたって一般化できるということを意味する。構成概念の分析に単純で直截的なものはない。調査担当者は、自分がやろうとしている目的に、どの方法が最も適しているかを自分で判断することが必要である。複数のフレーズや複雑なテキストの多くは、複数の単語や語句のような小さな単位と同じようにコード化することは難しい。何故なら、大きな単位は、通常、より多くの情報と、極めて多様なトピックスを含んでいるか

らである。従って、大きな単位のものは、コード分類担当者に対し、相互に矛盾する「キュー」(前述参照)を提起する可能性が高い。

コーディング計画の作成とテスト

1. コーディングの単位を決定する。単語、単語の意味（複数の意味を持つ単語には意味ごとに異なるコードを与え、意味論的に纏った単位を構成するフレーズには同じく意味ごとに異なるコードを与える）、文章（しばしば、共に現れる単語やフレーズに興味を持つ場合）、テーマ、場合によっては、テキスト全体のコードを与える。
2. それぞれのカテゴリーを決める。この場合、次の二つを明らかにすることが必要になる。すなわち、第1は、それらのカテゴリーが相互に排他的であるかどうか、第2は、それらのカテゴリーにどの程度の広がり（又は、狭さ）を持たせるべきか、ということである。
3. サンプル・テキストを用いて、そのコーディングをテストする。
4. 信頼性を評価する。
5. コーディング・ルールを修正する。
6. コード分類担当者が満足すべき信頼性を得るまで、上記ステップの3を繰り返す。
7. すべてのテキストをコード化する。
8. そのようにして得られた信頼性を評価する。コーディングが進むにつれて、コード分類担当者は、疲れやすくなり、間違いを犯しやすくなる。また、テキストのコード化が進むにつれて、コード分類担当者のコーディング・ルールに対する理解が微妙に変化し、信頼性を損ねる結果をもたらしやすい。

データ分析の解釈

データ分析は、情報を組織的に体系化し、簡潔に要約するとともに、論理的・統計的推論を行うためになされる。一方、解釈は、体系化された情報に意味づけし、結論を引き出すためのものである。すべての解釈は、ある程度、個人的であり、個別的である。このため、解釈のみならず、その背後にある理由付けも、すべて明示的に表現されなければならない。有用な解釈の方法には下記が含まれる。

1. 目標が達成されたかどうかを明らかにする。
2. 各種の法律、民主主義的理想、諸規則、あるいは、倫理原則にもとることがないか、明らかにする。
3. 設定されたニーズが実行されたかどうか、明らかにする。
4. 達成されたことの価値を決定する。
5. 重要なレファレンス・グループ⁶に、データのレビューを要請し、自己の評価が成功か失敗か、強いか弱いかについての彼らの判断を求める。
6. 結果を同じような組織や活動家によってなされた報告書と比較する。
7. 重要な変数について評価されたパフォーマンス・レベルを、期待値や標準値と比較する。
8. 結果をその結果をもたらした評価手順に照らして解釈する。

⁶ 訳注：自分の評価を比較評価してくれるグループ

解釈の作業に多面的な理解を与える一つの方法は、「ステークホルダー会議」を持つことである。ステークホルダーは、事前に結果を教えられ、評価計画、当該評価をもたらした質問リストや指標、基準といった関連する情報を提供される。その上で、この会議は、プレゼンテーションというよりは、議論のために開催される。この会議では、調査結果を、各人が一つひとつの評価を自分なりに解釈し、「これは何を意味しているか。」、「それはよいことか、それとも悪いことか、それとも、ナチュラルなものか。」、「それが意味するものは何か。」、「もし、何か行動を起こさねばならないとしたら、それは何か。」といった質問をしながら、それを全体として体系的にレビューする。

(2) 定量的データ分析

単純な統計的分析

測定の尺度

測定の尺度とは、測定の対象となる変数のタイプと、その測定方法を指す。測定の尺度が異なれば、異なる統計を使用するのが適切である。測定の尺度には下記のものが含まれる。

規範的尺度：相互に排他的で、論理的に網羅的なカテゴリー

参考事例：結婚の有無、性別、グループの会員権、加入宗教

階層尺度：上位階層か下位階層か

参考事例：学術書のランク、社会的階層、個人的判定に基づく変数

間隔尺度：標準的な判定評価単位において上位か下位か

参考事例：年令、資格、暦年、試験成績、IQ

比率尺度：絶対的なゼロ・ポイントをスタート・ポイントとする尺度間隔

参考事例：年令、教育年数、時間、長さ、重さ

記述的データの分析

中心的傾向の判定評価

中心的傾向を分析する目的は、一連の観測値を最も良く説明する一つのスコアやカテゴリーを報告することにある。平均値、あるいは、中央値はこの中心的傾向を測定するために最も一般的に使用されるもので、一つのグループを別のグループと比較したり、ある種の未知の行動を識別したり、あるいは、あるグループを標準的なグループと比較したりするために使用される。

平均値は、間隔のある変数に対して使用される。これはすべての観測値の算術的平均値である。すべての観測値（スコアや回答）を合計し、観測値の数でこれを除することで、これを計算することが出来る。この平均値は、観測値の中の、通常の分布から外れた値や、極端な値に影響を受けやすい欠点がある。あなたが集めたデータが、極端に小さい、あるいは、極端に

大きい値を含んでいる場合、そのデータは「歪んでいる」ことになる。

参考事例：15名の参加者が下記のスコアを取った。

-2, -1, 1, 4, 4, 4, 7, 7, 7, 7, 7, 8, 8, 8, 9.

この場合のスコアの中間値は $(3 \times X/n)$ は下記となる。

$$(-2)+(-1)+(1)+(4)+(4)+(4)+(7)+(7)+(7)+(7)+(7)+(8)+(8)+(8)+(9)/15 = 5.1$$

中央値は、順序のある変数に最も適した指標である。中央値とは、観測値の真ん中の値を指す。観測値の半数がこれより大きく、半数がこれより小さい。この中央値は、平均値ほど、極端な値の影響を受けない。

参考事例：

観測グループ 1 = 6, 8, 13, 18, 25 の場合、中央値 = 13 である。何故なら、半数がこの数字より上にあり、残りの半数がこれより下にあるからである。

観測グループ 2 = 1, 4, 7, 8, 10, 11, 21, 22 の場合、中央値は、中間の 2 数、つまり、8 と 10 を合算し、これを 2 で割って得る。よって、中央値は 9 である。

最頻値は、通常の変数に使用される概念である。これは最も頻繁に発生するスコア、もしくは、カテゴリーを表す。最頻値は、最もポピュラーな観測値あるいは価値を示すために使用することが出来る。この場合、分布は、単峰型であっても、双峰型であっても構わない。

分布 A		分布 B	
スコア	頻度	スコア	頻度
23	2	33	1
45	6	21	7
34	8	61	21
25	11	75	4
73	15	66	3
83	18	24	7
54	10	74	10
66	12	88	21

上記の場合、分布 A は、単峰型で 83 が、回答数 = 18 で単一の最頻値である。分布 B は、双峰型で 61 と 88 が、それぞれ回答数 = 21 で、最頻値は二つということになる。

平均値，中央値，及び最頻値を使用する場合

平均値を使用するのは、

- 分布がほぼ対照的である場合。
- 計数的値に関心がある場合。

中央値が使用されるのは、

- 典型的なスコアは何か、に関心がある場合。
- 分布が「歪んでいる」場合。
- あなたは保有しているデータが順序だったデータである場合。

最頻値が使用されるのは下記の場合である。

- 分布が「単峰型」又は「双峰型」である場合。
- あなたが、支配的な見解や、特徴、あるいは、傑出した性質を知りたい場合。

差異の検定

カイ二乗 (χ^2) 検定は、ノンパラメトリックな統計分析法として最も一般的に使用される手法である。このカイ二乗検定は、2組のカテゴリー変数（名義尺度データもしくは順序尺度データ）間の差異を示すものである。このカイ二乗検定には「一元配置」と「二元配置」がある。

一元カイ二乗検定の参考事例：標本グループに支持政党について質問をする。この場合、質問は事前に決められた選択肢（民主党、共和党、無党派など）の中から回答するようなものであることが必要である。一元カイ二乗検定は、この質問に対する標本グループの政党カテゴリー間の人気の差異を測定するものである。

二元カイ二乗検定の参考事例：これは、二つのカテゴリー変数を比較する場合に使用される。仮に、上記の参考事例と同じ標本グループが男女別に区分されたとした場合、「回答者の性別」という変数が生まれる。このカテゴリー変数が、政党の選択にどう影響しているかが比較される（あるいは、クロス比較）。この方法では、政党支持の男女別の差異が測定される。（例えば、男性の大多数は共和党支持であり、女性の多くは民主党支持である、というように）

一元と二元のカイ二乗検定は、どちらもカイ二乗値とその有意確率を提示する。カイ二乗検定はノンパラメトリック検定であり、母集団のパラメーターデータを仮説として必要としない。データは基本的にカテゴリー分けできるものであれば良い。

t-検定は、標本数が少ない場合でも、二組の平均値の差異を測定するために使用される。t統計値（t statistics）の有意性は調査者の仮説に依存する。もし、あなたが、二組の平均値差に有意性があるかどうかを明らかにしたいと考えてはいるが、どちらの平均値が大きいかわからない場合、両側検定（two tailed test）を使用すればよい。もし、あなたが、一方の平均値の

方が他方の平均値よりも大きいと言う仮説を検証しようとするならば、片側検定 (one tailed test) を使用すればよい。データに必要とされるパラメーターの仮定は、①正規分布している母集団から選別された標本であること、②分散が等しいこと。検定対象グループ内の従属変数の分布が等しいこと、そして、③データは間隔尺度で、従属変数は間隔尺度か比率尺度のデータでなければならない。

対応のある場合の t-検定

データ群の両方に各々のデータが含まれている場合、対応のある場合の t-検定を用いるのが適当である。この参考事例としては、各参加者が予備検定 (変数-1) と事後検定 (変数-2) を受験する場合の得点を比較することが上げられる。事前と事後の両方のグループに各人に対応したデータがある。

独立している場合の t-検定

各々のデータを、別の変数に基づいて特定のグループに割り当てる場合、独立している場合の t-検定を用いるのが適当である。例えば、男性と女性の読取の得点を比較する場合、性別に基づいて二つのグループに分ける。(データ・ファイルにおける各記録が、ある一つのグループと別のグループに割り当てられる、というのはこうした配分のことを指す。)

自由度：自由度 (degrees of freedom) は標本数に関連する。二組の独立した標本群が対象となっている場合、自由度は二組の標本数を合計したものから 2 を引いたものに等しい。すなわち、自由度 (d.f.) = $(n_1 + n_2 - 2)$ である。

分散分析指標は、特定グループの広がり、もしくは、分散を示す指標である。これには、範囲、分散、及び、標準偏差が含まれる。

範囲は、ある分布における最大値と最小値の差異を示すものである。

参考事例：あるグループの値が 3, 6, 8, 10, 14, 17 である場合、範囲は $14 (= 17 - 3)$ である。この値は最小値 3 から最大値 17 の範囲にある。

分散は、偏差値を二乗したものの平均値である。各データ値と平均値の差を求め、それを二乗し、それらを合計し、データ数マイナス 1 で割って計算する。

標準偏差は、平均値からのデータの広がりを測定するもので、どんな統計的検定においても必須のものである。これは分散の平方根によって求める。これによって、分散が同じ単位の測定値に還元される。標準偏差は「平均値+1 標準偏差」といった形で表現される。仮に、標準偏差が 11 で、値が 63 であったとすると、「平均値+1 標準偏差」は 74 であり。

「平均値+標準偏差×2」は 85 である。この数値は、標準偏差と正規曲線における百分率の関係を理解すると良く分かる。「平均値+1 標準偏差」と「平均値-1 標準偏差」で囲まれた領域は、当該分布のすべての値の約 69%を含んでいる。それ故、先の参考事例で示した 68%の値は 52 から 74 の間に収まっている。

標準偏差の意味を理解するもう一つの方法は、各値を百分率と比較することである。正規分布においては、各ケースの 97.7%が、「平均値+2 標準偏差」の範囲内に収まっていることが知られている。あるケースについての値が、「平均値+2 標準偏差」であると分かれば、このケースでは、97.7%の確率で記録されることが分かる。

良く使われる統計的手法を選択するためのガイド

			検定目的
データのタイプ	統計モデル	差異(グループ間)	関連性 (1 グループ内)
名義			
離散型	ノンパラメトリック	カイ二乗検定	一致係数
順序			
間隔			
		分散分析(ANOVA) (3 グループ)	
連続型	パラメトリック		ポアソン相関重回帰分析 判別分析*
比率		T-検定 (2 グループ)	

*もし予測されるべき変数がカテゴリー分類に適用ものである場合

2-9 調査結果を報告する (ステップ 9)

評価者は、調査結果を各種のステークホルダーや、当該調査結果に関心を持っている可能性のあるその他のオーディエンスに報告する責任を有している。ステークホルダーとのコミュニケーションは、調査結果を、意味のある、受容可能で、有用なものとするために、評価プロセスの全期間を通じてなされる必要がある。(この努力を継続することが必要である。)

報告プラン: ステークホルダーに対する報告プランを立案することが、調査結果をいつ、誰に、どのようにして報告すべきかを明らかにする上で役に立つ。

- ・意図的に報告すべきは誰か。
- ・当該情報はどの時点で必要とされるか。
- ・どんな情報が必要とされるか。
- ・どんな報告形式が好まれるか。

結果報告：多種多様な報告手段が使用できる。

- ・口頭での報告
- ・オーディオ・ビジュアル機器を利用した報告
- ・簡潔なコミュニケーション
- ・日誌又は新聞記事での報告
- ・幹部用サマリーでの報告
- ・図表やチャートを用いた報告
- ・パブリックミーティングで報告
- ・ニュースレター，広報誌，パンフでの報告
- ・個人的面談
- ・ポスターを利用した会議
- ・質疑の機会を用意する

評価の報告には通常，次のことが含まれる。

- ・全体的記述
- ・評価結果の要約と分析
- ・当該プログラムの説明
- ・結論の正当化に対する明示的説明
- ・プログラムの環境条件
- ・将来の所要の変更に対する提言
- ・評価の目的の説明
- ・使用した手段の説明

報告の要点

- ・報告は短く簡潔に，かつ，注意を引くように要点を突いたものとする。
- ・様式を巧妙に考案し，評価報告の内容が意図したオーディエンスのニーズに合うようにする。
- ・オーディエンスが知らないような専門用語は避ける。
- ・会話をするようなトーンを維持するようにする。
- ・長い文章と短い文章を適当に織り交ぜる。
- ・誤解を招きそうな考えや文章を明確に識別できるよう，大きな声で報告書を読み上げる。
- ・積極的口調の文章を書く。
- ・報告書は論理的構成とするよう心がける。
- ・ドラフトの作成に時間を掛ける。また，反応を聴取し，取り入れるとともに，校正を行う。

ネガティブな調査結果の報告

しばしば，あなたはネガティブな調査結果を報告するよう求められることがある。当該プログラムが所期の目的を達成できなかった場合がそれである。当該プログラムのマネジメントが間違っただけでなされているか，何らかの変化が求まっているか，といった状況が生じる。このような場合，評価は，そうしたネガティブな結果の原因を特定し，指摘することができる。こうした困難さを報告することは，将来の過ちを回避し，改善策を提案する上で有用である。しかしながら，ネガティブな調査結果は，失敗したという感覚を醸成するよりも，学習の成果と改善

点を向上させるという観点でなされるべきである。

ネガティブな評価結果は、以下に述べるような方法で報告されなければならない。

- ステークホルダーの感覚に敏感である。
- 最初にポジティブな調査結果を報告する。
- 「達成」とか「進歩している」とか「現在、取り組み中である」といったポジティブな表現を使う。
- 「反省している」、「対話をしている」、更には、「前向きに考えている」といった雰囲気を作る。
- ステークホルダー自らも問題解決の当事者であると自覚させるようにする。
- 評価プロセスの全般を通じて報告を行うようにする。
- ステークホルダーがネガティブな感情に対処するのを助けるようにする。

2－10 調査成果を実地に適用し、活用する (ステップ 10)

評価は、調査結果が実際に適用されるまでは完全に完了したとは言えない。

- プログラムの継続の可否を決定すること
- 現在進行中の各プログラムを改善すること
- 将来のプログラムを計画すること
- プログラムのステークホルダーに連絡をすること

評価者が自分の担当するプログラムの評価を行う場合、調査結果を実行する上で、直面する問題は少ないが、評価者自身が当該プログラムの実行を担当する人間ではない場合、調査結果が無視される可能性は大きくなる。ステークホルダーの懸念が評価プロセスに反映される場合、当該評価の調査結果が実際に使用される可能性は高くなる。

2－11 評価を評価する (最終ステップ)

評価を評価することは、評価の実際のプロセスとそれほど変わるところがない。それは、評価と同じ基準に適合したものでなければならず、また、元の評価と同じステップでなされなければならない。評価を評価する場合、下記の事項を考慮する必要がある。

- **概念の明確性** ー当該評価は十分焦点を当てたものとなっており、目的、役割、及び一般的アプローチの方法が明瞭に記述されているか。

- **評価対象の目的の記述** ー当該評価は、何が評価されようとしているかについて、徹底した、詳細な記述を含んでいるか。
- **正当なオーディエンスの認識とその代表意識** ー正当な聞き手が全員、当該調査に対して発言する機会を持ち、調査結果をレビューする機会を持っているか。
- **評価における政治的問題への感受性** ー当該評価は、それを破壊する可能性を持った政治的問題、対人関係問題、及び倫理的問題に対する感受性を備えたものであって、かつ、これらに満足に対応できていたか。
- **情報ニーズと情報ソースの明確化** ー当該評価は、所要の情報と情報ソースを明確に特定しているか。
- **データの包括性** ー当該評価は、取るに足らぬデータに捕らわれて泥沼にはまることなく、すべての重要な変数や課題についてデータを収集しているか。
- **技術的正確性** ー当該評価のデザインと手続きは、妥当性、信頼性、及び、客観性に関する科学的な指標に適法するようになされていたか。
- **適切な方法と分析** ー当該評価に適切な方法が採用されたか。それらは適切に使用されたか。データは慎重に分析され、解釈されたか。
- **費用への配慮** ー当該評価は、その他の変数と同様にコスト要素についても十分な配慮をしていたか。
- **評価を判定する明示的な基準と指標** ー当該評価は、評価対象を判定するために使用される指標や基準について、明示的にそれらをリストアップし、あるいは、議論を展開していたか。
- **評価結果に基づく判定と提言** ー当該評価は、単に調査結果を報告するに止まらず、当該データによって示唆される判断や提言を提案したか。（これらを報告文書に付記する）
- **それぞれのオーディエンスごとに個別に対応した報告書** ー当該評価は、適切なタイミングで、適切なフォーマットで、それぞれの適切な聞き手に報告されたか。

評価を評価すること：評価のアカウンタビリティの階層

プログラムのイベント		証拠の対応レベル
プログラムと方針決定の影響		プログラムをどの程度、また、どのようにして改善することが出来たか。 どの程度の情報を提供され、どの程度高品質な方針決定が出来たか。
実践とプログラムの変更		意図された利用がどの程度できたか。提言はどの程度実行されたか。
ステークホルダーの知識と態度の変化		意図されたユーザーは何を学んだか。
主要ユーザーの反応		意図されたユーザーは当該評価をどう受け止めたか。 信頼性、信用度は、重要性は、正確性は、潜在的な利用可能性については、どうであったか。
ステークホルダーの参画		誰が関与したか。全体を通じて主要なステークホ

	ルダーはどの程度関与したか。
評価活動	どんなデータが収集されたか。焦点や、デザイン、分析はどうであったか。当該評価で何が生じたか。
インプット	評価のための資源がどの程度十分に利用可能であり、かつ、良くマネジメントされていたか。当該評価を実施する上で十分な時間はあったか。

評価のプランニング用ワークシート

評価対象のプログラム、その目標、及びステークホルダーを特定する	評価の実施のフィージビリティを判定する	プログラムの利益を明確にする指標をステークホルダーと打ち合わせる	データ収集方法を明確にする	データ収集の手法を選択する
ターゲットの母集団を特定しサンプルを抽出する	データ収集担当者を決める	データの分析方法と解釈の方法を決める	調査結果をステークホルダーとどのように共有するか。	

参考文献：

- Archer, T. and Layman, J. (1991). "Focus Group Interview" Evaluation Guide Sheet, Ohio Cooperative Extension Service.
- Bennett, Claude F. and S. Kay Rockwell (1995). *Targeting Outcomes of Programs (TOP): An Integrated Approach to Planning and Evaluation*. Lincoln, NE: Cooperative Extension, University of Nebraska.
- Bennett, Claude. 1979. *Analyzing Impacts of Extension Programs*. Washington, DC: US Department of Agriculture, Science and Education Administration (ES C-575).
- Case R.; Andrews, M. and Werner, W. (1988). How can we do it? an evaluation training package for development educators. British Columbia, Canada: Research and Development in Global Studies.
- Contant, C. K. (1993). "Assessing What and Why: Designing and Using Evaluations Effectively for Local Level Programs." Paper presented at the Rural Nonpoint Source Pollution in the Upper Midwest Conference, March.
- Dillman, D. A. (1995). "Survey Methods." Class notes in AG*SAT Graduate Course in Program Evaluation in Adult Education and Training, University of Nebraska-Lincoln.
- Fink, A. (1995). How to Sample in Surveys. Thousand Oaks, California: Sage.
- Fraenkel J. R and Wallen, N. E. (1996). How to Design and Evaluate Research in Education. New York: McGraw-Hill Inc.
- Krueger, R.A. (1994). Focus Groups: a Practical Guide for Applied Research. 2nd edition, Thousand Oaks, California: Sage.
- Mueller, D.J.(1986). Measuring social attitudes. New York: Teachers College Press.
- Neito, R. and Henderson, J.L. (1995). Establishing Validity and Reliability (draft). Ohio State Cooperative Extension.
- Patton, M. Q. (1997). Utilization-focussed evaluation: The New Century Text. Newbury Park, California: Sage.
- Salant, P. and Dillman, D.A. (1994). How to conduct your own survey. New York, NY: John Willey & Sons, Inc.
- Wholey, J. S.; Harty, H. P. and Newcomer, K. E. (eds.). (1994). Handbook of practical program evaluation. San Francisco: Jossey-Bass Publishers.
- Worthen, B. R. and Sanders, J. R. (1987). Educational evaluation: alternative approaches and practical guidelines. New York: Longman.
- Yin, R. K. (1984). Case study research: design and methods. Applied Social Research Methods Series. Vol. 5. Newbury Park, California: Sage.

普及プログラムの評価実施のための資料

本資料は、前掲の「普及プログラムの評価方法」(How to Conduct Evaluation of Extension Program)の付属資料ではない。

ペンシルベニア州立大学で教材として使用されているもので、2004年9月にペンシルベニア州立大学を訪問したときに、Dr. Rama Radhakrishna から頂いたものである。

(調査部長 谷口敏彦)

1. 評価入門

評価は、伝統的な社会的リサーチに密接に関連するが、それとは明確に異なる方法論の領域の手法である。評価においては、伝統的な社会的リサーチに使用されるのと同様の手法が多く利用されるが、評価それ自体は、政策や組織との関連でなされるため、集団としてのスキルや、マネジメント能力、政治的な柔軟さ、多数のステークホルダー（関係者）に対する感受性、更には、一般的な社会的リサーチではそれほど使用されないその他のスキルを必要とする。ここでは、評価の考え方と、この分野におけるいくつかの重要な用語や問題について紹介する。

1-1 評価の定義

最も一般的に使用される評価の定義は、次のとおりである。

「評価（evaluation）とは、ある種の目的（object）が持つ価値又はメリットを体系的にアセスメント（assessment）することである。」

この定義は、完璧な定義とは言い難い。必ずしも、メリットや価値をアセスメントすることを目的としないタイプの評価というのがいくつもあるからである。いくつかの事例をあげると、記述的な作業、実施状況の分析、形成的評価などがこれに当たる。評価の情報処理機能とフィードバック機能をもっと強調した定義の方が、おそらく、より優れた定義と言えるだろう。例えば、次のような定義の仕方が可能である。

「評価とは、ある目的に対して、有用なフィードバックを行うために必要な情報の入手と評価のための体系的な手法である。」

どちらの定義も、評価とは体系的な取り組みであるとしている点、及びプログラムや政策、技術、人間、必要性、活動などのあいまいな用語を意図的に「目的」に掲げている点で共通している。後者の定義では、あるメリットの価値のアセスメントよりも、情報の入手やアセスメントに力点を置いているが、これは、評価に関わるあらゆる作業が、結果として得られるメリットや価値を評価するか否かに関わらず、データを収集し、これらをどこかに移動させたり、情報の妥当性について判断を下したり、そこから得たアセスメントの妥当性を判断したりすることに関わっているからである。

1-2 評価の目的

ほとんどの評価の最終目的は一般に、スポンサーや資金提供者、さまざまな顧客、管理責任者、スタッフ、及びその他の関係するオーディエンスに対して、「有用なフィードバック」を提供することにある。最も頻繁にみられる事例では、フィードバックは、方針の決定を助ける上で「有用である」と看做されている。しかし、評価とその影響の関係は、それほど単純なものではない。決定的に重要と思えた調査がしばしば短期的な決定を誤らせる場合もあるし、当初は決定にそれほど影響を与えないと思われていた作業が、状況の進展に応じて、あとから大きな影響を与えるようになって来る場合もある。こうした事実にも関わらず、評価の主たる目的

は、実証的なフィードバックによって、方針の決定や準備に影響を与えることにあるとする見方が幅広いコンセンサスを得ている。

1-3 評価戦略

「評価戦略」(Evaluation Strategies)というのは、評価に関する幅広い、支配的な見方を確立することである。評価という作業は、すべての集団の見方の中から適切なものを取捨選択して採用することであるが、「評価戦略」としては、評価する人間の中から、最も一般的なグループや集団を選択して対象とする。以下では、評価戦略のうちで、最も重要な4つのグループを取り上げる。

第一の戦略は、科学的実験モデルである。これは、おそらく、歴史的に最も多く利用されてきた評価戦略である。これは、科学、特に、社会科学に基づく価値や手法によって、収集された情報の不偏不党性や、正確性、客観性、更には、妥当性の望ましさを順位付けることである。この科学的実験モデルには、伝統的手法による実験設計と準実験的手法による設計、教育の結果としての目標の達成に焦点を合わせた調査、費用効果分析や費用便益分析を含む計量経済学的将来見通し、最新の理論的推論による評価の表明などが含まれる。

第二の戦略は、マネジメント志向システム・モデルである。これらのモデルのうち、最も良く使われている二つを上げると、一つは、PERT、すなわち、プログラム評価レビュー(Program Evaluation and Review Technique)であり、もう一つは、CPM、すなわち、クリティカル・パス分析法(Critical Path Method)である。どちらも、わが国では、ビジネスや行政の分野で幅広く使用されている。米国の国際開発庁によって開発された論理的なフレームワーク、すなわち、「ログフレーム」(Logframe)モデルと、このカテゴリーに含まれる一般的なシステム理論やOR(オペレーション・リサーチ)手法をこれらに含めることもできる。別の評価機関によって、これらとは異なる二つのマネジメント志向システム・モデルが開発されている。その一つは、UTOSで、UはUnits(単位)を、TはTreatments(処理)を、OはObserving Observations(観察結果の観察)を、SはSetting(環境)を表している。もう一つは、CIPPで、CはContext(内容)を、IはInput(インプット)を、最初のPはProcess(プロセス)を、二番目のPはProduct(製品)を表している。これらのマネジメント志向システムモデルは、より大きな組織的活動のフレームワークの中に当該評価を位置づけることによって、当該評価の持つ包括的性格を強調している。

第三の戦略は、いくつかの定性的・人類学的モデルである。これらは、観察の重要性、評価の対象である状況の現象学的な性質に留意することの必要性、及び評価のプロセスにおける人間の主観的な解釈が大切であると強調する。このカテゴリーには、評価の世界では「自然主義的評価」あるいは「第四世代の評価」として知られているものがある。すなわち、さまざまな定性的学説、批判的理論と美術批評的アプローチ、及びグレイザーとストラウスの「根拠のある理論」的アプローチなどである。

最後に、第四の戦略として、参加者志向型モデルがある。その言葉が示すように、これは、

評価の中心を参加者、特に、顧客や当該プログラム、もしくは、技術のユーザーに置くべきことを強調する。顧客やステークホルダーを志向するアプローチは、顧客志向型評価システムと同様、この参加者志向型モデルの例である。

これらのさまざまな戦略の中から、どれをどのように選択すればよいのか。この問題は、評価専門家の間でも大きな議論を呼んでいる問題である。議論は一般的には、さまざまな手法の信奉者の間で戦わされている。各論者はそれぞれの手法の優越性を論じたてる。現実問題としては、最も優れた評価者は、これら四つのカテゴリーの手法のすべてに通じていて、必要に応じて、その中から最も適切な手法を用いるということである。これらの手法には、本来、お互いに両立しがたい要素はない。それぞれの手法が何らかの価値ある評価を評価テーブルに提示してくれる。実際、近年では、異なる戦略によって、異なる見方に基づいて、異なる手法によって実施された評価の結論がある戦略がどのようにして統合すればよいか、という問題が次第に注目されるようになってきている。この問題に対しては、明らかに、単純明快な答えは一つもない。問題は複雑であり、必要とされる方法論も多岐にわたるからである。

1-4 評価のタイプ

評価には、評価される対象と評価の目的に応じて、さまざまなタイプがある。最も際立った対照を見せる評価のタイプは、形成的評価と総括的評価である。形成的評価は、評価される対象を強化し、あるいは、改善する。それらは、当該プログラムや技術を実施すること、その実施の出来栄え、それが組織の状況や、人間、手続き、更には、インプットなどの評価を検証することによって、評価の対象そのものをより良く形成することを助ける。総括的評価は、これとは対照的に、ある対象の影響やアウトカムを検証する。すなわち、それらは、当該プログラムや技術が実施されたあと、何が生じるかを明らかにし、当該対象事象が確かにそのアウトカムをもたらしたと言えるかどうかを評価し、中間アウトカムを越えた因果関係のあるファクターの全体的影響を明らかにし、また、当該対象事象に関わる費用を見積もることによって、当該対象事象を要約する。

形成的評価には、いくつかのタイプの評価がある。

- ① **ニーズ評価**は、そのプログラムを必要としているのは誰か、そのニーズの大きさはどれほどか、また、そのニーズにはどうすれば応えることができるかを決定する。
- ② **評価可能性アセスメント**は、評価が可能かどうか、また、ステークホルダーがその有用性を具体化する上でどのような助けをすることが出来るかを決定する。
- ③ **体系的概念化**は、ステークホルダーが当該プログラムや技術、ターゲットとなる層、予想されるアウトカムを明確に理解することを助ける。
- ④ **実施評価**は、当該プログラムや技術が忠実に実行されているかどうかをモニターする。
- ⑤ **プロセス評価**は、当該プログラムや技術が実施されるプロセスを代替的な実施方法を含め、調査する。

総括的評価も、次のように区分けできる。

- ① **アウトカム評価**は、当該プログラムや技術が、本当に、意図的に設定された所定の目標

としたアウトカムを目に見える形でもたらしたかどうかを調査する。

- ② **影響評価**は、それよりも広範なレベルで、あるプログラム、もしくは技術が意図されていたか、そうでないかを問わず、幅広い、あるいは、正味の影響を全体としてアセスメントする。
- ③ **費用効果分析及び費用便益分析**は、アウトカムを金額で費用と便益に換算してその効率を論じる。
- ④ **第二次分析**は、新しい問題に取り組むため、あるいは、それまでに使われなかった手法を使用するため、現存するデータを再検証する手法である。
- ⑤ **メタ分析**は、評価の対象となった質問に対して、全体的な、あるいは、総括的な判定を下すために、複数の調査から得られたアウトカムを統合することである。

1—5 評価のための質問と手法

評価者は、さまざまな質問をし、それらに対応するために、多くの手法を使用する。これらは、既に述べたように、形成的評価や総括的評価のフレームワークの一部とみなされる。

(1) 形成的評価のための調査の質問と手法

① 問題や論点の定義と範囲はどこまでか。また、質問は何か。

ブレインストーミングやフォーカスグループ、ノミナルグループ(名前を明らかにしてグループに参加させる手法)などの手法、デルファイ法、ブレン・ライティング(各自にそれぞれの意見を書かせる方法)、ステークホルダー分析、シネクティックス(多様なメンバーから成る小グループの自由な討論によって問題の発見や解決を図る方法)、水平思考(既存の思考法にとらわれずさまざまな角度から問題を包括的に扱う手法)、インプット・アウトプット分析、及びコンセプト・マッピング(さまざまな概念を一枚の紙の上に書き出し、地図化する手法)。

② 問題はどこにあるか。また、その大きさや深刻さはどの程度か。

ここで最も良く使用される方法は「ニーズ評価」であるが、これには、既存のデータ・ソースの利用、サンプル調査、評価の対象となる人へのインタビュー、定性的調査、専門家による証言、及びフォーカスグループを含めることが出来る。

③ その問題に対応するためには、そのプログラムや技術をどのように実施すればよいか。

ここでは、これまでにリストアップした手法のいくつかが適用可能である。シミュレーション手法のような詳細展開の手法、多重用途理論や実証的因果関係モデルのような多変数法、フローチャートのようなプロジェクトの計画と実施のための手法、PERTやCPM、更にはプロジェクト・スケジューリングの手法などである。

④ 当該プログラムや技術が如何に円滑に実施されているか。

ここでは、定性的な、また、定量的なモニタリングの手法、各種のマネジメント・インフォメーション・システムの利用、実施評価の手法などが適切な方法論である。

(2) 総括的評価のための質問と手法

① どのようなタイプの評価が可能か。

ここでは、適切な評価の設計を選定するための標準的なアプローチに加えて、評価可能性のアセスメントが使用可能である。

② そのプログラムや技術の有効性はどうであったか。

人は、望ましい効果が生じたかどうかをデモンストレーションするために、各種の観察手法や相関関係分析を選択し、また、観察された結果が当該プログラムや技術の介入によるものか、それとも、その他の要因によるものかを理論的に明らかにするために、準実験法や実験法による設計を選択する。

④ 当該プログラムの正味の影響は何か。

費用効果や費用便益をアセスメントするために、ここでは、意図された影響と意図されなかった影響のすべてを総括することが可能な定性的手法とともに、計量経済学的手法が適用される。

ここに述べた各種の手法と、ここでは触れられていない多くのその他の手法については、専門的なりサーチ手法に関する文献を参照すること。実に恐るべき数のツールが存在している。しかし、これらの手法を、改善して、アップデートし、かつ、変化する環境に適応させる必要があるため、方法論の分野での研究開発が評価という作業の中で、大きな位置を占めているのは確かである。

2. 指標についての考え方

2-1. 指標は、測定可能な特性を観測する。

指標は、目標値を表すために使用される。それらはまた、ある目標値がどのくらい良く達成されたか、あるいは、あるプログラムがどのように機能しているかを示すことの出来る測定可能な指数でもある。

2-2. 指標は、客観的であり、主観的でもあり、またその両方でもある。

客観的指標では、データは、人々が公然と行い、あるいは受容する事柄に関する直接的な観察や、自然現象の観察を通じて収集される。これらのデータの収集手順には、統計的な記録を通じた情報の収集、参加者に対する客観的なテストの実施、ある状況に関する第三者の観察などが含まれる。

主観的指標では、データは、プログラム参加者自身や、そのプログラムによって影響を受ける可能性のある他の人による自己報告の手順を通じて収集される。

階層の各レベルにおける客観的・主観的指標の例

SEE ⁷	客観的指標：平均寿命，損益計算書，大気や水の質に関する諸指標 主観的指標：個人の健康に関する人々の満足度，経済状況，大気・土壌及び水の清浄度
プラクティス	客観的指標：推奨されたプラクティスや技術のプログラム参加者による採用や活用に関する外部からの（直接的，「第三者」による，偏りのない）構造化された観察 主観的指標：推奨されたプラクティスや技術のプログラム・チームやプログラム参加者による採用や活用に関する報告等級付け
KASA ⁸	客観的指標：知識・態度・スキル，及び意欲に関するテストの成績や適切性が確認された諸指標 主観的指標：知識・態度・スキル，及び意欲に関する参加者自身による自己評価
反応	客観的指標：プログラム活動の内容の主題に対する参加者の注目度に関する外部からの構造化された観察 主観的指標：プログラム活動の内容の主題に対する参加者の興味に関する参加者自身による等級評価
参加	客観的指標：プログラム活動への参加や，プログラム活動を行う上での自主的なリーダーシップ等に関する外部からの構造化された観察 主観的指標：プログラム活動への参加やプログラム活動での自主的なリーダーシップ等に関する参加者自身による自己報告
活動	客観的指標：プログラム活動の頻度，期間，方法論，及び内容に関する外部からの構造化された観察 主観的指標：プログラム活動が実施された方法やその頻度，及び期間に関するプログラム・スタッフによる報告
資源	客観的指標：各プログラムに割り当てられた人員の時間支出に関する外部からの観察 主観的指標：各プログラムに割り当てられた人員時間支出に関するスタッフ自身による事後報告

2－3．階層の中での指標の使用

SEE，プラクティス，及び KASA レベルにおける指標は，コミュニティや社会のニーズのみならず，当該コミュニティ内に居住する個人のニーズに焦点を当てる。これらの指標は，一定のアウトカム目標を特定するために使用され，これらの目標値に関連するあらゆる変化を測定するためにも使用される。これらの指標はまた，プログラムのアウトカムや影響を特定するためのデータ収集の手順を策定するためにも使われる。

⁷ SEE：Social, Economic, and Environmental conditions の頭文字で，社会的・経済的・環境的条件をいう。

⁸ KASA：知識（knowledge）態度（attitudes），スキル（skills），及び意欲（aspiration）の頭文字を取ったもの。

社会的指標アプローチは、社会的・経済的・環境的条件を表すことができ、アウトカムはこれらの指標の変化によって測定できると仮定する。このため、乳幼児死亡率や世帯収入水準、水質汚染物質の PPM 濃度といった統計は、あるコミュニティや社会がその住民のニーズを満たすという課題をどれくらいうまく実行しているかを示す客観的指標であると理解されている。個々のプログラムの背後にある多くの要因がこれらの指標に影響を与えることがある。加えて、SEE レベルにおける指標はたいてい測定することが困難であり、変化に年月を要することもあり、また、通常は、追跡するのに費用がかかる。

プラクティス・レベルにおける指標は、顧客の行動を対象とすることもあれば、ある特定のプログラムによって定義される一連の活動を対象とすることもある。これらの指標の変化は、それぞれのプラクティスの導入時に予見されるように、ある特定のプログラムが実施されてから数日以内、あるいは、数か月以内に測定されることもあれば、変化に数年間を要する場合もある。

KASA レベルでの指標は、それぞれの活動が実施された直後にそのアウトカムを特定するというレベルから、数か月間をかけて生起するもっと継続的な変化を特定するというレベルまで、多岐にわたる。

反応、参加、活動、及び資源の各レベルにおける指標は、プログラムの実施に焦点を当てるものである。これらの指標は一定のプログラムの目的を特定し、特定のデータ収集のニーズを特定するとともに、当該プログラムがどのように実施されるかに関する書類を作成するために使用される。多くの場合、これらの指標はまた、プログラムを改善する目的でデータ収集の必要性を特定するためにも使用される。これらのレベルにおける指標のデータは、収集するのが最も容易で、かつ、最も費用がかからない。

2-4. 指標は、特定のプログラムの目標やアウトカムに適用される。

教育者らは、プログラムの改善や組織的なアカウンタビリティ（説明責任）を果たす必要性に基づいて、プログラムの指標を特定する。これらの必要性に基づき、プログラムの計画者たちは、彼らが測定しようとするアウトカムに関して、一つ、又は複数の指標を特定する。

さまざまな人的サービス機関や、若者、もしくは、家族に対するサービスを対象とした組織におけるプログラムのアウトカムやアウトカム指標の参考事例は、「ユナイテッド・ウェイ・オブ・アメリカ」⁹のウェブサイト、<http://www.unitedway.org/outcomes/examples.html>. に示されている。

⁹ 訳注：アメリカのボランティア団体で、全米各地に支部を有している。

3. ステークホルダー・インプット

1998 年の AREERA 法（農業における研究開発、普及、及び教育改革に関する法律）の第 102 章 C 項は、土地付与大学¹⁰がステークホルダー・インプットに関する情報を提供するように求めている。

同法は、「ステークホルダーを募り、彼らの参画を促すために取られた方策と、当該大学がステークホルダーとなり得る個人や団体を特定するために使用した手順と、彼らからのインプットを収集するために取った手順に関する文書」を情報開示するよう定めている。

クレムゾン大学では、このステークホルダー・インプットが普及プログラムの成功を左右する鍵となっている。クレムゾン大学では、このステークホルダー・インプットを得て、これを作業計画（Plan of Work, POW）に織り込むことが、長い間の歴史と伝統となっている。ここでは、ステークホルダー・インプットについて議論したい。これには、下記の問題がある。

ステークホルダーの識別、すなわち、POW の手順において、インプットを行うのは内部のステークホルダーなのか、外部のステークホルダーなのか、という問題と、ステークホルダー・インプットを求める際に使用された手順に関する問題、ステークホルダーへの POW に関する質問、及び外部ステークホルダーのプロフィールである。

① ステークホルダーの識別（内部、及び外部のステークホルダー）

<内部ステークホルダー>

内部ステークホルダーに含まれるのは、普及事業の管理運営者、各プログラムの管理運営者、カウンティ（郡）の普及事業責任者、普及事業の担当部局、各担当部局の関係者、専門家、教職員、学部長、普及（研究）部の教職員と管理運営責任者である。

<外部ステークホルダー>

外部ステークホルダーには、普及事業の諮問委員会、各産品グループ、地域のリーダー、福祉サービス提供者、企業・産業界、ファーム・ビューロ（農業団体）、商工会議所などの協力機関、農業局などである。

② ステークホルダー・インプットのために使用される手順と質問

POW の手順では、ステークホルダー・インプットを収集するために、NGT 法（Nominal Group Technique, NGT）が使用された。NGT 法は、多くの人々のグループが、比較的短期間のうちに、さまざまなアイデアを考え出す場合に使用される手順である。これは、問題を明確にし、その解決策を求め、優先順位を確定する上で有用である。

この NGT 法を実施する上で、いくつかのステップが取られた。まず、第一に、46 のすべてのカウンティに対して、それぞれのカウンティで、NGT 法を実施することが求められた。第二

¹⁰ 訳注：連邦政府によって土地を付与され、農業に関する研究、教育、普及を実施する州立大学

に、この NGT 法の実施に関与するすべてのカウンティ事務所とカウンティの管理責任者（郡の長）に対して、実施すべきステップに関する指示が与えられた。第三に、多様性を損なわないようにするため、各地域コミュニティの特徴を反映するような人々のクロス・セクションを正確に得ることが強調された。第四に、NGT 法を円滑に実施するための一連のルールと手順が提示された。第五に、それぞれのステークホルダーごとに回答を求める一連の質問が定められた。これらの質問には下記のものが含まれていた。

- 1) あなたのカウンティが今後 5 年間に遭遇する最も重要な課題は何ですか。それを、5 つあげてください。
- 2) 普及活動は、これらの課題のどれかに対応するものですか。
- 3) 課題や心配事に対応するためには、どんなパートナーシップを作ることが必要と考えますか。

これらの質問に対する回答は、カウンティごと、集団ごと、及び州全体に集約された。サウス・カロライナ州では、この手順を実施した結果、16 件のイニシャチブ（計画）と 70 件のプロジェクトが選定され、これらが後に、同州の公共サービス・農業局（PSA）の戦略目的に組み込まれた。

③ ステークホルダーのプロフィール

POW では、あらゆる外部ステークホルダーのうち、普及諮問委員会が重要な役割を果たしている。普及諮問委員会のメンバーは、サウス・カロライナ州の協同普及事業に対して、コンスタントにインプットを提供している。各諮問委員会メンバーのさまざまなプロフィールが多様な問題に対応する上で役に立っている。諮問委員会のメンバーは全員で 237 名であるが、このうち、48%が男性、52%が女性である。81%が白人、15%がアフリカ系、その他の人種が 4%である。年齢構成で言えば、35 才以下が 5%、36 才～55 才が 46%、55 才以上が 49%である。84%が既婚者、もしくは、特定のパートナーと同棲している者である。学歴では、15%が高校卒業者、18%は何らかの専門学校修了者、35%が大学卒業者で、残る 32%は、修士もしくはこれに相当する何らかの専門的学位を保有している。

デューク氏が 1999 年に行ったサウス・カロライナ州の伝統的、及び非伝統的なオーディエンスに対する土地付与大学の 21 世紀の公共サービスに対して、何を期待しているかについての調査では、同州のコミュニティが直面している重要問題は、次のとおりであった。

すなわち、成長と人口問題、環境、コミュニティと経済発展、農業生産性、家族問題、そして食料・健康・栄養問題である。

4. ステークホルダーとは誰か

4-1 ステークホルダーは、次のカテゴリーの人又はグループ

- 評価結果に基づき判断を下す立場にある。

- 評価の対象となっているプログラムの計画策定に携わっている。
- 評価によって影響を受ける可能性がある。
- 評価について知る法的権利を有している。
- 評価を支援又は依頼する立場にある。
- 評価の対象となっているプログラムを承認又は批評する立場にある。
- 評価のための情報を提供する立場にある。
- 調査の対象となっているグループに対して責任を負っている。
- 評価に要する費用を負担する。
- 評価の対象となっているプログラムを推奨している。

4-2 参考情報

ステークホルダーとなり得る者を分析する（評価のためのオーディエンス）

あなたの知っている人物全員をオーディエンスとしてリストアップすることに気を悩ますことはない。現実的に取り扱える人数以上に多くの人を参加させると、混乱を引き起こす可能性がある。重要なオーディエンスは誰であるかを良く判断すること。また、どの程度、参画させるのか、そのレベルをはっきりさせておくこと。

もし、特定のオーディエンスを評価の対象から除外する場合、そうすることによって、どのような結果がもたらされるかについて熟考すること。

一人のオーディエンスが他のオーディエンスにつながることもある。だから、数名の良く知っているオーディエンスと話をしたり、あるいは、彼らのことを考えたりすること。その上で、彼らから他のオーディエンスへとつなげていくこと。

各オーディエンス・グループのメンバーらと話をすることは、評価に対する彼らの考え方を知る上で良い方法である。

評価の途中で、オーディエンスを追加する必要があるかもしれない。

オーディエンスに参加させようと熱心になるあまり、彼らに何でも約束するようなことをしてはいけない。あなた方がどのような制約条件を抱えているかを彼らに良く理解させること。

評価の過程を通じて、あなた方がオーディエンスをより良く適応させようとする中で、彼らに戻らなければならない可能性がある。従って、最初に良好な協調関係を確立しておくことが大切である。

考えてみることに。この評価を駄目にする可能性があるのは誰か。この評価を成功させることができるのは誰か。それはあなたのリストに載っている人やグループか。

どんな評価も常に「結果」と、時には「副産物」を伴うことを覚えておくこと。従って、常に、これらの結果を予測し、影響を受ける可能性のある人々を含めるように配慮すること。

4-3 ステークホルダー

① 決定責任者や情報の利用者が適切な者は誰であるかをどのようにして明らかにし、組織化するか」－ パットン

- 1) その情報を利用したい人々と利用可能な人々、すなわち、その情報を欲しがっている人々を見つけ、結び付ける。
- 2) 本当に適切な人は誰であるか明らかにするに当たっては、正式な肩書きや権限は参考となる指針の一部となるに過ぎない。
 - ◆ 熱意があり、明確な意識を持っており、能力があり、関心を持ち、かつ、積極的な人間で、戦略的に好都合な位置にある人々。
 - ◆ これは、場合によっては、職階の低い人物の方が適切な人々であることを意味する場合もある。
- 3) 権限や提案要求等に関わりなく、最も重要な情報となるのは、使える情報である。
- 4) 結果を利用する人々に関心を抱く評価者こそ、全体としてみれば、良い仕事をする人間であるということを組織は理解している。

② ステークホルダーと4つのセッションを持つ（協力の仕方）

- 1) 問題に関するセッション
- 2) 方法に関するセッション
- 3) インストルメントを回収し、それらからデータを回収する作業を行う。
データ収集の前に模擬シミュレーション・セッションを持つ。
 - ◆ この種の情報に対してあなたならどうしますか。
(情報を収集する前に、この質問に対する回答を出しておくこと)
 - ◆ 人々が最終結果の報告に興味を抱くように、事前に、実際の結果がどうなるか推測してみる（それを文書にしておく）（要員及び資金）
- 4) 事前テストの結果を見て、それは本当にあなたが求めている結果であるかどうかを確認する。

③ 忘れずに！

- 1) これらの人々は、決定責任者として、また、当該評価結果のユーザーとして、プログラム活動のアウトカムを予測する能力を高め、思慮深さを深めることによって、決定の不確定性を減らすために、積極的に情報を求めている。
- 2) 人間は何かを手にするると、それを「自分のもの」にしたがる傾向がある。
- 3) 個人的要素は、利用の面にのみに当てはまるのではなく、評価のプロセス全体に当てはまる。

④ 結果の利用を防ぐ方法 - パットン

- 1) 評価者は、自身の関心、必要性、基準、及び創造性に基づき、すべての質問に答えること。
- 2) 個々人よりも、標準的な「オーディエンス識別法」を使用すること。
- 3) 決定責任者、すなわち、情報の利用者よりも、方針決定と情報そのものに注目すること。
(評価者は、選択に委ねられている)。
- 4) 評価やプログラムへの資金提供者は、単なる利用者に過ぎないと、機械的に判断すること。
- 5) 個人にではなく、組織における評価を対象とすること。(組織は情報を消費しないが、個人は情報を消費する。)
- 6) 最後に、利用者を組織化し、その後は、彼らを見捨てる。

⑤ ステークホルダーと協力して仕事をする際のガイドライン - パットン

- 1) 質問に耐え得るようなデータを持ってくることは可能か。
- 2) 一つの質問に対して複数の回答が可能である。言い換えれば、答えは、質問の言い回しによって予め決められるものではない。

＜経験則対判断＞

- 学生の知識の「向上」とはどのようなものか。
- 学生の知識における「差異」とは何か。

- 3) 決定責任者は、質問に答える上で役に立つ情報を「求めて」いる。彼らが知りたがっていない事柄については質問しないこと。
- 4) 決定責任者は情報を必要としている。答えをあらかじめ知っているわけではない。既に知っていることは何か、また、知らなければならないことは何かを、彼らが分かるように手助けすること。
- 5) 決定責任者は、他人のためではなく、彼ら自身のために質問に答えたいのである。

例：教師は生徒の学力レベルについては既に知っている。

- 直接的、かつ、個人的関心
- もし、他者も情報を求めているとステークホルダーが考える場合には、そうした他者を含める。

- 6) 利用者は、どのようにその情報を使用するかを特定することができる(将来的な利用について)。

変えることができないような状況について、なんらかの情報を引き出すような質問は、あれこれ考える価値がない。

⑥ 重要なステークホルダーを、どのようにして評価プロセスに参加させるか。

- 1) 適切な人々を判別し、彼らで「タスク・フォース」を結成すること。
- 2) 彼らからコミットメントを得ること。
 - タイム・スケジュールを定める。
 - 前もってコミットメントを得る
- 3) 彼らをすべてのステップに関与させること。

⑦ ステークホルダーとの最初のミーティング

ステークホルダーとの最初のミーティングにおいて、あなた方はまず、二つの情報を収集する必要がある。(プログラムについて、どのような説明が記述されているように、また、当該プログラムの目的が達成されようと否とに関わらず、それに加えて) 当該プログラムに関して、彼らはどのような質問を持っているか。彼らの評価に関する質問は何か。

<いくつかの代替的なアプローチ>

質問内容：評価という言葉聞いたとき、何を思い浮かべますか。(この質問により、あなたは、そのステークホルダーが評価の役割をどのように捉えているかを見抜くことができるだろう。)

空白部分を 10 回埋める：私はこのプログラムの_____について知りたい。

各質問には、成功に向けた、隠された（もしくは、それほど隠されていない）クライテリアがリンクされている。以下はその一例である。

質問 — 参加者は何か新しい情報を習得したか。

隠されたクライテリア — 私の定義では、「何か新しい情報を習得した」というのは、参加者全員が、当該プログラムの期間中に学んだ新しいアイデアを、少なくとも 3 つあげることができなければならないということを意味している。また、私のクライテリアでは、参加者全員が「私は新しいアイデアを学んだ」と報告することである。あるいは、私のクライテリアでは、参加者の大部分が事前テストよりも事後テストで高い点数を取得することである。

この例に関して留意すべき点は以下である。

ときには、クライテリアによって方法論が決められる（ミーティング後のアンケート、事前・事後テスト）。最終的に決定されるクライテリアは方法論を決定づけるが、最初に設定されるクライテリアはそうである必要はない。「大部分の生徒は知識が増えたことを報告する」ということには、「彼らがどのような方法で報告するか」については何も示さない。ミーティング後のアンケート、ミーティング後の電話、個人的な面談—その方法は何れでも良い。方法論について考えておき、将来のミーティングで、どのような方法を使用すべきか議論する準備を整えておくというのは良い考えである。そうすることによって、各ステークホルダーは、「私は『ハードデータ』しか受け付けない、私は知識の見方を受け入れる、など」、それぞれの反応を表明することができる。

クライテリアは個人レベルではなく、グループレベルで設定される。すなわち、一定の割合で、これは可能である。全員、大多数、60%、少なくとも半数、100%、など。これらのクライテリアが達成されたかどうかを判定できるようにするためには、あなた方は、各個々人に合わせた行動面での目標を書面に記録しておくことが必要である。すなわち、「Xを与えられた生徒はすべて、Yを行うことができる」というように。また、もしグループのクライテリアが60%に設定された場合、あなた方は、個々人の業績を検証することによって、それが達成されたかどうかを知ることができる。

ほとんど大部分のステークホルダーが非常に厳密な結果を得ることは困難であろう。彼らの評価に関する質問は次のように始められれば良い。

「そのプログラムは有効であったか」。この場合、あなた方は、彼らが言うところの「有効」とは何を意味するのかをあなた方に説明してもらうに当たり、彼らを手助けしなければならないだろう。次に、彼らが質問に対する回答を決める際、彼らにとってどのようなクライテリアが参考になったかを明らかにするとき、彼らを手助けしなければならないだろう。クライテリアを定めようとするときにたずねるべき良い質問というのは、以下のようなものであろう。

- 質問に対する答えを決めるとき、どのような証拠なら受け入れられるか。
- 「学ぶ」ということを、私たちは、どう定義できるか。
- 人々が学ぶことについて、あなたはどの程度まで計画を立てるか。
- 「生徒は学習した」と言い得るためには、あなたは、何名の参加者から知識の増加があったという報告が必要か。

これらのクライテリアのそれぞれに対する評価の質問はどのようなものになるだろうか。また、それらの質問はどのレベルに位置づけられるだろうか。

- 広告に500ドル以上は費やさない。
- 説明会のうち、少なくとも3時間は小グループのディスカッションで行われる。
- 5つの地域のそれぞれでワークショップが開催される。
- 参加者の大多数は40歳以上である。
- 教師はオハイオ州において特定の科目を教える資格認定を受けている。
- 少なくとも半数の顧客が、参加したことによる満足感を報告する。
- 講師は教室内の環境や設備を快適に感じる。
- 参加者の20%は、参加を通じて学んだことの中から、少なくとも2つのアイデアを実際に試してみるつもりであることを報告する。
- 2人の農業者が新しい農薬散布技術を実施する。
- 7人の農業者が演習用トウモロコシ畑において、適正な量の農薬を散布することができる。
- Xコミュニティにおいて、1990年までに、女性に対する犯罪は2%減少する。
- 生徒の80%が当該科目の知識を増やす。

5. 影響指標の特性

特定のプログラムが、ターゲットとするアウトカムをどの程度達成することが出来たかを示すために使用すべきプログラムの指標を当該プログラムに関与するスタッフが特定しようとした際、ある州のプログラム実施責任者は、指標の特性を次のように整理している。

- ① それらの指標は有効か。それらの指標は、当該アウトカムに的確に焦点を当てたものであり、当該プログラムの状況を適切に表したものであるか。それらの指標は観察可能で測定可能か。
- ② それらの指標は普遍的か。当該プログラムの大まかな全体像や、それがターゲットとするアウトカムの全体像を描き出せるような形で、相互にリンクされているか。それらは、TOPの枠組みの中で、十分なレベルをカバーしているか。
- ③ 各指標は、どのような特性、あるいは、変化が勘案されたかを適切に示しているか。それらの指標は、予想された変化の量を示しているか。また、それらの指標は、プラスとマイナスの両方のアウトカムを反映するものとなっているか。
- ④ それらの指標は、サンプルデータからより大きな母集団への一般化を可能としているか。
プログラム全体を正確に反映することができるようなサンプル集団からデータを得ることが可能か。
- ⑤ それらの指標は、あるプログラムの様々な活動の全体を通じて蓄積できる程度に十分幅広いものであるか。また、それらのデータは、場所、活動、及びアウトカムの多様性に対応できるものか。
- ⑥ それらの指標は、手ごろなものであるか。それらの指標に関するデータや許容可能な証拠を収集するためのリソースは利用可能であるか。

これらの質問は、我々が適切なプログラムの指標に着目する上で役に立つ。参加者に起きる変化を観察することによって、プログラムのスタッフは直感的に、彼らが実施しているプログラムの有効性に関する質問を行っていることに気づく。これらの指標に焦点を当てることによって、プログラムのスタッフは、彼らが観察したことを書類にし、報告を行うことができる。そのことによって、我々は、優先度の高い社会的、経済的、環境的問題に関して、今、どのようなアウトカムが起きているかを知ることができるのである。

<これらの指標に関する追加的参考資料>

指標に関する一般的な議論については、「ユナイテッド・ウェイ・オブ・アメリカ」が発行した「プログラム・アウトカムを測定する。一つの実践的アプローチ」を参照すること。

なお、この書籍からの有益な抜粋や発注の仕方等に関する情報は、同団体のウェブサイト：

<http://www.unitedway.org/outcomes/> を参照して頂きたい。

6. 指標プロジェクト

指標プロジェクト (Indicator Project) はアウトカムの測定と常に結び付けられる訳ではない。

6-1 指標プロジェクトの使用目的

1) 公的教育 (Public Education)

市民に一連の価値や考えを紹介し、彼らの行動を促すようにするために使用。

一般市民が状況を理解し、正しく反応することができるよう、彼らに情報を効果的に伝えるために使用している。

例えば、キッズ・カウント (子供の数を数える) プロジェクトは、サウスカロライナ州の子供が、他の州の子供と比較して、どのような状況にあるかについて、市民の理解を高めるのに役立った。ときには、人々は、自分達が慣れ親しんだものが「普通」であるのか、そうでないのかを理解する必要がある。

2) 政策のバックグラウンド情報 (Policy Background Information)

公的指導者が何に対して、また、誰に対して、注意を払うべきか、ということに影響を与えるために使用。

公的資金がどのように使用されるべきか、ということに影響を与えるため。

市民との間でどのような公的対話を持つべきか、ということに影響を与えるため。

例えば、「ジャックソンヴィルとパサデナの生活改善プロジェクト」では、政治家が何に注目したか、どのようなプログラムが長期的に資金を供給されたか、また、何が、市民たちをして、自分達が行政の方向性や公的資金の使い道に影響を与えていると、強く感じることをできるようにしたか、を明確に浮き彫りにしたことが報告されている。

3) 業績評価 (Performance Evaluation)

各コミュニティは、目標とリンクしたビジョンを策定し、コミュニティの特定の側面における変化の様相を測定するためにさまざまな指標が使用される。

(評価の)結果に基づいた、あるいは、アカウンタビリティに基づいたいくつかのプログラムが、現在、この方向で動いている。(例えば、ジョージア州、オレゴン州、ミネソタ州など)

「オレゴンベンチマーク・プロジェクト」と「ミネソタマイルストーン・プロジェクト」は、この分野における2つのリーダー的プロジェクトである。多くの州やコミュニティが、この両州の先導に従っている。

6-2 指標ステートメントの事例

1) ハワイの試金石

ビジョン・ステートメント

「ハワイのすべての家族は、自分たちの生涯の望みを叶え、家族のメンバー全員を十分に養い、自分たちのコミュニティに対して、思いやりのある安定した、健康的で、繁栄した環境の中で、その向上・改善に貢献する機会を与えられるべきである。」

目的Ⅲ：ハワイのすべての家族が良い健康状態を享受すること。

目的Ⅲのためのいくつかの指標

- 2015 年までに、ハワイでは、1 歳未満の乳児死亡率を 1,000 出生当たり 6.5 から 5.5 に低減することを期待している。
- 2000 年までに、ハワイの子供たちは全員、月齢 24 ヶ月に達した時点で、あらゆる最新の処方免疫処置を受けている。
- 住民 100,000 人当たりの糖尿病による死亡率を 10.4 人（1990 年）から 10.2 人（2000 年）に低減する。

2) 最も良い指標

特定のビジョン、価値、目標、及び範囲において、あなた方にとって基本的で中心的なものを反映している指標であること。

- あなた方が知る必要のあるシステムに適切に関連した指標であること。
- 一連のリンクを形成し、組織だった関係性を明らかにするような指標であること。
- 有効な指標でであること。それらの指標を含めることについて、妥当な合理性があること。
- 当該コミュニティ全体にとって、十分に理解可能で、信用するに足る指標であること。
- 首尾一貫しており、信頼できる指標であること。
- 他の各州の諸指標と比較可能なものであること。
- 先導的役割を果たせる指標であること—すなわち、問題の進展や進捗状況について、事前に警告を出せるような指標であること。
- 既存のデータが利用可能であるか、あるいは、直ぐに測定することが可能な指標であること。
- 長期にわたる変化が計測できるような指標であること。
- メディアの注目を集めることができる指標であること。
- 政策と適切な関連を有した指標であること。
- 実施された仕事の結果（アウトプット）に焦点を当てた指標であること。単に支出されたリソース（インプット）を示すだけの指標ではないこと。
- さまざまなニーズや問題のみならず、リソースや、資産、更には、進捗状況にも焦点を当てたものであること。

- 具体的活動に役立つものであること。
- 説得力があり、興味深く、かつ、エキサイティングなものであること。
- アクセスしやすく、かつ、費用が手頃であること。

資料：「コミュニティ指標ハンドブック」（アトキソン，ノリス，その他 9 1997 年発行）

3) 客観的指標と主観的指標

客観的指標は、観察可能な一定の事実に基づいたものである。

客観的指標は、しばしば、正式な記録や報告書の形を取っている。また、客観的指標は、特定の行動をカウントした観察を通じて得ることも可能である。

アンケートは、年齢、子供の人数、修了した学業といった一定の事実について質問するものである場合、客観的であると言える。

利点：客観的指標は、我々に、評価者や観察者の立場を離れた独立した立場で、世界がどのようなものであるかを、理解させてくれる。

主観的指標は、人々を取り巻く世界に対する彼らの認識に基づいている。

主観的指標は、通常、アンケートや面談という形をとる。

利点：主観的指標は、人々が自分たち自身について、どう感じているか、また、他者について、どう感じており、更には、人生における自分たちの経験について、どのように感じているかを、我々が判断する上で役に立つ。

主観的指標の事例としては、親としての役割について、彼らの日記に表現された気持ち、さまざまな危機に直面した時に、彼らが、如何に多様な局面に立たされているかということに関する彼らの認識、あるいは、彼らの子供が学校でなぜ良い成績を取れないのかに関する彼らの見方、が含まれる。

翻訳を終えて

米国の普及事業は、土地付与大学（land-grant university）と呼ばれる州立大学によって管理運営されており、国と県の行政部局が所管するわが国の協同農業普及事業とは、実施体制が異なっている。米国の普及事業は、連邦政府と州政府資金に加えて、カウンティ（郡）・農業団体等の地元が経費を負担していることもあって、これらの資金提供者に対して、普及活動の成果を説明することが求められており、普及活動の成果について厳しい評価環境に置かれているといえよう。このため、土地付与大学の普及部では、各普及プログラムに加えてカウンティごとに設置された普及所の活動成果を評価するための手法を開発し、また、普及部自身も積極的に評価を実施している。

この「普及プログラムの評価方法」（How to Conduct Evaluation of Extension Programs）は、普及プログラムの活動成果を評価するための基本的な事項を取りまとめた学習書であり、手引きである。わが国では、評価といえば、経済的効果や結果の格付けであると思われがちであるが、本書は、あくまでも評価結果を活用して、プログラムの運営改善や次期プログラムの計画立案に当たっての有用な資料を得るために、また、普及活動のアカウンタビリティを高めるために、評価をどのようにして実施すべきかを念頭に作成されたものである。

後段に収録した「普及プログラムの評価実施のための資料」は、ペンシルベニア大学の Dr. Rama Radhakrishna 准教授から頂いた資料を整理し、翻訳したもので、収録の順位や付された番号には特に意味はない。

調査部長 谷口敏彦