

評価の実施の改善

最良の評価を実施するためのガイドラインと バックグラウンドペーパー

目次

	ページ
まえがき	2
最良の評価を実施するためのガイドライン	3
評価の実施の改善:バックグラウンドペーパー	9
1. はじめに	9
2. 評価の定義	11
3. 評価の使用を改善するための評価の管理	29
4. 結論	50
参考文献	55
付録: 評価に関連するグループ	

まえがき

結果に焦点を当てることは、OECD各国において、近年の公共部門改革の中心的な要素となっている。評価は、結果指向の環境の中で重要視されている。というのも、評価は、公共政策の効率性、有効性、および業績といったことにフィードバックの必要性をもたらし、政策の改善、および革新にとって不可欠だからである。要するに、アカウンタブルな(説明のできる)行政に貢献するのである。

業績管理の仕事の一部として、OECD Public Management Service(PUMA)では、評価の使用法を改善するための重要課題、および行うべきことを明確にするために、メンバー国における評価についての研究を行った。この報告書には、「最良の評価を実施するためのガイドライン」およびバックグラウンドペーパー「評価の実施の改善」が収められている。

このガイドラインのねらいは、評価の使用を改善しようとしている場合にOECDメンバー国が考えるべき重要課題、および行うべきことを明らかにすることである。このガイドラインは、評価方法の問題点というよりは、政府の評価活動の管理と個々の評価の管理とに焦点を合わせている。強調されなければならないことは、評価をまとめ、そして行う方法はひとつではないということだ。方法の選択は、いくつかの要素に依存するであろう。その要素の中には、評価の目的、より広い業績管理の枠組みの中での評価の役割、ならびに制度上および政治上の考察も含まれる。評価を、いつ政策決定や業績管理のプロセスに取り入れるのが最も適切なのかを決めることが、このガイドラインの役割ではない。それは、メンバー国に委ねられるべきである。

このガイドラインは、PUMAの業績管理ネットワークおよび行政管理委員会(Public Management Committee)の双方によって検討され、是認されている。

この報告書およびガイドラインは、PUMAのSigurdur Helgasonによって書かれた。バックグラウンドペーパーは、Marjukka Ala-Harja(フィンランド大蔵省)およびSigurdur Helgasonによって準備された。事務局では、オーストラリア、カナダ、スウェーデン、アメリカ合衆国、およびECからの上級公務員、専門家からなるレファレンスグループからサポートを受けた。このレポートは、OECDの事務総長の責任のもとで出版されているものである。

このレポートに表されている見解は書き手のもので、OECDメンバー国の政府から委ねられたものでも、また必ずしもメンバー国の政府の意見を反映したものでもない。

このレポートおよびPUMAの業績管理と評価に関するその他のインフォメーションは、インターネット <http://www.oecd.org/puma> で入手可能である。

最良の評価を実施するためのガイドライン

ガイドライン

このガイドラインは、3つの セクションで構成されている。

評価から最大限を得る

このセクションでは、評価、評価の目的、主な担当者、および便益とコストについて明らかにする。

1. 定義、および目標
2. 主な参加者の特定
3. 便益およびコストのアセス(査定)

評価の枠組みの体系化

このセクションでは、公共部門において組織を改善すること、および評価の使用を改善することとの関連で実施すべきことを議論する。

4. 評価カルチャーの育成
5. 評価活動の戦略的な管理
6. 信頼の向上

効果的な評価の構築

このセクションでは、個々の評価を管理するために実施すべきこと、および優先度について議論する。

7. 意思決定過程とのリンクの確立
8. 正しい評価者の選択
9. stakeholder の関与、および評価結果のオープンな伝達

評価から最大限を得る

定義, および目標

＞**評価**は、公共的な政策、組織、あるいはプログラムの成果に向けた分析的なアセスメントであり、調査結果の信頼性や有用性を強調するものである。評価の役割は、情報を改善することであり、不確実なことを減らすことである。しかしながら、厳格な手法に基づいている評価でさえも、判断にかかるところが多い。事前評価（言い換えれば、政策のレビュー）と事後評価とは区別することができる。このガイドラインで議論される多くの実施例は、たとえ事前評価と事後評価とでは目標が異なっているとしても、両者に等しく適用することができる。

＞評価の主たる**目標**は、意思決定、資源の配分、およびアカウンタビリティの改善である。この目標は、国民に情報を与え、主要な意思決定のプロセスに情報を提供し、進行中の組織的な学習を元気づけることを通して達成されうる。

＞評価は、**より幅広い業績管理の枠組みの一部分**でなくてはならない。評価は、業績管理を補うもの、改善するものではあるが、業績管理と置き換わるものではない。

主な参加者の特定

＞成功する評価は、Commissoner（評価委員会の責任者または委員）のリーダーシップのもとで、**主要参加者**（評価者、使用者、stakeholder）**間の協力**に基づいて行われている。

＞Commissoner とは、評価を委託する機関である。Commissonerは、評価を計画し、その進展を絶えず監視し、評価報告書を受け取り、さらなる行動の決定をくだす。Commissonerは、省や中央政府の庁（例えば、大蔵省、あるいは独立した評価や会計検査の機関）であったりする場合もあるであろう。Commissonerが評価者となる場合もまた、あり得る。

＞**評価者**とは、データを収集、および分析し、さらに評価された題材の価値を判断する機関や個人のことをいう。

＞**評価の利用者**とは、政策決定責任者であったり、予算局であったり、会計検査院であったり、政策やプログラムの管理者やそのスタッフであったり、またサービスの利用者などであったりする。

＞stakeholder とは、評価される政策やプログラム、および評価結果に関心のある個人や機関のことである。stakeholder と評価の利用者は、しばしば同一であることもある。

便益およびコストのアセス(査定)

＞評価の便益は、そのコストや制約よりも重要であるべきである。コストも便益も、評価を慎重に管理することや、適切な評価者と評価手法を選択することにかかっている。

＞評価の主な価値は、**業績の綿密な研究**ができ、他の業績管理手段の有効性を独立してアセスできるというところにある。潜在的な便益は、大きな政策やプログラムでは最大級となる。

＞一方では、評価は、その実際の利用や有効性の割には、**コストがかかりすぎたり**、あるいは**時間がかかりすぎたり**することがしばしばあるということを、経験が示している。また評価には、意思決定を遅らせたり、怠惰さを正当化したりするために利用される危険性もある。

評価の枠組みの体系化

評価カルチャーの育成

＞評価に対する適切な**サポート**は、政治家、政策管理者、および中央管理庁（例えば、大蔵省）が、評価によって生ずる政策についてのアドバイスを有効に活用しようとする意欲を示すことである。

＞評価への**需要**は、内外の stakeholder によって発生し、具体化され、そして表現されることが必要である。stakeholderによる「オーナーシップ(当事者意識)」のない評価には、効果はなさそうである。内部の抵抗といったような評価への制度的な障害は、相互信頼を作ることを目指す協議会を通して軽減することができる。

＞政府は、変化している環境を改革したり適応したりすることを奨励する**評価カルチャー**をサポートすることができる。基本的に大切なことは、適切であり続けるためには、組織は成果についてのフィードバックから学び続ける必要がある、ということである。

＞**訓練と専門的な意見交換**、有能な評価者、情報通の Commissionerと、啓発された熱心な評価利用者、このすべてが、評価カルチャーに貢献するのである。

評価活動の戦略的な管理

＞**評価機関**は、異なった政策領域における需要と優先度を調和させるべきである。データの収集費用が高く、情報が限られているような、鍵となる政策領域においては、体系化され、制度化された評価が適切であろう。しかしながら、より柔軟なアプローチは、しばしばよりよい結果を生み出し、評価が机上の実習となってしまうことを防ぐこともある。多くの機関にまたがっている評価活動には、特別な注意が払われねばならない。

＞中央行政庁は、評価プロセスの管理に重要な役割を演じている。しかしながら、**実際の評価は**、政府のあらゆるレベルの各担当者に**分散化されて行われていることもある**。

＞それぞれの機関における評価スキルの開発は、**評価の手法や全体像**（例えば、内部からも外部からも評価者を集める）というような評価に必要な**領域**を確立し、各々の評価が、目標、焦点、信用性および当初想定した用途に関する課題を独自に組み合わせた設計となることを確立するのである。

＞評価に資金を提供する**特別基金**は、公共政策を評価する糸口として重要な役割を果たすことができる。しかしながら、他の業績管理のアプローチがより適切であるときには、特別基金は評価を行って改善することを元気づける役割も果たすであろう。

信頼性の向上

＞信頼性の欠如は、評価の調査結果の活用を妨げることになる。**信頼性に影響を与える要素は**、評価者の適切さと信頼性、評価者と評価される者との相互信用、専門的な打ち合わせ会議と stakeholder の参加、調査結果の伝達のプロセスである。

＞専門的および倫理的な基準、そして評価の**手法の質**（適切な基準、十分なデータと証拠、信頼性がある明確な調査結果といった周辺のことから）も、また評価の信頼性に影響する。評価手法のレベルの高さや、開放的で率直な意見交換は、評価に潜む弱さを表面化させ、修正する働きをもつ。その結果、信頼性を改善することもできる。

効果的な評価の確立

意思決定過程のとのリンクの確立

＞評価情報は、**政策の形成プロセス**においては政策介入の質を改善するための、また予算プロセスにおいては優先事項と資金の節約に対応するための、重要な要素になりうる。適切な評価は、政治上、予算上、管理上、およびその他の戦略計画上、重要な課題を指摘する。

＞評価の**目標**によって、評価の位置づけ、手法、使用法が決定する。提案される評価の使用法は、明確に定義されるべきである。評価は政策介入の特徴にあわせて調整されるべきであり、評価の手法はコストや時間といったような制約を考慮したうえで評価の目標と適合しなければならぬ。最初から評価への要求を政策に組み入れ、目標をはっきり定義することは、評価の有用性を増し、計画立案を容易にする。

＞**計画立案**は、評価の管理と質を改善する。Commissioner は、目標や基準の設定、データの収集とその方法を明らかにすることを含む評価の計画に責任がある。タイミングは重要である。しかし、意思決定のサイクルはしばしば予想できず、評価が最終段階に入る前に決定が行われてしまうことも、よくある。

正しい評価者の選択

＞組織自体による**自己評価**は、主たる目標が組織的な学習をしさらに業務の実施を改善することであるときには、適切である。しかしながら、時間やスタッフのスキルが不十分なこともあるだろうし、カバーしている問題の範囲が限られていることもあるだろう。また調査結果の信頼性も疑問視されるかもしれない。

＞**中央管理庁**による評価は、目標が予算の優先度の改善であるとき、また、評価者が意思決定のプロセスに密接な関連があるときに、ふさわしい。

＞**外部評価者**（例えば、調査団体、経営コンサルタント）による評価は、目標が、公共的な政策に新しいものの見方を与えることであるとき、あるいは、特別な評価スキルが必要なときに、ふさわしい。しかしながら、外部評価者は、評価される政策や機関の本質やカルチャーに対する理解が限られており、また、空論的な評価を提供する恐れもある。

＞**独立した評価**は、目標がアカウンタビリティや透明性の改善であるときに、ふさわしい。しかしながら、政策管理者、言い換えれば一般的に行政というものは、調査結果や提言を受け入れることに気が進まないものである。業積の会計検査は、しばしば評価と類似していることがある。会計検査の主要な特徴は、会計検査者の独立性と、改善することよりもアカウンタビリティに焦点を置いていることである。

Stakeholder の関与、および評価結果のオープンな伝達

＞スタッフを含む stakeholder は、評価委員会に指名されたり、あるいは運営グループやアドバイザリーグループを通じたりして評価に関与することができる。個人参加方式の評価手法は、コンセンサス(世論)や変更プロセスのための当事者意識を形成するときに有効である。評価の使用者とスタッフとの意見交換は、彼らのニーズと優先性についての理解を助け、感応性を改善する。stakeholder の参加は、コスト、時間の制約、そして、評価のプロセスを分捕ってしまう危険性を考えたうえで、行わなければならない。

＞評価の**調査結果**をオープンに**発表**することは、信用性を増し、調査結果に基づいて行動しなければならないというプレッシャーを生み出す。報告書と会合を誰にでも利用可能にすることは、調査結果に基づく意見交換を活発化するのに有益である。明確な基準に基づく判断や提言は、注意を喚起し、その後の行動に励みを与える。非難することよりも問題を打開することに判断の焦点をあてるべきなのである。

評価の実施の改善

1. はじめに

調査の結果に焦点をあてることは、OECD諸国における近年の公共部門改革の中心的要素である。評価は、機関やプログラムの調査結果に基づいてフィードバックをするための道具である。ここで議論される多くの問題は、他のタイプの評価（機関の業績の評価、マネージメントレビュー、業績会計検査）にとっても同じように適切であるかもしれないが、このペーパーでは主として「プログラム評価」について述べる。

プログラム評価の定義

プログラム評価は、次のように定義される。

プログラムの重要な局面やその価値を提示し、調査結果の信頼性と有用性を求める、体系的分析的なアセスメント。

プログラム評価は、意思決定することに対しよりよい情報を提供し、公共政策やプログラムのアカウンタビリティを高めることで、意思決定の内容を改善することを目的とする。意思決定の際に判断を決心させたり、判断を変更させたりすることを目的とするものではないが、プログラムの効果について、その設計や実施を改善するために使われるフィードバックの質を確かなものにする。

公共部門の改革の優先度を概説するとき、OECDの行政管理委員会（Public Management Committee）は、よい実績のある公共部門は継続的に政策の有効性を評価すべきであると指摘し続けてきた。管理のツールとしての評価への関心が高まってきていることのあらわれである。以前の評価の努力と比べ、今やもっと現実的な期待があり、以前ほど厳密でなくとも良い手法も広く受け止められ、組織内での評価の利用への理解が高まってきている。また、より体系的で、予算編成のプロセスと関連したアウトカム中心の評価に重点が置かれている。

評価結果を活用することはたやすいことではない。評価の歴史は、果たされない約束の歴史と特徴づけられる。それゆえ、評価の包括的な有益性について疑問が湧きあがっても不思議ではない。評価を、官僚制を作り上げるがほとんど結果を生み出さない新しい一時的な流行の管理方法、とみている者もいる。評価は、しばしば、本質的にあまりにも理論的で実際には機能しないとか、主要な政策の選択肢というよりもあまり重要でない問題に効果を持っている、というようにみられている。コントロールが強まることを恐れたり、アカウンタブルであることを望まない者もいる。プログラムの継続に主として関心を持つ者は、評価に脅威を感じるかもしれない。要は、業

積が継続的にきちんと測定され、それについて契約さえ取り決められている状況下では、評価の価値について疑問が湧き起こりうるのである。

これらの主張は論議されることができる。評価は、的確に使われ、包括的な業績管理の枠組みと統一されるときに、重要な役割を果たすことができる。評価は公共部門の効率性と有効性を改善することができるし、それゆえ公共部門の活動の基礎を強めることができる。質の高いプログラム評価は、業績測定、監視、会計検査を補足し、発展させることができる。現実的な期待を設定し、評価の問題点と直面し、さらに、実際のだが目的のある方法で評価を用いることにより、政府は、業務遂行やアカウンタビリティを高めることができる。この文書に示されている例は、この局面についていくつかのガイダンスを与えてくれるだろう。

この冊子では、行政管理におけるプログラム評価の役割を議論し、公共部門での利用を改善する方法を精査している。パート1では、評価を定義することを中心にした。パート2では、評価の利用法を改善しようとするときに検討すべき問題点を述べている。

2 . 評価の定義

2.1 概念

評価は、自分が納得するというより他を説得するものであり、説明するというより議論するためのものであり、確かなものにするというより信頼できるものにするものであり、強制されるものでなく自由に受けとめるべきものである。(House, 1980, p.73)

評価とは何かについて、一般的な合意がある訳ではない。概念は、いくつにも定義されており、ときには反対のことを定義している。このことは、専門分野(経済学、政策や管理の研究、統計学、社会学、心理学、等)、評価機関や実際の評価の現場の者が多彩なことや、評価の論点、ニーズ、そして依頼者の幅が広いことから伺われる。

評価はだんだん認められてきているが、実質的には、フィードバックや、inquiry(審問)も評価と呼ばれることがある。多くの言葉(review, follow-up, monitoring, audit, scrutiny, assessment)が評価活動に言及するために使われている。何人かの専門家は、評価という概念がこれらの言葉の持つ本来の意味をまったく失わせてしまうという懸念を表明している。なお、評価に加え、意思決定を改善するために使うことのできるフィードバックのしくみは他にもある。

共通の特徴が数多くあるとはいえ、すべてのフィードバックのしくみが評価を構成する訳ではない。評価と他のフィードバックのしくみの違いは、次の表の通りである。

評価と他のフィードバックのしくみとの相違点

・ 科学的研究	→ 情報の実際の使用に焦点を合わせた評価
・ 従来の会計検査	→ より広い見地から公共費の使用を研究する評価一併せてプログラムの目標が、適切に、効率的に、および効果的に達成されているかを問う（会計検査と評価の相違点はいくぶん曖昧なことが多いが、それについては後述する）
・ 監視（Monitoring）	→ 評価がしばしばそのための努力となり、問われているプログラムのより詳細な情報を得ようとするもののみになりがちなのに対し、よく機能する通常の監視制度は、しばしば成功する評価の不可欠な基礎となる。
・ 業績測定 （Performanse measurement）	→ より多くのものを求めて努力する評価。見出されたアウトカムを説明しようと試み、政治的介入の論理を理解しようと試みる。（しかし、良く設計された業績測定システムは、例えばアメリカの場合のように、しばしば評価の形のひとつとして受け入れられている。）
・ 政策分析（Policy analysis）	→ 事後のアセスメントに焦点を当てた評価。政策分析は、将来の政策の選択肢を研究している事前の評価と定義されることもよくある。

次の表では、相違点を強調しながら簡単に述べる。規模が違っていると評価へのアプローチも異なる。アプローチと適用する評価技術とにつながりがあることも少なくない。とられるプロセスあるいは当初意図した評価の目的に関連して違いが生じることも少なくない。

評価において強調される点

・タイミング	→ 事前、同時・中間、事後、あるいはこれらの組み合わせ
・フレームワークの柔軟性	→ 計画の目標に対するあるいは目標の適切性を含むアセスメント。意図されたあるいは意図されていなかった効果のアセスメント
・判断およびプレゼンテーション	→ プログラムが成功したか失敗したかを判断する。改善や変更を提案したり、調査結果を発表したり、また意思決定責任者に結論をまかせたりする。
・位置づけと自立性	→ 内部アセスメント(例、自己評価)や外部アセスメント、あるいはその組み合わせ(半独立の政府系評価ユニット)。
・専門知識あるいは協議	→ 結果の客観性を確保するために、科学的、技術的な手法を使った評価の専門家によるアセスメント。または、ねらいや手段(あるいは調査結果と提言さえも)が調査結果を最大限に利用し、大衆の参加を確保するために関係者と協議するパートナーシップ。
・使用	→ プログラムにより多くの洞察力を得るもの、またはプログラムの価値を決めるもの。プログラムの微調整を助け、あるいはその合理性を問い、より多くの抜本的変革を提案する。

違った局面を強調することにより、評価はいくつかのタイプに分類することができる。実際、評価のこういったタイプは、しばしば組み合わせられ、総括的、形式的な評価と共に取り上げられる。

評価のタイプ

・総括的評価	→ プログラムがある期間実施されてから、その有効性を研究したり、全体的な価値を判断したりするために行われる(事後評価)。これらの評価は、資源の分配を助けるためや、公共へのアカウンタビリティを高めるために典型的に使われる。依頼者は、政治家や他の意思決定責任者など、通常外部者である。評価結果の客観性や全体の信頼性は重要視され、それゆえ通常外部評価者が評価を行う。結果の問題点と全体のプログラムの適切性が指摘されることが求められている。
・形成的評価	→ 通常は、プログラムの実施中に行われ、より一層の洞察力を得、学習プロセスに貢献するために行われる(中間評価)。目的は、プログラムの管理、施行、および開発をサポートし、改善することにある。評価者も、依頼人と同様 概して内部者で、通常プログラムマネージャーである(自己アセスメント)。評価結果の客観性はしばしば主たる関心事項ではなく、結果を直接適用することの方がより強調される。運営上の疑問点、出来事の監視、そしてある程度の影響について指摘することが求められている。

プログラム評価

評価は、機関、機能、過程、政策、プロジェクトなどを対象とすることができる。本冊子では、公共プログラムのアセスメント、すなわちプログラム評価のアセスメントだけを取り上げる。「プログラム」は、通常、政策を行動に移すための、すなわち、政治上の望ましいゴールを達成する状況を作るための、一定の時間的スケジュールと予算を持つ、一連の組織的活動である。この言葉は、比較的狭い活動とか、切り替えを行った主要な政策プログラムを指している。しかしながら、「プログラム」の概念は、国によって違った理解がされており、全く使われないケースもあることを指摘しなければならない。

評価の場合でいうと、プログラム評価に一つの決まった定義というものはない。最もシンプルにいうならば、目標に対するプログラムの達成度のアセスメントである。プログラム評価は、プログラムのライフサイクルのどんな段階でも実行することができる、と言われることが増えてきた。事後のアセスメントの場合だけでなく、事前評価および同時評価でもそうである（同時評価の場合には、実際かなり行われている）。事前の評価、すなわち将来に関する分析は、またしばしば、政策分析、あるいは査定（appraisal）と呼ばれる。政策分析が、政策オプション、および起こり得る影響を探求する一方、評価は、実際の影響を研究し、政策の価値を判断する。しかしながら、概念は、互いに関連性があり、アセスメントプロセスは、相互に活用できる。

プログラム評価の定義の例

・プログラム評価の主要課題は、プログラムがある一定期間にどんな効果を生み出したかを判断しようと試みることであり、また、ある特定のプログラムが生み出した効果の大きさを見極めようと試みることである。（Hatry，1980）

・プログラム評価は、プログラムの設計、施行、有効性のアセスメントについて体系的な研究方法を適用することである。（Chelimsky，1985）。

・プログラム評価は、プログラムの全部または一部の体系的なアセスメントで、①マネージャー、および他の意思決定責任者がプログラムの妥当性、効率性、および有効性を評価することを助け、②新プログラムの設立や既存のプログラムの拡張をアセスし、③プログラムのための資金導入を、現レベルで継続すべきか、増大されるべきか、減少されるべきか、中止されるべきかどうかを決定する。（Doing Evaluations, A Practical Guide. Australian Ministry of Finance, 1994）。

・プログラム評価は、どのくらい良くそのプログラムが働いているか、また、プログラムが改善されたり廃止される必要があるかどうかを決定する権限のある機関によって行われる。（Mayne，

1994)。

・ 評価は、通常、分析的、体系的で、信頼性や問題意識があり、使用者主導という要素があるべきである。（Evaluating EU Expenditure Programmes. A Guide。European Commission, 1997）。

様々な定義に関わらず、いくつかの特徴は、すべてのプログラム評価に共通である。

プログラム評価は、体系的、分析的なアセスメントで、プログラムの重要な面とその価値を提示し、評価結果の信頼性と有用性を求めるものである。

評価の概念についての意見の一致がないため、混乱が起きることがある。しかし、意見の一致がないことは、社会科学の中での様々な暗黙の理論的枠組みを反映した、みのりの多い実用的な出発点と見ることもできる。しかし、この冊子の観点からいえば、広義の評価を適用すべきか、狭義の評価を適用すべきかどうか、ということは、特には重要ではない。この冊子の目的は、異なる定義であってもあてはまるよい実施の方法を明らかにすることである。

2.2 目標

評価の目標は、絶対真理の探求ではなく、洞察力と正当化された見解を政策プログラムに与えることである。評価は、政治的な意思決定にとって代わるものではなく、より多くの情報に基づき決定がなされるようにするためのものである。

管理は、しばしば計画/準備、予算編成、施行、および評価というサイクルで描写される。評価は、このサイクルを閉じ、フィードバックの環を完成するリンクである。評価された政策やプログラムは、改善され、拡張され、あるいは他のものにとって代わられたりする。

評価の主なねらいは、意思決定や資源配分、アカウンタビリティを改善すること、として特徴づけられるであろう。

評価の目標

・意思決定の改善	→ 評価は、政策の影響についての情報を提供する。それは 形式的なコントロールの仕組みから情報を通しての運営への、業務的なコントロールから戦略的なコントロールへの、さらにインプットだけのマネージメントからアウトカムによるマネージメントへのワンステップと見ることができる。評価は、公共プログラムの価値を判断したり、公共プログラムを改善または変更したり、さらに将来のプログラムを設計したりする際等には、政策決定責任者を助けることもある。評価は、彼らが成し遂げた仕事の理解を深め、新しい政策の開発に役にたつことによって、プログラムの管理や施行を学習するプロセスにも貢献する。それに加え、他の関心事項(科学的なもの等)についても、社会の機能に対するより優れた洞察力を得ることで十分であるとされてしまうかもしれない。
・資源配分の手助け	→ ほとんどのOECDメンバー国は、予算編成の圧力に直面している。評価は、資源配分を改善し、現存の資金をよりよく使用することにより、政策担当者と予算担当職員を助けることができる。評価は、望ましい業績に到達するためのより効果的、効率的な方法を見つけ出すことに役立ち、新しい優先分野の開発促進に予算を移し変えることを応援し、プログラムへの資金供給の正しい水準を見つけ出すことに貢献し、あるいは予算削減が必要なときの優先順位付けさえもするのである。評価は、予算編成のプロセスのなかでの改革と密接に関係してきた。例えばスウェーデン、オーストラリアがそうである。
・アカウンタビリティの向上	→ 評価は、公共政策のアカウンタビリティに貢献する。そして、それは、OECDメンバー各国でますます重要に感じられている。この傾向は、行政管理の大規模改革による自治権拡大の結果であり、いくつかの国で起きた公共部門の妥当性についての市民の関心によって引き起こされたものである。政府活動の影響を明らかにさせることによって、評価は、議会、および社会全般への透明性、およびアカウンタビリティの改善を計ることができる。それゆえ、評価は、進行中のプログラムの妥当性を強め、業績の維持を確保することができる。評価はまた、市民が自分たちのお金で何が達成されたかを知り、実行された政策の代替案を検討する助けにもなるのである。

これに加え、評価を特にタイムリーにするいくつかの要素がある。

評価をタイムリーにする要素

・公共部門の改革，戦略的行政管理	→ 評価は、近年の公共部門改革において脚光を浴びたツールの一つである。業績管理や戦略的行政管理の開発というシナリオのなかでも、しばしば登場する。業績を査定すること、特に自己評価を行うことは、公共管理を行う場合の標準になるべきであるということが、議論されている。進行中の公共部門改革をアセスすることの必要性が、公共部門改革の可能性を高め、より発展させ、達成された結果をたもち、公共の支援を強めるために、OECD各国で高まってきている。
・制度的必要条件	→ 公共プログラムに対して資金融資あるいは共同融資する機関は、しばしば、融資の受取人に評価を行わせたり、あるいは自分達自身で評価を行ったりする(欧州連合の Structural Fund および世界銀行など)。これが、政府がますます評価活動を制度化し、省やエージェンシーに自分たちの業績をアセスするように要求する真相である。

各機関は、いろいろな方法で評価を見るし、いろいろな目的のために評価を使う。目的が異なると、評価へのアプローチも異なる。しかし、より普遍的レベルでは、評価のゴールは組織的な学習と定義されるであろう。

明白に規定された評価目的の例.

・評価は、資源配分、プログラム改善、およびプログラムのアカウンタビリティを補うために、信用できる、タイムリーな、また客観的な発見、結論、および提言を行うことができる。(オーストラリア)
・(評価は、) 十分な量の、適切で信用できる情報をもたらし、それが EU の資金提供プログラムが、継続されるか否か、変更されるかどうか、変更すべきかどうか、中止するかどうかを決定するために役立ち、また、それに付随する予算配分を、増やすか、縮小すべきかどうかを決定するためにも役立つであろう。(EC, DG XIX , Assessment checklist)
・評価は、なにが本当に役立っているか、なにが役立っていないかに焦点をあてることによって、また、政府のゴールをコスト的により効果的に達成する革新的な方法を見出すことによって、その価値を高めつつあると見られている(カナダ, In Search of Results: Performance Management Practices) .

2.3 開発

OECDメンバー国のうち、評価が、政策決定過程において正式な役割を担っている国もある一方、ちょうど体系的な評価活動の可能性を思案し始めている国もある。アメリカ合衆国、スウェーデン、およびカナダといった国は、しっかりと制度化された評価システムをもち、評価に長い歴史がある。オーストラリア、オランダも、また、洗練された評価システムを開発してきた。

評価の実施の成熟度は、政策部門によって様々である。評価は、多くの国において、開発援助、労働市場政策、衛生、および教育といった分野で長い伝統を持っている。

歴史的な局面

国家行政における評価活動は、幾つかの段階を経て発展している。成熟度は、経験の長さ、活動の規則正しさ、および制度化されている程度によって特徴づけられる。発展の段階は、3 つに識別される(Derlien , 1990)。(公共プログラムのアセスメントは、例えばスウェーデンでは第一段階以前に行われているが。)

評価の最初の波は 1960 年代、1970 年代で、社会問題の解決のため、公共プログラムに着手した social-liberal の政府に主として関連していた。立案と計画の過程は、政策決定の改善のために導入された。好ましい財政状況、社会科学知識の地位の向上および普及が、これに寄与した。焦点はプログラムの改善で、プログラム管理者は、フィードバックの仕組みとして評価を使うことに関心を持っていた。

第2の波は 1980 年代で、(あまり目立たず急速でもなかったが)、ほとんどのものごとを変えようとしない保守的な政府が財政的に制約を課せられて公共プログラムを抑制しようとした試みによって推進された。評価は、政策の正当性を再考し、予算内での資源配分を合理化しようとするときに役立つものと考えられた。大蔵省、および会計検査院は、評価を開発することに積極的であった。

それ以来、評価は、ますます、議会の関心事項となった。公共部門の合法性の問題、政府のアカウンタビリティの強化の必要性は、評価といった新しい、より洗練されたコントロールの仕組みを求める。進行中の国際的な研究の準備段階での成果(Gray, 1997)では、新しい公共管理の開発が、評価の機能や実施と密接な関わり合いを持っていることを示唆している。評価は、代用される (本来的には新しい国内市場が評価者と思われる)し、補足される(私的独占の規制が、巨大な評価の需要を作っている)し、そして新たな課題(例えば内部契約のような新しいツールの評価)を提供する。

確立された評価の伝統は、評価を評価することの必要性を検討する傾向にある。この評価の評価 (meta evaluation) 活動は、アメリカがおそらく最も進歩しているが、アメリカでは評価を行った量から、いくつかの一般的結論を引き出すことができる。しかし、評価の評価は、依然としてかなり稀であり、評価の利用法などあまり良く知られてはいない。

評価と会計監査

評価の役割、およびその現在の発展状況を明らかにするためには、従来の会計検査との関係を研究することが役に立つ。評価と会計検査は、歴史的には別々の機関による別々の機能である。評価者、および会計検査官のそれぞれの専門的風土は、例えば新規雇用や社会化に関する場合のように、いろいろな点で異なる。目的も、異なっている。評価が、経済性、効率性、および有効性を促進することを目的としている一方、会計検査は、伝統的に資源について財政的に適正であること、およびアカウンタビリティを確保するために使われてきた。

会計検査は、多くの国で、議会の監督機能に対して長い間貢献してきた伝統を持つ。それは、公共機関の法律遵守、およびアカウンタビリティに重点をおいている。昔から、会計検査は、確立された規則、規制、手続き、あるいは命令が守られてきたかについて証拠を求める。会計検査の制度が独立性を確保することによって、信用できる意見、提言を提出することができるようになる。

評価は、会計検査としてではなく、はっきりした概念と、会計検査と同程度の高い目標を持つ、規則的な、組織化された活動として確立されている。評価は、会計検査と比べ、通常個人参加方式で、独立性に欠け、またかなりの程度まで stakeholder の視点を受け入れる。評価は、公共のアカウンタビリティのツールとしてより、プログラム管理のためのツールとしてのほうが多く使われている。

従来の会計検査と評価の間には明らかな違いがあるにも関わらず、これら2つの機能を相互に近づける必要があるのかどうか、あるいはすでに統合されているのではないか、という疑問が生じてきている。境界は、不透明になりつつある。会計検査は業績についての疑問にその視野を延ばしているし、従来の財務監査は業績会計検査の補完を受けている。例えばスウェーデン、イギリス(支出に見合う価値の会計検査) そしてオランダで遂行された業績会計検査は、プログラム評価と類似しているアセスメントの例である。方法、および適用される基準も、極めて類似している。ヨーロッパ評価協会(European Evaluation Society) のような共通のネットワークもまた、専門家の受けとめ方が統合されていることのあらわれである。

しかし、矛盾している現象も存在する。評価と会計検査の統合は、評価の多様性、競争性、社会科学ベースという考え方と、会計検査機関による評価という考え方に対する政治上の抵抗によ

って抑えられた。アウトソーシング、および発生主義会計といった新制度の導入を増やすことにより、財務監査は、よりその時々における監査になるかもしれない。会計検査機関が、主たる評価ユニットや行政管理のコンサルタントとなることを目指すべきではなく、公共支出に関し、誠実性、合法性、および効率性を持った独立した監視者としての現在の役目と責任を明らかにすべきことが、示唆されてきている(Pollitt & Summa , 1997)

制度化された評価

評価の発展の例として、評価の実施のいくつかの例を以下に挙げてある。

スウェーデンにおいて

評価は、1950年代、あるいはそれより以前に、政策決定をしようとしていた公共委員会によって始めて実施された。委員会のシステムは、stakeholder の視点を結び合わせて、調査結果を政策決定に持ち込んだ。1960年代から、評価は、政策決定者に情報を絶え間なく提供することができる活動として見られ始めた。専門の調査団体やエージェンシーが、評価を主な職務として設立された。

1980年代の初めには、評価活動を体系化し、それらをより確固に予算編成に連携させることによって、政策決定を改善する必要性が起きてきた。新しい予算編成と結びついた評価システムは、そもその始めから、いくつかのかなり実地的な調整を経てきている。現在では、エージェンシーの年次報告書が、エージェンシーから省に提出されるフォローアップ情報の主な供給源となっている。エージェンシーの活動の綿密な評価(3 年毎に 実施予定だった)は、各省によって特別な要請があったときにのみ行われている。

オーストラリアにおいて

評価が、オーストラリアの公共管理改革プロセスの欠くことのできない一部になったのは 1980 年代半ばにおいてであり、政府の評価戦略が打ち出されたときであった。評価は、法人やプログラムの管理、および立案へ体系的に統合された。全ての公共プログラム、あるいはプログラムの重要な部分は、3 ～5 年ごとに レビューされている。全ての主要な新政策の提案は、評価戦略を含み、各省は、現在は毎年の評価計画の提出を要求されている。そして、主要な評価の結果は、公表されることが期待されている。

近年、義務的な要求の持つ継続した有用性が再評価されてきた。要求ベースのアプローチが、

適切な実施原則に基づき、業績のモニタリングや評価を含む、統合された業績管理の枠組みによって取って代わられることはありそうなことだ。評価がより当り前になり、より広く受け入れられ、そして業績管理活動を各省の戦略やビジネスプランに直接、統合しようとする変化が認められる。また、評価の質、利用、および有効性により多くの注意が払われていることも認められている。

評価の計画、実行、使用の責任は、個々のエージェンシー、および部門の役目である。財政局は、評価戦略のコーディネーターであり、トレーニング、推進、ガイダンス、品質保証をつかさどり、時には、選ばれた主要な評価に参加する。会計検査院長が評価活動の作業をレビューすることもある。

オランダにおいて

会計検査院は、プログラム評価へより多くの注意を払うよう 1990 年代早期に各省に提言した。政府は、3段階の戦略をうちたてた。最初に、評価についての各省の責任を述べた規制の改正、第 2 に、年次報告書を省の予算と共に準備し、現在と将来の評価活動、結果、およびそれらの使用法に関する情報を提供、第 3 段階として、省レベルでの評価の管理とコーディネート、である。政府はまた、外部評価のみでなく、内部評価も、また事後評価のみでなく事前評価や中間評価も実施するべきであると指摘した。

後に、政府は、評価方法、および責務の概略を述べたのは、政策の責任者は政策の責任者であること自体で評価に責任がある。彼らは、管理を行うにあたり、中央集権的なアプローチでも地方分権的アプローチでもいずれでも選択することができる。内部の予算責任者は、評価の使用のコーディネート、推進、確認に責任があり、必要なアドバイス、ガイダンス、および調査専門技術のノウハウが提供されていることを確認する責務もある。彼らは、個々のプロジェクトの評価プログラムを作成し、政策の定期的な評価を促進し、評価の質、およびその使用を監視する。省の上層部は、究極的に評価の責任がある。他の責任者は、彼らの専門技術のノウハウ(人員や体制、会計検査、法律の制定)をもって支援する。大蔵省は、内部の予算責任者と同様、促進、およびコーディネートに対する責任を持っている。会計検査院は、評価の管理および体制の効率性をレビューし、レポートを公表する。

カナダにおいて

評価は、1977年以降、各部門(department)の管理の一部となっている。レビュー活動は、しっかりと制度化されている。各部門は、自分たちのプログラムが 確実に評価されることに責任がある。現在のレビュー政策では、重要性のある、あるいはリスクのある分野について確立された基準に従って、各部門に内部の会計監査を行うことが求められている。つまり、確立された基準により主要政策やプログラムの評価を行うこと; 依頼人優先のサービス基準も含め、部門および管理者がプログラムの業績の主要な面について信用できる情報があり、それを使うことを確認できる業績監視の実施体制を確立すること; 他のタイプのレビュー、例えば政策、規制、プログラム、運営についてのレビュー、クライアント調査、および特別研究などのレビューをすること; そして、内部監査、評価、業績監視や他のレビューを統合し、補足するというやりかたで使うこと を要求する。

評価は、しばしば 専門化された評価組織ユニットによって実行されている。財政庁事務局は、評価の政策、コーディネート、および監視に責任がある。会計検査院は、公共部門の業績レビューと、責任をもって成し遂げられた評価の適切さについて報告することに主要な役割を持っている。

フィンランドにおいて

スウェーデンにおけるように、政治的決定を準備する委員会を指定するという伝統もある。現在は、評価の使用は、いくつかの部門(例えば 開発援助および高等教育の部門) で十分に、確立されているが、一般的には、評価はアドホックベースで行われている。しかし、近年の調査によれば、行政管理の中で評価に多くの関心があることがわかった。近年の公共部門改革に関するより体系的で中身のある情報を得るために、12の個々の研究からなる主要な評価プログラムが 1995年から 1997 年に実施された。このプロジェクトのもう一つの目標は、行政のなかで評価に対する認識の強化方法を見つけることである。トレーニングコースが整えられ、手引き書が編集され、そして、省が自分たちの評価プログラムとその結果を年次報告に載せることを奨励する試みが行われた。

アメリカ合衆国において

アメリカ合衆国は、評価については長期にわたる広範囲な伝統を持っている。いくつかの州や地方自治体と同様に、連邦政府は評価ユニットを持っている。連邦レベルでは、評価ユニットは、通常個々の部門に置かれている。会計検査院(GAO) は、議会が行う行政庁の地方局の監視業

務を手助けするとともに、効果的な管理、およびアカウンタビリティを鼓舞しているが、極めて多くの業績会計検査、およびプログラム評価をも実施している。またプログラムの施行に責任のあるエージェンシーは、彼ら自身でプログラム評価を行う。エージェンシーは、事業計画のなかに、使用したプログラム評価の記述、および今後の評価のスケジュールを含んでいなければならない。さらに、議会は、しばしば、個々(通常大規模なものであるが)の評価を要求し、それを法制度の中に盛り込んでいる。

ECにおいて

EC では、開発援助、調査や技術政策、および構造資金(Structural Fund)を通して資金調達されたプログラム、といったEC政策の中心の政策分野において、制度化された評価を、かなり長い間、行ってきた。EC主導の Sound and Efficient Management(SEM 2000) の内容では、ECの政策全体を通して、より体系的、よりタイムリー、より厳格に EU 支出の評価が行われることが中心の目的となっている。Operational Service は、プログラム評価をきちんと実施することに責任があると述べている。職員、部門、あるいはユニットは、責任がある者として任命されねばならない。

Service 部のすべてが、毎年の評価プランを計画せねばならず、最初のものは 1996 年末に行われた。これらのプランは、集められ、ECにより記録され、ECの毎年の評価プログラムとして配布された。ECの Operational Service によって作られた評価報告書は、毎年の評価レビューの基礎資料として使われている。毎年の評価レビューは、毎年 1 月末までに、毎年の評価プログラムと共にECに提出される。さらに、Operational Service は、ECの支出を伴う活動を提案するときには必ず、事前評価報告書の提出を求められる。新しいプログラムを始めようとするときには、評価計画もそのプロセスの中に組み込まなければならない。

毎年予算を組んで資金供給される活動は、少なくとも6年に1度は評価されるべきことが推奨されている。Operational Service に加え、DG XIX (予算) および DG XX (財政コントロール)は、評価活動をサポートするいろいろな業務を行っている。評価カルチャーの一層のサポートのために評価者のネットワークが設立され、作業基準のマニュアルが編集され(European Commission , 1997)、トレーニングセミナー、デモンストレーションプロジェクト、ヘルプデスクも設けられている。

2.4 業績管理と評価

多くの政府が 政府組織の業績を管理するための包括的なアプローチを取り入れている。評価は、明らかに業績本位のための手段であり、それゆえに、より広い業績管理の枠組みの一部として見られるべきである。しかし、評価と業績管理の関係は、必ずしも明らかではない。

評価および業績管理

- ・評価は、多くの政府で、業績管理の導入以前に十分に出来上がっていて、新しいアプローチには、必ずしも評価が組み込まれなかった。
- ・新しい業績管理へのアプローチは、幾分、評価の失敗の認識の上に立って開発された。すなわち、評価に投入される資源に比較して、評価結果の実際の活用は制約があったのである。
- ・評価と業績管理は、異なる組織に責任があることもある。
- ・業績管理は、中心的な管理プロセスを改革することを含んでおり、これゆえ、評価を専門的方法として見ている各組織の管理者の責任は、しばしば外部の専門家や評価ユニットによって果たされている。

評価と業績管理の関係を考究し明確化することに失敗すると、浪費、重複を引き起こし、ねらいと結果の葛藤をひきおこし、組織と専門家グループに緊張が走ることとなる。大抵の政府は、評価を包括的業績管理の枠組みの一部と見ているが、統合の度合いは、評価の役割による。主なアプローチは、3つある。(下表参照)

評価と業績評価の統一へのアプローチ

- 1 . 評価は、業績管理と比べると、はっきりした役割をもち、異なった独立の機能とみなされている
- 2 . 評価は、異なった機能とみられているが、業績管理と統合される
- 3 . 評価は、異なった機能と見られてはいないが、個々の業績管理の手段と統合される

最初のアプローチは、たとえ評価が、業績測定と、業績管理を通して得られた他の情報によるところが大きいにせよ、評価は比較的業績管理から独立した存在であるとみなされている。基本的な見方は、業績管理が、評価に従わなくてはならない機関の他の管理プロセスと類似しているということである。評価と業績管理を過度に統一することは、業績管理の質を査定し、それを改良する有効な方法を提案するという評価の能力を蝕むこととなる。過度の統一は、評価機能の必要な独立性をも蝕むだろう。

第 2 の アプローチは、独立性の強調を少し弱くして、評価を包括的な業績管理の枠組みの中で使われている多くの手段の 1 つとみなしている。評価は、多くの他の手段よりも多くの点で優れているだろう。評価は、業績測定よりも業績のより詳細な研究を可能とする。また、より長期間をカバーすることができる。アウトカムや影響評価といった、実行中の測定を通して評価するのが難しい、あるいは費用のかかる要素も組み入れてしまうのである。評価は、プログラムや機関の業績の包括的なイメージを得るのに使うことができ、他のアプローチによって与えられるいくぶん断片化された映像と全く対照的である。このアプローチで、評価は、専門技術を確実に実施する異なった機能として整理され、他のプロセスでは得られなかった鍵となる情報を生み出すために、選択的に、および定期的に使われる。評価の実施にコストがかかるということは、評価の実施が最も重要なプログラムや機関に限られる必要があるだろうということを意味する。

第 3 のアプローチは、評価方法を個々の業績管理の手段に組み入れることの利点を強調している。この例は多くある。業績情報は、分析され、解釈され、評価される。広範囲な評価目標 (benchmarking) の研究は、しばしば、業績の違いを説明する、考慮すべき評価要素を含む。評価は、戦略的管理 (例えば環境の主要要素の評価、スウオット分析 (SWOT、strengths, weaknesses, opportunities, threats) の主要構成要素の 1 つだ。業績契約は、定期的に評価される。そして、質的管理のアプローチは、しばしば、自己査定を含む。ここでの基本的な見方は、評価は進行中の活動であるべきで、鍵となる管理プロセスに組み込まれるべきだということである。

これらの 3 つの アプローチは、相互に相容れないものではない。すなわち政府は、2つ以上の アプローチを使うことができる。第 2 のアプローチは、明らかに勢いを得てきている。これにもかかわらず、評価の独立性はある程度保たれており、これは会計検査オフィスにより行われる会計検査で顕著である。評価方法を鍵となる管理プロセスに組み入れることの利点は、認識されつつある。しかしながら、大抵の政府は、これを、より専門的な評価への置き換えというよりは、補足としてとらえている。

2.5 方法の課題

この冊子は、評価の使用法について集中的に述べているが、評価に関する主要な方法論的アプローチや課題についても手短かに取り上げることが適切だろう。方法論、および評価の使用法は、連結しているので、これらの疑問は、後程、再び出てくることとなる。

評価へのアプローチ

いかに新しい信頼できる知識を創るか、そしてそれをどのように使えば合法的にすることができるか、といった評価の知識の性質についての仮定を組み入れた評価へのアプローチは多様である。そのうえ、評価に選ばれるアプローチ、ツール、および手法の間には、しばしば関連性がある。4つの主要なアプローチとそれに関するいくつかの評価方法を、以下の表で述べる。(Pollitt, 1996)。

アプローチと手法

・実験的評価	→ これは自然科学の方法論を公共プログラムの技術へ適用するという実証的考えを組み入れたものである。実験的なデザインの中で、似たようなグループを無作為に選び、そのうちの1つにプログラムを適用した後でグループを比較する。プログラムの効果は、アウトカムにそれ以外の対抗する原因がないという仮定の上で成り立つ。
・实际的評価	→ これは、評価の結果の実用性にとりわけ焦点を合わせている。評価は、目標、および鍵となる政策決定責任者の実施中の作業に連動している。政治上の実行可能性、タイムリーさ、およびコストが重要視される。準試験的な計画(アウトカムの代替説明が無作為に何かに割り当ててしまうということではない方法でルール化することを追求したもの)、ケーススタディー、および様々な記述的な方法が、評価ツールとして典型的に使われる。
・経済的評価	→ これは、しばしば一般的な評価とはすこし違った形で行われる。ツールは、例えば費用便益分析および費用効果評価である。それらは、プログラムのコストの情報を評価の判断基準として取り入れる。
・自然主義的評価	→ これは 社会には社会性があり、個人の相互作用によってたえず変えられていることを主張することにより、実験的評価や経済的評価を否定する。それゆえに、評価は、正しく、客観的な答えを提供することはできないが、その一方、stakeholder の間でコンセンサスを作ることを促す役目をすることができる。適用される方法は、例えば、民族誌学、関係者の観察、ケーススタディーなどで、stakeholder の価値と仮定を描写し、明確化することと、評価プロセスのなかで彼らの交渉を支えることをねらいとしている。

評価を遂行する最良な方法というものはない。多様なアプローチや方法は、個々に成功要因を包含する。評価の役割は、学習、および意思決定のプロセスにより異なる(例えば、経済的・実験的な評価は予算編成の意思決定に使われ、また、自然主義的評価はプログラムの改善と設計に使われる)。重要な点は、評価される題目、望ましい評価情報、利用できる資源、結果の使用、これらに合うように 評価を設計することである。

方法に関する問題

方法に関する問題は、すべての評価アプローチに内在する。例えば、因果関係に関する問題は、一般的に社会科学に共通である。原因—結果の関係の確証は、関係のある変化の全てを管理することがほとんど不可能なので、めったに打ち立てることができない。実験的評価の設計は、実際に実行するのが不可能ではないにせよ、しばしば難しく、高額のコストがかかり、また非常に長くなる。たとえ実験的評価の設計が使われるとしても、実験の状況を超えた一般化ということには、通常、ならない。しかし、多くの評価で、例えばプログラムの施行や管理を評価するとき、確固たる因果関係を打ち立てることは、評価の主要な目的ではない。

評価の基準の選択が疑わしいこともある。それは、単に意図されている公共プログラムの目標がしばしばいくつもあったり、あいまいであったり、隠れていたり、進化していたり、さらに矛盾していたりさえするからである。別の難しい点は、どのように公式の目標が達成されるのかのみに焦点を置くべきなのか、より広い観点からプログラムのすべての効果を研究すべきなのかを決定することである。後者は、プログラムのアウトカムのより包括的な映像を得られるが、より複雑であって、また時間がかかる。

関連したアウトカムが十分に発揮されるには時間が十分に必要なので、プログラムが評価される適切な期間を設定することは難しいことだが、必要不可欠でもある。しかし、もし評価されたプログラムが評価の終了以前に変わったり、評価の結果が適用される前に変わったりしたら、その情報の有用性は減少するだろう。

評価の方法を選ぶことにも、また問題がある。データの収集や分析方法が多様なことが、評価されるプログラムに異なる見方を与えるからである。例えば、評価に量的方法と質的方法を使うことの長所と短所についてはかなりの論議がある。

プログラムのアウトカムを明らかにすることが難しいことは、よくある。以前に指摘したことだが、アウトカムを完全に、あるいは部分的に生み出す理由には、様々な理由(そのプログラム以外の

理由)が存在するだろう。これゆえ、アウトカムについて有効で信頼できる結論を引き出すためにどの証拠があれば十分か、ということを決めることは、難しいだろう。証拠を集める能力は、利用できる資源によってしばしば、制限される。結論に関する別の問題は、評価の結果が、それ以外の時期、場所、あるいは主題に一般化することができるかどうかということだ。これは、評価が、将来の政策決定に寄与すると期待されているときには、特に重要だ。

評価の方法に関する問題(ここではそのいくつかしかカバーしていないが)は、評価の限界が認識され、問題が適切に提出されていれば対処することができる。これには特別な知識とスキルが要求される。その知識とスキルは、評価を行うためにスタッフを教育したり、外部の専門家を任命することによって得ることができる。また、異なる方法を併用することは、方法論の問題を少なくするためにはもっとも有益なアプローチだ。なお、適切な品質管理メカニズム(技術的な質の確保についての後の項を参照)は、評価プロセスを導くために設定されることも可能である。

たとえ評価が決定的な答えを提供することができないとしても、評価は、業績についての論議に役に立つ情報を加えることができる。なにが起きたか、なぜ起きたのか、という信用できる情報は、何もないよりはましである。

3. 評価の使用を改善するための評価の管理

方法に関する問題に加え、もう一つの主要な課題は 評価の使用法である。評価は、それ自体としては知識ではなく、活動のための知識なのである。重要なことは、評価材料が意思決定に使われることである。もしそうでないなら、行政はすぐに興味を失い、評価はむなしい実習になるであろう。

評価が、意思決定に影響を与えるためには、質の良い供給と、熱心で誠実な需要が必要であると述べられている(Mayne 1994 ;Mayne , Divorski , Lemaire , 1998 ;Boyle , 1997)。この章では、評価「マーケット」のこれら 2 つの側面をいかに前進させるかについて議論する。すなわち、評価を使う機会をどのようにつくるか、ということである。

なにが評価の使用法を構成しているかを定義するのは、難しい。評価の使用法の 1 つの定義は、プログラム評価の用途と機関の評価の用途を区別する、ということである。プログラム評価は、個々の評価研究と関連があり、機関の評価は機関のなかでの一連の評価に対する累積的な効果(学習効果)と関連している。プログラムの用途は、さらに手段としての用途と、概念的な用途に分けられる。手段としての用途は、評価は具体的な行動(確立、修正または終了)をサポートしており、概念的用途では、直接の行動は取られず、評価の結果が考慮され、後に使えるような情報が獲得される。(Mayne , 1994)

プログラム評価の使用法

- ・大幅な政策変更
- ・プログラム改善(組織としての学習)
- ・予算配分の改善
- ・アカウンタビリティのサポート

いずれにせよ、評価結果の使用法は、説明しやすいものではない。利用する可能性のある者がいれば明らかにされるべきであり、予想される利用期間があれば決められるべきである。使用法には、具体的な改善が含まれるが、また、プログラム、組織としての学習、あるいはアカウンタビリティへのより深い洞察力のようなあまり具体的でないものも含まれる。

評価の使用法とその影響力を区別することも、また可能である。ここでの疑問は、評価結果を実際に利用するのにどのくらいの影響力が必要とされるかということである。経験によれば、評価は、政策の根本的変更や資源配分をサポートするよりも、現存のプログラムの改善により多く使われる。これは、スウェーデンのケースである。しかし、主要な政策変更は、何年もの間に集めら

れた累積的な評価の情報によってサポートされるということを指摘しておかなくてはならない。また、個々の評価で、プログラムの実質的な変更を提案しないものも多いだろう。

これらの要素のすべては、評価の使用法を査定することを複雑な問題としている。使用法に関しての体系的な研究はあまり多くはないが、終章でいくつかの例を挙げておいた。

評価は、政治的な範疇のなかで実施されるので、確実な使用法という問題は決して十分に解決されることはないであろう。評価は、1つの情報源にすぎず、意思決定に影響を与える他の要素と競合するものである。どの程度評価結果が決定に影響を与えたかを見つけ出すことは、ほとんど不可能だ。評価は、政治的優先順位が低い政策の測定には極めて大きな影響力があるかもしれないし、また全体的な政策の方向に広く合意がなされている場合もそうである。さらに言うと、政治家は公共的なことに介入するときには誠意を強調するが、いつもその結果を研究することに関心があるとはいえないかもしれない。評価を遂行し、その結果を議論することに不承不承である背景には、不十分な箇所がさらされることへの恐れもまたあるのかもしれない。政治家は、必ずしも政策目標を明確に述べたくないかもしれない、そして、それが評価を難しくするのである。意思決定過程に持ち込まれる資料の質にかかわらず、決定が行われ続け、そして今後も行われなければならないという事実は依然として残っている。これを心に留め、意思決定責任者との開かれた協議によって評価の使用法を促進することは、可能である。

評価に価値を見出すという業績指向の共通の認識は、一晩では生まれない。しかしながら、評価に重きを置き、行政、意思決定責任者および社会全般による評価の使用法を広げる、いくつかの具体的なステップをとることはできる。以下のセクションは、評価の使用法と価値を最大限に利用する実践のうちのいくつかに重点をおいている。しかしながら、異なった範疇においては、評価の利用を強化する方法も異なるかもしれないということは、指摘されるべきであろう。

最初の3つの課題は、プログラム評価を実施するに当たり、有効的な需要と助けとなる環境を作り出すことに焦点を置いている。次の3つの課題は、評価を実施するための効果的な枠組みの開発に焦点を合わせている。その次の9つの課題は、評価のプロセス自体の実際的な側面に焦点を合わせている。評価の結果の使用法をどのように推進していくか、ということである。最後の課題では、行政で評価を実施するという認識を作り出すに当たっての極めて重要な要素として、トレーニングとサポートを挙げている。

評価実施を改善するための課題

- 課題 1 トップからの支持を得る
- 課題 2 有効的な需要を生み出す
- 課題 3 現実的な期待しうる仮説を設定する
- 課題 4 評価活動の体系化
- 課題 5 予算編成プロセスとの連携
- 課題 6 適切な評価者の選択
- 課題 7 評価の計画
- 課題 8 評価の時期を適切に図る
- 課題 9 利用者のニーズを満たす
- 課題10 適切性の確保
- 課題11 stakeholder の関与
- 課題12 方法に関する質の保証
- 課題13 結論と提言の作成
- 課題14 調査結果の伝達
- 課題15 モニタリングまたは追跡調査
- 課題16 スタッフのトレーニングおよびサポートの必要性の認識

課題 1

トップからの支持を得る

マネージャーや政治家が評価の価値を理解し、その利用について現実的な期待を持つことは重要である。このことは、政府の公式な評価戦略の採用を通じて実現されるべきである。大蔵省や他の中央政府の機関の役割は、しばしばサポートと信頼性の確保には不可欠である。マネージャーは評価プロセスに関係すべきであり、必要に応じてインプットを用意するべきである。目標が主に現存のプログラムの施行の改善であるときは、プログラムマネージャーのサポートは不可欠である。

オランダにおいて

1990 年代初頭、政府が評価の機能を述べた戦略計画と、それを促進するための戦略計画を確立し、必要な責任と提言を定めた。政策の責任者も評価の運営に責任があるが、各省の長が究極的に責任がある。大蔵省と内部の予算責任者は、コーディネート(調整)、推進、利用の確保、そして必要なアドバイスやガイダンス、専門の研究者が用意されているかどうか確認する責任がある。

課題 2

有効的な需要を生み出す

プログラム評価の利用が不十分であることの背景にある理由の一つに、供給に追い立てられるという市場の特性があるということが示唆されている。評価への包括的な需要が強調される必要性があるのであろう。需要は、評価が実際に利用されるのかどうかについて問題提起し、明確に言い表し、プログラムの改善に導かなくてはならない(Mayne , Divorski , Lemaire , 1998)。顧客は、彼らの評価への関心を具体化し、それを知らせる必要がある。要求を促すために使われる誘因は、沢山ある。これらは、Sticks、Carrots、Sermons という言葉で描写されてきた(Boyle , 1997 ;Toulemonde , 1998)。

誘因

・Sticks	→ Sticksは、評価が実施され、適切な質問が聞かれたということを確認するために使われる。評価を強制的にし、また外部の stakeholder に評価の質問をする権利を与えることにも使われる。評価の質はまた、特別なメカニズムによって管理される。例えばフランスやカナダでの運営ないしレビューのグループがそうだ。政府ユニットの間で競争を増やすことは、彼らの活動と成果に関しての評価情報をより多く提供することを促すこととなる。
・Carrots	→ Carrots は、需要を開発するために使われる。評価のための資金の用途指定、および評価の努力への認識によって行われている(知名度、賞、促進、作業内容の改善、業績に関連した支払い)。評価によって増えた仕事は、インプットを縮小し、およびプロセスを管理することによって釣り合いをとることができる。
・Sermons	→ Sermons は、評価の役割を強調し、意思決定時の評価の使用を尊重する認識を向上するために使われる。会議、研修コースおよび訪問、ニュースレターおよびジャーナル、仲間集団ネットワーク、賞金などを通してこのメッセージは送られる。最も重要なのは、評価結果に基づく行動が、評価が本当に重要だということを証明するということである。

課題 3

現実的な期待しうる仮説を設定する

例えば評価結果のタイミングや範囲に関連した期待のように、現実的でない期待というのは、幻滅を引き起こすことがある。これは、例えば 1960 年代、1970 年代のスウェーデンなど、多くの国で見られた評価草創期の苦勞のケースである。プログラム予算の計上、および評価の導入は、各省による非現実的な期待、不十分な関心およびサポートのために困難にぶつかった。評価の本質的な限界を認識することは、重要である。

資源を評価につぎ込んだ政治家は、もちろん、評価結果が实际的にタイミング良く意思決定の改善に使われることを期待している。また一方、評価者は、現れるのに何年もかかるかも知れない累積的な評価結果の長い目で見た価値を強調するかも知れない。それゆえ政策決定責任者は、評価者が自分達はなににも与えることができないと言っている、という情報の提供を評価者に期待している。にもかかわらず、評価者は、より一層多くの評価を進め続けている。

需要側と供給側の双方が自分たちの期待を明らかにして評価が现实的に提供するものはなにかを確立することができなければ、両者の違いが不信や失意を引き起こすこともある。能力のある利用者は、ありがちな評価の限界を知って、现实的に満足しうる有効的な需要を作りださなければならない。評価の目標は、役に立つ情報を提供して、研究した政策に関する不確実なことを減らすことである。しかし、それは、「正しい」答えを与えたり、あるいは意思決定の判断を代替することはできない。さらには、評価を通して得られた情報のタイプは、評価者の選択(後の項を参照)にかかっている。評価の委員は、学識のある、鑑識眼のある買い手として行動せねばならず、プロジェクトに対し過度に楽天的な観点の、能力がない評価者を避けなければならない。

課題 4

評価活動の体系化

ある程度の評価の制度化は評価が政府機関の管理の中で真の役割を果たす前に行わなければならない、としばしば主張される。制度化の目的は、高品質の評価の提供、評価結果の利用、そして組織としての学習のための条件を作ることである。

重要なことは、評価の努力を、戦略的に、業績管理の枠組み全般の中に統合することである。様々なレビュー、および管理活動は、活動の重複やあまりにも重荷となるような報告を求められることを避けるために、論理的で筋の通った枠組みの中に組み込む必要がある。体系化の度合は、かなり多様である。

体系化の度合

	略式	体系的
・ 評価実施の必要条件	→ 自発的でフレキシブル	→ 強制的で定期的
・ 評価実施の責務	→ 分散化	→ 集中化
・ 評価者の立場	→ 内部	→ 外部
・ 報告	→ 政府の行政部門	→ 政府の法制部門
・ 評価の焦点	→ 意思決定と施行の改善	→ アカウンタビリティ
・ 主導性	→ 地方で主導、実施	→ 中央で主導、実施
・ 開放性	→ 低い	→ 高い

評価を行う方法は、意思決定の際の評価の役割に影響を及ぼす。異なる政策決定のプロセスおよび意思決定のプロセスの評価の統合は、それらの利用を確かなものにする鍵となる点である。(後の予算編成プロセスとの関連の項を参照)

付加価値、および例えば現在実施中の業績測定と比べてみた評価のコストは、査定されなければならない。評価の目的のためにお金を別立てしている政府もある(例えば EU の出資プログラムの評価は通常プログラム予算で財源措置されているが、この財源措置のほんの一部はしばしばこの目的のために別立てされている)。あるいは、評価プロジェクトに財源措置をするための特別のファンドを作った政府もある。

オーストラリアにおいて

評価を体系化する際に柔軟性と管理の間で起こる緊張は、そのプロセスを制度化するという最新の協議のもとで解決された。その一方で、個々の部門には、**どのように**仕事をするかの決定をかなり任せた。大蔵省は、主要な評価を選んで参加し、質的保証に対して支援することができる。

課題 5

予算編成プロセスとの関連

評価が効果的であるためには、政策の設計と施行の肝要な一部分でなければならない。予算編成プロセスとの正式かつ体系的な関連は、それゆえ重要なのである。しかしながら、7 カ国で

行われていた実際の管理を比較してみると、そのうち予算管理、会計検査、評価が適切に統合されていた国は一つもなかったのである(Gray & Jenkins, 1993)。

予算編成過程で評価を利用することに関する期待は、現実的であるべきである。予算編成は、政治的な選択を含み、そして、評価は、熟考すべき多くの要素のたった 1 つでしかない。それは、評価者と 評価される機関との間に建設的な関係を育むのに役立つ。評価を資源配分と関連づけることが、評価者と評価される機関との間の信用を弱め、したがって評価の質を落とすだろう、と考える人もいる。評価結果が予算カットの基準としてしか使われない管理主導および節約主導の評価では、抵抗が生じるであろう。評価プロセスの中に、評価される機関の マネージャーやスタッフを含めることで 評価結果と信頼性の創造のバランスを改善することができる。

多くの国 (例えば、オーストラリア、スウェーデン)、および EC もまた、予算編成過程との関連性を改善しようと試みている。以下に挙げられた例は、実際の解決法を具体的に説明したものである。

オーストラリアにおいて

オーストラリアでは、評価は、予算編成過程と関連付けられてきた。プログラムの予算管理は、プログラム評価を含み、意思決定の枠組みを提供する。評価は、プログラムの有効性を基準として新しい資源を再配分するための大事な手法となっている。

1994 年と 1995 年の予算編成プロセスにおける評価の使用法のレビューでは、評価結果は、新政策の予算要求の 77 パーセント に影響をあたえ、また 65 パーセントの節約を選択肢とする政策に影響を与えたことがわかった。評価は、内閣で審議でされた予算要求事項の 68 パーセントに影響を与えたと考えられた。これは、それ以前に行われた予算に関する調査の結果を超えるかなりの増加であり、これには、評価結果にかなり大きくサポートされているある一つの大きな一括法案、the Working Nation package(働く国民のための法案)では新政策の予算要求が大きな部分を占めていた、という事実も一部影響している。

オーストラリアの立場は、評価はそのプログラムを管理する者によって行われるべきであるというものであるが、だんだん全体のプロセスは資源の割当てに責任のある人達によって調整された良き実施原則に基づく業績管理と統合されたアプローチにより導かれていくべきである、というものになっている。大蔵省は、部やエージェンシーが行う評価に対する調整や指導の中心的な役割を担っており、ときには個々の評価に参加することもある。

スウェーデンにおいて

スウェーデンでは、1980 年代終りに新しい予算編成のプロセスが始められた。そのねらいの 1 つは、アドホックに行われていた事前の評価を正式な意思決定のプロセスにすることであった。毎年の業績情報の作成に加えて、エージェンシーは、3 年ごとに自分たちの活動の徹底的なレビューを遂行することを要求された。その目標は、政策のアウトカムの論議を容易にし、重要な政策決定によりよい基礎資料を提供することであった。

その評価システムは、より多くの柔軟性を取り入れるため、近年、改正された。予算編成過程で使われる主な書類は、主としてアウトプットに関する情報を含む年次報告である。徹底的なレビューが行われるのは、それぞれの省が求めたときだけである。予算編成過程で使われる概況説明資料は、大幅に改善された。しかしながら、評価の使用法に関する体系的な情報は、ないのである。一般的な印象は、エージェンシーの年次報告で与えられる情報は、主として省によるコントロールのため、そして政策の微調整のために使われるということである。

EC において

組織的なプログラム評価政策は、1996 年に採用された。Operational Service は、評価を遂行することに責任がある。その一方、予算編成と財政管理のための幹部会は、評価の質をサポートし、管理する。

予算編成と評価との間の重要なつながりは、すべてのプログラムは、そのプログラムのための評価計画を述べた財務諸表と共に提案されなければならないという要求を通じて形成される。新しいプログラムの建議は、事前評価に基づいていなければならない、プログラムの刷新に関する建議は評価によってサポートされていなければならないのである。要するに、「評価なくしては歳出なし」である。

課題 6

適切な評価者の選択

評価活動機関は、政府が、評価にあってほしいと求めている役割を反映する。制度的な調整は、評価から期待し得るものに影響を及ぼす。それらは、意思決定および全体の立場に関連する評価の焦点に影響を与える。この論点には、いくつかの局面がある。

まず第一に、政府機関(一つの機関だけとは限らない) に評価を実施せよ、あるいは評価を委任せよという正式な命令を出すことができる。第二に、政府は、評価の正式な命令とともに内外の組織の評価能力の開発を促進することができる。三番目の局面は、それぞれの評価のための適切な評価者を選択することである。

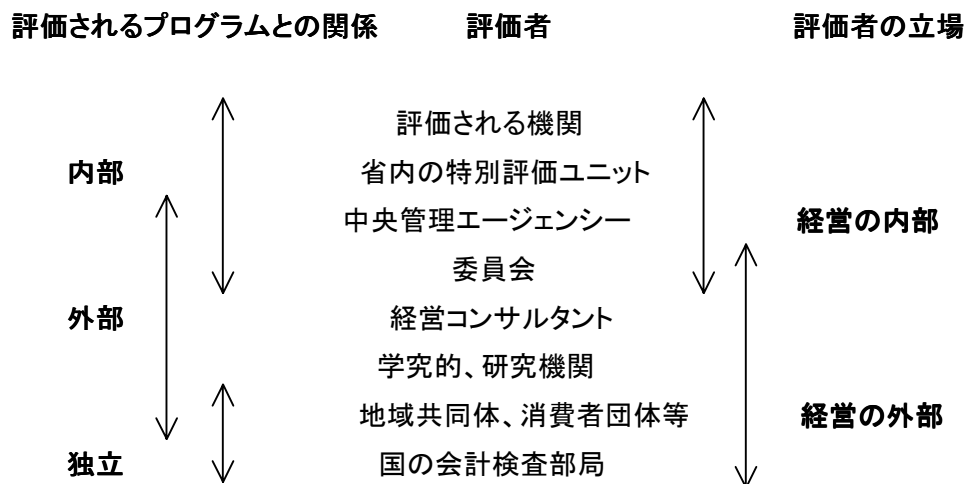
政府機関に評価の実施あるいは委任について正式に命令を出すことは、一般的に行われている。会計検査院に評価や業績会計検査を命ずることも一般的である。省内で評価ユニットを設定する際の必要事項の例がある。これは省をまたがって行うこともでき、また各省毎に行うこともできる。また、一般的評価命令と部門別評価命令の両方に、半独立の評価ユニットを設けている例もある

評価能力の開発は、評価実施の正式命令が下された機関に限られてはならない。多くの政府が、そこから評価者を選ぶ際のために、政府の内部と外部の両方を含む異なる範囲の評価者を持つことが大切だと考えている。

評価者の選択は、評価の有効性において重要な要素である。それは、それぞれの評価の目標を分析するために不可欠である。例えば、内部評価は、おそらく適切な問題点をあげ、実際的な提言を提供することができるだろう。しかし、内部評価に、批判的な情報の提供や、プログラムが継続されるべきか否かということと関連する問題点の指摘を期待するのは、非現実的である。独立した会計検査部局は、アウトカムに関連したより基礎的な問題点を指摘することはできる。しかし 彼らの問題点や結果そして提言の選択は、必ずしも政治的に適切ではない。半独立の評価者ならばこの二つの点をうまく組み合わせることができるかもしれない。

評価は、行為者がいれば実行することはできるが、評価を実施する正式命令がくだるのは、そのうちのほんの一部である。評価は行政内部では内部的に行うことができ、外部的には独立機関によって行われる。あるいは、よくあるケースだが、異なる行為者がいくつか一緒になって行われる。行為者のうち誰が内部者で、誰が外部評価者かを分類する程度は、何が期待されているかによる。自己評価以外はすべて、評価者は評価されるプログラムに対して外部者である。省内の評価ユニットは省に対しては内部者であるが、プログラムに対しては外部者なのである。会計検査部局は概して完全に独立している、などなどである。以下の表では、政府が正式な評価命令を下し、あるいは評価者として選ぶ可能性のある組織の範囲を示してある。またこの表は、どの組織が内部、外部、あるいは独立しているとして見られるかをも示している。何が期待されているかによって、これらのカテゴリーには重複する部分もある。

評価者の立場、および評価されるプログラムとの関係



評価が**評価される機関によって実施される**場合、そのアセスメントはその機関の基本的な情報の必要性に寄与する。内部評価者は、そのプログラムと自分たちが研究した問題点を心得ており、機関内の現在進行中の情報に容易にアクセスできるし、また、非公式の重要な情報も入手可能である。データが比較的簡単で、あまり費用がかからないこともある。自己評価は、経験からの知識を増やし、結果を利用に移すのを容易にする。また、要求が、あまり厳しくないこともしばしばである。というものの、プログラムの適切性やアウトカムに関する複雑な問題は、この形態の評価の対象にならないことがしばしばだからである。

組織内評価の問題点は、時間、スタッフのスキル、提示された問題点の範囲がかなり限られていること、客観性の欠除の可能性、そしてこれに伴う調査結果の信用性の欠除等に関することである。これは、結果的に機関外部で評価が利用されることを制限することにもなる。しかしながら、もし評価が、プログラムの全体的な適切性を評価するというよりも、プログラムの実施の改善に主なねらいを置いているのであれば、これは必ずしも問題にはならない。

カナダにおいて

部局およびエージェンシーは、プログラムの評価に責任がある。評価は、しばしば、部局に置かれた専門化された共同評価ユニットによって実行される。彼等は、時々、進行中のプログラムの論理的根拠を適切にレビューしないということで批判される。1996年に、大規模なプログラムレビューが行われ、この目的のために設立された特別の委員会を通じて、大臣によって行われた。このような大規模な評価は、はっきりした政治的リーダーシップを必要とする。

スウェーデンにおいて...

政府機関は、評価を行い、業績および評価結果について、報告することが求められている。1996年に国家財政の専門家グループが実施した研究によると、パブリックエージェンシーの自己評価の質は、結論が証拠に基づいていない等、一般的にかなり貧弱だということが発見された。それにより、独立したコントロールグループなどの、質の高い管理メカニズムの必要性が叫ばれた。

中央部局による評価は、政府全体の問題、および部門間の問題を指摘し、異なるプログラムの影響を深く研究し、現在行われているプログラムの適切性をより厳しくレビューすることができる。また、評価者の権限、および鍵となる意思決定のプロセスへのアクセスは、評価の使用法の程度を高めるであろう。評価結果に認められる信頼性、および指摘されたより政治的な問題の性格は、しばしば評価された機関以外から関心を持たれ、政治的な意思決定プロセスのなかでの評価の利用を促進する。

問題点は、評価されるプログラムからの距離によって持ち上がるかもしれない。中央部局は情報への直接的なアクセスには欠け、データの集積は時間や資源、さらに他の束縛要因によって限られるかもしれない。中央部局からの評価者には、矛盾する役割があることもある（例えば、個々の省や部局にアドバイスを与えるだけでなく、大蔵省にもアドバイスを与えなくてはならない、等）。評価者は、個々の評価に関する自分の役割を明確にし、客観的な情報と機関の命による影響があるかもしれない判断との区別をしなくてはならない（例、費用に見合う価値 value for money）。そうでなければ、信用性が危険にさらされることもあり得る。さらにプログラム管理者は、外部評価者が出した結果をあまり受け入れたがらないかもしれず、また提言の実行が容易ではないかもしれない。

プログラム管理者、あるいは中央部局は、しばしば、外部評価者（例えば、研究機関や、コンサルタント）による評価のコーディネーターの役を務める。このコーディネート機能には、評価者の選択、契約のプロセス、評価するプログラムの選択、そして評価の作業において評価者を手助けすることが含まれる。

フランスにおいて...

評価に関する省間の委員会があつて、内閣、エージェンシー、会計検査院からのリクエストに応えて行われる評価の数を決定する。独立した科学会議が、評価の方法論、および質について委員会にアドバイスをする。評価は、評価開発のための国家基金から資金が供給されている。

オランダにおいて

内部予算の幹部会、および大蔵省は、評価のコーディネート、評価の利用の促進と使用法の決定、ならびに必要なアドバイス、指導、および調査専門技術が提供されることを確認することに責任がある。

オーストラリアにおいて

大蔵省は、各省やエージェンシーが評価に熟知するようになるための支援、評価計画と評価活動の展開、専門技術や参考資料の展開、スタッフの訓練、および完了した評価の目録の発行、等に関して中心的な役割を果たす。大蔵省はまた、各省と共にいくつかの評価に参加もする。

評価が外部の評価者により実施されるとき、彼らの自治性、および評価スキルは、しばしば、提示された問題の範囲や結果の信頼性を改善する。次にこれは、関心のある観衆を広げる。一方、外部調査者は、実質的な理解や、評価される機関についての政治的、行政的なカルチャーの理解が不足していることがあるかもしれない。さらに、外部評価者のなかには、むしろ型どおりの評価を行うものや、顧客を喜ばせる結果だけを提示するものもある。学究的機関は、過度に理論的な研究を行う傾向があるかもしれない。

評価が成功するためには、いくつかの場所でいかりを下ろさねばならない、と述べる者もいる (Mayne , Divorski , Lemaire , 1998)。そのときになって初めて、異なる需要を満たすことができ、また多様性のある市場への情報を提供することもできる。非生産的な競争の危険性や評価の崩壊の危険性もある。しかし、異なる評価供給者を利用することは、評価の中でカバーされる問題や展望を広げ、それゆえ評価結果の利用の可能性も増える。供給を多様化することにより、利点を結一し、単一の評価者の限界を最小限にすることが可能である。これはまた、半独立の評価組織を作ることによっても可能である。

多くの異なる機関において評価能力を発展させることはかなり難しい仕事であるので、初期の段階で、評価活動を、明らかに評価の需要があるところ、なるべくならば資源の利用も可能なところで行うことが賢明である。始めはある程度まで評価の能力を集中することもやってみる価値があるかもしれないと示唆する人もいる。これは、スキルと他の資源が増えるのであれば、他の場所へと広げることでもある(Boyle , 1997)。

いくつかの研究、および経験によって示唆されることであるが、政府機関は、内部で作られた情報の方が受け入れやすいようで、一方、外部評価は威嚇的と見られることも少なくない(例、Rist ,

1994)。しかしながら、評価が受け入れられる必要性、および今後実施される必要性は、内部評価の信用性を高めるための適切な質的管理のメカニズムによって、バランスが保たれねばならない。これは、評価の透明性と公表を増やすこと、外部に委託するグループがいること、さらにレビューを行うことによって、達成することができる(後の技術的質の維持の項を参照)。

「メタ評価(meta-evaluation)」言い換えると「評価のための評価」が必要となる範囲と、その位置づけもまた、関連している。例えば、外部組織によって後でレビューされる予定の内部評価を促進することは、利点があるかもしれない。

課題 7

評価の計画

評価の準備の良さは、主要な成功する要素である。題目を選択し、範囲および目標を決定し、評価プロセスを計画し、使用する基準を選択することに、時間や努力を費やすことは価値のあることである。注意深く計画を立てることは、評価の管理をより容易にし、評価のアウトカムの質にかかわってくる。しかしながら計画は、厳密すぎてはならない。評価が必要と状況の変化に対応できる程度の柔軟性は、持つべきである。

Commissionerは、評価の利用者としてしばしば相談のうえ、評価計画に関する最終的な責任を取るべきである。目標、問題点、基準、利用可能な情報、データ収集の方法、そして分析の手法を明確に定めることに注意を払うべきである。タイムテーブル、予算、評価者、そして全体的な評価機関もまた、決定されなければならない。結果が提示され、利用され、監視される方法も考慮されなければならない

計画を立てることには、評価結果、および提言の利用者を明らかにし、また異なる行為者の責務も定め、どのように彼らが報告され、動機づけられたかということも含まれる。評価結果の使用法は、熟考される必要があり、これには、提言の実施に対する具体的な段取りも含む。(後のモニタリング監視やフォローアップの項参照)

評価の計画を注意深く立てることにより、評価結果に基づき行動するという最終責任も強化され、さらに進んだ行動が取られることを確実にするのを助長するかもしれない。評価の利用者を計画の段階で決めることは、評価の設計にかなりの影響を与える。利用者が必要とする情報は、指摘されなければならない、評価される側の関心も、彼らの関わり合いと協力を確実にするために、考慮されなくてはならない。

個々の評価の計画に加えて、それを編集し、部局のあるいは政府全体の評価計画を公表する

ことも有効かもしれない。(すなわち、オーストラリアのポートフォリオ評価プラン、EC の年次評価プログラム) これらは stakeholder に、どんな題目が研究されたのか、どんな領域と目標で行われたのかの情報を提供し、評価のコーディネートと実施に役立つ。

課題 8

評価の時期を適切に図る

評価使用法という観点から言うと、評価は政策と意思決定のサイクルに合わせる必要がある。タイムリーということが、評価でしばしば重要な問題になる。評価は、決定がなされようとしているときに情報を伝達するように計画されなければならない。しかしながら、意思決定が行われる時期はあまり予想しやすいものではなく、政策は評価のアウトカムが査定される前に改正されることも多いために、タイミングが難しいこともある。異なる人が 1 つの評価の利用者の場合、スケジュールが別々ということもよくある。評価には番狂わせがあるかもしれない、結果が遅すぎるより、早すぎる方がよいのである。

タイミングは、評価されるプログラム側の観点からも重要である。プログラムがある程度成熟した段階に達する前に、事後評価が行われることがないよう心がけることも不可欠である。しかしながら、プログラムの開始時に効率的な監視システムを置くことが、良い評価を行うための最高のサポートとなる。評価の最中に題材の分析に十分な時間をさくこともまた重要である。Commissionerは、何か問題が起きたらタイミングよく行動できるよう評価日程を注意深く監視する必要がある。

課題 9

利用者のニーズを満たす

評価は、特定の観衆にターゲットをおき、利用者の情報ニーズを満たさねばならない。威信、認められている信頼性、および利用者との関係は、評価の信頼性そして今後の評価の使用法に影響を与える。評価は、本質的には、政策決定のプロセス、および意思決定のプロセスを念頭においたものであるので、利用者の考え方や行動形態をよく知ることが重要である。政治的介入のプロセスを理解することは、健全な分析を行うことができるのと同じくらい重要である。評価の利用者と担当者のカルチャーについての意見交換も必要である。意思決定のプロセスの違いや、評価の利用者の多様性は、また、評価活動の多様化をもたす必要とするだろう。(例えばスウェーデンでは、多くの評価の供給者が異なる意思決定責任者に指示を与えている)。

しかしながら、評価は、特定の利用者への直接的なサービスをしのぐ役割があるだろう。評価は、他の責任も認識しなくてはならない。倫理的に容認できないような振る舞い、誤解、あるいは

依頼者のためであっても評価結果の誤った使用をすることを、受け入れてはならない。

課題 10 適切性の確保

特定のプログラムを評価することに利点が考えられるかとか、評価の結果どの程度信頼できる情報が集められるかを査定することは、重要なことである。評価は、予算編成、政治的、管理上、またはその他の戦略的理由によって重大な意味のある疑問を指摘し、プログラムを研究すべきである。これは、政府内の評価の初期の段階では特に重要である。というのも、確固たる姿勢の評価は、ほんの少しでも、評価カルチャーに重大な影響を持っているからである。評価が行うに値するためには、査定されるプログラムがまた適度に安定しているものでなければならない。どんな評価への要求も正直でなければならない。つまり、そこには少なくとも評価結果が使われる可能性がなければならない。

政治的な見方からすれば極めて興味のある問題を評価することが極めて難しいことは、しばしばある。例えば、アウトカムや全体的な政策の適切性への疑問である。評価者と利用者との相互の働きかけが十分でないために、あまり重要でない問題が提示されることもしばしばある。評価者を注意深く選び、できれば委託を受けるグループを作ることは、相互の働きかけの促進になるだろう。評価の commissioner は、評価の委任事項を決め、および評価を始める前にその事項についての幅広い討議のサポートをすることに十分な時間を確実にとらなければならない。

ある程度の柔軟性は、評価のプロセス全体を通して適切性を維持するために必要である。評価されるプログラムの変化や情報の需要の変化に応じて、評価を調整する必要もあるだろう。特定の問題の適切性は、評価が進むにつれて変化するかもしれない。それゆえに、評価契約は、あまりに厳密すぎではない。評価者は、先を見越した行動をとるべきで、他の見方が適切だと判明した場合には、評価をもととの問題に限定するべきではない。しかしながら、柔軟すぎる評価は、評価者によって分捕られてしまう危険性もあるため、ある程度のバランスは必要である。

フランスにおいて

ある評論によると、評価のための科学会議から委任された評価研究は、主として比較的重要性の薄い話題に限定されている。政治的に敏感な問題を指摘することに気が進まない官僚主義や、ある種の政治的な環のなかでの衝突の回避は、評価の適切性を蝕んでいる(Boyle , 1997 ; Toulemonde , 1998)。また、評価を行う学者と、情報を受け取る政治家との認識面での隔たりが、評価結果の使用を限られたものになっている。

論点 11

stakeholder の関与

Stakeholder とは、評価のプロセスまたは評価結果が彼らの役割や行動に関連性があるため、評価に関心を持っている人のことである。Stakeholderは、政策決定者（議会人、大臣、その他の意思決定責任者）、予算担当部局、プログラムの管理者、プログラムを施行する公務員、プログラムの恩恵を受けるグループ、あるいは、プログラムに関心があったりプログラムに影響を受けるその他の関係グループがなりうる。

評価に個人参加式のアプローチが望ましいかどうかについては、見解が異なっている。いずれの場合にも、Stakeholderを管理する戦略が必要である。Stakeholderを決めて、彼らを評価プロセスに巻き込むことは、通常、評価結果を使うことについての彼らの責任と意欲を増やすであろう。異なる Stakeholderによる異なったものの見方は、評価のプロセスを豊かにすることができる。

しかしながら、Stakeholderを含むことには、欠点もある。管理することが難しくなることがあるし、また、多岐にわたる関心のバランスをとることも難しくなることもある。ある種の Stakeholder（評価されるプログラムの管理者やスタッフも含む）は、特に彼等が評価結果が彼らの業績に批判的かもしれないと思っている場合には、参加し、情報を提供することに消極的になるかもしれない。Stakeholderとの関係をうまくすることは、個人参加的でないアプローチよりも費用がかかり、時間もかかることがある。さらに、Stakeholderが関わることにより、評価の信頼性も減るかもしれない。

これらの難点にも関わらず、Stakeholder達とのパートナーシップは、役に立つ。そうでなければ、評価は、潜在的な利用者と薄い関係しかない専門家によって行われる官僚的で形式的な手続きになってしまう危険性がある。

パートナーシップは、Commissioner、評価者、評価される者、潜在的な利用者、および他の関係グループといった、すべての関係者との間で打ち立てることができる。Stakeholderの関与はまた、運営グループや諮問委員会を作ることによって、評価の正式な部分とすることができる。協力の範囲は、例えば評価プロセスの段階によって、様々であろう。委任事項を明確にする上でのパートナーシップは、評価を最大限利用するために最も有効だとしばしば考えられている。Stakeholderもまた、適用する方法を決定し、結果の分析と解釈にも関係すべきだと考える人もいる。

評価者は、関係当事者との公正で信任の厚い連携を心がけなければならない。評価者は、評価のプロセスの中で提示される関心のバランスに気をつける必要がある。他より多くの注意を払う必要がある Stakeholderのグループは、そのサービスの利用者、あるいは評価されるプログラム

の受益者グループである。このことは、評価を自然と複雑化する。というのも、プログラムのアウトカムはプログラムの目標と関連してレビューされるだけでなく、顧客のニーズによってもレビューされなければならないからである。評価者はまた、どんな少数派グループであれその考え方を特に注意して保護しなければならないということも、しばしば示唆される。

スウェーデンにおいて

委員会システムは、今世紀の間、政治上の決定の準備に中心的役割を果たした。システムの相対的な重要性が近年減ったにもかかわらず、200 近くの委員会が毎年機能している。委員会は、政党、学界、商工組合、雇用主協会、消費者団体、環境団体などの代表者を集め Stakeholder の見解を組み入れる。さらに、委員会によって作られた報告書は、よく考えてレビューするために、かなりの数の公的、私的両方の関係団体にも言及している。この方法だと、社会で入手できる専門家の知識が、分析し政治的決定を準備するために組み入れられる。

課題 12

方法に関する質の保証

評価の方法の質は、適切な基準、十分な証拠、および今後の信頼できる明確な調査結果といった論点を包含している。方法の健全性は、信用性、および評価の利用にかなりの影響を及ぼすであろう。この側面には、多くの国で急速に公共部門の評価の市場が成長しているので、特に注意が必要とされる。

方法の厳格さと評価の有用性の間には、ある程度の二律背反性があるかもしれない。複雑な研究方法に支配され、認証を得られない限りなんらの結論も生み出せない評価は、政策決定者の注意を引く評価とはなりえないであろう。また、一方で科学的手法を、他方で政策に関して実用的な解決法を統合して使うということも、大変難しいと言える。費用と時間の問題も、最良の方法を使うに当たっての範囲を制限することがよくある。というのも、その方法が費用がかかり過ぎたり、時間がかかり過ぎたりして、現実的ではない場合があるからである。言い換えると、実用的な考察は、認識論的な考察とのバランスが取れていなければならない。

評価の方法は、評価の計画の段階で、可能な限りの程度まで、選ばれるべきである。これは、データ収集の方法、質的および量的な方法の使用法、サンプリング、因果関係、代替説明、および討議の論理を含む。評価の結論は、データからの調査結果に基づくべきで、提言は、結論に基づくべきである。評価者は、評価の仮定、不確実性、および他の方法上の限界について明確にする必要がある。可能な限り、事実は、意見や仮定から切り離されなければならない。

品質管理のプロセスには、様々なものがある。技術的に優秀な専門知識、および他の資源が、質の高い評価に寄与することは言うまでもない。評価のプロセスの開放性や透明性は、外部のレビュー者が評価の質を吟味することに結びつく。評価を公表することは、公表された評価の質に注意がより多くはられるようになるであろうから、品質管理のメカニズムにもなりうる(後の、結果の伝達の項参照)。評価の品質基準を設定するために、チェックリストを作ることも可能である(例えば、EC DG XIX Assessmeny checklist)。評価のための倫理規範もまた、評価の質を高めるために採用されている(例えば、アメリカやオーストラリアの評価協会)。評価の質を導き、あるいはレビューするための運営グループや完全に独立した外部のグループ(例えば、フランスの the Sientific Council of Evaluation)を結成することは、有用である。それに代わるものとしては、そう望むのであれば、会計検査院が評価のレビューについて責任あるいは権限を持つかもしれない(例えば、スウェーデンの国家会計検査院)。レビューは、内部評価の信頼性と質を高めるために極めて重要である。

課題 13

判断と提言の作成

評価において判断することと提言することの価値については、異なるいくつかの観点がある。判断は、Commissioner や評価の利用者に委ねられるべきだと考える人達もいる。評価の結論が一般的であればあるほど、違った問題に光をあてるために評価を利用することが容易になり、議論や学習方法の口火となる。

一方、判断および提言は、注意をひき、議論を引き起こし、議論を概説し、下すべき決定や行うべき行動を強く主張することが、示唆されるかもしれない。意思決定責任者は、時間が限られていることが多く、情報を背負い過ぎているかもしれない。調査結果や提言、行動の主な選択肢の概要を明確にすることは、彼らの注意を集め、その後の行動を刺激しそうである。また、単にプログラムの失敗に焦点を合わせるより、改善を示唆する建設的な方法は、抵抗を制限し、世論に必要な変化を起こすことに寄与するであろう。いずれにせよ、判断と提言が作られた基準を明らかにする必要がある。提言は、たった一点に焦点を置くよりは代替解決案を提示することでバランスを取ることができる。評価の提言に基づく実施計画を提出することもまた有益であろう。

課題 14

調査結果の伝達

評価結果の伝達に全く注意がはられなければ、多くの評価も何の効果もない。評価結果が効果的かつタイムリーなやり方で提出されることは不可欠である。

情報政策は、本来的にさまざまである。しかし評価の信頼性と利用は、しばしば開放性と透明性によって拡大される。オーストラリアでは、評価結果を誰でも利用できるようにすることが求められている。スウェーデンでは、一般市民は、伝統的に、評価報告を含むすべての政府資料にアクセスしてきた。評価結果は、特に批判的あるいは否定的な結論の場合には、政治的に敏感なことがよくある。調査結果が公表される前に調査結果に関する情報を前もって評価された者に示すことによって、不必要な論争は避けられるかもしれない。

一般的に言えば、開放性は、有用である。しかしながら、特定の状況によっては、透明性は、評価の使用法に積極的あるいは否定的に影響を及ぼすかもしれない。公表することは、外部からの圧力を作ることとなり、評価の提言の実施を促すかもしれない。しかし、批判された者のあいだに抵抗を生むかもしれず、彼らの協力しようとする気持ちや行動しようとする気持ちを削ぐことになるかも知れない。このことは、非難めいた口調を避けること、また学習のカルチャーを促進することによって、限定されるかもしれない。短所を報告することに集中する代わりに、評価は、指摘された問題を克服するために何がなされたかを報告すべきである。しかしながら、大事な点は、ケースによっては、公表することや、特にマスコミの目にさらされることが逆効果となる、ということかもしれない。

先に述べたように、評価を報告するタイミングは、意思決定のスケジュールに合わせるという観点において重要である。しばしば、いくつかの結果を前もって利用可能にすることが有用である。提言をするかどうかということは、結論を出すかどうかということと同じく、決定されなければならない。経験によれば、政治的な関連を伴う評価は、新しい政府が批判や提言を受け取ろうという気持ちがより強くなる選挙の直後に受け取られるのがよい場合がしばしばである。

正式な報告書を公表することは、評価結果を提示する最も一般的な方法である。報告書が明確であり、包括的で、幅広く利用できるものであることが必要不可欠である。評価報告の組み立てと主張は、わかりやすくなければならない。すべての stakeholder が報告書を全部読むことは考えにくいので、実務的な要約(executive summary)は有用である。様々な stakeholder のグループのために報告書の要約版を出版することも、意義がある場合があるだろう。

通常、報告書を出すだけでは不十分である。評価の結果に基づく意見交換をさせるために、そ

の後の説明や普及活動が必要である。様々な stakeholder とのミーティングが評価結果を得て、連携して調べさせることとなり、提言の実施を促進する。鍵となる意思決定責任者に対しては、特別な注意が払われるべきである。セミナー、ワークショップ、グループディスカッションも行われる。

課題 15

モニタリングまたは追跡調査

評価の影響を監視することは、評価の利用を促進し、評価の実行プロセスの手引きとなり、将来の評価の実行の改善に役立つ。追跡調査(follow-up)もまた、追加して評価する必要性の可能性を見極めるのに重要である。

追跡調査は、進行中のすべての評価の利用をレビューする、あるいはその後の追跡調査の必要条件を各評価に組み入れる制度的な枠組みによって容易になる。評価の使用法あるいは影響について何らかの結論が下される前に評価の実行のための時間を十分に取ることは、大切である。

評価結果を利用しない、あるいは提言に従わないことに対する制裁は、典型的には公表と催促に限られている。しかしながら、単なる一般の注目が、プログラムの変更を実行させる圧力を作ることもあるかもしれない。。

評価結果、および提言の遂行は、しばしば、サポート、および専門的な助言を必要とする。これは、財政的および技術的な支援によって容易になる。

課題 16

スタッフのトレーニング、およびサポートの必要性の認識

評価が成功するために、訓練された評価者、情報通の commissioners、そして啓発された利用者が必要である。多くの国では、公務員の学歴に、特別な評価スキルの用意はない。それゆえ トレーニングが重要なのであり、特に評価が導入され進展していくときには重要なのである。

方法の面、政治的な面、管理面、そしてコミュニケーション面についてのスキルは、必要である。commissionersと評価の利用者は、評価の本質的な知識を正しく頭に入れる必要があり、その結果、関連する質問をはっきり述べ、評価の目標を進展させ、プロジェクトを管理し、評価者を選択することができるようになる。彼等はまた、評価には、意思決定の改善の可能性とともに、限界があることも理解しなければならない。技術的な評価スキルに加え、現在進行中の公共活動のレビューの受けとめ方も重要である。外部評価者は、評価が行われる管理的政治的な範疇を理解

しようとする必要がある。

トレーニングおよびサポートが準備できるには

- ・ 機関のなかに、政府全体に、そして異なる分野の評価者と会計検査官との間にネットワークを作る。(EC の評価ネットワークなどのピアグループネットワーク、あるいはヨーロッパ評価協会、オーストラリア評価協会、アメリカ評価アソシエーション、イギリスの評価協会等のより正式なアソシエーション)
- ・ ハンドブック、優良事例集、ニュースレター、ジャーナルなどの出版
- ・ セミナー、ワークショップ、研修コース、および視察などの準備
- ・ 実験的プロジェクトの設定
- ・ ヘルプデスクまたは類似のもの、行政庁での設置

4. 結論

認識論的な考察と実際の考察は、評価においては調和されなければならない。健全な評価能力と調査結果を利用しようとする意思の間に調和が取れていることは、必要不可欠である。そのとき初めてプログラム評価が意思決定に影響を与えることができるのである(Mayne , Divorski , Lemaire , 1998 ; Boyle , 1997)。これらの必要条件は互いに依存しており、その結果一方を改善することがもう一方を促進するのである。これには、政治的意思と体系的な作業が必要である。

評価情報の使用法は、政策サイクルを通して適用され、単に、プログラムが施行された直後とか、政策の定期的なレビューの時期が来たからというときだけではないことが認められている。重要なことは、評価が提供しなければならないものを最大限に活用するために、政府がどのように評価活動を計画するか、ということである。組織的な学習は、通常、体系的な行動ではないが、評価はそれにかなり貢献することができる

評価の使用法のために厳密な枠組みを設定する必要はない。評価の目標および方法は様々で、評価結果の利用も様々である。機関の中での評価の主要な影響は、直接的なプログラム改善もまた求められるべきではあるけれども、組織管理への長期的な影響としてよく述べられている。

評価は、すべての決定に際して適切な訳ではない。ときには評価は、実用にはあまりにも費用がかかり、あるいは時間がかかる。ときには、活動のための選択肢がなにもないこともあるし、また、変化しようとする政治的意思が存在しないこともある。またときには、他の場所ですでに決定がなされてしまっている場合もある。評価が正しく使われないこともある。すでに行われた政治的決定の正当化のためだけに行われることもあるかもしれない。都合の悪い、あるいは政治的に微妙な調査結果がときには隠されたり、また、結果が選択されて利用されることもある。利用者を参加させることは評価の重要な部分ではあるが、評価が依頼者に分捕られてしまう危険性も常に持っている。

こうした束縛にも関わらず、評価結果の利用を増やすために不可欠な前提条件は、政策決定の改善に評価が使われると信じている人々の関与と励ましである。彼らは質の良い評価情報を作り出し続け、利用者とともに働き続け、常に変化している公共部門の環境下で適用し続ける、ということが必要なのである。

評価結果の利用例

評価の実際の利用、および効果的な評価戦略の示唆の例が、いくつか以下に挙げてある。評価結果が実際に利用される程度は、二つに解釈される。人は、評価結果のすべてが使われた訳ではなかったことに失望を感じるか、あるいは、評価の主要な部分は意思決定に影響を与えたことに満足を感じるか、どちらかである。後者の視点がおそらくより妥当なものであろう。

スウェーデン

評価の使用法についての包括的な研究がまだなされていないが、いくつかの所見を述べることはできる。評価の90%は説明資料として使われており、行動の基本的な再査定の基礎としてというよりも、主に政策の調整に利用されている。

長期的な学習効果はあるかもしれないが、より基礎的な課題を予算編成のプロセスに導入する意図は認められない。しかしながら、予算の概況説明資料としての質は、改善してきている。エージェンシーの年次報告には、主に運営上のデータが含まれ、主として内部管理に役立っており、また各省のコントロールをサポートしている。議会は、評価のことを最も知らない一団のままである。(Furubo & Sandahl , 1993)

カナダ

評価の政策および基盤は、20年以上にわたって開発されてきた。財政面その他での束縛はあるが、評価の継続的な利用は、その実用性の重要な指標である。極めて多くの評価が資源の再配分につながり、そして大多数の評価が、少なくともある程度利用されている。部局の評価計画と、評価委員会を組み合わせることが、評価の見落としをし難くしている。

研究によれば、評価の約 20%はプログラムの根本的な改革や終了に、40%はプログラムの修正に、30%はプログラムの理解を深めることに、そして残りの10%はプログラムの確認につながった。会計検査院によって行われた 1993 年の研究では、管理を助けるために最も重要な立場にある評価マネージャーが、運営上の問題を解決し、プログラムを改善しているということが分かった。資源配分を改善するプログラムに挑戦したり、議会にアカウンタビリティを提供するために大きいプログラムを評価することには、優先度ははるかに低くおかれている。(Mayne , 1994) .

オランダ

会計検査院によって行われた、評価の範囲、実践および利用の調査の研究において、インタビューされた公務員の80%以上が、評価結果は直接に利用されていると述べた。いくつかの評価をより綿密に吟味してみると、42の研究のうち20がはっきりと利用されていた。すなわち、利用した証拠が書き残されてあった。ほんの一部の評価だけが重要な変更を提案している。(Leeuw & Rozendal , 1994) .

効果的な評価戦略のための提案の例

アイルランド

アイルランドは、体系的な評価活動を開始しようとしているところである。Boyle によると (1997), 他の国からの教訓によれば、評価者の広がりが必要であると述べられている。(行政の中の半独立的な評価団体が良いモデルであろう)。評価の優先度が設定される必要がある。3年ごとに行われる包括的なレビューは、初期段階では難しすぎるかもしれない。利用者の見通しは、可能であれば、含まなければならない。トレーニングとサポートの開発が評価者と利用者が必要で、「ハード」と「ソフト」の両方のスキルが重要である。効果的な需要の蓄積 (sticks carrots と sermons、資源の指定、評価実施者による質問の具体化) が必要である。最終的には、予算編成、戦略的管理、および評価の間の連結が確立される必要がある。

アメリカ合衆国

Richard C. Sonnichsen は、評価情報の使用に役に立つ効果的な内部評価部局のためのモデルをデザインした。このモデルは、中央政府の 5 つの内部評価部局の調査と、彼自身の個人的経験に基づいている。効果的な内部評価部局の組織的な条件は、十分な資源、独立性、上級機関への報告、評価部局とその機関の間での政策と目標についての合意、組織の中での評価部局の長の十分な位置付け、組織的な評価の経験者、そして評価者がその機関のなかで信頼性、独立性、客観性がある者として受け入れられることである。

評価のプロセスの要素としては、部局の役割の明確化、協議事項の公表、自発的な評価の始動、効果的な報告書と評価結果の機関内への配布、提言、追跡調査のプロセスへの適用、結果と提言の公表にあたり評価スタッフが支援活動の役割を担うことが含まれる。これらのすべてが、組織的な条件、評価のスキル、および評価を成功させるのに必要な管理面での姿勢を適切に組み合わせる前提となる。(Sonnichsen, 1994)

オーストラリア

現在までのオーストラリアの経験は、評価のいくつかの重要な成功要素（最終結果であるアウトカムと調査結果を重視するいくつかの公共部門の改革、議会に支持された正式な評価の必要条件（業績管理への原則ベースのアプローチの正しい実践のために時間を割く）、何人かの主要大臣のサポート、大蔵省への明確な命令と評価の支援活動、省の長の支持、評価には価値があるとする十分な推進力と世論、評価者のネットワーク、アドバイスを提供する省内の中心的評価ユニット、立案、コーディネート、品質管理、健全な業績情報システム）を示している。（Mackay，1996）。