

諸外国における政策評価のチェックシステムに
関する調査研究
—報告書—

平成 21 年 3 月

総務省行政評価局

委託先：財団法人行政管理研究センター

本報告書は、平成 20 年度において、財団法人行政管理研究センターに委託して実施した調査研究の成果であり、本文中の見解にわたる部分は執筆者のものであって、総務省としての見解を示したものではない。

は し が き

諸外国においては、各府省の政策評価結果について、第三者的機関が検証したり、当該機関職員自らが評価を行ったりと、様々な形でチェックが行われている。このようなシステムについて実施の手順や視点などの把握は必ずしも十分ではない。

我が国において政策評価法に基づき行政評価局が行っている客観性担保評価活動を今後より充実させるための検討に当たって、各国における政策評価のチェック体制がどのような体制で、どのような基準で、どのような手順で、どのような視点で行われているのかなどを踏まえることは有益である。また、そのようなチェックを行うに当たって不可欠となる評価関連データの入手について、実際の各府省におけるデータの公表状況や公表されていないデータの入手手順について把握することも我が国の客観性担保評価活動に資すると考えられる。

したがって、本調査研究では、米国、英国、ニュージーランドの3カ国を取り上げ、各国でどのようなチェックシステムを構築しているかについての調査を行った。

本調査研究の実施に当たっては、財団法人行政管理研究センターに委託し、同センターでは、政策評価に関する有識者からなる研究委員会を設置して調査研究の実施、報告書の執筆に当たった。その構成委員及び執筆分担は次のとおりである。

- | | | |
|-----|-------|----------------------------|
| 委員長 | 山谷 清志 | (第5章担当・同志社大学政策学部教授) |
| 委員 | 佐々木 亮 | (第2章担当・財団法人国際開発センター主任研究員) |
| 委員 | 南島 和久 | (第3章担当・神戸学院大学法学部准教授) |
| 委員 | 源 由理子 | (第1章担当・明治大学大学院ガバナンス研究科准教授) |
| 委員 | 和田 明子 | (第4章担当・東北公益文科大学公益学部准教授) |

*委員は、五十音順

また、本調査研究の事務局として、財団法人行政管理研究センター行政研究所の武藤桂一主任研究員、鹿谷雄一研究員、馬渡剛研究員が参加した。

本調査研究の結果が、各方面において広く活用されれば幸いである。

平成21年3月
総務省行政評価局

目 次

第1章 評価の評価（メタ評価）－その概念整理	源 由理子 …… 1
第2章 アメリカにおけるメタ評価の現状	佐々木 亮 …… 27
第3章 イギリスにおける政策評価のチェックシステム －PSAシステムに対するチェックシステムを中心として－	南島 和久 …… 55
第4章 ニュージーランドにおける政策評価のチェックシステム	和田 明子 …… 75
第5章 政策評価の「チェックシステム」の可能性	山谷 清志 …… 103
米・英・NZのメタ評価に関する整理表 ……	123
参考文献一覧 ……	125

第1章 評価の評価（メタ評価）－その概念整理

源 由理子

1 はじめに

日本において行政機関が行う政策の評価に関する法律（平成13年法律第86号。以下「政策評価法」）が施行されてから7年たつ。「政策評価に関する基本方針」（平成17年12月16日閣議決定）では、政策を所管する各府省が自ら行う政策評価の目的は「その結果を政策に適切に反映させ、政策に不断の見直しや改善を加え、もって、効率的で質の高い行政及び成果重視の行政を推進するとともに、国民に対する行政の説明責任（アカウンタビリティ）を徹底するもの」と位置づけられている。そのためには、それら政策評価が適切な方法で客観的な価値判断が伴う評価である必要があるとして、その客観性を担保する役割のひとつを評価専担組織としての総務省が担っている。総務省は各府省による自己評価の「客観的かつ厳格な実施を担保するための評価」（政策評価法第12条第2項）を行うとして、①政策評価の実施形式について確保されるべき客観性・厳格性の達成水準等に関する審査と、②改めて政策評価が行われるべきものに関する評価実施の認定等の活動に取り組んでいる。これらは「点検活動」と呼ばれ、平成19年度に対象となった件数は計16府省の3,850件にのぼっている（総務省行政評価局2008:4）。これらの点検活動は、総務省が各府省の評価をさらに評価するという意味において「評価の評価」（meta-evaluation メタ評価）であり、総務省がメタ評価者として各府省の一次評価の質を確保するための評価活動を行うものである。

メタ評価という言葉はアメリカの評価研究の分野で生まれたもので、著名な評価研究者であるスクリヴェンが1969年に *Educational Product Reports* でこの言葉を最初に紹介したとされる（Stufflebeam & Shinkfield 2007: 649）。彼は「メタ評価は、一次評価者の責務である」とし、特に「評価結果が人々の福祉に影響を与えるときに倫理的かつ科学的な責務として行うもの」と定義づけている（Scriven 1991: 228）。またその後メタ評価の理論を発展させたスタッフルビームは、「評価者は自らの評価の品質を確保するためにメタ評価を受けなければならない」（Stufflebeam 2001: 183-209）とし、評価の品質確保の手段としてのメタ評価の役割を述べている。評価の品質は特に内部評価に対して問われるこ

とが多い。内部評価は客観性に欠けお手盛り評価になりがちなので外部のチェックが必要というのは十分に説得力があるのだが、広義のメタ評価は一次評価者が良い評価の基準とされる項目にそって自らがチェックをするという行為も含まれており、その項目は必ずしも厳格な評価設計のみを対象としているわけではない。そもそもメタ評価登場の背景には、一次評価だけでは本来の評価の目的遂行が難しいかもしれないという「評価行為」そのものが持つ危うさのようなものが関係しているのではないか。またスクリヴェンは、「評価者を誰が評価するのか？」の問いかけのもと「メタ評価はコンサルタント版ピア・レビュー (peer review) である」(Scriven 2009: iii-viii) と述べ、当然の手続きだとしている。

そこで本章では、諸外国における政策評価のチェックシステムに関する調査研究の導入部分として主にアメリカにおいて発展してきたメタ評価の概念整理を試みる。本稿では必ずしも総務省が行っている点検活動そのものを論じることを目的とはしていない。むしろ、評価論と評価技術の見地から現行の制度への何らかの示唆を得ることを目的とするものである。

2 プログラム評価とメタ評価

(1) プログラム評価におけるメタ評価の登場

メタ評価という言葉は前節で触れたようにスクリヴェンが教育のプログラム評価の分野で使い始めた言葉である。スクリヴェンは、「バイアスのかかった客観性に欠けた評価情報は、購入する価値がない教育資材を人々に押し付けることになり、その結果、子供や青年の成長に不利益をもたらすことになる」(Stufflebeam & Shinkfield 2007: 651) として、評価情報そのものが評価される必要性を説いている。またスタッフルビームは、「評価者 (evaluator) は評価のプロたるためには、メタ評価を活用し自らの評価を改善し、より適切な情報を顧客 (client) に提供する責務がある」(Stufflebeam & Shinkfield 2007: 649) とし、メタ評価を活用して評価の品質を高めることで評価者としての信用度を高める必要があるとしている。評価情報を使う顧客の側から見るとメタ評価を活用して質の高い評価情報を得たいという要求があり、また評価される対象 (たとえば教育評価ではプログラムのスタッフ、学生、教員など) は評価基準にそった客観的な評価をしてもらいたい権利があると主張する。このようにアメリカで発展したメタ評価理論の前提にあるのは、評価者はプログラム評価の専門家であり (professional evaluator)、評価情報のユーザー (顧客や国

民)に間違った情報を提供することにより生じる不利益への懸念から、評価の品質を保証するための取り組みを行うこと自体が評価者の責務と考えられていることである。したがって、評価の品質を確保するために、自らがメタ評価を行うことも重要だとされている。そこには評価専門家としての責任と評価情報のユーザーである顧客の視点の二つが併存していることに留意が必要である。

ここで、メタ評価がプログラム評価 (program evaluation) において登場したことに注目したい。プログラム評価とは、「ある社会介入 (ある政策のもと実施される施策や事業。以降、本稿では「プログラム」と呼ぶ) の計画、実施プロセス、インパクトならびに効率性を社会科学の調査研究手法を使って体系的に査定をしていくこと」(Rossi et al. 1999) で、1960年代半ばから特に社会政策を中心に発展してきたものである。1960年代のアメリカといえばケネディ、ジョンソン政権が、民権運動の高まりの中、失業・貧困問題に対処するため「貧困との戦い (the War on the Poverty)」政策のもと各種の社会・福祉プログラムを展開した時期である。それらのプログラムの有効性をチェックし、その結果をもってアカウントビリティの確保をめざしたのがプログラム評価の手法であり、1967年には会計検査院 (GAO) がプログラム評価を採用し有効性評価を開始した。時期を同じくして登場したメタ評価が、教育、福祉分野等のサービスの受け手である「顧客」の立場を重視していること、また社会科学分野の専門知識を前提としたプロフェッショナルによる評価を前提としていることは、このような社会状況を反映したものであるといえよう。プログラム評価は、ある社会介入が本来受益すべき人々に良い変化をもたらしているのかどうかという「評価情報」を生み出すことが目的で、その評価情報を生み出す過程や評価情報の質が適切であるかどうかを点検するのがメタ評価の役割ということになる。

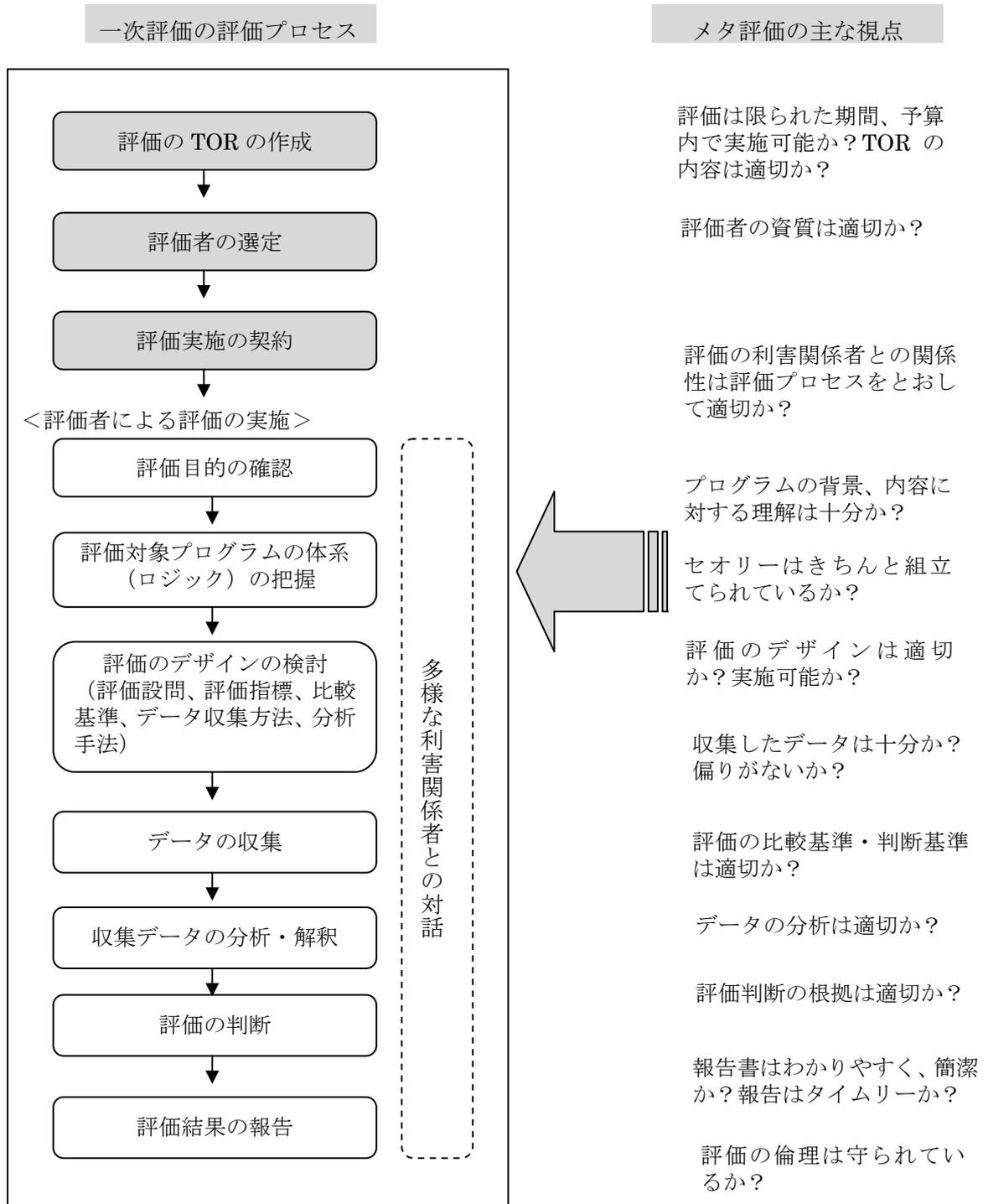
(2) プログラム評価の実施過程とメタ評価の視点

プログラム評価は評価対象となる政策やプログラムの内容を詳細に見ていくため、そのプロセスで様々なステークホルダーが関わる。たとえば、そもそもそのプログラムを計画したマネジメントとプログラムを現場で管理する人、実施に携わる人は異なることが多い。またそのほかにも、プログラムの受益者、資金を負担する納税者などがある。これらの人々のプログラムに対する関心は様々で、異なった関心に対応した評価を行うためにはそれなりのプロセスを踏むことが要求される。プログラム評価の評価プロセスでは、評価目的の確認、評価対象プログラムの内容把握、評価設問の設定、評価計画の策定、評価指

標・評価基準の検討、データ収集・分析、評価結果の報告といった一連の作業があるが、そのプロセスでは関係者との対話といったコミュニケーションが重要になる。プログラムで何が起きているのかに関する‘対話’は評価者が評価対象を正確に把握するためには不可欠な情報であるし、何を評価に期待しているのだろうかに関する‘対話’は、評価設問に反映されるであろう。また誰を、あるいは何を情報源とすることがより適切であろうかという‘対話’はデータ収集の計画作業に結びつく。これら一連の作業が適切に行われてこそ「質の高い評価」となり、関係者の役に立つ評価情報が提供できることになる。

また、多様なステークホルダーが関係する中で、あるプログラムの良し悪しに関する情報を提供する評価には、「評価者の倫理」も質の高い評価の必須条件である。全米評価学会では **Guiding Principles for Evaluators**（評価者のための指針）として評価者が遵守すべき基準を5つあげている。すなわち①体系的な調査の実施、②高い評価能力の提供、③誠実さ／正直さ、④人々に対する敬意、⑤公共的福祉に対する責任、である。メタ評価ではこれら評価者の倫理も含め、評価の実施プロセスがきちんと行われているかどうかを点検することが求められるのである（メタ評価の基準については後述。メタ評価の視点の概略を図1に示す）。

図1 プログラム評価の評価プロセスに対するメタ評価の視点



(出典) 筆者作成

注：一次評価の評価プロセスに関し、事業担当者による自己評価の場合は最初の三つのステップは該当しない。

TOR (terms of reference)：委託のための条件を定めた文書

プログラム評価に関してみる限り、評価行為の危うさというのは、このような評価のプロセスに関わる人々の多様な関心に対峙するという‘アート’の部分と、評価指標・評価基準やデータ収集・分析方法の検討を含む適切な評価軸の設定といった‘サイエンス’の部分の融合過程にあるのではないかと思われる¹。ここでいう‘危うさ’とは、「正しい評価」とされるためには多くの視点からの検証が必要であり、評価をしてもその結果が活用されない、あるいは的はずれた評価で終わるといった危険性があることを意味する。たとえば、評価はある価値や基準に沿って良し悪しを判断する根拠となる情報を提供することであるが、多様な利害関係者の誰の価値を優先するかによって評価結果の「正しさ」は異なってくるし、評価の目的によってどのようなデータを収集するのが「正しいか」も異なる。また、価値判断が伴うため評価される対象となった人々への配慮（守秘義務、誠実な対応等）も必要となる。方程式どおりの手順と方法で調査・分析を行えば適切な評価になるというわけでは必ずしもないのである。そのような危うさを補うのがメタ評価の基準と捉えることができるが、もっといえばその基準の基となる「質の高い評価」のガイドラインがまずあり、評価者がそれを遵守する、あるいはそれに従って評価を行えるだけの能力が備わっていること（＝プロの評価者）がまず必要ということになる。

3 メタ評価の基準

さて、メタ評価の対象となる評価事例は、ひとつである場合と複数の評価を統合的に扱う場合の二つのケースがある。ひとつの事例を対象とするメタ評価は、主に評価デザインの改善や評価結果の信頼性に焦点を当てて行うもので、当該評価の品質を確保するためのパワフルな手段として活用できる（Cooksy & Garacelli 2005: 31-32）。複数の評価をメタ評価する方法は評価統合（evaluation synthesis）と呼ばれ、類似のプログラムの評価結果を複数集めたメタ評価であり、その目的は、①ある社会介入に対する総合的な評価を行うこと、②一次評価で使われた評価方法の強みや弱みを分析することの二つであるとされる（Cooksy & Garacelli 2005: 32）。ただし、評価統合が厳密な意味でメタ評価といえるかどうかについては議論がある。ワイスは「複数の評価を統合してプログラム全体の効果

¹ 山谷は、政策評価は現場の実践的な「知」と「技（art）」の熟練がなければ不可能な作業であるとしている（山谷 2006: 7）。また、パットンも評価におけるアートとサイエンスのバランスの重要性を論じている（Patton 1997）。

を見る方法は、メタ評価というよりもメタ分析である」(Weiss 1998: 236) とし、本来のメタ評価が意味する「評価の評価」という要素は限定的としている²。

本節では、主にひとつの事例を対象とするメタ評価に焦点を当て、評価の品質確保のためにはどのような評価基準があり得るのかを事例をとおして考察する。

(1) プログラム評価のメタ評価チェックリスト (アメリカ)

アメリカで広く用いられている Program Evaluation Standards (PES) (プログラム評価基準) に基づいた Program Evaluation Models Meta-evaluation Checklist (1999 年) (プログラム評価のメタ評価チェックリスト) (Stufflebeam 1999) を参考に、「評価の有用性 (utility)」、「評価の実施可能性 (feasibility)」、「評価の正当性 (propriety)」、「評価の正確性 (accuracy)」の 4 つの基準を取り上げる。まず「評価の有用性」を見る項目は大きく分けて①評価者の資質に関するもの、②評価計画の妥当性に関するもの、③評価報告方法と評価のインパクトに関するものに分類できる。評価者の資質については、活用し得る質の高い評価者であるかどうか問われる。評価能力・専門性はもちろんのこと評価対象の利害関係者との関係を適切に保てることも重要なポイントになる。評価計画の妥当性は、そもそも評価計画に問題があれば活用し得る評価結果は得られないので、評価対象情報の範囲および評価設問や評価基準 (指標や水準) の妥当性を問うことになる。また、評価結果の有用性を高めるためには、評価結果をどのように伝えるのかがそのタイミングも含め重要である。報告書の書きぶりや報告方法がチェックされる。評価結果をどのように使っているのかといった「評価のインパクト」も基準としてあげられている。そもそも評価結果は使われなければ意味がないという前提のもと、この「評価の有用性」の観点から評価の品質基準としてある。

次に「評価の実施可能性」の基準では、評価の実施手続き、実施スケジュール、多様な利害関係者への配慮、政治的な排除への対応といった評価の実施が可能になるような配慮があげられる。また、評価の費用対効果の視点も含まれ、効率的にかつ遅れないように評価を実施することが求められる。いくら質の高い評価計画を策定できたとしても、情報収集が経費や時間の関係で実現できなかつたり、その過程で政治的な介入があった場合は

² 評価統合の事例として、米国エネルギー省の「耐気候構造化補助プログラム」のケースを巻末資料 1 にまとめたので参照ありたい。なお、第 2 章においても米国会計検査院による評価統合のプロセスがまとめられている。

評価の品質は確保できないのである。

三つめの品質基準である「評価の正当性」は、①プログラムの受益者ニーズの視点、②評価の契約内容、③評価倫理の遵守、④公正な価値判断、⑤評価結果の公開、⑥利害の衝突への対応、⑦評価予算の適切な管理などを含み、いわば評価関係者が評価実施にあたり守るべき基準をあらわしている。たとえ質の高い評価計画が実行できたとしても、その過程で情報源である人々に不利益がもたらされたり、偏向に満ちた価値判断が行われた場合はけっして質の高い評価とは言えないのである。

最後の基準である「評価の正確性」では、正確で厳格な評価を行っているかどうかを①評価対象プログラムの分析、②評価プロセスにおける軌道修正、③情報・データの収集方法と内容、④情報・データの分析方法（定量、定性分析）、⑤評価結論から検証する。また当該評価がメタ評価を行っているかどうかも評価の正確さを問う基準となっており、メタ評価が評価の一部として位置づけられている点が興味深い。以上述べたメタ評価の品質基準をまとめたものを表1に示す。これらの品質基準は、言い換えれば、質の高いプログラム評価を行うための行動指針でもある。

このチェックリストの特徴は、評価の活動ごとにチェック項目を設定するのではなく、評価活動全体を有用性、実施可能性、正当性、正確性という四つの評価基準で横断的にとらえた点であろう。評価基準のチェック項目ごとに詳細項目が10個ずつ設定され、チェック数によってレーティングされ（例：9-10 Excellent ～ 0-2 Poor）総合得点が算出される。

なお、参考までに、国連機関、援助機関、NGOなどが参加し、人道援助プログラムの評価を専門的に実施するネットワーク組織であるALNAP（the Active Learning Network for Accountability and Performance・パフォーマンスとアカウンタビリティのための学習ネットワーク）のメタ評価の基準（The Quality Performance: QP）を表2にまとめた。これは、ALNAP自身が、各参加機関が実施している評価を評価する基準として開発されたものである。QPはPESのものと異なり、評価活動のステップごとにチェックリストを構成している。人道援助プログラムという特徴から、援助の背景となる対象社会の問題の所在とその背景、それまでの実施機関の関わり方などに対するきちんとした調査が行われているかどうかが含まれている。評価者の評価対象とその背景への理解は、社会や顧客のニーズにかなった評価となっているか、つまり役に立つ評価かどうかといった観点から重視される。ALNAPによると、もともとはプログラム評価を念頭に開発されたものであるが、

評価項目の中にはプログラムのマネジメントのメタ評価にも使えるものが含まれており、必要な項目だけを抽出してメタ評価の目的に合った使い方ができるとしている（ALNAP 2005）。

表1と表2を比較してわかるとおり、一次評価の対象となっているプログラムの特徴や評価が行われる環境（社会・制度・文化等）によってメタ評価項目やその構成は少し異なっているが、両事例とも、評価実施の手続きや厳格な評価手法に加えて、評価者やTORの質、プログラムに対する理解度、利害関係者との関係性、報告書の質などの共通項目が含まれている。

表1 メタ評価の基準例（1） アメリカのプログラム評価メタ評価項目

評価基準	主なチェック項目
評価の有用性 (Utility)	評価者の資質
	利害関係者との関係性
	評価対象情報の範囲と選定の妥当性
	評価設問、評価基準の妥当性
	報告書の質、タイミング
	評価結果のインパクトの大きさ
評価の実施可能性 (Feasibility)	評価の実施手続きの妥当性、実施日程の妥当性
	多様な利害関係者への配慮、政治的排除への対応
	評価の費用対効果
評価の正当性 (Propriety)	プログラムの受益者ニーズの明確化、サービスの質の評価の適正度
	契約内容の適正度
	評価倫理の遵守
	公平な価値判断
	評価結果の公開度
	利害対立に対する配慮
	予算執行の適切性
評価の正確性 (Accuracy)	評価対象プログラムの内容把握の適切性、内容分析の質
	評価の軌道修正の適切性
	情報・データ収集方法の適切性
	収集情報・データの質、信頼性
	収集情報・データの分析方法（定性・定量）の正しさ
	評価結論とその根拠の妥当性
	メタ評価の仕組みの有無

（出典）Stufflebeam（1999）から抜粋・作成

表2 メタ評価の基準例(2) ALNAPのメタ評価基準

評価基準	主なチェック項目(各項目の適切性を評価)
Section1: TOR、評価のタイミング、 評価チームの適切性	<ul style="list-style-type: none"> 評価のコスト TORの内容(評価対象の記述、評価の焦点、評価方法、評価結果の活用方法、評価報告書のフォーマット、評価の実施機関) TORの合意過程における関係者の話し合い 評価メンバー選定のプロセス、評価メンバーの資質
Section2: 評価アプローチと評価方法 の妥当性	<ul style="list-style-type: none"> 評価アプローチ 評価方法、評価基準 制約に対する対応 利害関係者への配慮(守秘義務、尊厳)
Section3: 背景の分析の適切性	<ul style="list-style-type: none"> 対象地域の課題の分析、その評価分析への活用 実施機関の対象地域におけるオペレーションの分析
Section4: 評価対象プログラムの分析 内容の適切性	<ul style="list-style-type: none"> プログラム実施機関の政策、戦略の評価 プログラムのニーズ、問題の所存の分析 プログラムの目的の確認 プログラムの計画過程、実施過程 プログラムのモニタリング体制 プログラムの予算 人道主義プログラムの横断的課題に対する配慮(国際基準、調整機能、ジェンダー、保護活動、弱者への配慮)
Section5: 評価活動の適切性	<ul style="list-style-type: none"> 地域の現状を反映した評価分析 2次データの活用 受益者を含むステークホルダーとの対話 評価手法 評価項目ごとの根拠の提示(*1) 人道援助の活動に関する国際基準との比較 評価結果、提言の質
Section6: 報告書の質	<ul style="list-style-type: none"> 報告書のカバレッジ 報告書のフォーマット 報告書の入手経路 報告書の要約
Section7: 全体評価	(上記項目以外の課題、評価全体の良い点、悪い点等を含む全体的なコメント)

(出典) ALNAP (2005) から抜粋・作成

注(*1)「評価項目ごとの根拠の提示」とは、ALNAP で決めている一次評価の項目である、効率性 (efficiency)、有効性 (effectiveness)、インパクト (impact)、妥当性 (relevance/appropriateness)、持続性 (sustainability/connectedness)、網羅性 (coverage)、政策の一貫性 (coherence) をさす。

(2) 日本国内の事例－国際協力機構の二次評価票

日本国内においてメタ評価を実施している事例として、日本の政府開発援助を実施している国際協力機構（以下、JICA）の二次評価をとりあげたい。JICA のプロジェクト評価手法は、アメリカのプログラム評価の理論をベースに開発されてきた。JICA では、それらプロジェクト評価の結果の透明性と客観性を高めるために、2003 年度から、JICA が実施したプロジェクト終了時評価の評価（二次評価）を外部有識者事業評価委員会において実施している³。この仕組みは、有識者は外部であるが、JICA 自らが評価の品質を確保するために組織の内部にその機能を位置づけている例である。ただし、二次評価の内容は大きく分けてふたつに分かれ、評価票はメタ評価の項目と再評価の項目から構成される。本来メタ評価と再評価はその性質が異なるが、JICA では同委員会の外部性と専門性を活用して報告書から読み取れる再評価をとおしてプロジェクトの形成見直しへの提言を抽出したいという意図があるようである（国際協力機構 2008:80）。

評価票のうち、メタ評価に関する項目を表 3 にまとめた。JICA のチェック項目で特徴的なのは、はじめに「評価可能性」の評価がある点だ。評価可能性評価（*Evaluability Assessment*）は、評価対象となっているプログラムが評価可能であるかどうかを事前に審査し、活動の背景とプログラムがめざすものがロジカルに計画されているかどうかを確認する作業である。そもそもそれがきちんと設計されていなければその後の評価活動自体に支障をきたすことになる。終了時評価のメタ評価で評価可能性のチェックをすることにより、事前評価の質や、その後のモニタリング時でプロジェクトの見直しが適切に行われてきたのかの問題（マネジメント）に対する検証が可能になる。

「Ⅱ 項目ごとの二次評価」に含まれている各チェック項目は、JICA が 2004 年 2 月に作成した「JICA 事業評価ガイドライン」の中で説明されている項目に沿っており、チェック項目ごとの「適切性」や「明確性」が具体的に何を意味するのかは JICA の評価に従事する関係者間で共有されていると考えても良い（少なくともガイドラインに沿った評価が徹底されているという意味において）。メタ評価を行う前提として、メタ評価される側が何をもって「質の高い評価」とするのかの共通認識がない限り、形式的な評価に終わる危険がある。

³ 二次評価の対象案件数は年間約 70 案件にのぼるため、作業部会の業務を日本評価学会に業務委託して実施している。なお作業メンバーは学会内で公募により選り透明性を高めている。また JICA のプロジェクト評価は総務省の「事業評価」とは異なり、プログラム評価の手法を適用している。

表3 メタ評価の基準例(3) 国際協力機構(JICA) 二次評価の主なチェック項目

I 適切な評価が可能であったかどうかの前提条件(評価可能性)	
<ul style="list-style-type: none"> ● 事業計画(事前評価/PDM)の適切性(*1) ● プロジェクトのターゲットグループの設定 ● プロジェクト目標の達成度(実績)の検証可能性 ● 上位目標の達成度(実績)の検証可能性 ● プロジェクトの組み立て(因果関係)の検証可能性 ● 実績・実施プロセスの検証可能性 	
II 項目ごとの二次評価	
1. 評価フレームワーク	<ul style="list-style-type: none"> ▪ 調査団構成(中立性・公正性・専門性) ▪ 被援助国の参画度
2. 情報収集	<ul style="list-style-type: none"> ▪ 評価設問設定の適切性 ▪ 情報源と収集方法の適切性 ▪ 情報源の明確性 ▪ 収集実績
3. 分析	<ul style="list-style-type: none"> ▪ 実績 ▪ 実施プロセス ▪ 定性的因果関係の把握 ▪ 定量的因果関係の把握
3.1 現状把握・検証	
3.2 分析の方法	<ul style="list-style-type: none"> ▪ 客観的分析 ▪ 多面的分析 ▪ 貢献・阻害要因の分析
4. 評価(*2)	<ul style="list-style-type: none"> ▪ 「妥当性」の評価は十分か ▪ 「有効性」の評価は十分か ▪ 「効率性」の評価は十分か ▪ 「インパクト」の評価は十分か ▪ 「自立発展性」の評価は十分か ▪ 「評価全般(結論)」は適切にまとめられているか
5. 提言・教訓	<ul style="list-style-type: none"> ▪ 提言:抽出の十分性 ▪ 提言:根拠の明確性と内容の納得性 ▪ 提言の活用可能性
5.1 提言	
5.2 教訓	<ul style="list-style-type: none"> ▪ 教訓:抽出の十分性 ▪ 教訓:根拠の明確性と内容の納得性 ▪ 教訓の活用可能性
6. 報告書	<ul style="list-style-type: none"> ▪ 書き方全般(簡潔・明瞭・明快) ▪ 図表の活用度 ▪ 1次データの提示

(出典)国際協力機構(2008:110-112)から抜粋

注(*1) PDMとはプロジェクト・デザイン・マトリックス(Project Design Matrix)のことで、プログラム・セオリーの理論と業績測定指標・目標値を含んだプロジェクトの計画表である。JICAでは事前評価を経てすべてのプロジェクトにPDMを作成している。

注(*2)ここではJICAの評価項目である「5項目」に基づく評価を意味する。すなわち、妥当性(relevance)、有効性(effectiveness)、効率性(eficiency)、インパクト(impact)、自立発展性(sustainability)である。

ところで、メタ評価の文献や以上見てきたメタ評価基準には「客観性 (objectivity)」という言葉があまり見当たらない。できるだけ厳格な (rigorous)、正確な (accurate)、評価デザイン、データ収集・分析方法が必要とされている点は共通している。スタッフルビームとシンクフィールドは、「メタ評価報告書は一次評価の顧客が評価の妥当性、統合性、信頼性、費用効果、その適用性を判断することをアシストすることである」(Stufflebeam & Shinkfield 2007: 647) と述べており、客観性の議論よりもどちらかというところ、関係者への「説得性 (persuasive)」に重点が置かれているように思う。

4 メタ評価のタイプと対象

(1) 「形式的メタ評価」と「総括的メタ評価」

メタ評価も「評価」であるので、一次評価と同じく評価対象に対する評価結論 (evaluative conclusion) が必要となる。スタッフルビームは、「評価がプログラム実施のメリットを判断すると同様に、メタ評価は当該プログラムの評価実施のメリットを判断するものである」と述べている (Stufflebeam 1974: 73)。したがって、評価の目的がプログラム実施への意思決定に対する評価情報の提供と、アカウントビリティ確保に対する評価情報の提供の二つに分けることができるように、メタ評価も同様に二つのタイプがあるとしている。すなわち、①評価実施の改善に関する情報を提供する「形式的メタ評価 (formative meta-evaluation)」と、②評価全体の質に関する情報を提供する「総括的メタ評価 (summative meta-evaluation)」である。これは通常の評価において、形式的評価 (formative evaluation) と総括的評価 (summative evaluation) の二つのタイプがあることに呼応している⁴。

「形式的メタ評価」の目的は評価実施の意思決定に役立つマネジメント支援のメタ評価と位置づけることができる。評価目的、評価のデザインやデータ収集・分析方法、評価の実施プロセスについて定期的にチェックを行い、適切な評価の実施に役立つ評価情報をタイミングよく提供することである。したがって、「形式的メタ評価」のメタ評価者は必ずし

⁴ 形式的評価とは、プログラムの実施期間中に行われる評価でプログラムの開発や改善を目的としたものである。内部関係者が実施状況を継続的にモニターする形で行われることが多い。総括的評価はプログラムが終了した段階で行うことが多く、プログラム実施により生み出された効果やそれに至る実施プロセスとの関係に焦点を当てた評価である。

も外部第三者である必要はなく、内部の評価者もしくは関係者で行う方がより直接的に評価実施の意思決定に反映される。

他方、「総括的メタ評価」は評価のアカウントビリティを確保するために行うもので、一次評価の評価結果とそれに至るプロセスとの関係、その根拠となるデータやその分析過程の妥当性などをおして一次評価全体のメリットについて評価を行うことである。つまり評価結果の利用者の役に立つ評価であるかどうかを見るのである。評価結果の利用者の視点や信頼性の確保が求められるため外部第三者によるメタ評価が適しているとされる。

メタ評価を行うときに、これらの区別が混在していないかが問われる。たとえば、外部第三者によるメタ評価で、評価作業が終了した時点で評価のデザインをチェックし改善を提言しても、当該評価へのフィードバックは難しい。あるいは、評価デザインや実施の手続きをチェックすることだけで、評価全体の品質の良し悪しを言うことはできない。二つのメタ評価をどのように組み合わせて、誰が実施するのかという検討も必要になるであろう。

表4 「形式的メタ評価」と「総括的メタ評価」

	形式的メタ評価	総括的メタ評価
メタ評価の目的	評価実施の意思決定における活用	一次評価全体のアカウントビリティの確保
実施主体	どちらかといえば内部の関係者	どちらかといえば外部第三者
メタ評価の焦点	<ul style="list-style-type: none"> ▪ 評価目的、評価デザインに対する提言 ▪ 評価の実施プロセスのチェック ▪ 評価全体の改善への提言 	<ul style="list-style-type: none"> ▪ 評価全体の質、費用便益に関する判断 ▪ 評価結果とそれにいたるプロセスの妥当性に関する判断 ▪ 評価の有用性に関する判断
メタ評価の時期	継続的（一次評価実施期間中）	限定的（おもに一次評価終了段階）

（出典）Stufflebeam（1974: 77）を参照し作成

（2）メタ評価と業績測定

これまでメタ評価はプログラム評価において登場した理論であることを述べてきた。それでは現在、政策評価もしくは行政評価において広く使われている「業績測定」はメタ評価の対象となり得るのだろうか。総務省の客観性確保評価活動の例では、「事業評価」、「実

績評価」、「総合評価」の三つの評価方式を対象としている。実績評価は政策を決定した後
にあらかじめ達成すべき目標と目標値を設定し、終了時に達成度合い（アウトカム）につ
いて評価する方法である。いわゆる業績測定（performance measurement）である。それ
に対し、総合評価はプログラム評価に相当する評価で、特定のテーマについて掘り下げて
分析し問題点やその原因などを総合的に評価する方法である。事業評価は事前の時点で評
価を行い、評価で効果・費用を推計して事業採択の検討を行うことに焦点が置かれており、
必要に応じて事後評価も行う。

業績測定はあらかじめ目標の指標と目標値を設定し、そのとおりになっているかどうか
をモニタリングし測定することが主な活動で、マネジメントの道具としての役割がある。
プログラム評価のようにプログラムの形成、その効果との帰属性（attribution）、あるい
はプロセスに重点を置くのではなく、介入した結果や実績がすべてである。より深い調査
でプログラムの介入の効果や意義を探ろうとするプログラム評価とは一線を画すが、一方
で、ロッシらは‘広義の’プログラム評価では、業績測定はプロセス評価（プロセス・モ
ニタリング、アウトカム・モニタリング）に相当するとしている（Rossi et al. 1999）。

プログラム評価の理論によると、プログラムの実施段階ごとにニーズ評価、セオリー評
価、プロセス評価、インパクト評価、効率性評価があり、それぞれの評価の焦点は異なる
（章末資料2参照）。たとえば、セオリー評価ではプログラムの戦略や実施過程と目標を結
ぶセオリーの組み立てが中心となるし、プロセス評価ではプログラムが意図したとおりに
実施されているかどうかを指標や目標の達成度合いを中心に見ていく。‘広義’のプログラ
ム評価でいうところの業績測定であるアウトカム・モニタリングも同様に、アウトカムの
指標を定期的に把握し前後で比較することで成果を把握しようとする。（そのほかに費用対
効果、費用対便益、インパクト評価を含むすべての評価がメタ評価の対象となる。）

業績測定には、アウトカム指標の目標達成度を重視し戦略計画の策定に活用するものと、
アウトプットのモニタリングを中心に予算サイクルへのフィードバックや効率的なマネジ
メントを目的とするものの二つがあるといわれる。後者のように効率的なマネジメントを
目的とした業績評価の場合は、必ずしも外部第三者によるメタ評価を必要とするわけでは
なく、マネジメント強化の一環として組織内部でチェックする仕組みがまず必要である。
指標のモニタリングを適切に行うためには、組織内に情報マネジメント・システムなどの
仕組みを構築することも必要であろう。あらかじめ政策形成過程を経て選択されたプログ
ラムのプロセス管理を行う意味合いが強く、プログラムや政策の質に関する評価は難しい。

もちろんマネジメントの強化と政策の実現は切り離されるものではない。むしろ相互に補完しあって、社会へのインパクトを生み出すことがめざされるべきではないかと思われる。たとえば、プログラム評価と、マネジメント改善をめざすアウトプット重視の業績測定を組み合わせ（プロセス評価はインパクト評価に必要な情報源を提供する）、政策のインパクトや意義を全体的に評価する仕組みの中で、メタ評価との連携を取ることにより、本来のメタ評価の役割がより明確になるのではないだろうか。

5 メタ評価の実施手続きに見るメタ評価のポイント

メタ評価の実施手続きは、基本的には一次評価と同様である。外部第三者のメタ評価者がクライアントの依頼を受けてメタ評価を行うことを前提にした場合、表5に示すようなステップが主に必要となる。以下、スタッフルビームとシンクフィールドの文献を参考に、実施手続きをとおしてメタ評価のポイントと論点を考察する（Stufflebeam & Shinkfield 2007: 664-680）。なお、メタ評価の事例としてミシガン州政府「チャータースクール・イニシアチブのメタ評価」に関する概要を章末資料3にまとめたので参照ありたい。

表5 メタ評価の実施手続き

1.	メタ評価チームの設置（1人もしくはそれ以上の評価専門家を含む）
2.	メタ評価のステークホルダーとの話し合い・調整
3.	メタ評価の評価設問の設定
4.	メタ評価の評価方針、評価基準の合意
5.	正式なメタ評価契約もしくは協定書の作成
6.	関連する既存の情報の収集とレビュー
7.	必要に応じて新たな情報の収集
8.	情報分析と結果の統合
9.	適切な評価基準、評価方針に基づいた厳密な評価
10.	報告会、報告書などとおしたメタ評価結果の報告
11.	状況に応じて、クライアントや他のステークホルダーに対する評価結果の解釈と活用に関するサポート

（出典）Stufflebeam & Shinkfield（2007: 663）

(1) ステークホルダー間の話し合いと取り決め

まずメタ評価者（もしくはチーム）とメタ評価依頼者との間で契約を結ぶ前の評価の準備段階では、メタ評価チームの設置、ステークホルダーとの話し合い、評価設問の設定、評価方針・評価基準の関係者間での合意という極めて重要な項目が並ぶ。一次評価であれメタ評価であれ、評価するという行為は関係者に様々な軋轢を呼ぶ可能性がある。社会問題の解決のために行われる介入（＝プログラム、政策）は、異なる立場の関係者の多様な価値・考え方の中で行われるわけで、学校の成績表のようにあらかじめ設定された点数基準で判断ができるようなケースはほとんどない。であるからこそ、この準備段階におけるステークホルダーとの話し合いや調整が不可欠なのである。メタ評価を実施するにあたって、メタ評価の目的を共有し、プログラムや一次評価自体の利点や弱点（それらに対する彼らの認識）、メタ評価の評価基準などについて事前に意見交換をしていくことが評価活動全体の成否を決めるといっても過言ではない。またそこでの話し合いの結果は評価設問の設定や評価基準の合意過程にも反映されることになる。

さて、メタ評価の評価設問であるが、スタッフルビームとシンクフィールドによると、メタ評価の基本的な評価設問として、①どれだけ「正しい評価」たる基準を満たしているか（merit）、②どれだけ「（一次評価の評価結果が）顧客のニーズに合致した情報」を提供できるか（worth）、の二つを念頭において、評価対象独自の評価設問を設定すべきであるとしている（Stufflebeam & Shinkfield 2007: 677）。前者は評価方法の厳格さを、後者は評価結果の価値を問うものである。これらの設問は、プログラム評価の基本的な姿勢である「プログラムの介入は社会に対しどのような良い変化をもたらしたか。良い変化がもたらされなかったとしたらどのように改善すべきか。」という問いかけに対応するもので、その問いかけに一次評価が応えているかどうかメタ評価の視点となる。評価設問の検討後は、前掲表1～3にあるような具体的な評価基準を検討する作業に入るが、評価設問はメタ評価者を含め関係者間で評価の目的と方向性を再認識する過程として重要である。

この準備段階を経てメタ評価者と依頼者との間で何らかの取り決めや契約を交わすことになる。外部第三者にメタ評価を依頼しない場合とか小規模のメタ評価の場合は正式な契約締結などの作業を伴わないケースがあるが、非公式であれ「メタ評価者」と「メタ評価依頼者」あるいは「一次評価実施者」の間でメタ評価の目的、評価基準等について対話の場を設けることが必要である。

(2) 情報収集・分析、評価

メタ評価の方針が決定したあと、情報収集、分析、評価の判断といった評価活動そのものに入るわけであるが、ここでのポイントは一次評価で使った情報や既存の情報の入手であろう。事前の準備段階において、また何らかの取り決めの中でそれらの入手に関する合意ができていれば比較的容易である。たとえば、アメリカにおけるメタ評価の例では、取り決めの中で一次評価の評価報告書とその根拠となった情報資料を Web に掲載し、いつでもアクセスできるようにしたことがその後の作業を容易にしたという報告がある (Stufflebeam & Shinkfield 2007: 672)。メタ評価の根拠となる情報は既存の情報やデータがあまりに不備で評価設問にきちんと回答することができないと判断された場合は、必要に応じて新たに収集するが、基本的には一次評価で使った情報やデータを使うことで十分に対応できるとされている。またメタ評価がひとつの組織の内部機能として位置づけられた場合は情報の収集は比較的容易である。たとえば、3.(2)で紹介した国際協力機構 (JICA) の外部有識者評価委員会は二次評価 (メタ評価) の実施が委託内容に含まれており、表3 (前掲) に示すような調査票の項目を使って評価を行っている。そのような場合は、JICA からの評価報告書や関連資料の提供が前提にあり、情報収集に伴う困難は軽減される。

情報の分析過程では、収集したデータをもとに定量分析もしくは内容分析 (content analysis) 等の定性分析を行い、メタ評価基準ごとのレーティング等の根拠を示すことになる。ここで、「評価の正確さ」の視点からメタ評価を行った小学校の二つの読解力向上プログラムの比較分析の事例を紹介しよう (Stufflebeam & Shinkfield 2007: 674-675)。効率性を評価した一次評価でプログラム A とプログラム B を比較して、プログラム A の方が生徒一人当たりのユニットコストが低く効率的であるので、プログラム A が推奨されているとしよう。その評価の前提はふたつのプログラムの効果が一定であるというものであった。プログラム B が対象としている生徒は既に何らかのハンディを持っている子供で、全体の数からいけば少なく、しかしながらハンディを持った生徒全員がある一定の読解力を習得することをゴールに実施されておられる。一方プログラム A はより多くの子供に対し平均的な能力向上をめざすプログラムであった。メタ評価では、単に生徒一人当たりのユニットコストでふたつのプログラムを比較することは不適切とし、読解力向上プログラムを実施している学校の学生一人当たりのコストをプログラムごとに計算し (学校間比較が可能になる)、毎年、読解力の訓練を必要としない生徒の数を読解力プログラム全体にかか

ったコストで割ることによって、経年の変化を見ていく必要があるとした。その係数が小さくなればなるほど効率性が高いということになる。メタ評価の段階ではその経年データの入手は不可能であるが、メタ評価チームはプログラム B の特性と定性データからプログラム B の方がより効果的ではないかという問題提起を行っている。

この事例説明の中で印象的であったのは、メタ評価者の関心が主に「学校のマネジメント側の関心」に置かれていることである。すなわち、どのプログラムを実施したら生徒の福祉にプラスになるのか、という点である。そのためにメタ評価が一石を投じている。一次評価をしたプログラムの評価結果の質を問うことが、間接的にはプログラムの質の向上に結びつくことになるのである。各プログラムで何人が受講し、何人の読解力が向上したという目標達成の実績把握のチェックも必要であるが、それだけではプログラムの選択を含めた本来のプログラム評価もしくは政策評価の役割が限定的になってしまうことに留意が必要である。

(3) 評価結果の報告

メタ評価結果も一次評価と同様に報告書の作成、報告会の実施などをおして評価結果をフィードバックするが、ここでのポイントは、準備段階で相互交流を行ったステークホルダーへの評価結果のフィードバックであろう。可能であれば、再度集まり、評価結果の発表、意見交換などを行い関係者がメタ評価結果を‘納得して’受け入れることができるような対話の場をもつことが望ましい。ただしこの集まりはけっして「アドバイザー・パネル」のような諮問的な働きをもつのではなく、あくまでも「レビュー・パネル」として、それぞれの見地からメタ評価結果を具体的に改善するような建設的な話し合いを持てるのが理想的だとしている。

6 まとめ～メタ評価実践への示唆

本章では評価の評価（メタ評価）の概念整理を試みた。これまでの議論を踏まえ、まとめとして四点の論点を提示したい。

まず一点目に、プログラム評価の一環としてメタ評価を位置づける必要がある。メタ評価はプログラム評価の発展の中で生まれた。評価の品質管理を目的とするメタ評価の場合、その結果は、直接的には①評価の設計、実施方法の改善（形成的メタ評価）、②評価の質や

その有用性の判断（総括的メタ評価）に活用され、間接的にはプログラムや政策の質の改善に資することが期待されている。したがって、外部からの統制によるチェックというよりも、プログラム評価全体の流れの中に位置づけることが適当ではないかと考える（章末資料参照）。メタ評価は客観性を追求するについで評価の設計や実施方法の厳格さに焦点を当てがちになるが、プログラムや政策の質が適切に評価されているかどうか、評価結果が次の政策形成に活用されているかといった視点が実質的なメタ評価には重要である。それが、メタ評価結果を受けた評価の見直し、ひいては政策の議論につながるからである。

二点目としてメタ評価基準の設定についての議論がある。メタ評価の実務では評価基準が必要になるが、その前提として「質の良い評価」の議論が欠かせない。それは、多様な評価方式の目的に立ち返ることでもある。どのような評価が求められているのかの共通認識がない限り、評価のみならずメタ評価も形式的なものに終わってしまう危険性がある。

「良い評価」の基準については、アメリカでは全米評価学会（American Evaluation Association）が中心となり“Guiding Principles for Evaluators”（評価者のための指針）を作成し、随時改定作業を行っている。これがメタ評価の評価基準にも反映されていることは前述したとおりである。ただし、評価は実施される社会の文脈や文化、制度によってその内容も異なってくることから、アメリカの基準をそのまま使えるわけではない。日本国内の取り組みでは日本評価学会が2009年1月から「評価者倫理・スタンダード策定分科会」を立ち上げその検討を開始したばかりである。これらの基準は共通の品質基準として応用することが可能だが、各組織ではおのおのの組織目的に合致した評価のポリシーや具体的な評価手法の検討が必要となる。国の政策評価においては、政府統一の評価手法がないことや、方式も各府省の独自の裁量により自由に選択する柔軟な方式になっており、焦点が見えにくいという指摘がある（石橋2007: 30-40）。そのあたりがクリアされないとメタ評価が形式的な手続き論に終わってしまう。

第三点目にメタ評価の主体に関する議論がある。ひとつには一次評価を実施する組織自らがメタ評価の仕組みを構築することが可能である。外部委託によるコンサルタントの雇用や（日本の場合は独立性にやや問題があるが）、外部第三者によるメタ評価チーム（レビューパネルなど）を設けること（3.（2）JICAの例）が考えられる。国の政策評価では、総務省に政府内部の組織としてメタ評価を行う役割が課せられているが、評価のやり方チェックだけではなく、政策の議論を進める実質的なメタ評価の機能を発揮するための制度整備、関係者間の共通認識がこれまで以上に求められるであろう。なお、アメリカの行政

管理予算庁（OMB）の例のような政策チェックの仕組みは、厳密な意味でのメタ評価とは異なることを言及したい。OMB が使っている「施策の査定と格付けツール」（PART）は一部メタ評価の要素が含まれているもののそのほとんどは OMB による予算編成作業のための評価のツールであり（第2章参照）、メタ評価をとおして評価の品質を確保し、ひいては政策の質を高めるという目的とは異なるからである。

最後に、メタ評価者と一次評価者との間のコミュニケーションの重要性を指摘したい。メタ評価の実施過程では、通常の評価と同様、評価される側を含めて十分な対話が必要になることを指摘した。メタ評価は再評価ではないが「評価」である。その評価結果の活用度合いを高めるためには、評価される側と評価する側の関係性は重要である。これは歩み寄りという安易なコミュニケーションではなく、政策の質向上という共通の目的に向かって互いの専門性をオープンにする関係性である。その意味において、たとえば国の政策評価のメタ評価では、各府省間の評価リソースの人事交流をとおして共通知識を蓄積していくということも一案であろう⁵。

⁵ 南島（2007）は日本の政策評価制度の客観性の議論において、評価リソースを有効に活用するための各府省間のコミュニケーションのあり方を論じている。

米国エネルギー省「耐気候構造化補助プログラム」¹⁾のメタ評価 (評価統合)

1. 概要

本メタ評価 (評価統合) は、米国エネルギー省の貧困世帯向け「耐気候構造化補助プログラム」を対象に、1993年から2002年までに16の州で実施された37個の評価調査に基づき2003年2月に実施されたものである。メタ評価実施機関は、エネルギー省の傘下にあるオークリッジ国立研究所 (Oak Ridge National Laboratory: ORNL) である。なお、本メタ評価の前に2回にわたり (1997年、1998年) 同様のメタ評価が実施されている。

本メタ評価の目的は、①耐気候構造化補助プログラムの対象となった世帯の燃料別省エネ率の比較と、②天然ガス使用による省エネ率の全国平均予測を計量分析により行うことである。

2. 評価方法の概略

- ① 州レベル評価調査の現状把握とメタ評価データ収集票の作成
- ② 天然ガスの省エネ分析に活用する州レベル評価調査の選定とデータ収集
- ③ データの分析
 - ◇ 28個の州レベル評価報告書に基づく耐気候構造化以前の暖房ガス消費量 (独立変数) と省エネ量 (従属変数) による単回帰分析
 - ◇ ミネソタ州を事例とした「燃料オイル」と「プロパンガス暖房システム」のモニタリングデータを使用した分析。ミネソタ州が開発した省エネ測定ソフトウェアである ASAP (Achieved Savings Assessment Program: 省エネ達成度査定プログラム) を使用。
 - ◇ 6個の州レベル評価報告書に基づく電力暖房世帯の省エネ分析。
 - ◇ 4個の州レベル評価報告書に基づく暖房目的以外の電力省エネ分析。

3. 結論 (要旨)

- ① 耐気候構造化以前の暖房ガス消費量と構造化による省エネ量の関係性の強さから、耐気候構造化をした方が省エネが進むことが明らかである。その省エネ率は21.9%で、暖房使用のみに限ると30.8%である。
- ② 電力の省エネ率は10.5%、暖房使用のみの場合は26.7%であり、ガスの省エネ率と同じような効果が認められた。
- ③ 本メタ評価対象期間における、耐気候構造化補助プログラムによる天然ガスの省エネ率全国平均は30.8%である (1989年は18.3%)。

注: 1) 耐気候構造化補助プログラム (The National Weatherization Assistance Program) は、低所得世帯に対する省エネ支援政策である。省エネをとおし、環境問題のみならず貧困層のエネルギー消費費用軽減による生計向上も目的としている。1976年の開始以降620万世帯が対象となった。

(出典) Berry, L. & Schweitzer, M., *Metaevaluation of National Weatherization Assistance Program Based on State Studies 1993-2002*
http://weatherization.ornl.gov/pdf/CON_488.pdf

章末資料2（参考） プログラム評価の種類

評価の視点	評価の焦点	
ニーズ評価	プログラムのサービスと顧客のニーズが合致しているかどうかを明らかにすること	<ul style="list-style-type: none"> ▪ 新しいプログラムの顧客は誰か ▪ サービスが顧客に届いているか ▪ サービスは顧客のニーズに合致しているか
セオリー評価	プログラムがどのようにデザインされているか、そのデザインが目的を達成するために妥当であるかを明らかにすること	<ul style="list-style-type: none"> ▪ プログラムの目的は何か ▪ プログラムはどのように目的を達成しようとしているのか ▪ セオリーの組み立ては妥当か ▪ プログラムの戦略は妥当か
プロセス評価	プログラムが意図されたとおりに実施されているかを明らかにすること	<ul style="list-style-type: none"> ▪ 計画どおりに実施されたか ▪ 誰が関わっているのか ▪ プログラムに対する関係者の認識の変化はあったか ▪ プログラムの成否に影響を与えた要因は何か
インパクト評価	プログラムの成果や効果が上がっているかどうかを明らかにすること	<ul style="list-style-type: none"> ▪ 効果はプログラムの実施によるものか ▪ サービスを受けた顧客はどう変化したか ▪ プログラムを継続もしくは拡大する意味があるか（一般化）
効率性評価	プログラムが効率的に実施されているかどうかを明らかにすること	<ul style="list-style-type: none"> ▪ 効果に対して費用は適切に投入されたか ▪ 他のプログラムと比較して効率性は高いか、低い

（出典） Robson（2000） および Rossi et al.（1999）を参照し作成

ミシガン州政府 チャータースクール・イニシアチブ¹⁾のメタ評価

1. 概要

本メタ評価は、ミシガン州教育局の委託を受けてウェスタンミシガン大学評価研究所（以下、「評価研究所」）が実施した「ミシガン州チャータースクール・イニシアチブ」の評価（1999年1月）を対象とし、1999年5月に第三者（Kenneth H. McKinley, Ph.D.）がメタ評価を実施したものである。メタ評価の結果は、教育局、評価研究所、ならびに評価対象となったチャータースクール各校にフィードバックされた。

1. 評価方法の概略

① メタ評価の視点

総括的なメタ評価として、評価のステークホルダー（州政府教育局、評価研究所、チャータースクール）に対し、チャータースクール制度の評価とそれにいたるプロセスに焦点を当てて評価が行われた。

② メタ評価の評価基準

プログラム評価のメタ評価項目を使い（本章表1参照）、評価の有用性、評価の実施可能性、評価の正当性、評価の正確性についての評価を行った。

③ 関連する既存の情報の収集とレビュー

<関連ドキュメント類>

- ◇ 教育局による当初の評価設問の内容（委託候補先に対する提案書提出依頼に記載）
- ◇ 評価研究所のプロポーザルの内容
- ◇ チャータースクールによる「自己評価手法」の内容
- ◇ サンプル校におけるフィールド調査記録
- ◇ インタビュー調査の日程およびインタビュー項目
- ◇ 研究所のウェブサイト上の関連情報
- ◇ 一次評価報告書（Evaluation of Michigan Public School Academy Initiative by Horn and Miron 全159頁）の内容

3. メタ評価の提言項目（要旨）

① 評価の有用性に関するもの

【利害関係者の特定】

- ◇ 直接的な利害関係者（州教育局、州議会、チャータースクーラー学生、教師、親、マネジメント含む）以外にもミシガン州以外の州教育局や州議会や、連邦政府の教育政策担当者なども重要な利害関係者として含めるべき。
（教育政策全体のあり様にも関係するため）

【評価報告書の内容・質】

- ◇ 議論の中心となる用語の明確な説明が必要。（具体的に再定義が必要となる用語のリストを提示）
- ◇ 州教育局が提示した評価設問（4分野 24項目）ごとに評価結果が明示されている頁を入れたリストをつけるべき。
- ◇ チャータースクールが導入された背景や歴史的変遷の解説を加えるべき。
（報告書の読み手が内容をより理解できるようになるため。）
- ◇ 定量・定性データの標記、解釈の方法について訂正すべき箇所の指摘。
- ◇ プログラム評価ではなく教員評価にあたる箇所が含まれていることの説明を加えるべき。

② 評価の正確性に関するもの

【情報・データの収集方法／情報源の適切性】

- ◇ チャータースクールのインパクトについて、公立学校の代表者からの意見を取り入れるべき。
- ◇ チャータースクールの認可プロセスについて、大学等の認可機関からの意見を取り入れるべき。

注：

- 1) チャーター・スクール (Charter School) は、アメリカ合衆国で 1990 年代から増えつつある公募型学校で、チャーター (Charter) と呼ばれる特別認可、あるいは達成目標契約により認可された学校である。認可された場合、公的な資金の援助を受けて学校が設立されるが、運営は設立申請を行った民間のグループが担当する。

(出典) McKinley, K., *Metaevaluation Report of the Evaluation of the Michigan Public School Academy Initiative*, May 1999

(<http://www.wmich.edu/evalctr/charter/reports/metaeval.html>)

Horn, J. & Miron, G., *Evaluation of the Michigan Public School Academy Initiative, Final Report*, January 1999

(http://www.wmich.edu/evalctr/charter/reports/wmu_finalrpt.pdf)

第2章 アメリカにおけるメタ評価の現状

佐々木 亮

1 メタ評価の種類

メタ評価は、現在は次の二つの意味で使われているとされる (Bustelo, 2002)。

(ア) 評価の質的管理 (評価デザイン批評)

(Evaluation quality control (Evaluation design critique))

(イ) 評価結果の統合

(Evaluation Synthesis)

さらにスクリヴェンは、評価の結果が客観的証拠 (いわゆる Evidence) に基づいて、論理的な結論が導出されているかを検証する意味でもメタ評価が使われていると理解しており、次のように名づけている (Mathison 2008: 250)

(ウ) 独立的立場からの評価結果の再検証

(Independent confirmation of evaluative conclusion)

上記のうち、(ア) は、何らかのチェックリストを使用して、評価報告書を構成する一般的な項目がカバーされているかとか、評価活動において行うべき活動項目や遵守すべき注意事項が遵守されたかどうかなどをチェックすることを具体的内容とする。

上記のうち、(イ) は、いわゆるメタ分析 (Meta-analysis) に近いものである。ただしメタ分析が、類似の複数のインパクト評価で明らかになった効果サイズ (Effect size) を、合成して平均的な効果量を計算することを目的とするのに対して、メタ評価は類似の複数の評価結果をレビューして、総合的な評価的結論 (「やって良かった/悪かった」「やるだけの価値があった/なかった」等といった価値的結論) を導出することを目的とする。つまり、メタ評価も評価の一種であるから、価値に触れる用語 (Value-laden words) を使って明快に評価的結論 (Evaluative conclusion) を出すことが求められる。

上記のうち、(ウ) は、厳格な方法・デザイン (科学的あるいは客観的な方法・デザインとも言う) によってエビデンスが得られたかどうか、そして得られたエビデンスによって論理的・合理的な結論が導出されたかどうかを検証する作業である。

2 メタ評価のためのチェックリスト

前節で解説したメタ評価の3種類のうち、(ア)評価の質的管理としてのメタ評価のための基準はどのようなものであるべきだろうか。この点に関してスクリヴェンは次のように述べている (Scriven 2009)。「(各種の) 評価の体系的なアプローチは、自動的にメタ評価のための体系的な基礎を提供するものだ。明らかな例は、CIPP モデルチェックリスト (CIPP Model Checklist) (Stufflebeam 2002)、プログラム評価基準 (Program Evaluation Standards (PES)) (Sanders & Joint Committee 1994)、基幹評価チェックリスト (Key Evaluation Checklist) (Scriven 2006) などのチェックリスト群である。ただし、それらの本来的な有効性にも関わらず、単純に形式的に適用され重要な点を見失う結果になることもある。」¹ なお、評価の質的管理の作業の一部として、(ウ) 独立的立場からの評価結果の再検証も含まれる場合が多いのが現実のメタ評価の運用である。

ここでは、「基幹評価チェックリスト」 (Scriven 2008) を掲載する。また、「評価者の基準原則によるメタ評価チェックリスト」 (Guiding Principles Checklist for Evaluating Evaluations) の概略も述べる。また、「プログラム評価チェックリストのメタ評価チェックリスト」 (Stufflebeam 1999) については、本報告書の第1章を参照されたい。

¹ “A systematic approach to evaluation automatically provides a systematic basis for metaevaluation. Obvious examples are the checklist approaches, e.g., CIPP (Stufflebeam 2007), the Program Evaluation Standards (PES) (Joint Committee, 1994), and the Key Evaluation Checklist (KEC) (Scriven, 2007). ... However, you’re then pinning your effort to the validity of these checklists and even if they are intrinsically valid, they may be applied ritualistically, and thus miss important points.”

BOX 1 「基幹評価チェックリスト」 (Key Evaluation Checklist)

スクリヴェンが 1980 年に作成して提案して以来 (Scriven 1980. The Logic of Evaluation)、広く用いられている基幹評価チェックリストであり、メタ評価にそのまま利用可能である。

A: 準備 (Preliminaries)	
I. 要約	通常は項目 11~15 の要約を書く。
II. 序文	評価を依頼された目的と背景を書く。(i)統制評価、形成評価、帰属的評価のいずれの目的で実施するのか。(ii)等級、順位、側面描写等のどれを総合評価として求められているのか。(iii)評価依頼者、利害関係者 (ステークホルダー)、その他で報告書を読むのは誰か、を記述する。
III. 方法	どのような評価方法や評価モデルを用いて評価を実施するのかを記述する。例) 実験デザイン、マッチングデザイン、フォーカスグループ、インタビュー等。
B: 基礎 (Foundations)	
1. 背景と状況	政策/施策/事業の歴史や背景を書くとともに、次の項目を説明する。(i)川上のステークホルダー (政策/施策/事業の資金負担者等)、(ii)関係する立法や政策変更、(iii)公式のプログラム・セオリー等、(iv)類似の介入の結果、(v)以前の評価結果その利用結果 (もしあれば)
2. 政策/施策/事業の概要	政策/施策/事業の正確で完全な記述を行う。主要な目標と予想される成果、活動内容、活動を取り巻く環境や状況を説明する。それらは、公式のプログラム・セオリーとかなり違うことがあることに留意する。また、当該評価で使用することが避けられない専門用語があれば、ここで解説する。
3. 受益者 (Consumers or impactees)	受益者は、(i)サービス/生産物の利用者 (川下の直接的 '被影響者') と(ii)サービス/生産物の利用から間接的に影響を受ける人々 (川下の間接的 '被影響者') に分けて説明する。政策/施策/事業の実施に携わる職員等は受益者ではないので記載する必要はないことに留意する。
4. 投入資源 (Resources)	政策/施策/事業実施のための投入資源を説明する。投入資源は、財務的、物理的、人的、そして社会能力的な資源に分けて記述する。社会能力的資源とは、知識、技術、職員・ボランティア・地域住民・支援者等の '良い意志' のことである。
5. 評価基準 (Values =Evaluative criteria)	政策/施策/事業を評価するために用いる価値基準である「評価基準」を特定して記載する。なお、「評価基準」は変化しないものではなく、その時代やその場所によって変化することに留意する。「評価基準」は以下の情報源から得られた情報を統合して特定する。 「評価基準」は、「価値側面」(Criteria of Merit) と「価値水準」(Standard of Merit)で構成される (「価値側面」は、後述の '過程、成果、効率性、比較優位' などのことで、「価値水準」とは、それぞれの価値側面における '優、良、可、不良、不可' のそれぞれの水準の状態の具体的な記述で

	<p>ある)。</p> <p>(i)受益者のニーズ査定結果、(ii)定義的情報 (評価対象の政策/施策/事業の上位目標および目標が設定されている場合で、十分に論理的な場合はそれを援用する)、(iii)論理的情報、(iv)法律的情報、(v)倫理的情報、(vi)個人的・組織的目標、(vii)誠実さ (コンプライアンス) に関する基準、(viii)準法規的基準、(ix)専門的基準、(x)熟練者による基準、(xi)歴史的/伝統的/文化的基準、(xii)科学的情報、(xiii)技術的情報、(xiv)市場的情報、(xv)政治的情報、等。</p> <p>得られた情報にしたがって、評価基準を整理して記載する (縦横のマトリックスにするとわかりやすい)。さらに、それぞれの価値側面に加重点 (Weights)を決めて併せて記載しておくことが勧められる。さらに、必要に応じて、基準点切り(Bar)として、特定の価値側面に関して受け入れ可能な最低基準の設定をしておく (いわゆる「足切り」)。そのほか、より洗練された方法として、段階付け(Stepping)も用いることができる。</p>
C : 部分評価 (Subevaluations)	
それぞれの部分評価では、(i)事実を特定する、(ii)特定された事実に評価水準を適用して価値判断を下す、という二つの作業を行う。	
6. 過程 (Process)	政策/施策/事業の成果 (アウトカム) が得られるまでに起こった全ての重要な物事を明らかにし、それらの意義や重要性を評価して記述する。具体的には、(i)目標 (ゴール)、(ii)計画 (デザイン)、(iii)実施過程、(iv)管理 (マネジメント)、(v)活動・手続き、(vi)知識習得、(vii)態度変容、(viii)他に分けて評価するとよい。
7. 成果 (Outcome / Impact)	<p>受益者が被った成果 (アウトカム) を明らかにして、それらの意義や重要性を評価して記述する。成果には、(i)直接・間接、(ii)意図したもの・意図しなかったもの、(iii)直後・短期・長期の別があることに留意しつつ、(i)通常の結果、(ii)社会的成果、(iii)環境的成果に分けて評価するとよい。また、重要な成果のいくつかは、政策/施策/事業の実施中にも発生することに留意する。</p> <p>なお、非常に重要な成果は予期していなかったものであることが多いので、いわゆる仮説検証の手法は適さない。したがって、あらかじめ設定された目標にとらわれずに、成果を特定する洗練された作業が要求される。</p>
8. 効率性 (Cost Efficiency)	金銭的成本と、非金銭的成本の双方を明らかにして、それらの効率性を評価して記述する。開始/維持/アップグレード/終了・閉鎖などの発展段階に分けて特定して評価するとよい。
9. 比較優位 (Comparative analysis)	より少ない投入資源で類似の便益*をあげられる他の政策/施策/事業 ("Critical Competitors") と比較する。5. ~7. は、政策/施策/事業と受益者の価値観等に基づく「評価基準」の対比で評価が行われたが、8. では当該政策/施策/事業と、他の政策/施策/事業の比較で評価が行われることになる。比較により、当該政策/施策/事業の意義・値打・重要性を明らかにする。*ここで言う「便益」とは、過程、成果、効率性の各段階で得られた積極的な (プラスの) 価値判断のことである。
10. 持続性 (一般化可能性)	政策/施策/事業が、今後とも自立的に持続するかどうかを明らかにする (持続性)。また、他の時間/場所/職員/受益者/環境でも類似の便

(Sustainability / Generalizability)	益をあげられるかどうかを明らかにする（一般化可能性）。後者は、キャンベル（Campbell, 1969）の「外部妥当性」とほぼ同じ概念である。(i) 財務面、(ii)（人的な）能力面、(iii)組織・制度面、(iv)社会面、(v)環境面、(vi)政治面、(vii)その他の条件に分けて評価するとよい。
D：結論と示唆 (Recommendations and Implications)	
11. 総合化と結論	<p>5つの部分評価を総合して総合評価を記述する。加重・合計、基準点切り、段階付け等を施して、単一の等級や順位を決定して記述する。総合化が難しい場合には、少なくとも「側面描写」（5つの部分評価を並列的に記述する）を行う必要がある。</p> <p>通常は、受益者のニーズに対する政策／施策／事業の成果（現時点の実績と将来時点の予想）にもっとも焦点を当てて記述する。また、評価委託者やその他のステークホルダーの情報ニーズ（委託者の組織目標ほどの程度達成された等）に応える結論も記述する。</p>
12. 教訓と提言 （可能であれば）	<p>提言の記述は、通常考えられているほど簡単なことではない。提言を書かなくとも、評価結果を記載することが要求されている評価報告書としては何ら問題はないとされる。</p> <p>提言は、ミクロレベルの提言（Micro-recommendations）とマクロレベルの提言（Macro-recommendations）に分けられる。ミクロレベルの提言は、管理運営や資器材選択に関するマイナーな改善提言であり、通常は非常に有用である。</p> <p>一方、マクロレベルの提言は、政策／施策／事業の拡大・中断・廃止などに関する提言であるが、このレベルの提言を行うには、(i)当該分野に関する豊富な知識、(ii)追加資源を要するような多大な作業努力、(iii)政策レベルの意思決定者にアクセスできる位置にいること等の高度な条件が要求される。</p> <p>また、今後の政策／施策／事業の計画立案に参考になる「教訓」もここで記載する。ただし、無理に書かない。</p>
13. 結果責任の所在 （可能であれば）	この記述は、万が一求められた場合で万が一それを特定するだけの能力がある場合に実施する。評価者は気軽にこの作業を引き受けてはならない。
14. 報告と今後の支援	<p>政策／施策／事業の実施者に対する場合と、一般市民に対する場合と、議会関係者に対する場合等では、評価報告の体裁や用語のレベルはまったく違うものになるだろう。それぞれの対象に対してどのような報告を行うべきかを記述する。</p> <p>また、評価が終了したあとに、受益者、利用者、職員、資金提供者等に追加で行う支援サービスについても記述する。</p>
15. メタ評価	メタ評価とは「評価の評価」のことである。当該評価結果に関して、その利点と制約を明らかにして記述する。通常は、当該評価を実施した評価者とは別個の独立した品質管理部署がメタ評価を実施する。ただし、投入資源がない場合に評価実施者自身が行ったとしても、実施する価値は高い。メタ評価の視点は、有効性（Validity）、有用性（Utility）、信頼性（Credibility）である。
（出典） Scriven (2004) <i>Key Evaluation Checklist</i> （一部、筆者（佐々木）が変更）。	

BOX 2 「評価者の基準原則によるメタ評価チェックリスト」

(Guiding Principles Checklist for Evaluating Evaluations)

全米評価学会で承認されている「評価者のための基準原則」を利用して作成したチェックリストである。以下の5項目で構成される。この基準原則は、各分野のメタ評価の基準を含んでいないので、プログラム評価チェックリスト (PES) などと併用することが勧められる。

- ◆ 系統的調査スタンダード (Systematic Inquiry)
- ◆ 作業能力 (Competence)
- ◆ 誠実な態度 (Integrity / Honesty)
- ◆ 人々に対する敬意 (Respect for people)
- ◆ 一般・公共福祉に対する責任 (Responsibilities for General and Public Welfare)

(出典) <http://www.wmich.edu/evalctr/checklists/guidingprinciples2005.pdf>

3 アメリカ会計検査院 (GAO) によるメタ評価の実践状況

(1) GPRA、GAO、OMBの関係

政府業績成果法 (GPRA)、およびその普及を推進してきたアメリカ会計検査院 (GAO)、そして GPRA と予算審査を連動させようと試みるアメリカ議会予算局 (OMB) の関係は複雑である。廣瀬 (2006)、伊藤 (2006) がこれらの関係をうまくまとめているので、別途参照されたい。

1993年に成立した「政府業績結果法」(Government Performance and Results Act of 1993、通常 GPRA と呼称される) は、アメリカ連邦政府の各省庁に、戦略計画 (Strategic Plan) と業績測定 (Performance Measurement) およびプログラム評価 (Program Evaluation) を義務付けた法律である。各省庁は、連邦議会との協議を経て、3年間をターゲットとする戦略計画を策定する。戦略計画を踏まえて「年次計画」(Annual plan) を策定するが、その中で、業績指標 (Performance Indicators) を選定して業績目標 (Performance Target) を設定することが義務付けられている (龍・佐々木 2004)。なお、年次計画は OMB と連邦議会に提出されて承認を得る必要がある。そして、各省庁は、業績目標の達成度合を測定して、「年次業績報告書」(Annual Performance Report) を策定し、大統領と連邦議会に提出することになっている (廣瀬 2006)。GPRA に基づく戦略計

画と業績測定は、1994年度に試行期間が始まり1998年度から本格実施された。

GPRAの導入をOMBとともに推進したのはGAOであり、GAOは、戦略計画および各省庁が業績測定を実施するための各種指針等の策定、連邦議会の要請に基づくGPRAの実施状況に関する研究報告書の策定、GPRAそのものの有効性に関する評価の実施等を行っている。

さらに最近の動きとしては、OMBは、GPRAによって策定される戦略およびその業績と、連邦政府の予算を連動させるために、PART (Program Assessment Rating Tool)を導入したことが注目されている(伊藤2006)。

(2) GAOにおけるプログラム評価の実際

GAOは、1921年の設置以来、伝統的な会計検査を実施してきたが、1960年代後半から徐々に活動範囲を拡大し、1970年の立法府改革法や1974年議会予算法によりプログラム評価が正式にGAOの活動として認められるに至った。現在では、GAOの活動の大半(約90%)が、議会の要請によるプログラム評価の実施で占められている(廣瀬2006)。残りの10%はGAOが独自の判断で実施するプログラム評価であるが、議会の要請によるプログラム評価同様、報告書は議会に提案して承認を得るため、報告書の記述からはGAO独自の判断で実施したものがどれかかは判断できない。プログラム評価報告書は、通常の評価項目(プロセス評価、インパクト評価、費用便益分析等)で構成されていることが多い。

現在、GAOが行うプログラム評価は、「取り決めのための電子支援ガイド」(Electronic Assistance Guide for Leading Engagements (EAGLE))というシステムにより報告書作成手続きが電子的に管理されている。同支援ガイドは、議会要請の受付から報告書公表までの業務を7段階に分けて、それぞれの段階の手続きを詳細に定めている。7つの段階とは、(i)業務受付、(ii)業務開始、(iii)任務分担の設計、(iv)メッセージ合意、(v)第一担当者の承認(報告書原案の作成後)、(vi)各省庁のコメント拝受、(vii)報告書の発行、である(黒田2003)。以下にその概要を示すが、詳細については、黒田(2003)によくまとめられているので参照されたい。また参考までにGAOの組織図およびプログラム評価担当部署も示す。

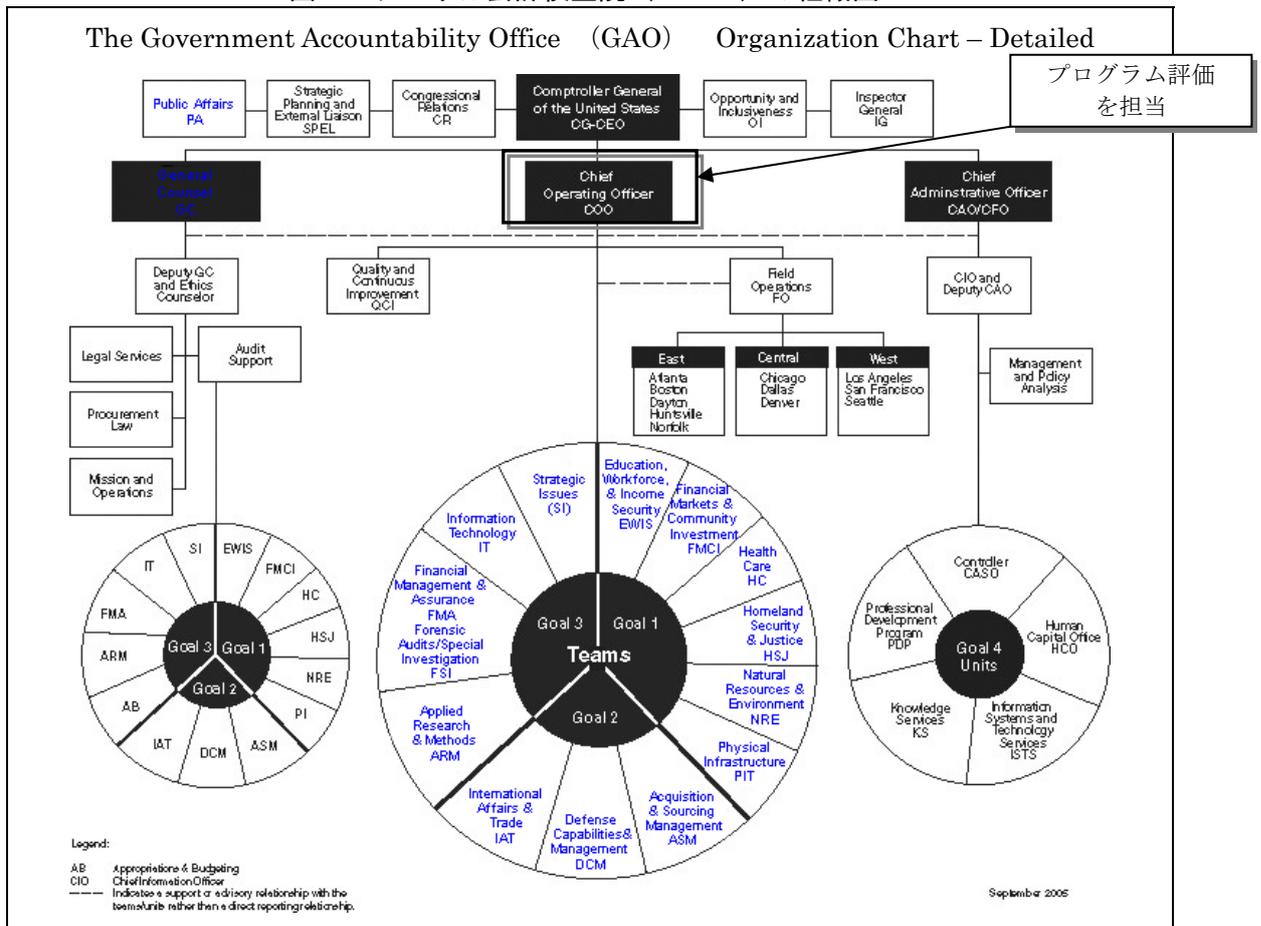
表 1 GAO「取り決めのための電子支援ガイド」の概要
 (Electronic Assistance Guide for Leading Engagements (EAGLE))

段階	具体的作業項目
(i) 業務受付	1-1 要求の受付 1-2 業務受入会議(EAM: Engagement Acceptance Meetings)への議題登録 1-3 受入にあたっての要望の評価とほかの主要な決定 1-4 業務受入会議の決定の内部伝達 1-5 業務受入会議の決定についての要望者との話し合い
(ii) 業務開始	2-1 業務の開始、2-2 要求者との会合、2-3 業務設計の開始の承認 2-4 関係省庁への通知
(iii) 任務分担の設計	3-1 業務設計、3-2 設計についての合意 (Design Matrix の策定を含む) 3-3 評価方法、報告書、データについての合意を得るための要求者との会合 3-4 報告書発行日の承認
(iv) メッセージ合意	4-1 データ収集と分析、4-2 メッセージの作成、4-3 メッセージについての合意、 4-4 文書形式の合意、4-5 要求者との意思疎通
(v) 第一担当者の承認 (報告書原案の作成後)	5-1 報告書原案の作成、5-2 報告書原案の更新、 5-3 報告書原案の承認 (第一担当者による報告書原案の審査と承認)
(vi) 各省庁のコメント拝受	6-1 第二担当者の同意 6-2 各省庁への送付 6-3 業務審査会議に対して各省庁への送付を周知
(vii) 報告書の発行	7-1 各省庁のコメントの反映 7-2 各省庁のコメントの扱いの審査 7-3 最終承認、7-4 要求者への事前通知、7-5 最終プロセス (印刷、インターネットへの掲載等)、7-6 報告書発行後の作業 (少なくとも年一回、報告書が実現した便益や改善効果についてレビューする等の作業)

(出典) 1. 黒田 (2003) 資料 1 より転載 (一部筆者が修正)。

2. Electronic Assistance Guide for Leading Engagements (EAGLE)

図1 アメリカ会計検査院（GAO）の組織図



(出典) <http://www.gao.gov/about/workforce/orgchartdet.html>

(3) GAO内部におけるメタ評価（報告書の質的管理）の実際

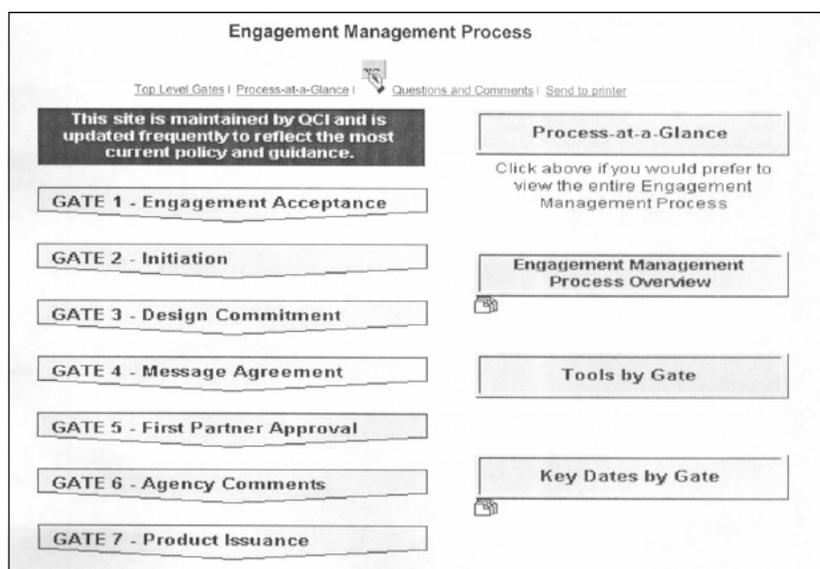
GAOは毎年、1,000冊におよぶ報告書（Performance Audit Reportを含む）を公表するが(Smith 2005)、報告書の質的管理はどのように行われているのだろうか。少し古くなるが、金本（1990）が次のように記載している。「報告書の内容の正確性を確保するために、GAOはきびしい内部チェックのシステムを持っている。この内部チェックは何層にもわたって行われるが、最も重要なのはGAOでreferencingと呼ばれている手続きである。すべての報告書は他の調査官（evaluator）の検査を受けるが、この検査は報告書の一行一行についてデータの裏付けと論理性をチェックするものである。また、この検査では調査に使ったすべての内部資料や作業メモが検査者に渡され、それらを用いて検査が行われる。このreferencing以外にも、管理職によるチェックや経済的問題に関しては内部のエコノミストによるチェックが行われ、通常は、10～12ヶ月の調査に1～3ヶ月を費やし

て内部検査が行われる。」

この内部チェックの手続きを具体的に定めたのが、既出の「取り決めのための電子支援ガイド」(EAGLE)であると言える。一連の作業の中で、メタ評価(報告書の質的管理)に関する作業は2回行われる。第1回目は、「5-3 報告書原案の承認(第一担当者による報告書原案の審査と承認)」である。第2回目は、「6-1 第二担当者の同意(第二担当者の報告原案の審査と第一担当者への同意)」である。さらに、メタ評価ではないが、関係する各省庁へ報告書原稿を送付してコメントを拝受するとしている(ただし自動的に反映されるわけではなく、別途コメントの取り扱いを審査するとしている)。

では2回行われるメタ評価ではどのような基準が用いられているのであろうか。EAGLEはGAO内部向けのイントラネットであり、外部からアクセスすることはできない(図2参照)。ただしEAGLEは、これに関して公表された二人の担当者が用いる共通の基準は、「GAOの主要価値」「専門的基準」であり、「GAOの主要価値」として、(i)説明責任(Accountability)、(ii)誠実(Integrity)、(iii)信頼(Reliability)があげられている(黒田2003)。

図2 「取り決めのための電子支援ガイド」(EAGLE)のトップページ



(出典) Patrick (2006). Figure 3.4

「専門的基準」として何が用いられているかは明らかではないが、GAOが行うすべての評価は有名な「GAOイエローブック」(GAO Yellow Book. 正式名称 Generally

Acceptable Government Accountability Standards (GAGAS) に従うことになっており (Sherman 2009)、同イエローブックが定める基準がその主要な部分になっていることは想像に難くない。ただし今回は現地調査を実施しなかったため、GAO 内部で利用されている EAGLE を直接参照することはできなかったことから、今後の現地調査で確かめられる必要がある。

表2 GAOの Yellow book で定められている基準
(Chapter 2 Auditors' ethical responsibility からの抜粋)

- 一般的基準
 - 独立性(Independence)
 - 専門的判断(Professional Judgment)
 - 高い能力(Competence)
 - 質的管理及び質的保証(Quality Control and Assurance)
- 倫理的基準
 - 公共の福祉 (The public interest)
 - 専門家としての振る舞い (Professional behavior)
 - 誠実 (Integrity)
 - 客観性 (Objectivity)
 - 政府の情報・資源・職位の適切な利用 (Proper use of governmental information, resources, and position)

(出典) GAO (2006)

(4) GAOにおけるメタ評価 (評価統合) の実際

GAO が手がけたプログラム評価のいくつかは Synthesis report である。つまり、いくつかの評価報告書をレビューして統合した報告書である。著名な評価研究者であるエレノア・チェリムスキーが、GAO のプログラム評価・手法局の局長 (Director) だった時期に、GAO のプログラム評価・手法課 (Program Evaluation and Methodology Division) によって、The Evaluation Synthesis と題する報告書が作成されている。その中では「評価統合」の意義を次のように述べている。

「評価統合 (Evaluation synthesis) は、特別のプログラムに関するクライアント (Clients) (議会等の評価委託者) の情報ニーズを満たすために行われ、いくつかの別々の報告書から、確認された事実を統合する体系的な手続きである。それは、いくつかの共

通の評価質問をあらかじめ設定して、違う時間、違う場所で、違う人々によって作成された評価報告書から評価結果を収集して統合する作業である」(GAO 1992: 1 & 6)

そして、GAO は、今までどのようにプログラムが実施されどんな効果が実現したのかわかりたいという議会の要求に対して回答を出すために、評価統合を用いてきたとも述べている。なお、評価統合の対象となるもとの報告書は、GAO が作成した評価報告書、各省庁が作成した評価報告書であり、「可能な限り多数の評価報告書を収集する」としている (GAO 1999)。評価統合により、ある特定の介入行為 (プログラム) がもたらす平均的なインパクト幅を推定することができる。また、個別の介入行為 (プログラム) はそれぞれの場合でいろいろな外部要因から影響を受けるが、複数の評価結果を統合することによりそれら外部要因による影響が緩和されて、より一般的なインパクト幅を推定することが可能になるのだ。

複数の評価報告書を収集して、比較したり総合的に分析することには、ひとつの評価報告書をメタ評価するのとは別の意味がある。複数の報告書を統合分析することで初めて可能となる利点があると指摘し、主に以下の利点を列挙している。

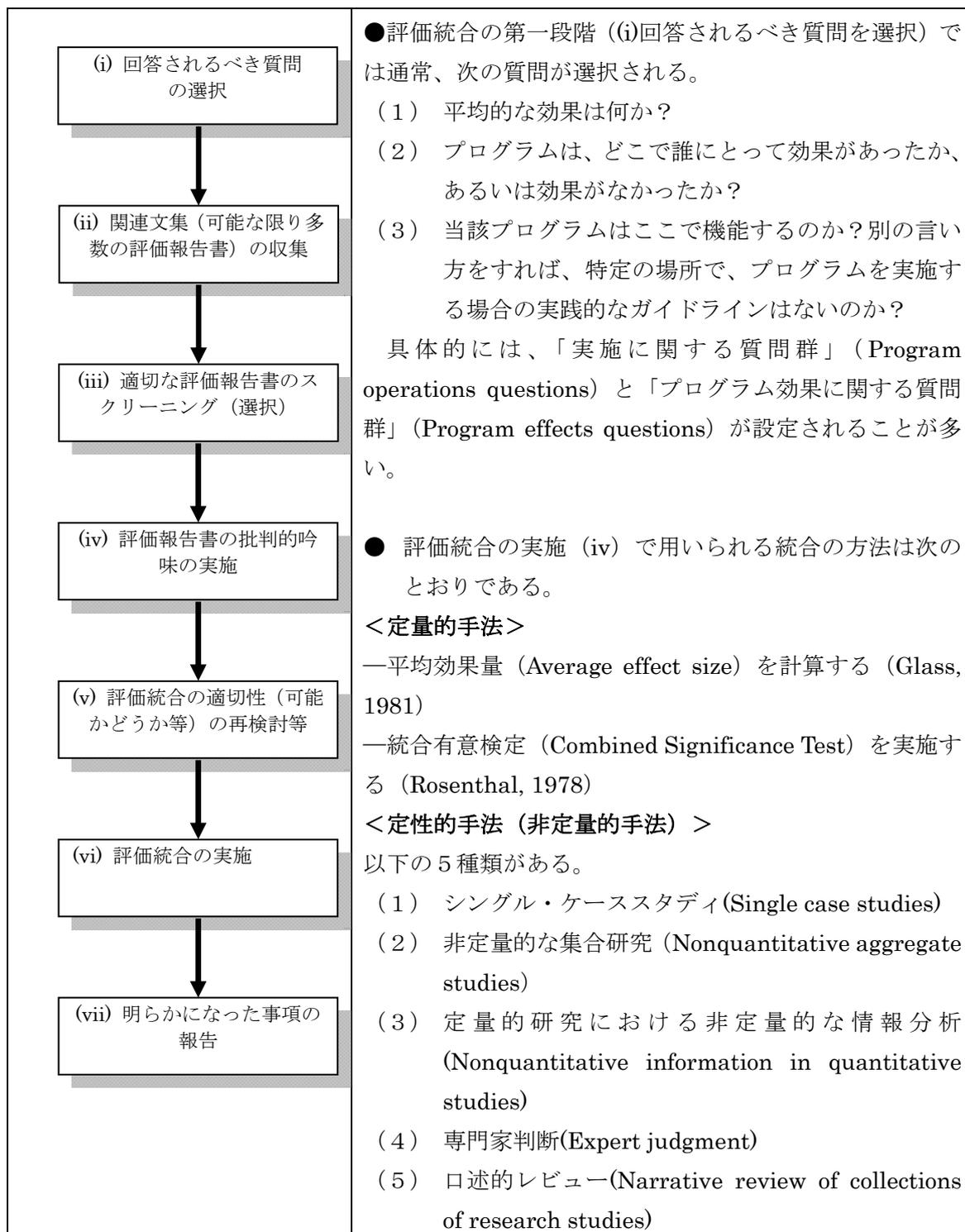
表3 メタ評価 (評価統合) の利点

- | |
|--|
| (ア) 受益者 (Impactees) のタイプと介入のタイプを一致させて効果を説明する |
| (イ) 特に重要な介入行為を峻別して効果を説明する |
| (ウ) 矛盾する結果を説明する |
| (エ) 重要なアウトカムに関して「絶対的効果」のみならず「相対的効果」も説明する |
| (オ) 介入行為による効果の安定性を査定する |
| (カ) 適用したリサーチデザインについて査定する |

(出典) GAO (1999)

次のページに、評価統合として適用される作業段階について図示した。また、必要な評価質問および定量的・非定量的統合手法についても列挙した。なお、評価統合でも「価値判断が行われるので、評価者は伝統的な口述によるレビューに傾きがちになるだろうがそれは間違っている」と指摘している。評価統合により、(表4で列挙したような) 統合手法を駆使して、可能な限りパワフルな回答を得るべきだと述べている。

表4 GAOによる評価統合 (Evaluation Synthesis) の手順



(出典) GAO (1999)

4 アメリカ行政管理予算庁（OMB）によるメタ評価の実践状況

（1）GPRAによるプログラム評価の義務付け

1993年のGPRAの導入により、各省庁には、戦略計画の策定・実施・モニタリングだけでなく、プログラム評価の実施も義務付けられた。

（2）OMBによるメタ評価：PARTの導入

一方、OMBでは、「施策の査定と格付けツール」(Program Assessment and Rating Tool (PART))という仕組みを用いて、GPRAで義務付けられた戦略策定やプログラム評価の実施が適切に行われているかどうかを評価するとしている。各省庁で実施されたプログラム評価の結果を評価するということは、すなわちメタ評価であり、その基準がどのようになっているかを把握することが本研究にとって重要である。OMBでは、各省庁を担当する職員が、メタ評価の作業にあたる。下表にPARTで定められている設問と配点についての概略を述べる。

表5 PARTの共通設問と配点基準

セクション1：プログラムの目的とデザイン（設問数5問、ウェイト20%） プログラムの目的とデザインが、明確かつ確固たるものかを評価する。
セクション2：戦略計画（設問数8問、ウェイト10%） プログラムが長期および年次に関する妥当な目標と指標を持っているかを評価する。
セクション3：プログラム・マネジメント（設問数7問、ウェイト20%） 財政面や改善努力を含め、省庁のプログラム・マネジメントを格付けする。
セクション4：プログラムの成果とアカウンタビリティ（設問数5問、ウェイト50%） 目標・指標に照らして、または他の評価を通じて、プログラムの実績を格付けする。

（出典）左近（2008）

「施策の査定と格付けツール」(PART)の導入経緯とその運用については、左近（2008）が適切に解説しているので別途参照されたい。以下に概略を述べる。

導入以来10年余りが経過したGPRAに対する主な批判は、多数の指標と膨大なデータ（指標値の経年変化等）を産出した一方で、それらが予算要求や予算査定と必ずしも関連しているわけではないため、「政策コストと成果の関係（費用対効果）が見えにくく、結果

として政策資源の配分や決定や日常のマネジメントにおいて十分に情報が活用されていない」という点だったとされる（左近 2008）。こうした状況を改善するために導入されたのが「施策の査定と格付けツール」（PART）である。毎年、各省庁が 5 分の 1 のプログラムを評価対象として選定して、まずは省庁自身が PART を適用して評価を行うことにより評価報告書を作成して OMB に提出する。OMB では、予算審議官（Examiner）が審査および格付けを行い、その結果に基づいて OMB 内部で予算調整を進め、最終的に議会に提出する大統領予算にもその結果を添付することになっている（左近 2008）。

つまり、PART は、プログラムの目的・デザイン・戦略計画・予算の出来具合および成果が出ているかが評価されるので、プログラム評価のツールであると言える。さらに、プログラムの成果に関して行われた評価結果を、再び OMB がチェックする作業があり（後述の質問 2.6 と質問 4.5）、その点に注目すればメタ評価の手法であるとも言える。本解説では、後者のメタ評価に重点を置いて解説する。

なお、PART は 4 つのセクションに分かれており、ウェイトもそれぞれ違う。たとえば、「セクション IV:プログラムの成果とアカウントビリティ」に対するウェイトが 50% と非常に高くなっている一方で、同セクションの平均点は 4 セクションの中でもっとも低くなっていることが注目される（他のセクションが 76-87 点であるのに対して、セクション IV は 49 点（FY2009））。

次のページに「予算編成に用いた PART の共通設問と配点基準」を掲載した。2005 年度用のものであるが、現在の 2008 年度用のものと違いはない。この中で、プログラム評価に関する記載があるのは、質問 2.6 と質問 4.5 である。どちらも、各省庁自身によってプログラム評価がなされているかどうかに関して査定することになっている。したがって、これはメタ評価（評価の質的保証）にあたる作業だと言える。

表6 予算編成に用いたPARTの共通設問と配点基準（2005年用）

<p>セクション I:プログラムの目的とデザイン 20%</p> <p>(プログラムの目的とデザインが、明確でかつ確固たるものかを評価)</p> <p>1.1 プログラムの目的は明確か。</p> <p>1.2 プログラムは、現存する問題・利害・ニーズに対応するものか。</p> <p>1.3 プログラムは、連邦政府、州政府、地方政府や民間における他の取組と重複しないようにデザインされているか。</p> <p>1.4 プログラムのデザインにおいて、プログラムの効果や効率性を損なうような欠点はないか。</p> <p>1.5 プログラムには、その資源が意図した受益者に到達するように、あるいは、プログラムの目的に直接的に対応するように、効果的な目標が設定されているか。</p> <p>セクション II:戦略的計画 (Strategic plan) 10%</p> <p>(プログラムが、長期および年次に関する妥当な目標と指標を持っているかを評価)</p> <p>2.1 プログラムには、限られた数の具体的な長期的業績指標（プログラムの成果に関するものであり、かつプログラムの目的を意義ある形で反映するもの）が設定されているか。</p> <p>2.2 プログラムには、長期の業績達成に向けて、野心的な目標やタイムフレームがあるか。</p> <p>2.3 プログラムには、限られた数の具体的な年次業績指標（プログラムの長期目的の達成に向けた進展状況を示しうるもの）が設定されているか。</p> <p>2.4 プログラムには、年次の業績に向けて、ベースラインと野心的な目標があるか。</p> <p>2.5 すべてのパートナー（交付金の受け手、委託先、費用分担者、政府における他のパートナー等）は、プログラムの長期的および年次の目的の達成に向けて、業務を実施しているか。</p> <p>2.6 適切な範囲・質を有する独立した評価が、定期的にもしくは必要に応じて、当該プログラムの改善や、有効性、問題・利害・ニーズに照らした妥当性等の評価のために、実施される予定であるか。</p> <p>2.7 予算要求はプログラムの長期および年次の業績目標の達成状況に応じたものとなっているか。プログラムに必要な資源は、その予算において、完全かつ透明性のある形で提示されているか。</p> <p>2.8 プログラムは、その戦略的計画における課題に対して、意義ある対策を講じてきたか。</p> <p>セクション III:プログラムのマネジメント 20%</p> <p>(財政面や改善努力を含め、省庁によるプログラムのマネジメントを格付け)</p> <p>3.1 省庁は、時宜を得たかつ信頼性のある業績情報を、定期的に収集しているか。また、省庁はこれらの情報を、プログラムを管理して業績を改善することに活用しているか。</p> <p>3.2 連邦政府のマネージャーやプログラム実施のパートナーは、プログラムの費用、スケジュール、業績等に関して責任を果たしているか。</p>
--

- 3.3 資金は時宜を得て出されたか。また、意図した目的のために費やされたか。
- 3.4 プログラムは、その実施に際して効率性（費用対効果）を測定しかつ実現するための手続き（例えば、市場化テスト、コスト比較、IT 関連の改善、適切なインセンティブ等）を持っているか。
- 3.5 プログラムは、関連する他のプログラムの調整や協力が効果的になされているか。
- 3.6 プログラムは、財政面でのマネジメントを実施しているか。
- 3.7 プログラムは、マネジメント上の欠陥に対して、意義ある対策を講じてきたか。

セクション IV:プログラムの成果とアカウンタビリティ 50%

（目標・指標に照らして、または他の評価を通じて、プログラムの実績を格付け）

- 4.1 プログラムは、長期の業績目標の達成に向けて、適切な進展を示しているか。
- 4.2 プログラム（およびそのパートナー）は、年次の業績目標を達成したか。
- 4.3 プログラム実施の効率性（費用対効果）は、毎年の業績目標の追及に際して、改善していると言えるか。
- 4.4 プログラムの業績は、政府および民間等による他の同様の目的を持つプログラムと比べて好ましい状況と言えるか。
- 4.5 適切な範囲・質を有する独立した評価は、当該プログラムが有効であり、その目的を達成していると評価しているか。

注：セクション内のウエイトは OMB との協議により省庁ごとに設定される。

（出典）左近（2008）の訳を採用（一部、筆者（佐々木）が変更）。また、OMB（2008）も参照した。

（3）PARTにおけるメタ評価の詳細

次に、設問 2.6 と設問 4.5 をより詳細に見てみる。まず、質問 2.6 の具体的な質問は以下のようになっている（OMB 2008: 30-32）。とくに注目される点としては、評価の「質」を確保するために、厳格な手法が用いられることが推奨されていることと、独立性が重視されていることである。

表 7 設問 2.6 の詳細

2.6 適切な範囲・質を有する独立した評価が、定期的にもしくは必要に応じて、当該プログラムの改善や、有効性、問題・利害・ニーズに照らした妥当性等の評価のために、実施される予定であるか。

目的：パフォーマンス情報に関するギャップを埋めるために、バイアスのない評価を定期的にあるいはニーズがあるときに参加できることを保証する。

YES のための要素：YES のためには以下について証拠を示し明確に説明できることが必要である。

✓ プログラムは、以下に記載された基準を満たすような評価を実施すべきである。

— 高い質

— 適切な範囲

— バイアスがない（中立である、独立である）

— プログラムの改善を支援するために定期的に実施される

(1) 質 (Quality)：評価は、プログラム効果に関する情報を提供するために、十分に厳格でなければならない。省庁は、（選択した評価手法によって）最も厳格な証拠を産出する評価手法を選択したことを説明できねばならない。たとえば、ランダム化比較試験などは、インパクトを特定することに特に向いている。他のタイプの評価アプローチ、たとえばよく計画された擬似実験デザインなどは、プログラムのインパクトに関して有用な情報を提供するだろう。

(2) 範囲 (Scope)：評価は、プログラムの一側面ではなく、プログラム全体の効果を総合的に評価していなければならない。

(3) 独立性：独立であるためには、利益相反（Conflict of interest）がなく、かつバイアスがない（＝不偏不党の）者によって評価が行われなければならない。もし省庁やプログラムが第三者に評価委託する場合には、その第三者は十分に独立的でなければならない。評価が、省庁の検査官（Inspector General）あるいはプログラム評価担当部署によって行われる場合には、それらは独立であるとみなされる。

(4) 頻度：プログラム評価の情報の鮮度を保つために、定期的に更新されねばならない。

エビデンス/データ：エビデンスとして、プログラム評価計画（実施日程）、および評価概要（評価のタイプ、範囲、質、独立の評価者を選択する基準等）が提出されねばならない。

（出典）OMB（2004）

次に、設問 4.5 の具体的な質問は以下のようにになっている（OMB 2008: 55-60）。基本的に質問 2.6 を踏襲しているが、事後評価に限定しており、実際に厳格かつ客観的な手法に

よって評価が実施されたかどうか注目している。

表 8 設問 4.5 の詳細

<p>4.5 適切な範囲・質を有する独立した評価は、当該プログラムが有効であり、その目的を達成していると評価しているか。</p> <p><u>目的</u>：独立かつ総合的な評価が、プログラムは効果的かどうかを判断していること。この質問は、特に、定量的な実績測定方法を決定することが本質的に難しいプログラムにとって重要である。</p> <p><u>YES のための要素</u>：YES のためには以下について証拠を示し明確に説明できることが必要である。</p> <p>✓ 独立かつ総合的な評価が、プログラムは効果的かどうかを判断していること。評価は、質問 2.6 で定義されている質的基準、範囲基準、独立基準の 3 つの基準に従うことが求められる。</p> <p>この質問では「Not applicable」は取り得ない。全てのプログラムは、質、範囲、独立の各要素を満たすように評価を実施せねばならない。詳しくは 2.6 を見よ。</p> <p><u>エビデンス/データ</u>：エビデンスとして、学術機関、リサーチ機関、政府契約、その他の独立機関、GAO、あるいは査察官 (Inspectors General) が関係した評価の結論の議論の要約が提出されるべきだ。なお、質問 2.6 とちがって今後実施されるであろう評価計画はエビデンスにはなりえず、実際に実施された評価結果だけがここではエビデンスに該当する。</p>
--

(出典) OMB (2004)

(4) PARTにおけるメタ評価の詳細

さて、質問 2.6 で言及された「厳格な評価」とは何を指すのであろうか。OMB の考えるプログラム評価は、「何がプログラムの効果に関する強力なエビデンスを構成するか? (What Constitutes Strong Evidence of a Program's Effectiveness?)」で詳細に解説されている (OMB 2004)。それよると次の 5 種類があげられており、エビデンスが強力である順番 (=厳格である順番) に並べられていると考えられる。

表9 OMB「何がプログラムの効果に関する強力なエビデンスを構成するか？」の概略

<p>●ランダム化比較試験 (Randomized Controlled Trials (RCT))</p> <p>ランダムアサインメント (無作為割当) によってプログラムを適用するグループと適用しないグループに分けることにより、純粋にプログラムの効果を測定する評価デザイン。</p> <p>●Direct Controlled Trials</p> <p>効果に影響を及ぼすと考えられる外部要因を評価者が直接管理することができる場合に、直接管理して影響が及ばないようにしてから、プログラムの効果を測定する評価デザイン。宇宙開発や兵器開発で用いられる。</p> <p>●Quasi-Experimental</p> <p>ランダム化を用いないけれども近似的に比較グループを構築して、プログラムの効果を測定する評価デザイン</p> <p>●Non-Experimental Direct Analysis</p> <p>プログラムを適用したグループだけを観察する評価デザイン。代表的なものは「事前・事後比較デザイン」</p> <p>●Non-Experimental Indirect Analysis</p> <p>独立した専門家パネルを設置して専門家が直接観察により評価を行う。</p>

(出典) OMB (2004)

上記の図にあるように OMB のガイドラインは、ランダム化比較試験をベストな方法として推奨していることが分かる (GAO 2005: 27)。

なお、OMB のメタ評価者 (OMB examiner) と省庁側の評価担当者 (Federal evaluation officials) の間で、評価の質をどのように査定するかという一連の議論が行われている。まず、2004 年夏に OMB は Interagency Program Evaluation Working Group を立ち上げて一連の議論を行った。省庁側の評価担当者は OMB が「厳格な評価」を過度に狭く定義していることに懸念を表明している (GAO 2005: 28-29)。

その後、2005 年春に再び集まり、彼らをもっとも適切だと思う評価アプローチについてまとめた。それが次の表 10 である。

表 10 省庁側の評価担当者の見方：プログラム効果の評価のためのデザイン

- プロセス／アウトカムのモニタリング・評価
事前に設定された目標（値）と比べる。
- 擬似実験モデル：単一グループ
実施グループのみで、事前事後比較を行う。
- 擬似実験モデル：比較グループ
近似した比較グループを設定して、比較を行う。
- ランダム化比較試験
ランダム化により介入効果を測定する。

しかし、OMB は 2007 年の段階でも、ガイダンスの大幅な見直しは行っておらず、プログラム効果を評価するためのベストなデザインとしてランダム化比較試験を推奨している。

5 アメリカ教育省（DOE）によるメタ評価の実践状況

アメリカにおけるメタ評価の利用として最も注目される動向の一つがアメリカ教育省による「エビデンス（科学的根拠）に基づく研究」である。以下に概略を述べる。なお、この動向については、田辺（2006）がよくまとめているので、別途参照されたい。

（1）NCLB法と「エビデンス（科学的根拠）に基づく研究」

ブッシュ政権はさまざまな改革に取り組んだが中でも重視したのが教育改革である。アメリカの教育レベルの長期的凋落傾向に危機感を抱いた同政権は、限られた予算を有効に活用するために、確かに効果があると認められた教育施策を明らかにしてそうした施策に重点的に予算配分することを目指した。

こうした発想を具体化したのが 2002 年に超党派の支持を受けて成立した「子供を一人も落ちこぼれにしない法案」（No Child Left Behind Act of 2001 : NCLB 法）である。アメリカの教育水準を引き上げることを目的とし、学区や学校に大幅な裁量権を認めることと引き換えに、子供の学力向上という「成果」を厳しく問う内容であった。そして「成果」を挙げた学区や学校には予算を増額するとともに、「成果」をあげた取り組みを全米の学区や学校に普及することを目指していた。

ではどうやって「成果」を明らかにするのであろうか？ここに NCLB 法の特徴がある。

NCLB 法は、「エビデンスに基づく研究」(Evidence-based Research) を重視している。そして、「エビデンスに基づく研究」で有効性(つまり「成果」)が確認された施策のみに連邦予算を配分するように求めている。

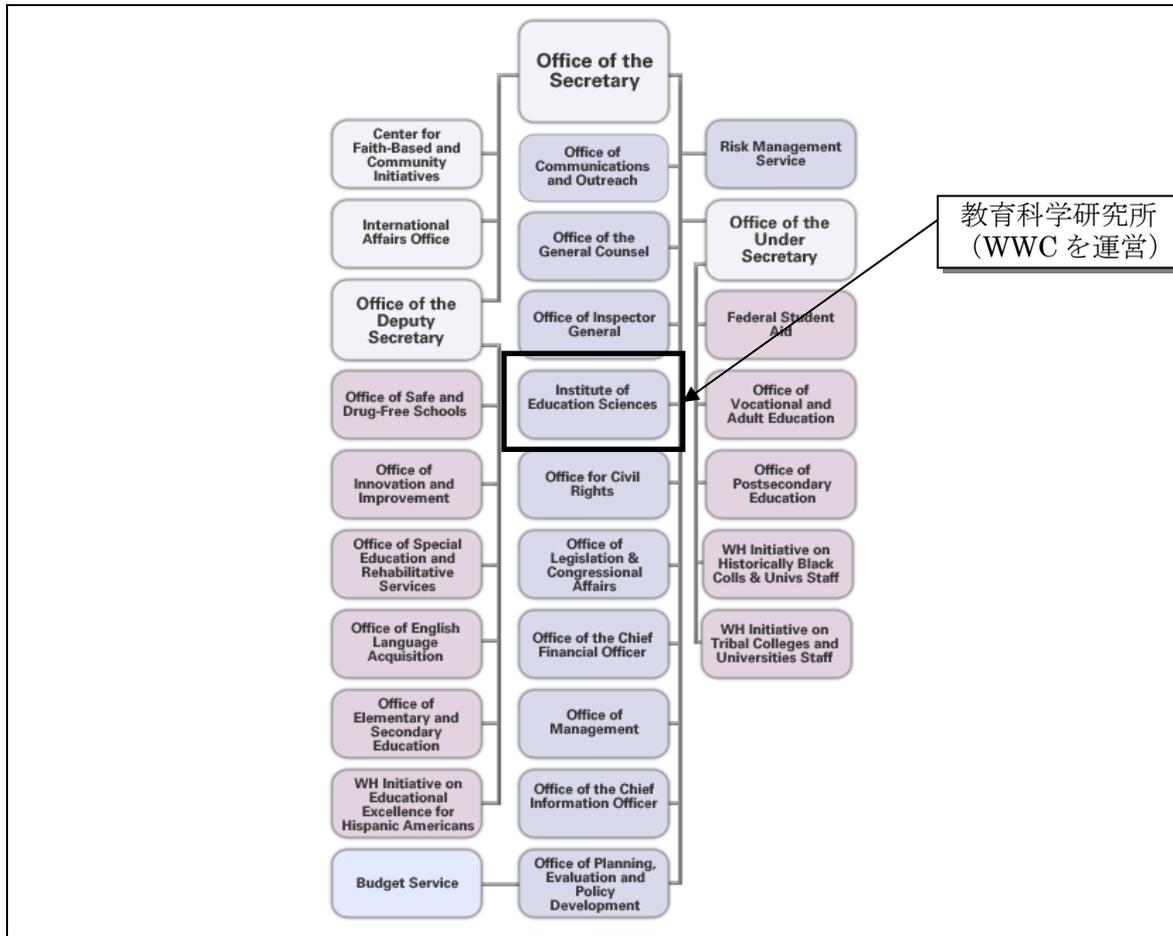
それでは「エビデンス」とは何を指すのであろうか。NCLB 法 9101 条では、「教育活動や施策に関する妥当で信頼できる知識を得るための厳格・体系的かつ客観的な手続きの適用を含む研究」であって、具体的には、観察や実験に基づく実証研究、仮説検証を伴うデータ分析、実験・疑似実験デザインを用いた研究などが含まれるとされた(田辺 2006)。ただし、同条項の定義は必ずしも明確ではない一方で、連邦予算は「エビデンスに基づく研究」のみに配分されることが求められたことから、教育現場で疑念と混乱が広がったとされる。

この状況を受けて、アメリカ教育省により、「厳密なエビデンスによって裏付けられた教育実践の識別と実施：ユーザーフレンドリーガイド」(Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide) が策定されて公表された。同ガイドでは、エビデンスを産出するための方法や制約について詳細に解説されているが、そのなかで最も厳格なエビデンスが得られるのは、ランダム化比較試験(いわゆる RCT)だと明記された。

(2) アメリカ教育省によるメタ評価の仕組み：What Works Clearing House

エビデンスに基づく教育を普及させるためには、蓄積された情報を体系的に整理して、教育関係者に提供する仕組みが必要である。その必要を満たすために、教育省が創設したのが、What Works Clearing House (WWC) という仕組みである。邦訳すれば「何が機能するのかの情報拠点」ということになる。WWC は、教育省傘下の教育研究所(Institute of Education Science) がキャンベル共同計画(Campbell Collaboration) と共同で開発・運営を行っている(教育省の組織図(図 3)を参照)。WWC は、教育に関わる政策、施策、活動、製品の有効性について、最新で質の高いレビューを提供し、エビデンスに関する信頼できる情報源となることを目的としている(田辺 2006)。

図3 教育省の組織図



教育科学研究所
(WWCを運営)

(出典) U.S Department of Education のウェブサイト (<http://www.ed.gov>)

(3) WWCのハンドブックに記載されたメタ評価の手順

WWCでは、ランダム化比較試験などを利用して作成・提出される膨大な評価報告書のなかから、特定の施策に関する報告書を選定して収集し、それぞれの評価報告書に関してメタ評価を行う。そして基準を満たした複数の報告書に記載された有効性の度合いの組み合わせにより、当該施策の一般的効果を「好ましい効果」(Positive effect) から「否定的な効果」(Negative effect) まで6段階で結論するとしている。そして選定から最終結論を得るまでのメタ評価の過程を公開し、教育関係者に対して、何が機能し何が機能しないのか、の情報を提供している。教育関係者はそれを参照して、機能するとされた施策を、自分の学区・学校で適用するための判断材料にすることになる。

WWCの「手続き・基準ハンドブック」では、メタ評価のプロトコルを次のように定めている。なお、同ハンドブックでは「メタ評価」という用語を使用しているわけではなく、

レビューという一般的な用語を使用していることに注意する必要がある。

表 11 「手続き・基準ハンドブック」に記載されたメタ評価の手順

<ol style="list-style-type: none">1. レビュー対象分野の特定<ol style="list-style-type: none">(1) レビュー対象分野の特定(2) 調査報告書の検索(3) (レビュー対象分野としての) 的確性の評価2. 調査報告書レビュー作業<ol style="list-style-type: none">(1) レビュープロセス (The review process)(2) エビデンスの基準 (Evidence standards)<ol style="list-style-type: none">(i) 研究のデザイン (Study Design)(ii) サンプル集団からの脱落度合 (Attrition)(iii) RCT による同質性の確保 (Establishing Equivalence in RCTs)(iv) 対抗要因 (Confounding Factor)(v) 基準を満たさなかった理由 (Reasons for Not Meeting Standards)(vi) 訂正・修正 (Correction and Adjustments)3. レビュー結果のとりまとめ<ol style="list-style-type: none">(1) ドラフト報告書の作成(2) クオリティ保障レビュー (Quality Assurance Review) の実施(3) 独立評価スタッフと外部専門家によるレビュー(4) 最終報告書の作成と公表

(出典) Institute of Education Sciences, Department of Education (2008)

以上の手続きに基づいて、個別の評価報告書に関してメタ評価を行うわけであるが、複数の報告書に記載された有効性の度合いの組み合わせにより、当該施策の一般的効果を以下の通りの6段階のいずれかを用いて結論するとしている。

表 12 WWCの効果レーティング

<p>好ましい効果 (=正の効果) (Positive Effects)</p>	<p>好ましい効果を示す強力なエビデンスがあるとともに、それを上回るような反対のエビデンスがない。</p> <ul style="list-style-type: none"> 二つ以上の研究が統計学的に有意な正の効果を示しており、少なくともそのうちの 하나가、WWCエビデンス基準を満たしている。 統計的に有意な負の効果を示す研究がない。
<p>部分的に好ましい効果 (=部分的に正の効果) (Potentially Positive Effects)</p>	<p>好ましい効果を示すエビデンスがあるとともに、それを上回るような反対のエビデンスがない。</p> <ul style="list-style-type: none"> 少なくとも一つの研究が統計学的に有意な正の効果、あるいは本質的に重要な正の効果を示している。 統計的に有意な負の効果を示す研究がないとともに、本質的に重要な正の効果を示す研究と同数あるいはそれよりも少ない数の研究が中間的な効果 (Intermediate effects) を示している。
<p>混合的な効果 (Mixed Effects)</p>	<p>一貫しない効果のエビデンスがある。</p> <ul style="list-style-type: none"> 少なくとも一つの研究が統計学的に有意な正の効果あるいは本質的な重要な正の効果を示すと同時に、少なくとも一つの研究が統計学的に有意な負の効果あるいは本質的に重要な負の効果を示している。ただしそうした負の効果を示す研究の数が、そうした正の効果を示す研究の数を超えていない。 少なくとも一つの研究が統計学的に有意な効果あるいは本質的な重要な効果を示すと同時に、それらの研究の数よりも、中間的な効果を示している研究の数の方が多い。
<p>効果は明確ではない (No Discernible Effects)</p>	<ul style="list-style-type: none"> 正の効果に関する肯定的なエビデンスがない (No affirmative evidence of effects)。 統計的に有意あるいは本質的に重要な効果を示す研究がない (正の影響でも負の影響でも)。
<p>部分的に否定的な効果 (=部分的に負の効果) (Potentially Negative Effects)</p>	<p>否定的な効果に関するエビデンスがあるとともに、それを上回るような反対のエビデンスがない。</p> <ul style="list-style-type: none"> 少なくとも一つの研究が統計学的に有意な負の効果、あるいは本質的に重要な正の効果を示している。 統計的に有意な負の効果を示す研究がない。あるいは統計学に有意な負の効果あるいは本質的に重要な負の効果を示す研究の数が、統計学的に有意な正の効果あるいは本質的に重要な正の効果を示す研究の数よりも多い。
<p>否定的な効果 (=負の効果) (Negative Effects)</p>	<p>否定的な効果に関して強力なエビデンスがあるとともに、それを上回るような反対のエビデンスがない。</p> <ul style="list-style-type: none"> 二つ以上の研究が統計学的に有意な負の効果を示しており、少なくともそのうちの 하나가、WWCエビデンス基準を満たしている。 統計的に有意な正の効果を示す研究がない。

(出典) Institute of Education Sciences, Department of Education (2008)

6 結論と考察

アメリカで提案されて普及した「メタ評価」の本来の意味は、厳格な方法・デザインによって質の高いエビデンスが得られたかどうか、そして得られたエビデンスによって論理的・合理的な結論が導出されたかどうかを検証する作業ということである。また、その意味が拡大し、メタ分析に近い「評価統合」という意味合いで用いられる場合でも、まずは、複数のもともとの評価報告書の質がチェックされることになる。本章では、アメリカにおけるこうしたメタ評価の実践を、GAO、OMB、そして最近最も注目されているアメリカ教育省の事例を用いて分析した。日本における政策評価の実践に対しては以下の示唆を得ることができる。

(1) 「日本版メタ評価チェックリスト」の検討・策定

本章で詳細に解説したいくつかのチェックリストを踏まえて、日本の状況に合うメタ評価チェックリストを総務省が検討・策定して公表することが勧められる。総務省自身が使用するほか、他の省庁でもそれらを利用することにより、自身の評価報告書の質をみずからチェックすることができるようになる。また、公益法人や学校法人などでも広く利用されることが見込める。このことを通じて、日本における政策評価の質が高まると考えられる。

(2) エビデンスの質に関する検討

一方で、何が質の高いエビデンスなのかの議論は、アメリカでも未だに論争が続いており、エビデンスの質やそれを得るための評価デザインの厳格さの順位を示すようなリストを作成して公表することには慎重な議論を要する。OMBやアメリカ教育省の例に見られるようにランダム化比較試験（実験デザイン）が最も望ましいという見解が強くなっている一方で、それに反対する評価研究者も多数おり (Donaldson & Christie 2005)、今後の議論の展開を見守る必要がある。

(参考) アメリカにおける評価人材育成

GAOにおける職員の採用基準は変化してきたとされており、以前(1960年代まで)は、会計士の資格を持つものが多かったが、1970年代以降は、公共行政学、公共政策学、経済学、心理学、社会学等を専攻した者に採用対象を拡大してきた(黒田2003)とされる。

アメリカでは大学院教育によって評価人材が供給されている。とくに、公共政策学大学院が主な人材供給源となっている。公共政策学修士課程(M.P.P: Master of Public Policy)で教えられる科目は、どの大学院でもほぼ同一である。通常はコア科目と選択科目の2段構えになっている。コア科目としては、以下の5つの科目で構成されていることが普通である。なお、一部の大学ではこれに加えて、費用便益分析が必修である場合がある。

- ミクロ経済
- 財務管理
- 統計学(基礎編)
- マネジメント
- 公共政策(基礎編)

上記科目は、通常、公共行政学修士(M.P.A. Master of Public Administration)と公共政策学修士(M.P.P. Master of Public Policy)の2つの専攻で共通のコア科目であり、公共政策学修士課程ではさらに以下のような選択科目を履修することになっている場合が多い(ただしこの限りではない)。

- 公共経済・財政
- 統計データ分析(重回帰分析、多変量解析)
- プログラム評価
- 政策分析
- 特定分野の選択科目(教育、保健、行政一般、国際開発等)

さらに、National Association of Schools of Public Affairs and Administration (NASPAA)という団体が、一定の基準を満たした公共政策・公共行政大学院に対して「認証」(Accreditation)を行っており、その認証制度が質的保証の役割を果たしている。主要な公共政策大学院はすべてNASPAAの認証を獲得しているが、認証を得るためには一定の基準が設定されている。そのなかで、カリキュラムに関する基準があり、それを満たすためには上記のようなスタンダードな科目を揃えることが必須になっている。こ

うした事情を受けて、主要な公共政策大学院の科目構成はほとんど同一となっているのだ。また、それぞれの科目で使用される教科書もほぼ定まったスタンダードなものがある。たとえばプログラム評価であれば、Rossi, Lipsey & Freeman (2006). *Evaluation: A Systematic Approach*が幅広く用いられているし、別の教科書を使う場合（例えばWeiss, C. (1999). *Evaluation 2nd edition*）を使う場合でも内容はほぼ同じである。したがって、どの大学院を卒業しても同じ技術と同じ知識を身につけた人材が輩出されることになっている。その知識の例が、ロジックモデリングや実験デザイン・疑似実験デザインであり、公共政策大学院を卒業して修士号を保持する者で、これらの概念と技法を知らないものはいないと言っても過言ではない。そのように同じ技術・同じ知識を身につけた者が入省してくることが、GAOによるプログラム評価の質を保つ大きな要因になっている。

（参考文献）National Association of Schools of Public Affairs and Administration (NASPAA) (2008)、および主要な公共政策大学院のホームページ

第3章 イギリスにおける政策評価のチェックシステム —PSAシステムに対するチェックシステムを中心として—

南島 和久

本章ではイギリスの政策評価のチェックシステムについて記述する。

最初に本章で議論するイギリスの「政策評価」概念について触れておきたい。日本語（政府の用語法）と英語の「政策評価」は必ずしも一致するものではない。日本語の政策評価には、費用便益分析を中心とする「事業評価方式」、民間企業の目標管理をモデルとする「実績評価方式」、アメリカの会計検査院（GAO）が行っているような政策研究・調査型の「総合評価方式」の3つの異なる評価方式が含まれている。これに対し本章で議論する用語法では、大きくいえば民間企業の目標管理制度を模した業績測定（performance measurement）的な評価とその他の評価とが明瞭に区別されている。

本章は主に前者の「業績測定」的な評価のカテゴリーの、「PSA」（Public Service Agreements：公共サービス協定）や「SDA」（Service Delivery Agreements：サービス提供協定）およびそのチェックシステムを取り上げるものである。

1 イギリスにおける評価

（1）イギリスにおける「評価」

イギリスにおける政策評価について目立った存在であるのは PSA である。PSA についての研究は稲継（2001）や児玉（2001）、あるいは平松（2002）などに詳しい。PSA は 1998 年に導入された各省庁と大蔵省（HM Treasury）の間で合意された公共サービス達成目標のことであり、予算編成とも密接な関係のあるシステムのことである。PSA の制度官庁は大蔵省であり、したがって、PSA の制度運用状況を確認しようとするれば、各省の運用実態と大蔵省のチェックシステムについて論及することとなる。

なお、イギリスにおける政策評価としてはこのほかにも政策レビューや規制評価（影響分析：IA）、会計検査院（NAO）のレビューなどが知られている。日本との比較、あるいは国際比較（とくにアメリカとの比較）を考えるならばこれら各種の評価まで含めた総合的な比較が必要となる点には、あらかじめ注意を喚起しておきたい。

(2) 政治的な文脈－NPM

前置きとしてもうひとつ触れておかなければならないのが、イギリスの政策評価の議論をしようとする場合、しばしば NPM (New Public Management) に言及されることが多い点である。NPM は管理をめぐるドクトリンの集合であるといわれ、論者によっては政策評価の理論的背景ともいわれているものである¹。

NPM とは、公的部門の運営管理にかかる行政改革運動のことであり、20 世紀最終四半期におけるニュージーランドの公的部門改革にその起源を求めることができる (参照、和田 2007)。NPM は第一に経営学や企業経営の手法を、第二に市場重視の視点から近年の経済学 (近代経済学) のアプローチを摂取しようとするものであるので、しばしば「市場メカニズム」と「管理者主義」の組み合わせとして説明される (Hood 1999)。より具体的な改革としては、サッチャー保守党政権時 (1979-90) の規制緩和・民営化、FMI (Financial Management Initiative)、執行エージェンシー化、メージャー政権時 (1990-97) のシテンズ・チャーターや PFI (Private Finance Initiative)、ブレア労働党政権 (1997-2007) 以降の PPP (Public Private Partnership) や予算制度改革や PSA を中心とする評価システムなどをあげることができるだろう。

ところでこのうち本章が主題としている PSA は、保守党から労働党への政権交代期に登場した点に注意を喚起したい。NPM 論ではしばしば、大きな政府や福祉国家の見直しという視点が強調され、サッチャー政権からブレア政権までの改革の連続性が強調される。確かにブレア政権は労働党政権でありながら、サッチャー改革に由来する多くの遺産を継承した。だが、評価システムを考えるにあたって重要な点は、これが NPM の理論的系譜にあったか、改革が継続したかという点よりも、1997 年に保守党から労働党への政権交代が起きたこと、そしてこれに伴って歳出にかかるゼロ・ベースでの見直しがはかられた点の方が重要ではないかと思われる。少なくとも本稿では、PSA を中心とする評価システムについて、その理論的背景よりもこの歳出の全面的な見直しが中期支出計画たる CSR (Comprehensive Spending Review) として実現したことおよびその道具立てとしての PSA の位置づけがあった点に関心を寄せたいと考える。

¹ NPM と政策評価との関係を強調するのは会計学や経営学分野からのアプローチである。これは目標管理等が経営学や管理会計論で論じられること、これを公的部門に導入しようとするものであることなどを背景としている。

表1 イギリスの評価システムの導入経緯

1979年	・サッチャー政権発足
1990年	・メージャー政権発足
1991年	・「市民憲章」制定 ・「グリーンブック」(評価ガイダンス)の初版を公表
1995年	・「資源会計・予算(RAB)」(発生主義会計型財務管理システム)の導入を公表
1997年	・ブレア政権発足(政権交代)、「グリーンブック」の改訂版を公表
1997年	・大蔵省「包括的歳出レビュー(Comprehensive Spending Review)」を実施
1998年	・「サービスマスター・新憲章プログラム(SFNC)」を公表 ・経済財政戦略報告書(Economic and Fiscal Strategy Report)を公表 (3か年の歳出総額の決定及び歳出計画・管理に関する新システム導入方針を明示) ・最初の包括的なSRとなる「1998 CSR(1999-2004年)」を公表 ・「公共サービス協定(PSA)」を公表 ・財政安定化規律を議決・制定(透明性、安定性、責任性、公平性、効率性の5原則の設定)
1999年	・「アウトプットおよび実績に関する分析(OPA)」を公表 (各府省が達成すべき実績(一部に具体的数値目標を含む)を明示) ・「省庁別投資戦略(DIS)」を公表 ・「政府の近代化(Modernizing Government)」構想を公表
2000年	・中央省庁の1999年度決算をRABで作成 ・RABの導入にかかる「政府資源および政府会計に関する法律(GRA)」を制定 ・「2000SR(2001-2004年)」を公表 (新規の歳出3か年計画。2000、2002、2004の3回のレビューはたんに'Spending Review'という) ・PSAの改訂を発表 ・「サービス提供協定(SDA)」を公表(OPAとPSAの一部を統合) ・DISを改訂(このときからPSAの補足説明たる技術覚書(TN)の公表も開始される)
2001年	・中央省庁の2001年度予算をRABで作成(RABの完全施行) ・「シャーマン報告」(大蔵省の監査およびアカウントビリティに関する勧告) ・大蔵省、内閣府、会計検査院、監査委員会、統計局「業績情報の枠組み」発表
2002年	・「2002SR(2003-2006)」を公表、「シャーマン報告」への政府回答を公表
2003年	・「グリーンブック」の第3版を公表
2004年	・「2004SR(2005-2008)」を公表
2005年	・会計検査院、PSAの外部評価を公表(第1回)
2007年	・ブラウン政権発足、「2007CSR(2度目の包括的なSR; 2008-2011)」を公表 ・「新しい業績管理フレームワーク」の導入(「省庁戦略目標(DSO)」の設定)
2009年	・RABの会計基準をUK GAPP(国内基準)からIAS/IFRS(国際基準)へ変更

(出典) 児玉(2001)、総務省行政評価局政策評価官室(2002: 3-4)、平松(2002)他関係資料をもとに作成。

2 PSAのフレームワーク

PSAはブレア政権下で導入され、ブラウン政権下でも継承されている。評価制度の全体像は複数のシステムが複合している。制度の基盤となっているのはRAB (Resource Accounting and Budgeting) と呼ばれる公会計システムである。その上に大蔵省が行うCSRやSR (Comprehensive/ Spending Review) と呼ばれる歳出の見直しと中期予算計画とを合わせたものが稼働している。CSRやSRに関連して各省庁側はPSA (Public Service Agreements) と呼ばれる公共サービス協定を大蔵省との間で取り結んでおり、これを補完するのがSDA (Service Delivery Agreements) という目標管理にかかるサブシステムである。これらのシステムの導入経緯については表1を参照されたい。本文ではこの順序にしたがってそれぞれの中身を具体的に見ていきたい。

(1) RAB (資源会計 (発生主義会計)・予算)

RAB (Resource Accounting and Budgeting) は英語名称からも読み取れるように、「資源会計 (発生主義会計)」と「予算」を組み合わせた統合パッケージである。RABは発生主義会計に基づき、財務諸表等は一般的会計基準 (UK GAAP) に準拠して作成されている²。

RABはメジャー政権下の1995年に『納税者の税金に対するよりよい会計システムのあり方』 (*Better Accounting for the Taxpayers Money: Resource Accounting and Budgeting in Central Government*) という白書の中で提唱された。その後このアイデアはブレア労働党政権に引き継がれ、1998年から導入が始まり、2000年の「政府資源および政府会計に関する法律」 (Government Resources and Accounts act 2000) の制定をまって2001-02会計年度より本格的に導入されている。

RABのアウトラインについてはHM Treasury (2001a) に詳しい。同システムの要点は発生主義会計と予算および成果管理との連結にあるといえ、政府の方針を省庁の計画や予算に落とし込み、効率的・効果的なサービス提供をはかり、その結果をつぎの予算に反

² 参照、東(2009)。なお、東によれば2009-10年度よりUK GAAPから国際会計士連盟のIASB (国際会計基準審議会) が設定する国際会計基準 (IAS/IFRS) に準拠することになるという。2008-09年はUK GAPP準拠とIAS/IFRS準拠の2種類の財務諸表が用意されるとしている。なお、民間企業は2004年の改正会社法よりIAS/IFRS準拠の財務諸表が作成できるようになっている。UK GAAPからIAS/IFRSへの変更は、「2000年に行われたRABの導入以来の大改革」 (東2009: 137) であるという。

映していくという財務管理にかかるプラットフォームを提供しているものである(図1参照)。

図1 RAB概念図



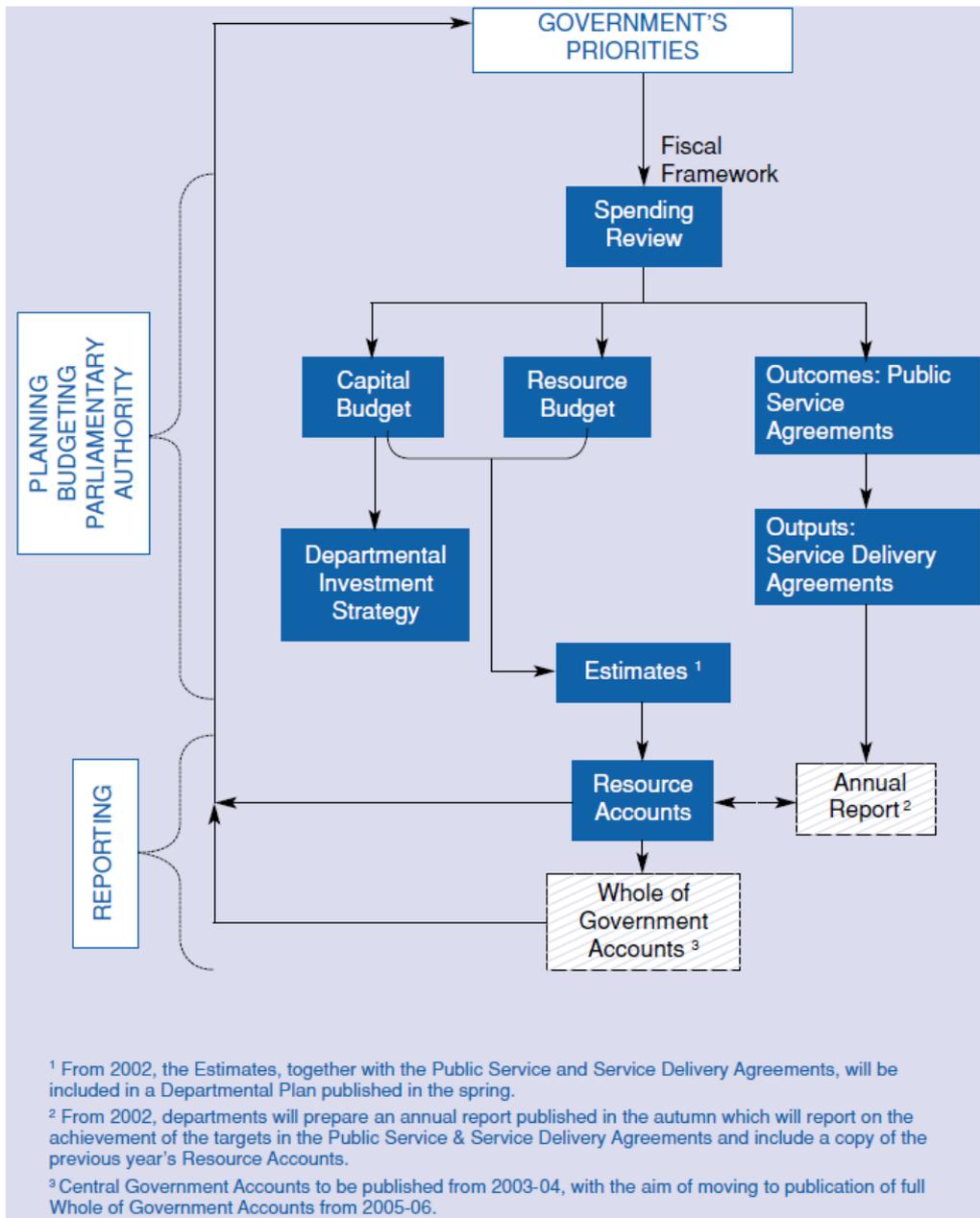
(出典) HM Treasury (2001a: 3)

図1では、国民へのサービス提供をめぐって、政府の優先政策の設定、政府の計画や予算の決定、そして報告が循環構造をなすべきことが理念として掲げられている。RABはこのためのツールであり、プラットフォームとなるものである。

より具体的な RAB の全体像は図2に詳しい。「政府の優先政策 (Government Priorities)」を中心としてこの下に3カ年の計画となる「SR (Spending Review)」が置かれており、さらにその下に予算管理 (Budget) と目標管理 (PSA および SDA) が置かれている。この図をみれば、RAB の中心は予算管理の部分であり、「資源会計 (Resource Accounts)」がその中核に位置づけられていることが分かるだろう。

なお平松 (2002) は RAB が導入されたことについて、「RAB の下では政府活動のアウトプットや業績の分析・評価も今まで以上に行うことになる」と述べている。RABの上では、民間企業の管理会計の各種の手法を活用する環境がいつそう整うことになる。後述するイギリスの評価システムもシャーマン報告以降の内部統制への展開も、この点と密接に関係しているといえるだろう。

図2 RABシステム



(出典) HM Treasury (2001a:5)

(2) CSRとSR

「CSR」と「SR」の違いは、ブレア政権の発足時およびブラウン政権の発足時が「CSR」、ブレア政権の継続中は「SR」であったことから、政権移行時、あるいは長期の抜本的な見直さないし政策目標の設定が「CSR」であり、定期的な見直さないし中期目標の設定が「SR」ということができる。

ところで、「Spending Review」は日本語に置き換えにくい言葉である。先行研究では、「包括的歳出レビュー」（稲継）や「支出審査」（児玉）、あるいは「包括的支出レビュー」（平松）という訳語があてられている。注意を要するのはこれがたんなる「歳出見直し」とどまらず、「計画」的な側面も有していることおよび後者の点がより重要であることである。この意味では SR は独立行政法人の中期目標終了時の評価・分析に新しい中期目標の設定を加えたものという表現に近いだろう。こうした用語法をふまえ、本章では「Spending Review」を「SR」と略記し、個々の SR については「1998CSR」や「2000SR」などと表記することとする。なお、これまでに公表された SR の一覧は表 2 の通りである。

SR のアウトラインについては稲継（2001）および児玉（2001）に詳しい。特徴は以下の 3 点である。

第 1 に、SR は 1997 年に導入され、最初の SR である 1998CSR が 1998 年に出された。SR 自体は包括的な支出見直しであると同時に、3 会計年度を対象とするローリング型の中期予算計画である。また SR そのものは図 2 で示された通り、政府の方針に基づいて行われるものである。ただし、表 2 をみて分かるように、実質的には 2 年に 1 回のレビューとなっている。（最後の 2007CSR は抜本的な見直しなのでこの限りではない。）

第 2 に、SR は図 2 で示された通り、省庁の目標となる PSA や SDA を包括する位置にある。なお、2007CSR より新たな枠組みとして省庁レベルの戦略計画（Departmental Strategic Objectives: DSO）が導入されている。東（2009）によれば、各省庁は CSR の期間において 1 又は 3 年の業務計画を策定するが、その中で DSO にかかる目標・指標、政策手段、実施機関、DSO と PSA の対応関係等の設定を行うこととされている。また、年 2 回の報告書を公表し、指標の実績の達成度を検証することとされている。

第 3 に、SR は財務規律をふまえている。1998 年に制定された財政安定化規律により、ゴールデンルール（政府借入は公共投資のみに限定。義務的経費での借入は不可）や持続的投資ルール（政府借入額は GDP に連動）などがルール化されている。これをふまえ省庁の予算には上限が設定されている。

表2 SRの一覧

SR の名称	計画期間	SR のタイトルと内容
1998CSR	1998 -2002	「英国における現代の公共サービス：改革への投資」 <i>(Modern Public Services for Britain: Investing in Reform)</i> 第1部 CSRの概観、第2部 政府の重点政策 (Key Objectives)、 第3部 省庁別及び省庁横断レビュー (全25章構成)
2000SR	2001 -2004	「目的のための儉約：すべての人のための機会及び安全の構築」 <i>(Prudent for a Purpose: Building Opportunity and Security for All)</i> ※(変更点) 1998CSRの第3部の「省庁別及び省庁横断レビュー」を細分化 第1部 概観および安定した力強い公共サービスの提供に向けて、 第2部 政府の重点政策 (Key Objectives)、第3部 各省別レビュー、 第4部 分権型の地域 (スコットランド、ウェールズ、北アイルランド)、 第5部 省庁横断レビュー (全37章構成)
2002SR	2003 -2006	「すべての人のための機会及び安全の構築：投資の積極化、英国の公平化」 <i>(Opportunity and security for all: Investing in an enterprising, fairer Britain)</i> ※部構成を簡素化 第1部 概観と重点政策、第2部 省庁別レビュー、第3部 省庁横断型レビュー (全31章構成)
2004SR	2005 -2008	「すべての人のための安心・安全と機会：英国の未来に向かっの投資」 <i>(Stability, Security and Opportunity for All: Investing for Britain's long-term future)</i> ※部構成は廃止され、章のみの構成へ変更 第1章 概観、第2章 効率、第3章 よりよい公共サービス (以上が総論) 第4章～第6章は重点政策、第7章～第23章は省庁別及び省庁横断レビュー 及び分権型の地域 (全23章)
2007CSR	2008 -2011	「英国人の希望との接点」 <i>(Meeting the aspirations of the British people)</i> ※Pre-Budget Reportと統合。章構成は大幅に見直され省庁別及び省庁横断型レ ビューは付録に収録。総論部分及び重点政策のみで構成。全7章。

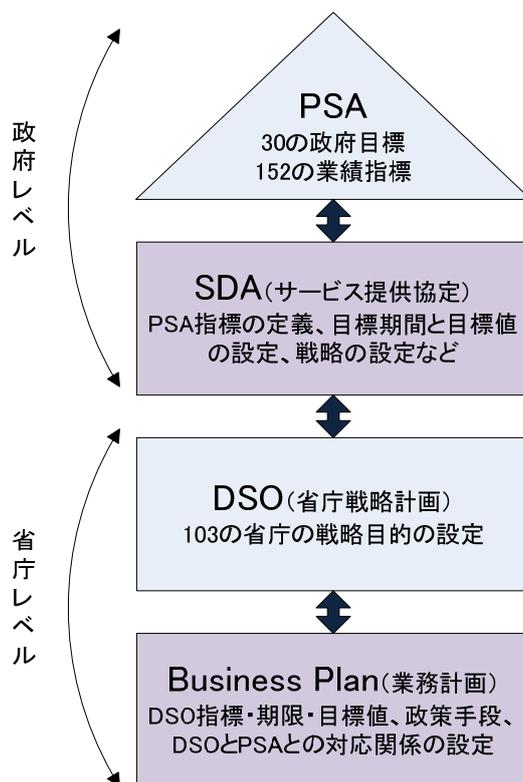
(出典) 筆者作成

(3) PSA (公共サービス協定) のシステム

PSAはSRをふまえたものであるが、省庁側はPSAにおいて今後3年間の目的・目標を示し、省庁内の戦略計画や業務計画に反映させることとなる(図3参照)。PSAは各省庁と大蔵省との間で毎年12月頃に協議され、政府の方針として決定される。平松(2002)は「PSAは立法によるものではなく、大蔵省と各省庁との間でかわされた合意に過ぎない

ため、その内容は議会の承認を必要とするものではない。PSA の主導権は大蔵省にあり、議会にとってみれば予算審議のための「参考資料」的なものとして情報を与えられるに過ぎないのである。」と指摘している。

図3 PSAシステム



(出典) 東 (2009: 146) の図及び HM Treasury 資料 ('Outcome Focused Management in the United Kingdom') などをもとに作成

1998年のPSAでは600もの指標が設定されていたが、2000年にはこれが160指標まで整理された。2007CSRの対象期間では152指標が設定されているので(1つの省庁では15前後)、2000年以降は安定的に推移しているといえる。なお、1998年(600)から2000年(160)への指標の整理に際しては枠組みが変更されている。当初は、PSAを支えるものとしてRABの発生主義の考え方をふまえた「アウトプットおよび実績に関する分析(OPA)」(1999年公表)が実施されていたが、2000年段階では旧PSAの一部とOPAとがPSAをふまえたアウトプットレベルのSDA(サービス提供協定)に統合されている。

表3 PSAの設定例（複数省庁にまたがる政策の場合）

【ねらい (Aim)】

薬物乱用が引き起こす被害を減らし、健康で安心して過ごせる社会をめざす。

【目標と業績指標 (Objectives and Performance Targets)】

目標Ⅰ 若者が社会で活躍できるようにするため、若者の薬物乱用からの離脱を支援する。

業績指標： 25歳以下のAクラス薬物を使っている人々の割合を2005年度までに25%（2008年までに50%）にする。（内務省のPSAとの共管事項）

目標Ⅱ わたしたちのコミュニティを薬物に関連する反社会的・犯罪的作用から守る。

業績指標： 薬物乱用者中の再犯者の水準を2005年までに25%にする（2008年までに50%）。（内務省のPSAとの共管事項）

目標Ⅲ 薬物問題を抱えた人々がこれに打ち勝ち、健康で犯罪にかかわりのない生活が送れるようにする。

業績指標： 薬物の不正使用の問題を抱える人々の薬物取扱プログラムへの参加者を2004年までに55%にする（05年までに66%、08年までに100%）。（保健省のPSAとの共管事項）

目標Ⅳ わたくしたちのまちでの薬物の使用を排除する。

業績指標： Aクラス薬物利用を2005年までに25%にする（2008年までに50%）。（税関のPSAとの共管事項）

【誰がサービスの提供に責任を持つか】

当該PSAは、内閣官房長官、保健大臣、教育雇用大臣、内務大臣及び国庫局長官との調整に基づく違法な薬物に対する政府活動を取り扱うものである。内務大臣は「業績指標Ⅰ」の解決を主導するが、保健大臣及び教育雇用大臣もまたこれを支援するものとする。内務省の主たる業務は薬物防止助言機関の効果的な運営及び薬物アクションチームの効果的な支援を通じたものとする。内務大臣は、「業績指標Ⅱ」の提供に対して責任を負うものとする。保健大臣は「業績指標Ⅲ」の提供に責任を負うものとする。国庫局長官は内務省の支援を受け、「業績指標Ⅳ」の提供に責任を負うものとする。スコットランド、ウェールズ及び北アイルランドの反薬物関連事業は、各地の自治体が責任を負うものとする。

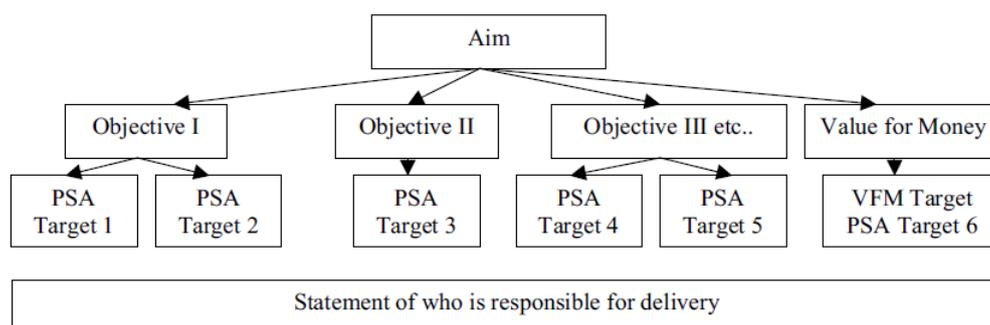
（出典）HM Treasury 資料（'Outcome Focused Management in the United Kingdom',p.4）より訳出

(4) PSAのチェックの視点

PSAは政府全体の「ねらいと目的」を明らかにし、「具体的なターゲット」(業績指標)を通じてそれをどの程度、いかなる期限で達成しようとするのかを明らかにするものとされる。また、PSAは「アウトカム志向」とされている。これは「国民が何を求めるのか」に焦点があるためと説明されている(HM Treasury 1998)。なお、実際のPSAは具体的な数値目標を掲げているもの、抽象的な方針にとどまっているもの、あるいはいわゆるアウトプットレベルにとどまっているものなど幅がある(具体的な設定例については表3を参照)。

PSAで設定される指標(PSA Target)と省庁の設置目的(Aim)や活動目標(objective)を省庁の側から整理すると図4のようなイメージとなる。目的が目標に分割され、指標(PSA Target)はその下に配置される。また、表4にもあるが、「だれがサービス提供に責任を持つのか」という点が確認されるべきものとされている(HM Treasury 1998)。

図4 PSAの構造



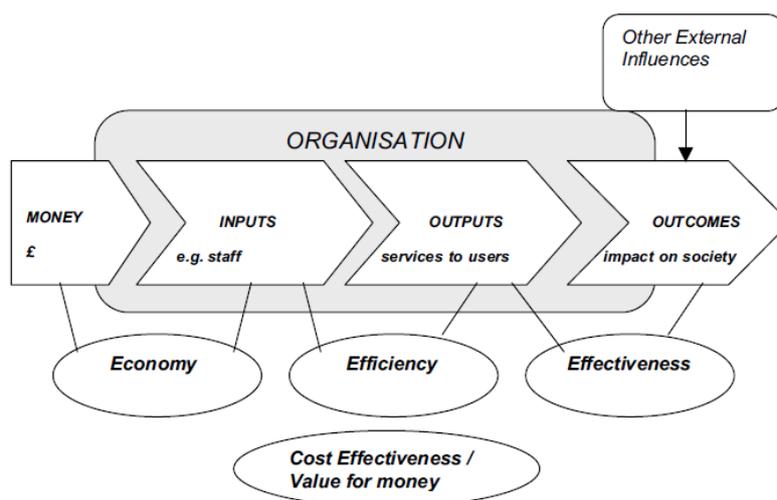
(出典) HM Treasury 資料 ('Outcome Focused Management in the United Kingdom', p.3)

なお、PSAの業績目標の設定にあたっては「SMART」であることが求められている。SMARTとは以下の頭文字を取った造語である。これらはPSA全体にかかるスローガンである。

- ① Specific (活動方針は具体的か)
- ② Measurable (目標は測定可能に設定されているか)
- ③ Achievable (その事業目標は無理がないか)
- ④ Relevant (目的と手段の関係は適切か)
- ⑤ Iimed (実施時期や期限は明示されているか)

また、PSA では「支出に見合った価値 (Value for Money: VFM)」も重視される。VFM とはいわゆる「3E」(経済性 (Economy)、効率性 (Efficiency)、有効性 (Effectiveness)) のことである。児玉 (2001) は「経済性とは、ある投資 (Investment) に対してどれだけの投入 (Input) が得られたか、効率性とは、ある投入に対してどれだけの産出 (Output) が得られたか、有効性とは、ある産出に対してどれだけの成果 (Outcome) が得られたか」ということである。VFM は、ある投資に対してどれだけの成果が得られたかということであり、経済性、効率性、有効性を掛け合わせたものと等しい。」と説明している。これを図示したものが図5である。

図5 VFMの概念



(出典) HM Treasury 資料 ('Outcome Focused Management in the United Kingdom', p.16)

(5) SDA (サービス提供協定) のチェックの視点

SDA (Service Delivery Agreements) は上でも触れたとおり、2000年のPSAをアウトカム志向とするという方向性の中で整理されていったものである。PSAをアウトカム志向というならば、SDAはPSAの下に整理されるアウトプットレベルである。省庁側からいえばSDAはPSAを詳しく説明したものであり、実質的にはアウトプットレベルの具体的な説明を充実させるためのものといえる。具体的なSDAの設定例は表4の通りである。

表4 SDAの例

<p>PSA ターゲット（業績指標）： 2002年までに刑事訴訟制度上の事件被害者及び証言者の満足度の5%増を目指すものとする。また、以後最低限の業績として当該水準を維持するものとする。</p>	<p>SDA： このサービスを提供するためにわれわれは以下のことを行うものとする。</p> <ul style="list-style-type: none"> ・事件被害者及び証言者（エスニック・マイノリティも含む）に対する支援の実施。治安判事裁判所における事件被害者及び証言者については2001年3月までに40%を適用範囲とする。 ・司法プログラムの活動の大半（非難されやすいあるいは脅迫のおそれのある証言者に対する各種の支援方策）を2002年春までに実施する。 ・事件被害者のための憲章を見直し、刑事手続きにかかる事件被害者の役割に関する協議を2001年7月までにおえる。
---	--

（出典）HM Treasury 資料（‘Outcome Focused Management in the United Kingdom’, p. 6）より訳出

大蔵省の説明によれば、各省大臣はつぎの5点について大蔵省との間で合意するとされている。SDAはこれを具現化する方法のひとつである³。

- ①省庁の上位レベルの目標達成する姿をどれだけ明確化できるか。
- ②省庁等の責任の範囲においてどれだけ業績が向上するのか。
- ③省庁の提供するサービスがどれだけ国民のニーズに近づけるか。
- ④省庁の人的資源またはIT資源がどの程度の変化をもたらすのか。
- ⑤政策形成が向上するような、『政府の近代化』白書に沿った、あるいはとくにエビデンスに支えられた政策となるような方向へのステップがふまえられているか。

また、SDAの主な要素としては以下の7点が掲げられている。SDAの各要素とその内容については表5にまとめた通りである。

³ ‘A Guide to the Service Delivery Agreements’
（http://hm-treasury.gov.uk/guide_to_the_service_delivery_agreements/htm/）

表5 SDAの7つの要素

SDAの7つの要素	各要素の主な内容
①アカウントビリティ	誰が合意事項の提供に責任があるか
②主な結果の達成	主要業績目標の設定、業績目標の達成方法、業績の測定方法など
③業績の改善	欧州品質経営財団（EFQM）の卓越モデル（Excellence Model）や内閣府が支援する Better Quality Service プログラムといった管理手法の利活用、上位の VFM 目標を捕捉する目標の設定、類似業績間での業績の差異の査定、政府調達に関する取り組みや虚偽および錯誤の取り扱いの改善など
④顧客志向	顧客志向のためのテストの設計、国民アクセスの改善、優良サービス提供事例の表彰、文書への即時回答など
⑤人的資源管理	公務員制度改革の計画項目の測定、病気欠勤や病気退職を減らすことなど
⑥電子政府	各省庁の電子政府計画など
⑦政策と戦略	『政府の近代化』白書で述べられた政策形成や調査分析のあり方についての実施方策など

(出典) 児玉 (2001: 187-188) 及び'A Guide to the Service Delivery Agreements'(注 3)から作成

(6) TN (技術覚書) - 関連データの取り扱い

TN は SDA が導入とともに 2000 年から導入されている。これは、「業績目標の測定方法について、データの収集方法や用語の定義を定めるものであり」、「各省庁と大蔵省との間の目標の正確な測定方法に関する合意」であるという (児玉 2001: 188; 197)。

たとえば大蔵省の『大蔵省の技術覚書：公共サービス協定 (PSA) 2005-08』(2007 年) は、全 42 頁で 10 の章、1 つの付録で構成されている。その中身は、大蔵省が所管する 10 の PSA ターゲットについてそれぞれの章が設定されており、それぞれの章は「定義とデータ」、「ターゲットへの接近法」、「データの出典」などの項目が共通的といえるが、統一されているわけではなく、政策の特性に応じた説明のための項目だてがなされている。

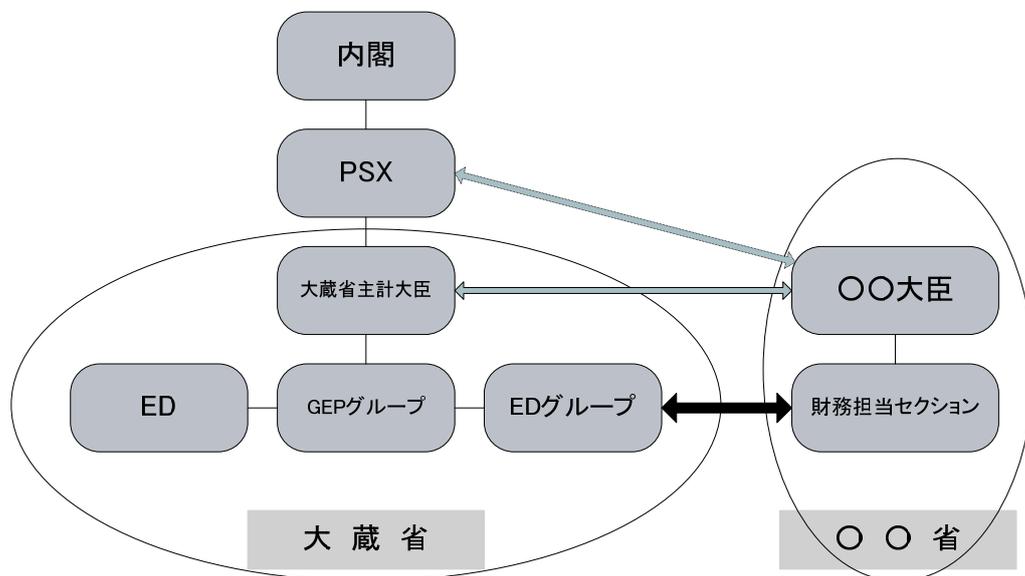
3 PSAのチェックシステム

(1) 政府部内のチェック体制と手順

PSAのチェックシステムについては稲継(2001)および児玉(2001)に詳しい。CSR2007時点においても基本的にこの態勢は維持されている。

PSAの編成は予算編成過程で行われる。大蔵省は「ガイダンス・ノート」(論点メモ)や予算関係のマニュアルをはじめいくつかの関連文書を出している。これらの強制力はないが、非公式な指導の際に用いられる。稲継(2001)の説明によれば、「ガイダンス・ノートの作成にあたっては、各省との協議(negotiation)はないが、実行可能性について各省と諮問・相談(consultation)する場合がある。その上で大蔵省主計大臣、さらには公共サービス歳出閣僚会議(PSX)の承認を得る。ガイダンス・ノート自体が明確な法的根拠を有しているわけではないが、PSXの承認を得ているということで、権威付けはなされており、各省庁はこのガイダンス・ノートに沿って、PSAs ターゲット等の作成にあたることになる。」とされている。

図6 CSR・PSA作成過程概念図



(出典) 稲継(2001)

また稲継はより具体的な政府部内でのチェックシステムについても説明している(図6)。まずガイダンス・ノートで注意喚起されるのは「ねらい」(Aims)、「目的」(Objectives)、

「業績達成目標」(Performance Targets)であり、これらの項目について各省側と大蔵省が検討を行うという。具体的なチェックの手順は、最初に各省の担当者が大蔵省の歳出担当各課(ED)と相談しながら草案を練り、ここで固められた草案が大蔵省の一般歳出政策課(GEP)に集められ、そこでチェックが行われる。最後にGEPに集められてチェックが終わったものが大蔵省主計大臣に渡され、大臣からPSXに報告される⁴。PSXでは個々のPSAターゲットのうち重要な点について各省大臣と協議が行われる。なお、稲継が調査を行った当時はCSRおよびPSAについて専門に扱う部局は存在していないということであったが、2009年版の大蔵省の組織図を見る限り、この体制に大きな変更はない。

かつてはこのほかに公共サービス生産性パネル(Public Service Productivity Panel)というチェック体制も存在していた(児玉2001:203)。同パネルは「チェックリストを作成し、各省庁が事業計画を作成する際、目的が長期に短期の達成目標と結びついているかどうか、政府の横断的な目的を組み入れているかどうか、機会の均等や多様な登用に関する目標があるかどうか」(同上)についての検証を行う品質管理機関であったが、2006年に廃止されている。

(2) PSAに関するガイドライン

PSAに関するガイドラインとしては大蔵省、内閣府、会計検査院、監査委員会、統計局の連名の『正当な枠組みを選択する：パフォーマンス情報の枠組み』(2001年)が重要である。この全文については総務省行政評価局政策評価官室(2002:11-51)に全文訳が掲載されているので、ここではその特徴に触れるにとどめたい。

同ガイドラインは、業績情報—とくにアウトプット、アウトカムなどの考え方—に関する一般原則がまとめられている。同ガイドラインは枠組みや考え方、基礎概念の説明のためのものであり、厳重なルールではないことが明記されている。

同ガイドラインは全8章で構成されている。第1章および第2章は序論であり、第8章はまとめであるから、本論は第3章から第7章までである。第3章から第7章の内容は以下の通りである。

⁴ 2009年3月現在のPSXの参加メンバーは次の通り。(Chancellor of the Exchequer (Chair); Secretary of State for Justice and Lord Chancellor; Secretary of State for Business, Enterprise and Regulatory Reform; Chief Secretary to the Treasury; Minister for the Cabinet Office; Chancellor of the Duchy of Lancaster)
(<http://www.cabinetoffice.gov.uk/secretariats/committees/psx.aspx>)

第3章ではインプット、アウトプット、アウトカムが概念が解説されている。先に触れたVFMも図5とともにここで説明されている。

第4章では業績情報体系の要件として、

- ①「重点性」
- ②「適切性」
- ③「バランス性」
- ④「強固性」(※政策が組織や人の変化で揺らがないこと)
- ⑤「統合性」(※管理プロセスに組み込まれていること)
- ⑥「費用対効果性」

の6点が必要であることが述べられている(※印はいずれも筆者。以下同じ)。

第5章では業績測定の成功条件が触れられている。ここで触れられている条件は、

- ①「(組織目標と業績目標との) 妥当性」
- ②「誤ったインセンティブの回避」
- ③「帰属性」(※外部要因の強さや行政責任の所在のこと)
- ④「明確な定義」(※データ収集・分析の前提として)
- ⑤「タイムリー性」
- ⑥「信頼性」(※または正確性)
- ⑦「比較可能性」(※他のデータや他のプログラムとの比較可能性)
- ⑧「検証可能性」

の8点である。

第6章では実際の業績測定として、SMART基準が重要であること、業績測定に際しては組織文化や主観、あるいは視座設定に注意を要するべきものであること、優良事例の発見を伴ったり理知的に行われるべきこと、分かりやすくかつ正確な情報提供たるべきこと、最新情報であるべきことなどについて触れられている。

第7章では専門知識の情報源として会計検査院、監査委員会、統計局、業績情報パネル⁵などが紹介されている。

なお、いわゆる内容にかかる客観性審査の項目としては、ここに掲げられたうち第5章

⁵ 業績情報パネル(Performance Information Panel)は、内閣府、統計局、大蔵省、会計検査院、監査委員の代表者からなる、大蔵省に設置される組織で、PSAおよびSDAの質に関する助言、SRの指針の作成などの活動を行う。

が参考になるだろう。

(3) その他の機関の関与

PSA を中心とする評価システムについては、大蔵省をはじめ、PSX、内閣が予算編成システムとの関係もあって密接な関係を持つことを述べてきた。このほか、上述の評価ガイドラインの第7章では、会計検査院、監査委員会、統計局および業績情報パネルが関与する旨が述べられていた。

このうち会計検査院は業績評価や PSA に対しての活動を活性化させている。評価ガイドラインにおいては業績測定局を中心に優良事例の確定・普及および費用対効果に関する技術面での支援を行っているとされている。

ところでこの評価ガイドラインが出された同年、「シャーマン報告」(HM Treasury 2001b) と呼ばれる改革案が出された。その内容は平松 (2002: 241-252) に詳しいが、シャーマン報告とは、大蔵省のもとでシャーマン卿 (元大手会計事務所会長) が出した「中央政府の監査とアカウントビリティのレビュー」という報告書のことである。平松は、①監査主体の多元化と連携、②監査対象の拡大、③PSA に対する外部評価の3点が特徴であるとされている。

第1の監査主体の多元化と連携については、行政の内部統制強化論および外部監査強化論である。内部統制は民間企業において近年注目を集めるリスク管理の手法だが、公金の番犬としてこの内部統制と外部監査を導入すべきこと、またそのガイダンスを大蔵省と会計検査院が行うべきことが勧告されている。

第2の監査対象の拡大とは、民間委託やパートナーシップの導入によって拡散した公金の使い道について監査対象の拡大によってこれを捕捉すべしという内容である。とりわけ執行的 NDPB や民間企業への監査対象の拡大に論及している点が特徴的である。

第3の PSA に対する外部評価は、これらの3点のうち本章にとってはもっとも重要なものである。PSA は予算編成にからむ政府の「内部」の議論であるが、シャーマン報告は PSA に外部からの「認証」(validation) が必要であり、その役割を NAO が果たしうることを主張している。また、この PSA の外部評価に関しては統計局等の他の機関との連携も示唆している。

このシャーマン報告に対し、政府は全面的にこれを受け入れた (HM Treasury 2002a)。これに基づき会計検査院はこれまでに4回の PSA に対する外部評価 (認証評価) を実施

している。初回が 2005 年 3 月、第 2 回が 2006 年 3 月、第 3 回が 2006 年 11 月、第 4 回が 2007 年 11 月である。この外部評価はデータシステムの質的な保証を中心とするものであり、日本の場合と比較すると、会計検査院よりも総務省の業務に親和性が高いのではないかと考えられる。

4 小 括

本章では PSA システムを中心にそのチェックシステムを概観してきた。最後に今後の研究課題等に触れて稿を閉じたい。論点は 3 点である。

第 1 に、本章で論じたのは PSA システムを中心とするチェックシステムであった。しかし冒頭に述べたように「評価」は業績測定的な評価だけではなく、その他の評価についても見ていく必要がある。この中で重要と思われるのは、大蔵省の出しているガイダンスである『グリーンブック』（総務省行政評価局政策評価官室（2003）に訳文所収）である。『グリーンブック』は‘appraisal’（事前評価）と‘Evaluation’（事後評価）に関するガイダンスであり、これまでに 3 回の改訂を経ている。こうした業績測定以外の評価については規制評価を含め広がりがある。‘Regulatory State’（Hood *et al.* 1999）とも表現される評価先進国のイギリスでは、評価文化の広がりや蓄積がある点には注意が必要であり、その広がりの中で PSA システムも把握し理解していく必要があるだろう。

第 2 に、シャーマン報告以降の動向についてである。シャーマン報告以降の動向については、国内では十分な研究蓄積がない。外部評価（認証評価）の蓄積についても何らかのレビューが必要となるものと思われる。とりわけ内部統制と外部評価のあり方については、政策の質をどのように確保するかという試行錯誤の中で検討されているので、とくに重要ではないかと思われる。

第 3 に、このうちとくに外部評価と連携することが予定される内部統制についてはとくに重視すべきものと思われる。PSA に対する認証評価についてはもちろんのこと、シャーマン報告のコンセプトにもあるように、外部評価と連携する内部統制システムの構築は現代政府の信頼の問題にかんがみ重要な論点である。日本の年金機構や独立行政法人でもこうした論点が出始めているが、比較研究も視野に入れながら、今後の研究課題となるものといえるだろう（参照、南島 2009）。

第4章 ニュージーランドにおける政策評価のチェックシステム

和田 明子

はじめに

ニュージーランドにおける政策評価のチェックシステムの位置づけを正確に理解するためには、同国の政治・行政システムの全体像を把握することが欠かせない。なぜなら、同国における政策評価とそのチェックシステムは、それ単独で行われているものではなく、80年代の行政改革後の行政システム全体の中でその位置づけがあるものだからである。

そこで、本章ではニュージーランドの政治・行政システムの概要を説明した上で、政策評価とそのチェックシステムについて説明することにした。

1 ニュージーランドの政治・行政システム

(1) 概要

ニュージーランドは、オーストラリア大陸の東南海上に位置する島国である。北島と南島を主要島とし、面積は約27万km²、日本の約4分の3である。人口は約400万で、ヨーロッパ系（pakeha）が約77%、先住民族マオリ（Maori）が約15%、アジア系や太平洋諸島系などがその他を占めている（Statistics New Zealand 2008: 87-88）。

政治・行政システムは、基本的にイギリスの流れを汲んでいる。イギリス女王を国家元首とする立憲君主制で、議院内閣制を採っている。イギリスと異なる主な点は、一院制であること、小選挙区併用の比例代表選挙制度（Mixed Members Proportional）を採っていること¹などである。比例代表制下で国民党（National Party）と労働党（Labour Party）の二大政党が第三位以下の少数政党と連立を組むことにより、政権交代を繰り返している。

大臣数は通常25名前後であり、省庁数は2009年1月現在35である（表1参照）。省庁のうち、首相内閣庁（Department of the Prime Minister and Cabinet）・政府サービス委員会（State Services Commission）・財務省（Treasury）は中心省庁（central agencies）と呼ばれ、それぞれ政策調整、人事・組織、財務の各側面から省庁間の横断的調整を行っ

¹ 1993年の選挙制度改革で、従来の単純小選挙区制から変更された。

ている。

なお、政策評価に一定の役割を持つ会計検査院（Office of the Auditor-General）は、政府から独立し国会に対してアカウンタビリティを持つ国会関係機関（Office of Parliament）である。

表1 ニュージーランドの省庁（2009年1月現在）

農林省	Ministry of Agriculture and Forestry
公文書館	Archives New Zealand
建築住宅庁	Department of Building and Housing
自然保護庁	Department of Conservation
矯正庁	Department of Corrections
法制局	Crown Law Office
文化省	Ministry for Culture and Heritage
税関庁	New Zealand Customs Services
防衛省	Ministry of Defence
経済開発省	Ministry of Economic Development
教育省	Ministry of Education
教育監査局	Education Review Office
環境省	Ministry for the Environment
漁業省	Ministry of Fisheries
食品安全庁	New Zealand Food Safety Authority
外務貿易省	Ministry of Foreign Affairs and Trade
政府情報安全局	Government Communications Security bureau
保健省	Ministry of Health
歳入庁	Inland Revenue Department
内務庁	Department of Internal Affairs
法務省	Ministry of Justice
労働庁	Department of Labour
国土情報庁	Land Information New Zealand
マオリ政策省	Ministry of Maori Development
国立図書館	National Library of New Zealand
太平洋諸島政策省	Ministry of Pacific Island Affairs
首相内閣庁	Department of the Prime Minister and Cabinet
科学技術省	Ministry of Research, Science and Technology
重大詐欺捜査局	Serious Fraud Office
社会開発省	Ministry of Social Development
政府サービス委員会	State Services Commission
統計庁	Statistics New Zealand
運輸省	Ministry of Transport
財務省	The Treasury
女性政策省	Ministry of Women's Affairs

（出典） <http://www.ssc.govt.nz> より作成。

省庁の役割は担当大臣に対して「自由で率直な提言 (free and frank advice)」を行うことであり、どのような政党の大臣に対しても客観的根拠に基づいた政策提言を行うが、最終的に決定するのは大臣である。省庁職員の政治任用は行われておらず²、省庁のトップであるチーフ・エグゼクティブ (chief executive) は政府サービス委員長 (政府サービス委員会のチーフ・エグゼクティブ) が、その他の職員は当該省庁のチーフ・エグゼクティブが、採用する (後掲図 1 参照)。

(2) 行政改革

ア 経緯

1970 年代のオイルショックと主要輸出相手国イギリスの EC 加盟を契機としてニュージーランドの経済・財政状況は悪化し、その結果、84 年の総選挙で誕生したロンギ労働党政権が行政改革を開始した。現在のニュージーランドの行政システムの原型はこの時にできたものである。

1984 年以降に行われた行政改革は、2 ないし 3 つの段階に大きく分かれる。第 1 段階は、84~90 年の労働党政権とそれに続く 90~99 年末の国民党政権の時期である。民営化を促した 86 年国有企業法 (State Owned Enterprises Act 1986 No. 124)、人事面を改革した 88 年国家部門法 (State Sector Act 1988 No.20)、財務面を改革した 89 年財政法 (Public Finance Act 1989 No. 44) 等により現在の形が定められた。民間との「イコール・フットイング」を原理・原則とするこれらの改革をイギリスの行政学者 Hood は「New Public Management (NPM)」と名付けた (Hood 2009: 205-214)。第 1 段階の改革の特徴を一言で表すと、「アウトプット (output)」の達成責任 (accountability) を重視し、その結果として「効率性 (efficiency)」を重視するものであった (後述)。

第 2 段階は、1999 年末~08 年のクラーク労働党政権による修正期である。04 年に国家部門法と財政法が改正され、省庁以外の公的機関 (クラウン・エンティティ) の通則法であるクラウン・エンティティ法 (Crown Entities Act 2004 No.115) も制定された。焦点を「アウトプット」から「アウトカム (outcome)」へ、また「効率性」から「有効性/効果 (effectiveness)」へと少しずつ修正する改革であったと一言で表される。

2008 年 11 月の総選挙ではキー国民党政権が誕生し、第 3 段階の行政改革が開始されるのかが注目されるが、2009 年 3 月現在、現状の行政システムに大きな修正はない模様で

² 大臣スタッフの政治任用は行われている。

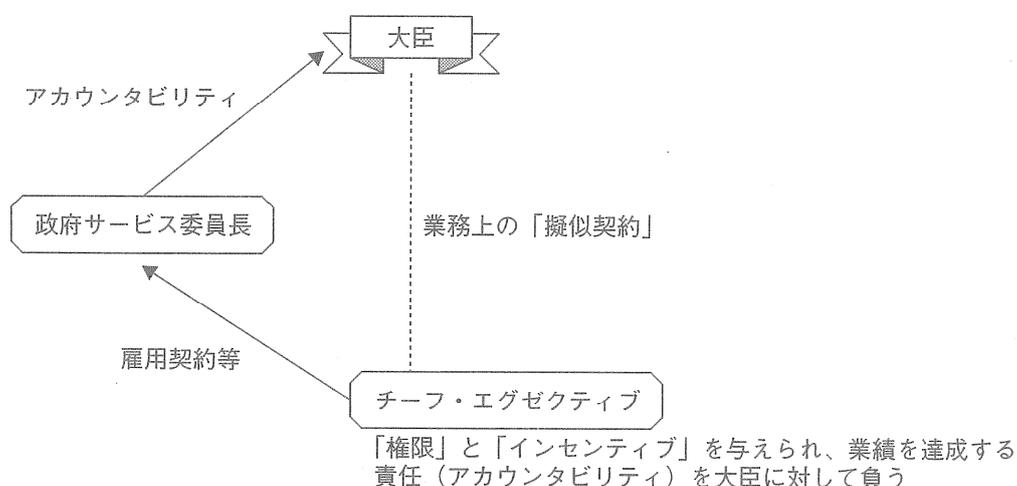
ある（現地調査①～⑥）。

イ 改革の概要

（ア）基本的フレームワーク

ニュージーランドの行政システムの基本的フレームワークは、次の通りである（図1参照）。

図1 省庁チーフ・エグゼクティブの位置づけ



（出典）和田（2007: 136）

日本の事務次官に相当する省庁のトップは、民間企業と同じ「チーフ・エグゼクティブ」の呼称で呼ばれ、省庁の最高運営責任者と位置づけられている。具体的には、担当大臣と業務上の「疑似契約」（以下、アカウンタビリティ・ドキュメントと呼ぶ）を結び、そこで明示された業績（performance）を達成する責任（accountability）を大臣に対して負う（業績主義）。業績を構成する主要な項目は「アウトプット」である。省庁のチーフ・エグゼクティブは大臣に対してアウトプットを達成する責任を負う一方、大臣は省庁からアウトプットを「購入」し国民に対してアウトカム（成果）を達成する責任を負う。省庁がアウトプットを達成するのに必要なインプットに関する権限、すなわち職員の任用権、組織編成権、予算執行裁量権³などは、それまで政府サービス委員会や財務省の中心省庁が担ってい

³ 改革前は財務省がインプットによるコントロールをしていたが、改革後はアウトプットの範囲内で各省庁による裁量が認められるようになったということである。

たが、各省庁のチーフ・エグゼクティブに移譲された（権限移譲）。

各省庁のチーフ・エグゼクティブは政府サービス委員長と5年以下の雇用契約を結び、毎年人事評価を受ける。評価の結果により毎年のボーナスが支給されるほか、任期中の評価が良くなければ雇用契約の延長が行われない（インセンティブ）。政府サービス委員長は、有能なチーフ・エグゼクティブを採用するアカウンタビリティを大臣に対して負うということになる。

省庁運営の最高責任者であるチーフ・エグゼクティブには権限とインセンティブが与えられる代わりに、厳しく業績の達成が求められるというフレームワークになっている。

（イ）アウトプットを基礎とするマネジメント

省庁マネジメントの基礎は「アウトプット」である。アウトプットとは、「省庁等が提供する財・サービス」と財政法でも正式に定義されている。省庁のチーフ・エグゼクティブは「アウトプット」の達成に責任を負い、予算も「アウトプット」を基本単位として編成される。

アウトプットには、指標とその目標値が設定される。つまり、省庁チーフ・エグゼクティブは、アウトプットに配分された予算額の範囲内で、アウトプット指標の目標値で表される「量」と「質」のアウトプットを大臣に届ける責任を負う。以上のような「アウトプット」を中心としたマネジメント・システムを採用しているのには、いくつかの理由がある。

第一に、コントローラビリティ（controllability）の観点である。アウトプットは、「省庁等が提供する財・サービス」であるので、インプットさえ与えられていれば省庁が必ず達成できるはずのものであるが、アウトカムは、省庁のアウトプット以外の要因によってその達成が左右され必ずしも省庁の努力のみで達成できるとは限らない。省庁に言い訳の余地を与えてしまうアウトカムの達成ではなく、コントローラビリティのあるアウトプットの達成を求めることで、省庁に業績を達成する責任（accountability）を厳しく課したのであった。

第二に、市場の財・サービスとの比較可能性である。既述したように、省庁の活動を「アウトプット」として整理し、量・質の指標と発生主義で算定した予算額を明示することによって、省庁はどのような財・サービス（アウトプット）をどれほどの量と質（アウトプット指標）でどれくらいのコスト（予算額）により提供できるのかが明らかとなる。つま

り、市場における同等の財・サービスとの量・質・コスト面での比較が可能になる。大臣は市場の財・サービスの方が良ければそちらを「購入」することが可能である、という競争環境（疑似市場）を作ることによって、省庁業務の効率化を強く促したのであった。

以上の「アウトプット予算」採用の背景には、1970年代における PPBS (planning, programming, and budgeting system) の失敗がある (Gill 2008)。政策評価結果と予算編成を結びつけようとする PPBS が失敗に終わったことにより、ニュージーランド政府は、より厳格な歳出項目となり得るアウトプット (財・サービス) に対して予算を付与するシステムを 80 年代の改革で導入したのである。

(ウ) アウトカムへの移行

1999 年末以降の第 2 段階の改革では、上記 (ア) の基本的フレームワークに変更はなかったが、上記 (イ) のマネジメントの焦点が「アウトプット」から「アウトカム」へと少しだけ移行している。これは、「アウトプット」に対するアカウントビリティだけでなく、「アウトカムの達成に向けてどのように取り組んでいるか (managing for outcomes)」に関するアカウントビリティを求めるものとして整理されている。つまり、アウトプットがベースであることに変わりはないが、その上でアウトカムとの明確で論理的なつながりについてもアカウントビリティを求めるということである。アウトカムとのつながりがないままにアウトプットだけ達成しても意味がないからである。

これに伴い、予算書をはじめとするアカウントビリティ・ドキュメントにも、アウトプットだけでなく何らかの目的 (インパクト、アウトカムまたはオブジェクティブ⁴) を明示することが義務づけられた (改正後の財政法第 15 条・第 40 条・第 45 条)。

1999 年末以降アウトカムが少しずつ重視されてきている理由は、次のように説明される。99 年末以降のクラーク労働党政権は、84 年から 15 年間続いた「効率化」重視の改革を批判し、国民にとっての「効果 (effectiveness)」を重視する改革を促した。国民にとって大事なものはアウトカムでありアウトプットだけ達成しても意味がないという批判は、90 年代から学界でも見られた (Gregory 1994: 18-21 など)。これに対して、「効率化」重視の改革を先導した財務省も、改革の結果として 90 年代には財政状況が大きく改善していたこと

⁴ インパクトは「アウトプットや活動 (activity) がアウトカムにもたらす貢献」、アウトカムは「社会・経済・環境の状態 (状態の変化を含む)」、オブジェクティブはそれらの定義に当てはまりにくい何らかの目的である (改正後の財政法第 2 条および Treasury 2008b: 22)。

や「アウトプット」を基礎とするマネジメント・システムもおおむね定着したことなどから、「アウトカム」への焦点の移行を受け入れたのである。

(3) ニュージーランドにおける政策評価の位置づけ

ニュージーランドの政策評価は、大きく「業績測定 (performance measurement)」と「評価 (evaluation)」に分けることができる。「業績測定」とは、アウトプットやその目的 (インパクト・アウトカム・オブジェクティブ) に指標を設定し達成度を報告するものであり、日本の「実績評価」に近いものである。「評価」とは、実施過程の評価である「formative evaluation」やプログラムの有効性を測る「impact (or outcome) evaluation」などに代表されるものであり、政策・施策の正式な評価 (formal evaluation of policy and programmes) であるとされる (State Services Commission & Treasury 2003: 8-9)。日本の「総合評価」に近いものであると位置づけられる。

以下では、「業績測定」と「評価」のそれぞれのチェックシステムを順を追って見ていくことにしたい。

2 「業績測定 (performance measurement)」とチェックシステム

(1) 「業績測定」

ニュージーランドの政策評価は、「業績測定」が主である。それは、先に見たアウトプット・ベースのマネジメント・システムから必然的に導かれることである。

具体的には、各省庁は年度当初に予算書、SOI (Statement of Intent : 業務計画書)、アウトプット計画書 (output plan) というアカウントビリティ・ドキュメントを作成し、アウトプットと何らかの目的 (インパクト・アウトカム・オブジェクティブ) に関する指標と目標値を明示する。また年度末には年次報告書 (annual report) を作成し、SOI に示された指標の達成度を報告する。

それぞれのアカウントビリティ・ドキュメントの主な違いは次のとおりである。

ア 政府予算書

省庁・大臣間のアカウントビリティ・ドキュメントである SOI・アウトプット計画書・年次報告書と異なり、予算書は大臣・国会間のアカウントビリティ・ドキュメントである。

予算書は省庁毎ではなく大臣の担当分野（portfolio）毎に編成されており⁵、「アウトプット群（output class）」とそれに対する指標・予算額も、後述する SOI とは異なり大臣の担当分野毎に明示される。「アウトプット群」とは、複数のアウトプットをグルーピングしたものである。何らかの目的（インパクト・アウトカム・オブジェクティブ）も明示される。

イ SOI

SOI（Statement of Intent：業務計画書）とは、各省庁の向こう3年間以上の意図・目的・計画（=intent）を包括的に記した書類である。当該省庁の活動とその目的、省庁の置かれた環境や直面するリスク、人材をはじめとする組織内部の現状と課題など、当該省庁の全体像を把握するのに必要な情報が盛り込まれたものと考えてよい。SOI は向こう3年間以上の計画である中期計画部分と当該年度の計画である年次計画部分から成り⁶、中期計画部分の中心として目的（インパクト・アウトカム・オブジェクティブ）と指標が明示される。また、年次計画部分の中心として、アウトプット群・指標・目標値の一覧表である「予測サービス業績表（Statement of Forecast Service Performance）」が明示される。予算書と異なり「省庁」毎に作成される。予算案と一緒に国会に提出される（改正後の財政法第39条）。

ウ アウトプット計画書

アウトプット計画書は、予算書・SOI・年次計画書と異なり法定される書類ではないが、アウトプットのより詳細な説明資料として作成されることが通例となっている。「アウトプット群」ではなく「アウトプット」毎に指標と目標値が示される。また SOI に示される「目的」とのつながりも示される。国会には提出されないが、予算書や SOI の詳細資料という位置づけになる。

⁵ ニュージーランドの大臣は首相から通常複数の担当分野（portfolio）を割りふられ、予算は担当分野に対して付与される。これを「Vote」という。つまり、大臣が担当分野の目的を達成するのに必要な予算が国会で承認（vote）され、その予算で大臣は様々な省庁からサービスを「購入」する。「大臣」「省庁」「Vote」の数は一致しておらず、09年1月現在、大臣数は28、省庁数は35、Vote数は69である（詳細は、和田2007: 12; 35-36; 101を参照のこと）。

⁶ 年次計画部分は、2008年度から予算説明資料（Information Supporting the Estimates）に掲載するという運用がなされている。

エ 年次報告書

年次報告書は、SOI で明示された事項の達成度を報告する書類である。このうち、アウトプット群の達成度を示す「サービス業績表 (Statement of Service Performance)」は財務諸表とともに会計検査院の監査意見が表明される。

SOI・アウトプット計画書・年次報告書の作成にあたっては、財務省と政府サービス委員会からガイドラインが出されており、その中で目的・アウトプット・指標に関する指針も適宜示されている⁷。目的・アウトプット・指標専門のガイドラインも多数示されており、財務省・政府サービス委員会からだけでなく会計検査院からも公表されている⁸。

各省庁ではそれらのガイドラインを参照し、財務省・政府サービス委員会・会計検査院のアドバイスを適宜受けながら、目的・アウトプット・指標を設定し達成度を測る。

(2) チェックシステム

「業績測定」のチェックを行う主な機関は、会計検査院・財務省・政府サービス委員会である。

⁷ SOI のガイドライン (最新版) :

Treasury & State Services Commission, *Statement of Intent: Process and Production Guidance for Departments*, 2009.

Treasury & State Services Commission, *Statement of Intent: Technical Guide for Departments*, 2008.

Treasury & State Services Commission, *Performance Information Measures and Standards in the SOI and Annual Report*.

アウトプット計画書のガイドライン (最新版) :

Steering Group Management for Outcomes, *Managing for Outcomes: Output Plans Guidance for Departments*, 2002.

年次報告書のガイドライン (最新版) :

Treasury, *Preparing the Annual Report: Guidance and Requirements for Departments*, 2008.

全て <http://www.ssc.govt.nz/display/document.asp?navid=339> に掲載されている。

⁸ 財務省・政府サービス委員会のガイドライン (最新版) :

State Services Commission & Treasury, *Performance Measurement: Advice and Examples on How to Develop Effective Frameworks*, 2008.

会計検査院のガイドライン (最新版) :

Office of the Auditor-General, *Report of the Controller and Auditor-General: Reporting Public Sector Performance 2nd edition*, 2002. など。

全て <http://www.ssc.govt.nz/display/document.asp?navid=339> に掲載されている。

ア 会計検査院

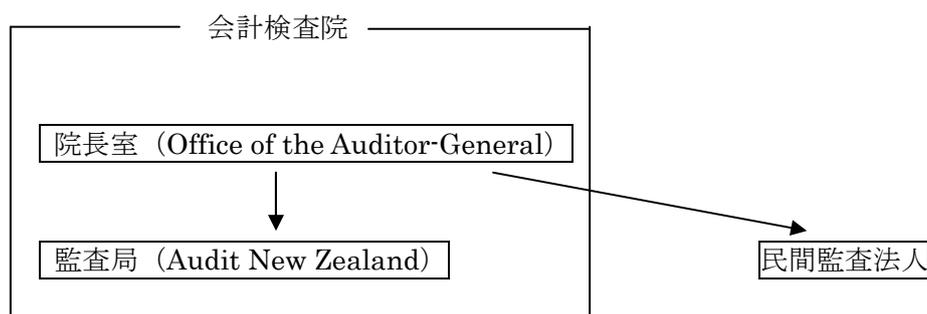
(ア) 組織・体制

会計検査院は、政府から独立し国会に対してアカウンタビリティ（報告義務）のある組織である⁹。省庁だけでなく、クラウン・エンティティ、国有企業、地方自治体を含む公的機関の監査人となっており、次のような業務を行っている。

- ・年次監査（annual audits）・・・財務諸表（financial statements）のほか、サービス業績表にも監査意見を表明する。ニュージーランドは、中央政府レベルではじめてサービス業績表という非財務業績（non-financial performance）にも監査を始めた国である（Boston 1996:303）¹⁰。
- ・業績監査（performance audits）・・・会計検査院が具体的なテーマを定め、3E（economy（経済性）・efficiency（効率性）・effectiveness（有効性））の視点などから監査を行う。
- ・調査業務（inquiry）・・・国会・政府・国民などから提起された問題について調査を行う。
- ・国会支援・・・国会の委員会に対し、上記業務から得られた知見に関するブリーフィングを行う。

以上のように量的にも質的にも幅広い業務をこなすため、会計検査院は次のような組織・体制をとっている（図2参照）。

図2 会計検査院の組織・体制



(出典) Office of the Auditor-General (2008a: 16) をもとに作成。

⁹ 根拠法令は、2001年公監査法（Public Audit Act 2001）である。

¹⁰ 地方政府レベルではオーストラリアの西オーストラリア州などが実施していた。

会計検査院は、「会計検査院長室 (Office of the Auditor-General)」と「監査局 (Audit New Zealand)」という 2 つの組織から成る。監査局は年次監査を行う事業実施部局であり、会計検査院長室はその他の業務、すなわち業績監査・調査・国会支援などの業務と全体に関わる業務（監査計画の決定、監査基準の設定、監査人の割り振り、監査人の業績評価、会計検査院全体の品質保証など）を行う。そして、年次監査は、監査局だけでなく民間監査法人にも、会計検査院長室から委託される。このうち、省庁の年次監査は、監査局が主に担当している。

(イ) 実施手順・ツール

会計検査院は、年次監査の一環として、「サービス業績報告書の監査基準」（後述）に基づき省庁の SOI と年次報告書をチェックする。財政法に基づき正式に監査意見を表明するのは、年次報告書のサービス業績表、すなわち「アウトプット群」とその指標・達成度のみであるが、それらの適正性を判断するためにも事前の計画書である SOI のチェックや「目的」（インプット・アウトカム・オブジェクティブ）の指標・達成度のチェックが必要とされている。

SOI と年次報告書の具体的なチェック手順は次のとおりである。

a. SOI のチェック手順

- (i) 監査人は SOI 原案を各省庁から入手し、
 - ・省庁のサービス業績（アウトプット群・指標・目標値）を評価するのに、SOI が適切な (adequate) ベースを提供しているか
 - ・予測サービス業績表、すなわちアウトプット群・指標・目標値が適正 (appropriate) かをチェックする。
- (ii) 上記二点に問題があると考える場合は、当該省庁の幹部と協議する。必要な場合は、事前に会計検査院長室に協議する。
- (iii) 省庁から受け取る最終版の SOI について上記二点を再度チェックし、問題がある場合は当該省庁の幹部に報告する。必要な場合は、事前に会計検査院長室と協議する。
- (iv) 予測サービス業績表の適正性に関する監査人の判断は、年次報告書のサービス業績表に監査意見を表明する際、考慮しなければならない。

b. 年次報告書のチェック手順

- (i) 監査人は、(サービス業績表から) 重要な (material) アウトプット群と指標を選択する。
- (ii) それらを監査するのに必要な証拠を集める。
- (iii) アウトプット群に対するコストの配分方法もチェックする。
- (iv) サービス業績表、すなわちアウトプット群・指標・達成度の適正性を判断するために、SOI と年次報告書全体を熟考する。必要な場合は、会計検査院長室に協議するか省庁幹部に報告する。
- (v) 監査人は、最終的に
 - ・省庁のサービス業績 (アウトプット群・指標・目標値) を評価するのに、サービス業績表が適切な (adequate) ベースを提供しているか
 - ・サービス業績表が省庁のサービス業績を正しく反映しているか (たとえば、目標値に対する達成度が正しく報告されているか、など)という観点から監査意見を表明する。
- (vi) 重要な (material) 瑕疵がある場合は、「限定意見 (qualified opinion)」、すなわち「限定付適正意見」「不適正意見」「意見不表明」¹¹のいずれかの表明を考えなければならぬ。より具体的には次のとおりである。
 - ・重要なアウトプット群・指標・目標値・達成度が省かれている場合や不適正 (inappropriate) な (関連性・信頼性・理解容易性・比較可能性の基準 (後述) を満たさない) 場合、監査人は「不適正意見」又は「限定付適正意見」の表明を考えなければならぬ。
 - ・重要なアウトプット群・指標・目標値・達成度が適正であるが記載の仕方が悪い場合、監査人は省庁幹部への報告を考えなければならぬ。
 - ・重要な指標の達成度が証拠に関する基準を満たさない場合、達成度を記録するプロセスに欠陥がある場合、重要な指標の達成度が実証できない場合は、監査人は「意見不表明」か「限定付適正意見」の表明を考えなければならぬ。
 - ・法令に基づかずに作成された場合、監査人は法令違反を会計検査院長室に報告すること

¹¹ 「限定付適正意見」とは、一部不適切な事項はあるが全体に対してそれほど重要でないと考えられる場合に表明される監査意見、「不適正意見」とは不適切な事項が全体に重要な影響を与える場合に表明される監査意見、「意見不表明」とは不適切な事項が監査意見を表明できないほど重要と考えられる場合にとられる行為である。

や省庁幹部に報告することを考えなければならない。

・重要な指標が正式な手続きを踏まずに変更された場合、変更後の指標の方がより不適正である場合、変更が年次報告書等の中で適切に説明されていない場合、監査人は会計検査院長室への報告や限定意見の表明を考えなければならない。

・指標が文章による説明で代替された場合、監査人は「不適正意見」又は「限定付適正意見」の表明や省庁幹部への報告を考えなければならない。

・文章による補足説明が、不十分である・矛盾している・誤解を招く場合、監査人は「意見不表明」か「限定付適正意見」の表明や省庁幹部への報告を考えなければならない。

(ウ) チェックを行う前提となるデータの入手権限

会計検査院によるチェックは「監査」の一環であり、日本の会計検査院の検査に該当するものである。このため、日本の会計検査院の検査と同様のプロセスで、必要なデータを入手することができる。

(エ) 政策評価のチェックの基準・視点

「会計検査院監査基準 (The Auditor-General's Auditing Standards)」¹²の中で「サービス業績報告書の監査基準 (The Audit of Service Performance Reports)」が示されており、SOI と年次報告書をチェックする際の基準・視点が示されている。「サービス業績報告書の監査基準」は、後述する 2008 年 6 月の指摘を受けて現在改訂中であるが、改訂案における主な基準・視点は次のとおりである¹³。

○監査人は、次の事項を監査する。

- ・ SOI・年次報告書の作成が省庁全体のマネジメント・システムと結びついているか
- ・アウトカム・インパクト・アウトプット群と指標・目標値の設定プロセス
- ・報告書の具体的内容

○より具体的には、次の視点から監査を行う。

- ・ SOI・年次報告書など外部向けの報告書と同じ指標を内部管理でも使用しているか

¹² <http://www.oag.govt.nz/2008/auditing-standards/index.htm>

¹³ パブリック・コメントを経て 2009 年上半期中に最終決定される予定である。

- ・目的・アウトプット群・指標・目標値の設定プロセスが適切 (adequate) か (たとえば、適切な協議 (consultation) が行われたか)
- ・SOI・年次報告書が法令に基づき作成されているか
- ・アウトプット群・指標・目標値は「適正 (appropriate)」(後述) か
- ・SOI・年次報告書の内容 (特に指標の監査可能性) に関して、監査意見に影響を与える課題があるか。

○「適正性 (appropriateness)」の要素は、次の4点である。

- ・関連性 (relevance)
 - －省庁の戦略目的、過去の業績、現在置かれている環境などに基づき、示されているか
 - －目的・アウトプット群・指標・目標値の間の関係が明確で論理的か
 - －利用者のニーズに応え、内部の意思決定にも有用か
 - －財務情報 (歳出など) とリンクしているか
- ・信頼性 (reliability)
 - －誠実に提示されているか (測定可能であるか、提示すべきものを提示しているか、どのような利用者でもおおむね同じ結論に達するか、誤りがないか)
 - －中立的か (アウトプット群・指標・目標値の設定にバイアスがないか)
 - －包括的でバランスがとれているか (まとめるべきところは基準を示した上でまとめられているか、重要な側面を全てカバーしているか)
- ・理解容易性 (understandability)
 - －わかりやすい形式になっているか
 - －視覚に訴えるよう図表が使われているか
 - －報告事項が明確・論理的に分類されているか
 - －各パートの間に一貫性があり、つながりがわかりやすいか
 - －省庁の戦略目的、過去の業績、現在置かれている環境などに基づき、示されているか
 - －目的・アウトプット群・指標・目標値の関係が明確・論理的でわかりやすく示されているか
 - －内容が明確で正確か

—読みやすく、平易な英語で書かれ、利用者に合った用語が使われているか（略語・俗語・専門用語の説明が妥当か）

・比較可能性（comparability）

—当年度の目標値に対する達成度、さらには年度間の傾向について、利用者が類似点・相違点を把握し進捗状況を追跡できるか

（オ）チェックによる具体的な指摘事例

2008年6月には、各省庁の「業績測定」の質が不十分であるという報告書が会計検査院から公表された（Office of the Auditor-General 2008b）。具体的には、

- ・アカウントビリティ・ドキュメントが、外部利用者（大臣、国会議員など）のニーズに
 応えておらず¹⁴十分に活用されていない。同時に、内部管理にも使われていない、
- ・アウトカムとアウトプットの論理的で一貫したつながりが明らかでない、
- ・適正性の要素（関連性・理解容易性・信頼性・比較可能性）の基準を満たす指標と目標
 値が設定されていない、

というのが主な点である。

以上の点を改善するために、同年11月に「サービス業績報告書の監査基準」の改訂案が公表された。従来の監査基準では「指標の適正性」と設定された指標に対する「達成度報告の正確性」を主な基準としていたが、「業績測定」の質を上げるためには「何を測るのか」つまりアウトプット設定の適正性と目標値の適正性、さらにはそれぞれの省庁が置かれている環境も考慮する必要がある。そのような観点から（エ）でみた監査基準の改訂案が示されたのである。

イ 財務省

（ア）組織・体制

財務省は、予算調製を担当するとともに、経済・財政に関する政策提言をする省庁である。財務省の達成すべきアウトカムとして「経済成長（Economic Growth）」「マクロ経済の安定（Macroeconomic Stability）」「国家部門の業績改善（improved State sector

¹⁴ アカウントビリティ・ドキュメントの第一義的外部利用者は大臣・国会議員であるが、メディア・研究者・市民などによる利用も想定されており、それぞれの利用目的が異なるため、それらのニーズ全てに応えるものにはなっていないということである。

performance)」が挙げられており、それにそって次のような組織体制がとられている¹⁵。

- ・ 経済業績局 (Economic Performance Group)・・・経済に関する政策提言
- ・ マクロ経済局 (Macroeconomic Group)・・・財政に関する政策提言
- ・ 国家部門業績局 (State Sector Performance Group)・・・国の予算を使用する省庁等の効率性・有効性の改善に関する提言
- ・ 総務局 (Organisational Performance Group)・・・広報・財務・人事・情報・法務などの総務業務

このうち、各省庁の「業績測定」をチェックするのは国家部門業績局である。

(イ) 実施手順・ツール

財務省は予算調製を行う省庁として、政府予算書および予算書と一緒に国会に提出される SOI をチェックする。

ニュージーランド政府の会計年度は7月～翌年6月であり、予算案は通常5月に国会に提出される。国会提出に向けた予算調製の一環として、予算書・SOI に示される目的・アウトプット群・指標のチェックが行われる¹⁶。

なお、アウトプット計画書と年次報告書については、財務省のチェックは任意である。アウトプット計画書は予算書・SOI の詳細資料として財務省のチェックを受けることが多いが、年次報告書は予算関係資料ではないため、必要に応じて¹⁷チェックを受けることとされている (Treasury 2009)。

(ウ) チェックを行う前提となるデータの入手権限

財務省の行うチェックは予算調製担当省庁としての権限に基づくものであり、日本の財務省と同じように予算調製のプロセスにおいて必要なデータは入手される。

¹⁵ このほか省内の独立ユニットとして、国有会社監視助言局 (Crown Company Monitoring Advisory Unit: CCMAU) がある。

¹⁶ 具体的な予算プロセスについては、毎年財務省から「予算プロセスガイド」(Budget Process Guide) (<http://www.treasury.govt.nz/publications/guidance/budget/process/>) が公表されているが、その中では目的・アウトプット群・指標の具体的なチェック手順は示されていない。

¹⁷ 年次報告書のチェックは、予算書・SOI のチェックに資するために財務省が行うというのではなく、財務省をはじめとする中心省庁のアドバイス・確認を得たい場合に省庁から申し出て行われるものである (Treasury, 2009)。

(エ) 政策評価のチェックの基準・視点

財務省のチェックの基準・視点は、「予算書（とそれに付随する SOI）に示される目的・アウトプット群・指標・目標値が、国会における有意義な議論を可能にするよう設定されているか」である（現地調査②（財務省））。なぜなら、予算書（と SOI）は国会で審議されるものであり、予算書（と SOI）は、国会が審議をするにあたって省庁の活動と予算を的確に把握できるように、作成されることが最も重要な点だからである。

財務省は、各省庁の指標の達成度を用いて予算査定をすることはない。それは、「目標値を達成できなかったから予算額を増額してほしい」というように、財務省の望ましい意図とは逆のインセンティブとなる可能性があるからである（現地調査②（財務省））。先にも述べたように、ニュージーランドでは PPBS の失敗から政策評価の結果を直接予算案に反映させることはできないと考えており、ニュージーランドで「指標」を使い「業績測定」が行われているのは、これまで説明してきたように予算査定とは違う趣旨によるものなのである。

ウ 政府サービス委員会

(ア) 組織・体制

政府サービス委員会は、各省庁のチーフ・エグゼクティブの採用業務を行うほか、行政システム（public management system）全体に関することについて各省庁にアドバイスを行う立場にある省庁である。

省庁全体の業績を改善するため、政府サービス委員会は「政府部門の開発目標（Development Goals for the State Services）」を 2005 年に導入した。「時の政府に仕え国民のニーズに応える世界一の専門性を持った政府部門を構築する」という大目標のもと 6 つの開発目標から成っており、毎年その達成度を公表する¹⁸。政府部門の開発目標は、政府サービス委員会の達成すべき「アウトカム」でもあり、同目標にそって政府サービス委員会の組織・体制は、担当大臣である政府サービス担当大臣（Minister of State Services）とチーフ・エグゼクティブである政府サービス委員長（State Services Commissioner）以下、次のように編成されている。

- ・業績マネジメント局（Performance Management Branch）・・・政府部門の開発目標と各省チーフ・エグゼクティブの採用

¹⁸ <http://www.ssc.govt.nz/display/document.asp?DocID=6545>

- ・行政システム改善局（System Improvement Branch）・・・各省庁への具体的アドバイス
- ・情報技術局（Information and Communication Technologies Branch）・・・電子政府
- ・人材能力開発局（People Capability Branch）・・・公務員の能力開発
- ・総務局（Corporate Services Branch）
- ・委員長室（Commissioner's Office）

このうち、各省庁の「業績測定」をチェックするのは行政システム改善局である。

（イ）実施手順・ツール

政府サービス委員会は行政システムに関することについて各省庁にアドバイスを行う省庁であり、同じ中心省庁である財務省とともに多くのガイドラインを省庁向けに公表している¹⁹。特に、SOIは、1999年末以降の行政改革で、何らかの目的（インパクト・アウトカム・オブジェクティブ）を明示するために政府サービス委員会が導入を促したものであるため、行政システム改善局の業績専門家チーム（performance specialist team）が、財務省とともに各省庁の相談窓口となっている。

ただし、政府サービス委員会の行う「業績測定」のチェックは会計検査院や財務省の行うチェックに比べると実施手順が定式化されたものではなく、必要に応じて省庁の相談のりアドバイスを行っているというのが実状である²⁰。

（ウ）政策評価のチェックの基準・視点

政府サービス委員会の目的・使命は「政府部門全体の業績を向上させる」ことであるため、各省庁へのアドバイスも「各省庁の業績を向上させる」という視点から、政府サービス委員会の公表する数々のガイドライン²¹に依拠し、各省庁のアウトプット・アウトカムの設定および指標の適切性に対して、なされている。

¹⁹ 脚注 7 および 8 参照。

²⁰ 今回の現地調査で政府サービス委員会を訪問先に含めなかったのも、「政府サービス委員会を調査すべき」という声がニュージーランド側から聞かれなかったためである。「業績測定の調査にあたっては会計検査院と財務省を訪問すべき」との声が大勢を占めた。

²¹ 脚注 7 および 8 参照。

エ 3 機関の協力体制

以上見てきたように、会計検査院はサービス業績表（アウトプット群・指標・目標値・達成度）を監査する立場から、財務省はアウトプット予算を編成する立場から、政府サービス委員会は省庁全体の業績を改善する立場から、それぞれ「業績測定」のチェックを行っているが、互いに連携・協力もしている。

財務省と政府サービス委員会は各種ガイドラインを共同で作成・公表している²²ことからわかるように、中心省庁の一員として常に連携している。

会計検査院と2つの中心省庁は必要に応じて連絡をとってきたが、既述した2008年6月の会計検査院の報告書の中で、各省庁が会計検査院と中心省庁から異なるアドバイスを受け混乱しているため会計検査院と中心省庁間の連携を強化する必要性が指摘された。そこで、現在では、会計検査院・財務省・政府サービス委員会とアドバイスを受ける省庁とが一同に介するミーティングを定期的に行うようになってきている（現地調査②および⑤）。また、財務省・政府サービス委員会が共同でガイドラインを作成する際、会計検査院に協議することも多くなってきている²³。

財務省・政府サービス委員会・会計検査院は、「業績測定」の分野に限らず80年代以降のニュージーランドの行政改革を先導してきた機関であり、今後とも望ましい役割分担と協力体制に向けて改善を繰り返していくものと思われる。

3 「評価 (evaluation)」とチェックシステム

(1) 「評価」

ニュージーランドでは、先に定義した意味での「評価 (evaluation)」は「業績測定」ほど広くは行われていない。アメリカにおける「プログラム評価」とはかなり違った様相である。これは、繰り返しているようにニュージーランドの省庁のマネジメントの基礎が「アウトプット」にあり、「政策 (policy)」・「施策 (programme)」といった切り方をマネジメント上はあまりしないこと、そしてこれまで説明してきたような理由により特に財務省が「評価」にあまり関心を向けてこなかったことなどによる。また、「評価」には多大な労

²² 本章に示される各種ガイドラインの著者を参照のこと。財務省と政府サービス委員会の共著が多い。

²³ たとえば、State Services Commission & Treasury (2008) や Treasury & State Services Commission (2007) など。会計検査院に事前協議したことが本文に明記されている。

力・資源を要するが、ニュージーランドのような小国ではそのための資金や人材が不足していることも、理由として挙げられる（現地調査⑤（会計検査院））。

しかし、1999年以降マネジメントの焦点がアウトカムに少し移行したのに伴い、アウトプットとアウトカムとの因果関係を明らかにする「評価」の重要性が再認識され、多くの省庁で「評価」が行われるようになってきた。各省庁では、評価担当部署（evaluation unit）を設置して「評価」を行うことが一般的である。たとえば、社会開発省（Ministry of Social Development）では社会調査評価センター（Centre for Social Research & Evaluation: CSRE）を省内に設置して、同省の実施した政策・施策の「評価」を行っている。また、労働庁（Department of Labour）では5つの局の中にそれぞれ評価担当者・部署を設置し「評価」を行っている²⁴ほか、労働市場政策局（Labour Market Policy Group）が他の4つの局の「評価」の設計（design）に対してアドバイスをする役割も担っている（Turner 2003:40-41）。このように、社会保障・福祉・労働・教育といったソーシャル・ポリシー（social policy）分野の省庁では、特に「評価」に力を入れている²⁵。

しかし、「評価」は「業績測定」のように義務化されているものではなく、あくまで各省庁の自主的な活動である。「評価」に関するガイドラインでも、全てに正式の「評価」をするのではなく重要なものにしばって行うこと、また全ての省庁が評価担当部署を設置すべきとは考えていないことが明記されている（State Services Commission & Treasury 2003）。先ほどふれたように、「評価」に要する資源・人材がニュージーランドでは十分になく、より効率的で効果的な方法が望ましい場合があると考えられているのである。

なお、ニュージーランドで「評価」という場合、通常「事後」の評価を指す（State Services Commission & Treasury 2003:8）。「事前」の評価は「政策分析（policy analysis）」と呼ばれることが多く、事後の評価である「評価」を「政策分析」（事前評価）につなげるという政策サイクルが想定されている。

²⁴ 大規模な調査が必要になるなど当該評価部署で「評価」を行うことが困難な場合には民間に委託することもある。

²⁵ ソーシャル・ポリシー分野では関係省庁の行う「評価」を横断的に調整する委員会（Social Policy Evaluation and Research Committee）が設置されている。同委員会のメンバーは、社会開発省・科学技術省・財務省・法務省・教育省・保健省・労働庁・女性政策省・太平洋政策省・マオリ政策省・統計庁・内務庁・首相内閣庁・政府サービス委員会・歳入庁などであり、これらの省庁では「評価」が行われていると解釈することができる。また、環境省・運輸省・経済開発省・国立国会図書館でも「評価」は行われている（State Services Commission & Treasury 2003）。

(2) チェックシステム

ア 組織・体制

「評価」を各省庁が自主的に行っているため、その「チェックシステム」も各省庁が自らの判断により構築している。

イ 実施手順・ツール

たとえば、社会開発省の評価担当部署である前述の社会調査評価センターでは、「評価」の質を確保する品質保証（quality assurance: QA）システムとして、次のようなチェックを行っている（現地調査③（社会開発省））。

(ア) チェック・シートによるチェック

「チェック・シート」（表2参照）を用いて、同センターの作成する全ての調査・評価レポートをチェックする。「チェック・シート」は「目的・読者」「構成」「文体」「レファレンス」「分析」の5つの柱から成り、各柱には複数のチェック項目が設けられている。また、定量的分析（quantitative analysis）を使ったレポートは定量的分析に詳しい者がチェックするというように、5つの柱はそれぞれ適切な技能を持った者がチェックしなければならない。最終的に、評価部署の長がチェック・シートにサインし承認する。

以上のチェックに加え、外部に公表するレポートはさらなるチェック・シート（'pink peril'）でチェックする。さらに、内閣に提出するレポートはさらなるチェック・シート（'purple peril'）でチェックする。

表2：チェック・シート（社会開発省社会調査評価センター）

A3018149

CSRE QA CHECK SHEET

The purpose of this check sheet is to help those reviewing CSRE reports and related outputs to meet CSRE's standards for its reports. Please use the QA sheet for all research and evaluation reports. R&E reports which will be released externally also require a completed pink peril form.

Instructions to Author

Please complete report title and description and arrange reviewers.

Each section of the QA sheet will be signed by a person appropriate to review that section. For example quantitative analysis must be reviewed by a reviewer with quantitative skills. The reviewer could be internal or external to your team. One reviewer may sign off more than one section of the QA sheet. After all reviewers have signed the QA sheet the unit manager will sign.

Instructions to Reviewer(s)

Please check each section. Provide feedback if required. Sign approval when all feedback has been responded to.

Section 1: Purpose and audience

Who the intended audience is for the document and what is its purpose are critical to assessing whether the document is at the correct standard. For example, reports to be released to the public require a different style from internal documents for CSRE analysts.

Section 2: Report structure

The way the report is organised is very important in determining how easy it is for people understand the important findings and conclusions of the analysis.

Look for: key information is at the beginning of the document (purpose, findings, conclusions), whilst detailed analysis, tables and technical method are towards the end.

Look for: are there any sections redundant to the overall analysis or discussion.

Section 3: Style, look and voice

Consistent, clear and easy to understand style helps to communicate the key information in the document as well as identify the document as a CSRE product.

Look for: a low proportion of passive sentences, concise sentences, judicious use of technical terms.

Look for: font type, size, headings, table and figure layout.

Section 4: Referencing and documentation of information sources

Referencing of information sources is important to provide confidence in the quality of the analysis itself, but also to enable updating of analysis, quality assurance and to follow up with the analysts involved.

Look for: an EDRMS reference or name(s) of analysts involved.

Look for: a summary of method, references for tables and charts, dynamic links to excel workbooks.

Look for: in text referencing, a list of references.

Section 5: Analysis, tables and figures

CSRE stands on providing robust analysis where findings and conclusions are supported by the evidence.

Look for: summary of method, appendix of technical notes or reference to a separate technical document.

CSRE Report Description	
Report title:	
Brief Description:	
Purpose of report:	
Intended audience:	
Version:	

Report Review			
Item	Description	Comment/Recommended Action	
Section 1 : Purpose and audience		Reviewer :	
Audience	Is this document suitable for this audience?	Yes <input type="checkbox"/>	
		No <input type="checkbox"/>	
		N/A <input type="checkbox"/>	
Purpose	Is it clear what the document is trying to achieve?	Yes <input type="checkbox"/>	
		No <input type="checkbox"/>	
		N/A <input type="checkbox"/>	
Approved	Signature:	Date:	
Section 2 : Report structure		Reviewer :	
Flow	Is the document structured so the purpose, findings and conclusions are obvious?	Yes <input type="checkbox"/>	
		No <input type="checkbox"/>	
		N/A <input type="checkbox"/>	
Length	Is the document an appropriate length to communicate its message?	Yes <input type="checkbox"/>	
		No <input type="checkbox"/>	
		N/A <input type="checkbox"/>	
Approved	Signature:	Date:	
Section 3 : Style, look and voice		Reviewer :	
MSD naming and syntax conventions	Does the document meet MSD's syntax and naming conventions? (refer to MSD style Guide)	Yes <input type="checkbox"/>	
		No <input type="checkbox"/>	
		N/A <input type="checkbox"/>	
Spelling and grammar check	Is spelling correct throughout the document? Is the text in plain English?	Yes <input type="checkbox"/>	
		No <input type="checkbox"/>	
		N/A <input type="checkbox"/>	
CSRE format style guide	Is the document consistent with the CSRE style guide?	Yes <input type="checkbox"/>	
		No <input type="checkbox"/>	
		N/A <input type="checkbox"/>	
Approved	Signature:	Date:	

Section 4 : Referencing and documentation of information sources**Reviewer :**

Provenance	From the document is it possible to know who undertook the analysis?	Yes <input type="checkbox"/> No <input type="checkbox"/> N/A <input type="checkbox"/>	
Source information	Are you able to easily identify the source of the information used in the report? Are all references to other studies referenced?	Yes <input type="checkbox"/> No <input type="checkbox"/> N/A <input type="checkbox"/>	
References	Are all references to other studies or documents properly referenced?	Yes <input type="checkbox"/> No <input type="checkbox"/> N/A <input type="checkbox"/>	
Data consistency	Do numbers cited in the text match the tables and figures?	Yes <input type="checkbox"/> No <input type="checkbox"/> N/A <input type="checkbox"/>	
Approved	Signature:	Date:	

Section 5 : Analysis, tables and figures**Reviewer :**

Data collection	Are the sources of data clearly stated? Is the date of information gathering stated? Has respondent privacy been protected?	Yes <input type="checkbox"/> No <input type="checkbox"/> N/A <input type="checkbox"/>	
Methodology	Is the methodology clearly documented? Is the date of information gathering stated?	Yes <input type="checkbox"/> No <input type="checkbox"/> N/A <input type="checkbox"/>	
Analysis	Is the method of analysis clearly stated? Is the analysis technically robust?	Yes <input type="checkbox"/> No <input type="checkbox"/> N/A <input type="checkbox"/>	
Data to findings	Are the findings consistent with the information presented?	Yes <input type="checkbox"/> No <input type="checkbox"/> N/A <input type="checkbox"/>	
Findings to conclusions	Do the conclusions logically follow the findings?	Yes <input type="checkbox"/> No <input type="checkbox"/> N/A <input type="checkbox"/>	
Peer review	Has the analysis been peer reviewed by a technical expert?	Yes <input type="checkbox"/> No <input type="checkbox"/> N/A <input type="checkbox"/>	
Approved	Signature:	Date:	

Final Approval

Team Manager	Signature.....	Date
Unit Manager	Signature.....	Date

(イ) 外部の専門家による評価

従来とは異なる新しい手法を用いて「評価」を行った場合には、当該「評価」について外部の専門家による評価（external peer review）²⁶を受ける。

(ウ) 外部の評価機関による評価

社会開発省社会調査評価センターの行う「評価」は、ニュージーランド経済調査研究所（New Zealand Institute of Economic Research）²⁷による評価を年1回受けている。センターの行っている「評価」のうち、より重要な「評価」、すなわち内閣に提出する「評価」レポートを中心に当該評価の手法・プロセスが適切かどうか²⁸を評価する。

なお、「評価」レポートの外部への公表を義務づけることが最大の品質保証であるとの声も現地調査ではきかれた（現地調査③（社会開発省））。「評価」レポートの内容に問題があれば、直ちに外部から指摘されるというのである。

以上、社会開発省の事例をあげたが、労働庁や教育省（Ministry of Education）でも外部の専門家による評価は行われているという（State Services Commission & Treasury 2003:21）。このことから社会開発省以外の省庁でも、(ア) チェック・シートによるチェック、(イ) 外部の専門家による評価、(ウ) 外部の評価機関による評価、などの方法を適宜組み合わせることにより、チェックを行っていることが想定される。

省庁のチーフ・エグゼクティブに業績達成責任が厳しく課されているニュージーランドでは、チーフ・エグゼクティブが自主的に「評価」をチェックするシステムを設けることは十分に予想されることである（現地調査③（社会開発省））。政策提言は「アウトプット」の一つであり、質の高い政策提言というアウトプットを大臣に届けるためには、「評価」を行うことが必須になるからである。

²⁶ たとえば、社会調査を専門とする国内外の大学の研究者が評価する。

²⁷ ニュージーランド経済調査研究所は民間非営利の研究所で、（政府を除いた）民間ではニュージーランド最大級のエコノミストを擁する機関である。国内外の政府及び民間部門のクライアントに対して様々なコンサルティング業務を行っている（<http://www.nzier.org.nz>）。

²⁸ たとえば定量的分析を行っている場合には、定量的分析を採用することが適切なケースであるか、また（適切な場合には）当該定量的分析が適切に行われているかなどをチェックすることである。

4 その他の政策評価とチェックシステム

(1) 規制インパクト分析

省庁が新たな「規制」(regulation)を提案する際は、各省庁は規制インパクト分析(regulatory impact analysis)を行うことが原則として義務づけられている²⁹。そのうち、経済成長に重大な影響を与える規制については、政府の経済・財政アドバイザーである財務省がその適正性をチェックする。

規制インパクト分析をチェックする機能は、従来経済開発省(Ministry of Economic Development)が担っていたが、2008年11月より財務省経済業績局に移管された。

(2) 費用便益分析

費用便益分析(cost benefit analysis)は各省庁が新たに政策を立案・選択する際に有用であるとして、財務省からガイドラインが公表されている(Treasury 2005)。財務省が各省庁の予算を査定する際にも活用するとされているが、各省庁における費用便益分析の実施および財務省におけるそのチェックは、「規制インパクト分析」の場合と異なりあくまで自主的なものである。

5 まとめ

ニュージーランドでは、「業績測定」と「評価」という2つの政策評価が主に行われているが、中心を占めるのは「業績測定」である。各省庁は「業績測定」を行い、そのチェックを会計検査院・財務省・政府サービス委員会が行う。また、1999年末以降は「評価」もかなりの省庁で行われるようになってきたが、あくまで各省庁の自主性によるものであり、そのチェックシステムも各省庁の判断で構築されている。

「業績測定」のチェックを行う会計検査院・財務省・政府サービス委員会はそれぞれ明確に異なる立場からチェックを行っているが、同時に連携・協力もしている。日本においても、関係機関の役割分担を明確にした上で連携して業務を行うという体制をよりいっそ

²⁹ 規制インパクト分析に関するガイドライン(*Guidelines on the Regulatory Impact Analysis Requirements*)は財務省から公表されている(Treasury 2008a)。

(<http://www.treasury.govt.nz/publications/guidance/regulatory/impactanalysis>)。

う整備していく必要があるのではないだろうか。

ニュージーランド現地調査（2009年2月24日～27日）	
①	2009年2月24日 15:00～ヴィクトリア大学行政大学院（School of Government, Victoria University of Wellington）でのインタビュー Rob Laking 氏（Senior Lecturer） Derek Gill 氏（Visiting Senior Fellow）
②	2009年2月25日 10:00～財務省（Treasury）でのインタビュー Jonathan Ayto 氏（Principal Advisor, Public Sector Management） Louise Lennard 氏（Senior Analyst, State Sector Performance Group）
③	2009年2月25日 15:00～社会開発省社会調査評価センター（Centre for Social Research & Evaluation, Ministry of Social Development）でのインタビュー Jeff Sheerin 氏（Manager, Knowledge Sharing & Communications） Daniel Soughtton 氏（Analyst）
④	2009年2月26日 9:00～内務庁（Department of Internal Affairs）でのインタビュー Janine Hamilton 氏（Manager, Risk & Audit Services） Ian Paterson 氏（Retired Officer from the Audit Office）
⑤	2009年2月26日 15:00～会計検査院（Office of the Auditor-General）でのインタビュー Ann Webster 氏（Assistant Auditor-General, Research & Development）
⑥	2009年2月27日 13:00～ヴィクトリア大学政治学部（Department of Political Science, Victoria University of Wellington）でのインタビュー Stephen Levine 氏（Professor）

第5章 政策評価の「チェックシステム」の可能性

山谷 清志

はじめに

「二次評価は、二度評価することではない」、政策評価のチェックにはそうした意味が込められているが、この言葉を理解できる人は評価実務にかなり詳しいはずである。

そもそも、政策評価のチェックシステムを考える際の重要なポイントは、およそ6点ある。すなわち、①チェック機関の組織・体制、②チェックの実施手順・ツール、③政策評価に関わるデータの開示状況、④チェックの基準と視点、⑤チェックによる具体的な指摘事例、⑥チェックをめぐる理論の状況の6点である。ただし、これらを国際比較によって考えるとき、いくつか事前に確認していなければならないことがある。

たとえば①に関しては比較政治制度論・比較政府論の観点が必要である。各国政府におけるチェック機関（その代表はアメリカの会計検査院 GAO）の制度上の配置、その政府の改革を導いた理念・理論に関する理解と、アナクロニズムの誤りを避けるためこれら制度・理念・理論における時代適合性に関する認識が必要であろう。

他方、②・③・④・⑤はミクロ・レベルでの視点であり、そもそもチェックシステムと言うとき、「何を、どのようにチェックするのか」の実践知を導く現場の経験と、この実践知をもとに発展させた理論が必要になってくる。アメリカ、イギリス、ニュージーランド各国の経験は、この点において多いに参考になるのであるが、これについては本報告書の各章でふれている。

ところで、米・英・新の各国政府で行われている「政策」を「評価」したものをチェックするとはいったい何を意味しているのであろうか。言い換えると‘policy evaluation’ ‘performance measurement’ と呼ばれる活動をチェックすることによって、各国政府が何のために、何をしようとしているのかを把握する必要がある。これはメゾ・レベルでの認識であり、上記⑥の議論である。

いずれにしても、「政策評価とそれに対するチェックシステム」という言葉で意味している概念と実践活動が、実は各国それぞれにおいて微妙に違っている。ここでは、こうした一連の議論をふまえて、わが国の政策評価制度との関連で残された論点について考察をす

すめたい。

1 チェックシステムの概念と実務

(1) チェックシステムの本質をめぐる議論

そもそも、政策評価のみならず、およそあらゆる政府活動に対するチェックシステムという言葉でわれわれが思い浮かべるのは、法令や標準、基準、ガイドラインを設定し、それらにしたがって評価活動が行われているかどうかを確認する作業であろう。

わが国の政策評価システムも、「行政機関が行う政策の評価に関する法律」（平成 13 年法律第 86 号）、およびそれに先だつ「政策評価に関する標準的ガイドライン」（政策評価各府省連絡会議、2001 年 1 月）によって設定された大枠の中にチェックシステムが置かれている。その大枠においては、一方で評価の 3 方式（総合評価・実績評価・事業評価）を前提に、評価内容に対して行うチェックがあり、また同じ内容のチェックでも総務省行政評価局が行う「複数の府省に共通するそれぞれの政策についてその政府全体としての統一性を確保する見地から行う評価」（統一性確保評価）、「複数の府省の所掌に係る政策についてその総合的な推進を図る見地から行う評価」（総合性確保評価）がある。他方、評価の形式・やり方に対するチェックとしては「各府省の政策評価についてその客観的かつ厳格な実施を確保するために行う評価」（客観性担保評価）がある。この統一性確保評価も総合性確保評価も政府全体、府省横断などの視点においてチェックシステムを構築しており、ともに行政実務から得られた経験知や行政のプロフェッショナルリズムの専門的知識、あるいは統一性と総合性を考えると公務員であれば当然もっている常識によって実施可能であろう。

しかし、政策評価のチェックシステムの議論に一番深い関わりを持つ客観性の担保については、悩ましい事情がある。それが外部者の視点で行う外部評価を意味する活動なのか（外部審査）、それとも第三者による評価であるのかははっきりしないからである。たとえば、わが国の政策評価制度が外部委員会を活用している実態は、各府省の外部有識者会議や総務省の政策評価・独立行政法人評価委員会を、第三者委員会として全く利害関係の無い人を連れてきているのではなく、各府省が知りうる範囲での外部専門家を活用する実態になっている。したがって、たとえばある省の政策評価外部有識者のメンバーが総務省の政策評価・独立行政法人評価委員会のメンバーでもある実態があり、結果として「全く利害関

係のない第三者」を任命していない。その意味で、わが国の政策評価制度における実際を見れば、評価の無関係の第三者による「独立性」は想定していないことになる。

そもそも、評価理論において外部評価とは評価の任務にたえると信用（accreditation）され、鑑識眼（connoisseurship）があると認められたとき機能する（Mathison 2005: 150）。したがって、議会や国会の委員会が外部者として機能しているのは、たとえフィクションであれ立場上信用と鑑識眼が法律によって正統化されているからである。逆に、評価を任せるほど信用できない、鑑識眼が疑われる人を連れてきたとき、しかも行政内部の要綱・要領等によって任命しているとき、外部評価は機能しないと思うのが常識であろう。この常識から考えると、各府省が鑑識眼ありと信用できる専門家を連れてくるのは最低限の条件であり、また欧米政府が外部のシンクタンクやコンサルタントをおおいに活用するのも同じ理由による。時おり耳にする批判、すなわち「外部の人を委員に選任し外部性を主張するとしても、人選の原案を作るのが政策とその評価を担当する当の行政機関であって、なおかつその行政機関から委員手当や謝金が出るのであれば第三者性は失われる」という批判は、この意味での的外れなのかも知れない。ただ、民主主義の理想から推奨されることがある公募方式は、評価についての理解もなく、政策内容に関しても無知な素人が来た場合、議論は混乱を極めるリスクがある。また、政策に関して個人的に重大な利害関係を持っている人が来ても排除できないため、問題かも知れない。

なお、「素人の健全な常識」に頼り、行政の外から連れてきた素人によるチェック方法を考える場合もあるが、これについては問題が多い。外部の第三者が政策の執行プロセスに関わる評価に口を出し、重箱の隅を突つつくような細かな注文を付けると、政策実施は滞り、関係者や受益者の不利益、トラブルを招くだけでなく、いたずらに手間暇かかりすぎる事態になる。そこで、あらかじめ設定したポイントについて結果の事後チェックを行う方法が選択されるが、素人には「何でそのポイントを選んだのか」理解は難しい。したがって、的外れの質疑を繰り返したセレモニーの後、「こうなっています」と言う行政機関の説明を形式的に事後了承することになる。この場合、行政機関は形式的な事後了承をもって外部の有識者に見てもらったという正当化に外部有識者会議（素人）を使い、アカウントビリティを果たしたことにする。つまり、わが国の現状では、外部の第三者を使って政策評価をチェックすることに意味があるとは思えない実態が多いのである。

チェックシステムを制度化する対応としては、アメリカ連邦政府の会計検査院（GAO: General Accounting Office, 2004年7月から Government Accountability Office）や、イ

ギリスの会計検査院（NAO: National Audit Office）の存在が考えられるが、わが国の総務省行政評価局や会計検査院の制度的改革が必要になり、この議論は現実的ではない。したがって、チェックシステムの議論をするときに重要なことは、わが国の制度的制約を考えたときに現実的であるか否かであろう。

表 1 政策評価のチェックシステムの類型（筆者作成）

	「客観性」を確保する制度と方法	専門性で代用する場合
外部から	A： 第三者委員会、議会の委員会、監査委員会によるチェック。 <u>結果</u> 重視。	B： 評価の専門家、政策評価について知見を持つと考えられるシンクタンクやコンサルタント・NPO、専門学会（日本評価学会）を活用。
内部で	C： 政策評価担当課、総務課、財政課、官房などがコンサルタント・シンクタンクに委託。その際、TORの事前チェック・事後検査。手続・ <u>プロセス</u> でチェック。	D： 研修。職員を大学院に派遣で学位取得。資格（評価士）取得。有識者の叡智を借用。他の自治体との相互交流、職員交換。専門家を任期付きで採用、評価システムをチェックしてもらう。

（2）チェックシステムの現実的アプローチ

そこで、わが国の政策評価においてチェックシステムを考える際に不可欠な現実的アプローチは、およそ3点ある。

その第1はチェックを担当する制度・組織、そしてスタッフについての考え方である。そしてわが国では長期的展望は別としても、前述の第三者評価制度や GAO のような議会附属組織への制度変更は今のところ非現実的なので、「外部評価」を専門性で代用する方途を考えるべきであろう（表1のB）。すなわち、評価の専門家や政策の内容に関わる専門家の活用である。評価の専門家の資格認定を行っている専門学会の活用を考えても良い（日本評価学会）。

コンサルタントの活用を「外部評価」と見なす例も欧米では多いが、わが国ではそこまで考えない。理由は行政側にコンサルタントを「(行政の) 下請け業者」として低く見なす気風が残っていること、コンサルタントの側にも「商売」を考え客観性や自律性を軽んじることが少なくないこと、良質なコンサルタントを雇うには費用がかかるのに行政側の予算が不十分で「安かろう・悪かろう」に委託せざるを得ないこと、またコンサルタント側にもコンサルタントを雇う側にも社会全体で合意された資格確認手続がないので「委託→外部評価と見なす」がうまく機能しないためである。実態としては、評価業務に詳しいス

スタッフが、透明性が高く質的にも高レベルの TOR (terms of reference : 委託のための条件を定めた文書) を作成することによって、内部で客観性を確保ことが可能になるが (表 1 の C)、ゼネラリスト育成型の行政組織では難しい。委託時の TOR を巧く活用できなかった場合、コンサルタントから出てきた報告書は「ゴミ」になってしまう。なお、表 1 の A 欄の場合、無能な第三者を連れてきた場合、最悪のチェックシステムができあがる。

ところで専門性を備えていることと、「有識者」であることとはイコールではない。有識者は世間一般では「立派な常識人」であるが、専門家ではないからである。専門家 (プロフェッショナル) と言われるようになるのは、‘capacity-building’ (能力開発) や研修を受ける、学位・資格を取得する、ピア・レビューによって同業の玄人から見て有能と判断されることである。したがって、専門家でプロであると素人と玄人の両方を納得させることができれば、仮に行政の内部者であっても一定の信頼を得ることができる (表 1 の D)。そして、評価の専門家とは具体的に言えば、評価手法ごとに何が必要か、どのようなポイントを押さえていなければならないか、助言できる人であろう (その例として表 2 を参照)。

表 2 評価手法ごとのチェック・ポイント例

評価の種類	チェックすべきポイントの例
プログラム評価	プログラムは政策の目標と手段についてロジカルな思考に基づいて、正しく構築されているか。
セオリー評価	プログラム・セオリーの因果関係の確認。
プロセス評価	評価プロセスは正しいセオリーで実施されているか。
アウトカム評価	評価対象はプログラムのアウトカムなのかアウトプットなのか。
業績測定	設定された目標数値の適切さ。無理に定量評価にしていないか。
アウトプット測定	インプット指標、アウトカム指標と混同していないか。
プロジェクト評価	費用と便益の定義は正しいか。
評価サイクルの設計	ニーズ・アセスメントからはじまり、evaluatability assessment、formative evaluation、summative evaluation で終わる評価サイクル実施は適切か。

(筆者作成)

チェックシステムの現実的思考の第 2 は、あらかじめ設定した厳格な手続や工程表に基づいて評価実施を確認する「プロセス・チェック」だと割り切ることである。この割り切りはアメリカで見られるようなメタ評価の専門家がないとき有効であり、しかしこれは監査に似てくる。すなわち、

- 執行部 (たとえば内閣や経済財政諮問会議) の要望 (予算・政策への反映) に応

えているか

- 基本計画・実施計画・実施要領にもとづき評価が行われているか
- データは信頼できるか
- タイムリーか
- わかりやすく、活用しやすい評価結果か
- 法律や政令で定められた通りであるか（かつて外務省は総務省に客観性担保評価で未着手案件と未了案件の重複を指摘された）

第3の現実的アプローチは、チェックが何のために行われ、何に使うのかをあらかじめ決定しておき、この決定から逆算してチェックシステム、チェック項目を構築することである（後述する JICA の課題別研修・第三者検討委員会の事例がこれである）。ここではまず、数多く存在するチェックシステムを分類する仕事から始めるべきであるが、その場合でもまた空理空論では困るので、実際面からアプローチする必要がある。

ところで、実際のチェック活動を分類するとどのようなタイプに分けられるであろうか。その一例は、2009年3月末に各府省に通知された2008年度総務省行政評価局「政策評価の内容点検」である。すなわち政策評価の質の向上と実効性の確保を目的として、各府省が行った政策評価の内容を点検し、改善を求めた総務省の活動事例を取り上げると、以下のように分類される（下線は筆者）。

表3 総務省の内容点検事例－2008年度

I 公共事業（19件）	*継続のものは含めない
1. 便益算定の前提となる <u>需要予測の妥当性</u> に疑義があるもの 2. 便益算定に際しての <u>評価方法の妥当性</u> に疑義があるもの 3. 便益算定に用いられている <u>データ等の信頼性</u> に疑義があるもの 4. <u>マニュアルの適用の妥当性</u> に疑義があるもの	
II 一般政策（30件）	
1. 目標の達成度合いが低調であるにもかかわらず、その <u>原因分析が行われていない</u> と考えられるもの 2. 設定されている指標が専ら政策の執行の状況をとらえており、 <u>政策効果に着目した指標の設定が必要</u> と考えられるもの 3. あらかじめ設定した指標による <u>効果の測定が行われていない</u> もの 4. 判断 <u>基準・指標等の設定について改善が必要</u> と考えられるもの 5. <u>測定指標等の状況と評価結果の結びつきの説明について改善が必要</u> と考えられるもの	

（出典）総務省資料を基に筆者作成

この事例で見る限り、公共事業に関しては事後的チェックではあるにしても、方法や考え方は「形成的メタ評価 (formative meta-evaluation)」であり、評価の実施にあたってそのデータの集め方やデータそのもの、評価分析方法、マニュアルを対象にしている。一種の「評価の指南役」である。また、一般政策については結果や成果に関係した議論であり、総括的メタ評価 (summative meta-evaluation)、つまり「評価のご意見番」に近い使い方をしている。

政策評価に関する方法論の認識が深まり、専門家が育ち、アメリカの大学院教育のように評価の専門化を目指す人びとがほぼ同種のテキストを使い、類似のカリキュラムでリテラシーの標準化と知的レベルアップが可能になれば、表3の公共事業に対する事後的「形成的メタ評価」は必要なくなるであろう。そのときには一般政策に対する総括的メタ評価についての本格的議論が可能になり、国民的レベルでの政策リテラシーも向上するはずである。しかし現実はそれとほど遠い。

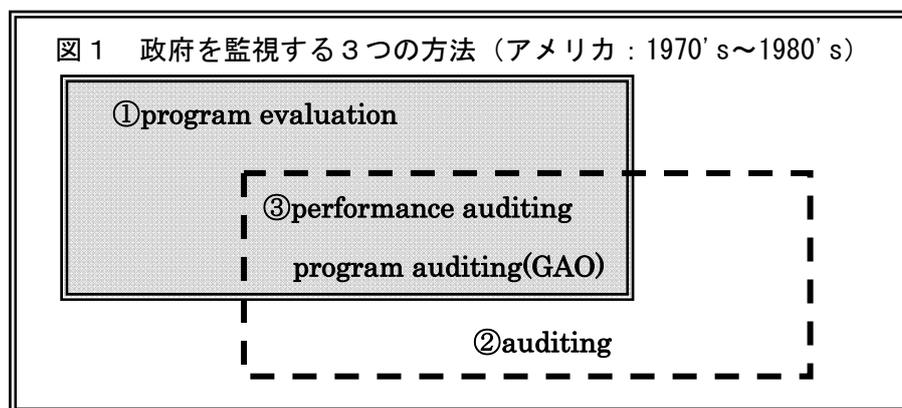
2 チェックシステムのプロトタイプ

わが国では総務省行政評価局が中心となった政策評価チェックシステムが機能しているが、その理論構築についてはあまり議論が無い状況にある。一方、政策評価の先進国のアメリカではかつて、政策評価（わが国の総合評価にあたるプログラム評価）が使われ始めた1970年代から1990年代はじめにかけて、実務においても理論においても多くの議論が見られた。この時期には予算編成プロセスと評価プロセスとは別立ての議論であるという認識のもとで、予算編成にはパフォーマンス測定を使い、「評価の評価」はメタ評価の議論に収斂させて1990年代はじめに一定の決着を見ている。

(1) 評価をチェックするふたつの異質な方法

もともと1970年代から1980年代にかけて政府活動、政策やプログラムをチェックする方法はふたつ考えられていた(図1)。それは‘evaluation’と‘audit’であり、両者の関係について実務から出た議論が存在した。すなわち、GAOのプログラム評価が1980年代にある程度定着を見てから1992年にオズボーンとゲーブラーの *Reinventing Government* が刊行され、翌1993年にG P R A (Government Performance and Results Act) が作られるまでの15年ほどの間、アメリカ政府では図1の①と②が重なり合う③の

部分を理論的にどのように考えるべきかが論争になっていたのである。つまり、評価と監査の違いは何で、それぞれ何をめざして行われ、それを議会付属の会計検査院として監査を主たる業務としてきたGAOは、どのように使い分けているのかをめぐる論争である。



(筆者作成)

論争は結局、実務主導で‘auditor’が‘evaluation’を行う形の折衷的なスタイルで決着され、一方の‘auditing’は‘performance auditing’（業績検査）と名付け、これをGAOはプログラム検査（program auditing）と再定義し、プログラム評価（program evaluation）と区別して実施していた。こうして事実上監査と評価は「異種混交状態」になったのである（山谷 1997）。なお、‘performance auditing’は後に‘performance measurement’に接近し、‘program auditing’は‘evaluation research’の方法をより反映させたため‘program evaluation’に吸収されたと考えて良いであろう。監査と評価の違いは表4に見られるように、目的、関心事、対象、担当機関の独立性において違いが顕著である。異なる活動であるにもかかわらず、「異種混交状態」ではあえて両者の異質性を意識せず、また組織対応としても違いをつけないで実施してきたのであろう。

これにならって、評価をチェックする方法にもふたつあると考えられている。一つは「評価の評価（meta-evaluation）」であり、他は「評価の監査（auditing evaluations）」である。

表4 監査と政策評価、業績測定の関係

	業績監査 (performance audit)	評価 (evaluation research)	業績測定 (performance measurement)
主たる目的	アカウンタビリティの確保、統制	プログラムの背景にある学問分野や専門領域への知的貢献、プログラムに知的情報を反映	マネジメントの支援
主たる関心事	不正や乱費の調査・予防、規則や手続きの遵守 (compliance)	プログラムのインパクト、環境への影響、プログラムの改善 (政府外部の活動結果が対象)	内部管理、プログラム改善、組織活動の効率化に要する情報の産出、予算への反映
対象	組織構造、執行手続き (ただし財務監査ではないので財務情報は除かれる)	プログラム目標、プログラムのセオリー、プログラム構成要素、アウトカム (成果)、インパクト	プログラムの実施過程、アウトプット、組織活動 (パフォーマンス)
担当機関の独立性	強い (監視・コントロール機関としての独立性)	弱い (政策作成者、プログラム管理者に対する情報提供機能重視)	不要 (内部管理)
明確な責任基準	有 (権威ある機関が決定する)	無 (専門家としての一般的な倫理基準は有る)	無 (組織活動の効率化、節約について責任を持つ)
ルーツ	会計学 (accounting)、財務監査 (financial auditing)	社会科学 (social sciences) に基づく評価理論	経営学、マネジメント理論
プロフェッショナルリズム	強い (公認会計士・公認内部監査人としての 'auditor')	弱い ('evaluator' は専門職として定着していない)	ある (経営者、マネージャーとしての責任、マネジメント・コンサルタントとしての能力)
成功の判断基準	顧客ニーズに応える (分かりやすさ、低コストなど)、迅速なフィードバック	学問として、科学としての妥当性	効率化、節約による行政経営改善
タイミング・期間	定期監査	期間と報告タイミングは自由に設定	定期
政策評価への取り組み	弱い	積極的 (事前評価、中間評価、終了時評価、事後評価)	弱い

(筆者作成)

(2) わが国のチェックシステムの分析

きわめて単純化して言えば、「評価の監査」は評価の実施プロセスに対する形式審査であり、評価の運用実態に対するコンプライアンス確認とコスト分析を中心としたチェックである（総務省行政評価局の行う客観性担保評価活動のうち、「評価のやり方点検」がこれに相当する）。評価そのものの実質的内容には踏み込まない。この実質的内容を議論するのが「評価の評価」（以下「メタ評価」と言う）であり、これは評価の内容の適否、評価の結果についての実質審査を行う活動である（総務省行政評価局の行う客観性担保評価活動のうち、「評価の内容点検」がこれに相当する）。

前者の評価結果を監査の視点でチェックする活動である評価の監査は、おそらく監査一般と同じように評価活動に関わる資源の適切な使用、外部委託に出したときの入札の方法、仕様書（TOR、すなわちどのような仕事をいつまでにしたいのか、何を報告書として求めているかの条件記載）の適切さ、法令遵守、合規性などを中心とするのであり、形式審査であるため、評価の専門家が行う必要性は低い。ただし、「外部」評価と言える体制を作らなければ、内部監査に疑問を抱く人から「お手盛り」の批判を受ける。

他方、評価についての評価、メタ評価は評価方法に深く踏み込んだ実質審査になる¹。たとえばデータの正しさ、分析手法の適切さ、選択した評価手法の妥当性、評価報告書の有用性（「報告書は役に立ったか」）などを、評価結果を見ながら検証する作業になる。場合によっては費用便益分析、産業連関分析、ロジカル・フレームワークなど分析方法の選択の適否、あるいはその前提となる方法であるトラベルコスト法、ヘドニック・アプローチ、仮想市場評価法、コンジョイント分析、階層化意思決定法、バランススコア・カードなどの活用実態の調査分析が必要になるかも知れない（伊多波、2009）。測定指標や目標値の妥当性チェックを含めたこれらメタ評価の方法を遂行できるようになるには、専門的知見を持つことが重要な前提であり、したがって専門能力に欠ける人には不向きである。

ところで、図1の③に類似した活動も、わが国には存在する。たとえば、多すぎる評価件数の再検討（2005年までの1万件前後を2006年以降に4000件前後まで削減）、問い合わせや確認作業（確認のための問いの例としては、事業評価ではなく実績評価を選択した理由、記述に終始し総合評価方式と呼べない評価もどきのモノをなぜ「総合評価」と呼ぶ

¹ 客観性担保評価については、評価の実施の必要性の認定作業を経ても各府省が評価を実施しない場合、総務省が評価を行うスキームが存在するが、これはメタ評価ではなく、評価の代行であると考えられる。

のかなど)、定量分析の強調、予算への反映件数の確認などである。わが国のメタ評価においてこの種のチェックシステムが比較的有効なのは、担当する行政評価局がなじんだかつての行政監察の手法だからである。現場で綿密な調査を行い、インタビューをして入手される情報をもとに提示された結論は、「あるべき評価の姿」に近づける有益なアドバイスになる。その意味で、このカテゴリーの活動は外形的な審査ではある一方で、評価そのものにもアプローチしている折衷的な活動になっている。ただし、評価を監査のマインドでチェックすることになるので、細部にわたることが多く、メタ評価とは言えないかも知れない。

(3) メタ評価の種類

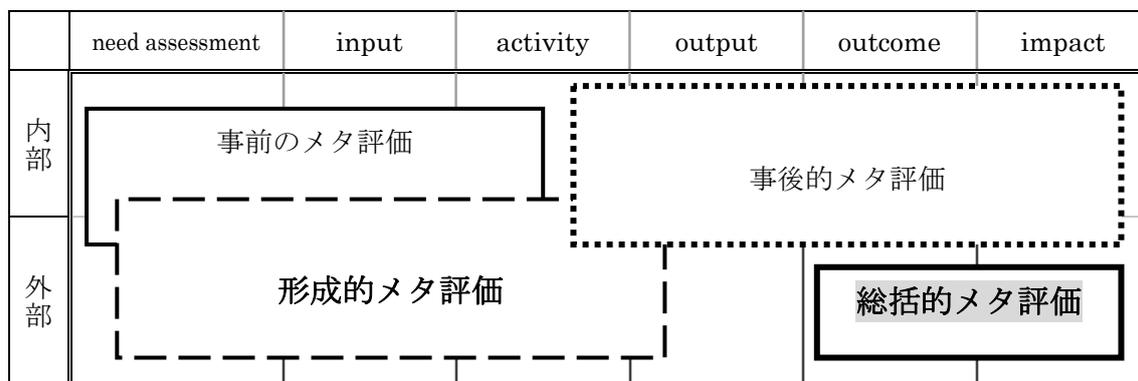
メタ評価についてはふたつの分類が可能である。その一つは事前・事後の時間的視点に立つ分類である (Hanssen et al. 2008)。組織の内部、外部にはそれほどこだわらない。

- 事前のメタ評価 (PME : proactive meta-evaluation)。評価を実施する前に、評価担当者の評価活動を支援する。
- 事後のメタ評価 (RME : retroactive meta-evaluation)。終わった評価を (外部者や他機関が) 判断するのに役立つ。

もう一つの分類は機能に注目する分類である (Stufflebeam, 2001)。

- 形成的メタ評価 (FME : formative meta-evaluation)。正しい評価を、確実に実施するため行う。
- 総括的メタ評価 (SME : summative meta-evaluation)。類似の評価との優劣比較、評価に対する深く掘り下げた調査、評価の有用性、フィードバックの状況が関心事である。

図2 4つの評価の位置関係概念図



(筆者作成)

「事前のメタ評価」は評価の事前チェックであり、対象は評価の計画（基本計画・実施計画・実施要項）、評価デザイン、評価スケジュール、選択予定の評価手法である。開発中の新しい評価手法を使おうとするときに必要になる、現場の評価担当者の「コーチ役」である。各府省にある政策評価の取りまとめを行っている課・室の役割に近い。また「事後のメタ評価」とは評価が出してきた報告の妥当性、有用性を「事前のメタ評価」と同じ項目を見て判断するのであり、もし妥当性に欠け有用でない結論が出たときは、その原因を評価プロセスに求めるため、評価手法やデータの内容にまで踏み込んだ調査をする。各府省の政策評価外部委員会・有識者会議の役割かもしれない。

「形成的メタ評価」は事前のメタ評価に似ているが、より客観性が高く、評価実施担当者とは距離があり、一段高所の外部から見た評価活動の円滑な運用に力点があるプロセス・チェックに近い。評価可能性（evaluability）、インプット評価、活動のモニター、アウトプット測定、アウトカム評価、インパクト評価などから構成される評価サイクルを見る。詳細に行うにはコストと時間がかかるため、ガイドラインや基準、チェックリストを作成して、これらに対する準拠状況のチェックで代用することもある。形成的メタ評価を事後的に行えば、総務省行政評価局の統一性確保評価、総合性確保評価、客観性担保評価の実践に似てくる。

最後の「総括的メタ評価」は事後のメタ評価とほぼ同じであるが、外部からの視点が強く求められ、まさにオーソリティに「総括」を求める場合が想定されている。たとえば評価結果が新たな政策に反映されているかどうか、仮に反映されていなければそれはなぜかを問うような時、この評価が使われる。あるいは、公共事業の再評価（時のアセス）によって中止や休止になった事業が、1998年から2008年まで全評価対象のたった2%に過ぎ

ないのはなぜかと問うときにも有効であろう。

もつとも、実際にはわが国では総括的メタ評価を行う処まで進んでいないのも現実である。可能性として考えられるのは、X省がQ県で行っている干潟埋め立て事業が住民の反対運動を引き起こし、なおかつQ地方裁判所で「バブル崩壊で需要予測が破綻した」ため工事差し止めの判決が出たにもかかわらず、その事業がX省の政策評価対象に上ってこないとき外部有識者が疑問を呈する、あるいは総務省が客観性担保評価のスキームを発動するという形での実施はあり得るかも知れない。

いずれにしても、これら4つのメタ評価に共通するのは、評価も「やりっ放しではいけない」「必要なモノについては評価しなければならない」というプロフェッショナリズムから発する意識であり、それは評価専門職の義務感である。

3 理論検証の事例研究

どの種類のメタ評価がいかなる場合に可能であるのか、事例に重ね合わせて見ることによってより具体的にイメージできるかも知れない。いくつかの事例を仮定してみたい。

(1) 某P県の独立行政法人・病院機構の評価事例

まず最初に、某P県の独立行政法人・病院機構の評価事例をつかって、さまざまなメタ評価の可能性、あるいはどのメタ評価が向いているのかを仮想してみよう。

表5 P県の独立行政法人・病院機構の評価事例

評価の種類	評価ポイント (evaluation の対象)	メタ評価の種類
「大きな政策」の評価	① 知事に指示された、不良債権60億円の5年以内での解消	総括メタ
	② 政策医療の継続 (儲けの少ない診療科目もやる)	総括メタ
小さな政策の評価	③ 知事が提示した中期目標の各項目の達成	事後メタ
プログラムの評価	④ 医師研修プログラムの評価を受ける (某大学医学部による)	事後メタ
業績測定 (達成度のチェック)	⑤ 蛍光気管支鏡による検査件数 07年の22件→08年85件	形成メタ
	⑥ 患者満足度: 他病院とベンチマークで比較	形成メタ
	⑦ 自閉症専門外来の充実: 確定診断待機患者数。 2008年4月838人→2009年1月646人。	形成メタ 事後メタ
	⑧ 待ち時間の改善	形成メタ

医療の質の確保	⑨ (財) 日本医療機能評価機構の病院機能評価を受ける	総括メタ
	⑩ 診療内容の標準化：クリニカルパスの適用率	形成メタ
	⑪ 「地域医療支援病院」の認定を受ける	総括メタ
事業評価	⑫ 女性専用外来棟の P F I による建設とその稼働率	形成メタ
経営評価(政策評価ではない)	⑬ 業務のアウトソーシング率	—
	⑭ 「高度専門医療」の提供で、診断単位を向上させる	—
	⑮ 同種同効薬品・機材の集約化による費用削減	—

(筆者作成)

この P 県の独立行政法人評価委員会では、県そのものに政策評価制度がないため（行政評価はある）、政策評価の議論もこの独立行政法人評価で行われることがある。地方自治体の現場ではこのように、政策評価、事務事業評価、経営評価などさまざまな評価が同じ土俵で進められるところに特徴がある。表 5 でも、知事選挙で取り上げられた大きな政策に対する評価、独立行政法人・病院機構に知事が提示した中期目標（政策）の評価、プログラム評価、業績測定、医療の質という専門評価、事業評価、そして経営評価が混在している。そして、これらの評価について外部評価委員会が話を聞いて 5 段階で評定し、「評価」したことにする。その評定する際に必要な情報としてメタ評価情報があるが、このメタ評価情報は先のメタ評価の 4 分類に当てはめるとそれぞれ表 5 の「メタ評価の種類」欄に記載しているようになる。

この表 5 の評価は某 P 県の独立法人評価委員会に提出された事後評価資料であるので、ここでは事前のメタ評価は存在しないが、毎年 6 月にその年度の評価項目等の検討が行われるので、その時が事前のメタ評価になる。また⑬～⑮は経営に関する評価であるため、政策評価とは関係ない。

(2) JICA 研修事業に対する外部評価委員会の事例

もう一例 JICA の課題別研修・第三者検討委員会の事例をとりあげる。ここでは研修事業を外部の第三者の目でチェックしようと試みている。委員会のメンバーは評価の専門家が 3 人、教育の専門家が 1 人、経営の専門家が 1 人であり、彼らが JICA 事業を実施の妥当性、客観的検証が可能か、有効性、効率性、総合評価の視点でチェックしている。

表6 JICAの課題別研修・第三者検討委員会による
チェックの視点とネガティブな指摘の例

<p>1. 案件実施の妥当性</p> <p>①案件目標は3年で達成できるとは思えない。</p> <p>②（自動車安全に関する基準・認証制度）研修の具体的なイメージが詰め切れていない。</p>
<p>2. 客観的検証の可能性</p> <p>①案件目標達成を判断すべき指標がない。</p> <p>②案件目標の達成を判断できるのは案件終了後相当年数経ってから。</p> <p>③継続案件であるが、なぜ継続すべきなのか、具体的な説明が弱い。</p>
<p>3. 有効性</p> <p>①案件の公式的目的（政策形成能力の向上）と実際の研修内容（工業用水利用・排水再利用）と食い違う</p> <p>②成果がアクションプランや報告書の作成になっているが、それでは成果の証明にならない。</p>
<p>4. 効率性</p> <p>①案件目標と研修の構成に整合性があるとは思えず、効率的ではあるが、目標達成は困難ではないか。</p>
<p>5. 総合評価</p> <p>①政策形成・実施の能力習得の講師がいるのか。</p> <p>②研修生の選考がまずい（研修を理解できる能力の確認になっていない）</p>

（出典）JICA 資料を基に筆者作成

なお、JICA「課題別研修」第三者検証委員会のメタ評価は、以下のⅠ～Ⅶのプロセスで行われる。

- Ⅰ. 各省が案件を提案する（165件）。
- Ⅱ. JICA内の作業部会で検討する。
- Ⅲ. 2008年10月、第1回第三者検討委員会は1～5で評定。評定のうち4と5は採択、3は条件付き採択、2は案件再形成、1は不採択。第三者委のコメントを付けⅣの段階に進める。
- Ⅳ. JICAが最終コメント付き評定（A、B、C）を行う。
- Ⅴ. 外務省が上記ⅢとⅣに基づいて各省と協議を行う。
- Ⅵ. 案件採択を提案する（165件中17件）。
- Ⅶ. 第三者検証委員会、C評定案件について再協議する。

このように各省提案の事業を評価したものを再度外部の委員会でメタ評価し、これをもとに採択・不採択を判断するのであるが、この外部評価委員会のメタ評価が実質的に機能

している理由は、事業内容が研修に限定されており単純であること、採択するかしないかというめざすべき判断基準が最初にあって、この基準を裏付けるため5つの視点をおいているからである。評価の評価をあらかじめ想定した評価手順が置かれているとき、メタ評価はうまく機能するのである。

結論

政策評価については施行後3年を経ての見直し（附則第3条）をきっかけに、2005年に大きな変更が見られた。この時には法改正はせず、政策評価に関する基本方針の改定と、新ガイドラインの制定で対応したが、この前提になっていたのは①2005年6月の「骨太方針」による予算との連携強化、②同年8月「内閣の重要政策に関する評価の徹底について」（総理大臣の閣議での指示）、③同年6月「政策評価制度の見直しに関する決議」（参議院本会議）であった。

このトピックの中で重要な新しいミッションは予算との連携強化である。これはそもそも政策評価を所管する官庁が総務省になっていたため、当初はそれほど強く意識されていなかったもので、制度設計も予算・決算連携に向いていなかった。この予算・決算との連携強化は現実問題として政策評価の変質と、政策そのものの構造（つくり）の変化を必要とし、したがって新たな理論展開が必要になる。しかし、制度的な見直しに対応した新たな理論展開はまだ十分でない。1960年代～70年代のPPBS（planning programming budgeting system）の挫折の反省が見られないまま再度同じ轍を踏もうとすることに、研究者の一部が躊躇し反発するからである。

要するに、政策評価とそのチェックシステムを取り巻く体制には、後付けで仕事が追加され、政策評価・業績測定、そしてこれらのメタ評価の理論的洗練もままならない中で、慣れない追加業務を遂行しなければならないのである。一般にわが国では評価に対してはそれほどのエネルギー費やせず、実際の費用も用意していない。本来業務で手一杯だからである。まして、メタ評価においてはさらにこの事態は深刻である。残された方法としては、選択と集中、省エネであろう。あるいは限られたスタッフやコストの中では、メタ評価機関が自ら出向くパトロール型のチェックシステムを展開するのではなく、行政相談や公益通報制度のように国民からの情報提供を待つ動き出す火災報知器型のチェックシステムの構築が有効で、効率的かも知れない。

最後にいくつか提言を試みる。

提言の第1は、総務省行政評価局とその政策評価・独立行政法人評価委員会（以下「政
独委」）、各府省の政策評価担当課室と各府省政策評価外部委員会あるいは外部有識者会議
などとの「役割分担」である。各府省には府省内部で政策評価をメタ評価したり評価を監
査したりする政策評価担当課室があり、その課室の「ご意見番」「指南役」的な役割を持つ
外部有識者会議が存在する。また総務省にも行政評価局と政策評価・独立行政法人評価委
員会とが存在する。実はこれらの関係があまりはっきりしないまま、評価の評価、評価の
監査、評価のプロセス・チェックが行われているのではないかという懸念がある。そこで
これら4機関それぞれの役割を確認し、相互の関係を整理するべきであろう。整理のガイ
ドは先に示した事前のメタ評価、事後のメタ評価、形成的メタ評価、総括的メタ評価の4
つであり、その個人的意見の案を示せば、以下のようになるであろう。

表7 メタ評価関連機関の整理表

機関	チェックシステム	具体的作業
各府省・政策評価担当課	事前メタ評価 形成的メタ評価	<ul style="list-style-type: none"> ・基本計画・実施計画、ガイドライン、評価マニュアルの作成 ・基本計画・実施計画、ガイドライン、評価マニュアルの順守状況のチェック ・評価業務の執行状況モニター。 ・評価サイクルのチェック ・評価結果の予算・政策への反映支援 ・会計課と連携してアメリカの PART のような仕組みを作る（可能であれば）
府省・政策評価外部委員会	形成的メタ評価 事後メタ評価、	<ul style="list-style-type: none"> ・基本計画、実施計画の事前チェック ・評価可能性（evaluability）の確認 ・評価結果の妥当性の確認 ・評価結果の予算反映の可否チェック
総務省行政評価局	形成的メタ評価 総括的メタ評価	<ul style="list-style-type: none"> ・各府省の基本計画、実施計画の確認 ・各府省の評価対象・評価手法の妥当性、評価結果の妥当性確認 ・各府省評価結果の取りまとめ状況と評価形式の事後チェック ・評価サイクルの確認 ・手続手順に関わる問題状況の指摘

総務省の政策評価・独立行政法人評価委員会	総括的メタ評価	<ul style="list-style-type: none"> ・総務省行政評価局を通じて、各府省の評価対象・評価手法の選定、評価結果の妥当性確認 ・各府省独立行政法人における事業評価との連携
----------------------	---------	---

(筆者作成)

第2の提言は総務省内における役割分担である。行政評価局は札幌市・仙台市・さいたま市・名古屋市・大阪市・広島市・福岡市に管区行政評価局を、また高松市には四国行政評価支局、さらに那覇市には沖縄行政評価事務所をもっており、これら管区行政評価局等と本省行政評価局との役割分担を考える必要があるのではないかと。じつは各府省においても地方出先機関が存在し、そこで政策評価とその準備作業を行っている場合が少なくないからである。総務省管区行政評価局と各府省の出先機関との連携は、政策評価の質を向上させるためには必須であるが、これまで表だった動きは「政策評価の統一研修」の他にはあまり見られていない。また、各管区行政評価局はたとえば「学者懇談会」に代表されるように近くの地方有名大学との連携があり、場合によってはその有名大学・大学院の卒業生を職員として採用していることが少なくない。こうした財産を、政策評価の形式に詳しい大学教員、政策評価の実質的方法、分析手法に詳しい大学教員との連携をはじめとして使うべきではないか。

さらに、各府省が地方自治体（都道府県や政令市）に政策評価のデータ出しを求め、これが各自治体の負担になっていることも少なくない。アメリカで‘inter-governmental evaluation’という議論になっているが、これについてわが国では議論される場は見られない。かつての行政監察的な視点でチェックする必要があるかも知れず、これはやはり、管区行政評価局の仕事であろう。

提言の第3はこの大学教員に関わることであり、アカデミズムとの連携を強化することである。その際、個別に大学教員との連携を図るだけでなく、日本評価学会・日本公共政策学会・日本行政学会に代表される学会との、「調査もの」を含めた連携が必要であろう。アカデミズムの側でも、かつてのような批判のための批判と言った姿勢はなくなっており、実学思考が高まって、大学院生をはじめとして実務に学ぶ姿勢が強い。公務員・企業人・NPO関係者の社会人大学院生も多い。この状況は評価の将来的にはかなり有望であろう。

最後に、総務省の政策評価・独立行政法人評価委員会内部の連携を提案したい。政策評価・独立行政法人評価委員会は政策評価と独立行政法人評価のふたつの分科会に分かれているが、合同の会議の他はあまり連携がない。しかし、たとえば独立行政法人内部で事業

費が 10 億円以上の国の補助事業もしくは、ODA 事業を実施している場合であれば、当然政策評価の対象になってくる。また、独立行政法人に大臣が提示している中期目標は政策評価の対象である。独立行政法人評価と政策評価がこのように一部重なり合っていることを考えると、両部会の連携が必要であろう。

いずれにしても、たとえば「公共事業の一般入札化の割合」とか「〇〇省改革」など、政策評価の対象ではないものをわざわざ評価したり、逆に、本来政策評価の対象になるものをはずしたりしている状況、あるいは評価に使う基礎データに異論が出る状況は、政策評価そのものの劣化だけでなく、政策評価に対する失望を招くのである。その点で言えば、評価の監査的チェックは政策評価の現状改革、最低限の品質保証になりまだまだ不可欠なのである。

米・英・NZのメタ評価に関する整理表

※ p～、pp～は、本文の記述に対応するページを示す。

国名	チェックの対象とする政策評価の評価主体・評価方式	政策評価のチェック機関	政策評価のチェックの実施の枠組み・手順	政策評価のチェックを行う前提となるデータの入手権限	政策評価のチェックの基準・視点	備考
米国	GAOが行うプログラム評価 (Program Evaluation)	会計検査院 (GAO)	GAOが公表する報告書は、「取り決めのための電子支援ガイド」(EAGLE)により作成手続きが電子的に管理されている。その中で①第一担当者による審査と承認、②第二担当者による審査と同意と2段階のチェックを行う。【p36】	調査に使ったすべての内部資料や作業メモが審査者に渡される。【p35】	<ul style="list-style-type: none"> 「GAOの主要価値」：説明責任、誠実、信頼【p36】 「専門的基準」：内部情報のため明らかでないが、主要部分は“GAO Yellow Book”に定める基準と共通と思われる。【pp36-37】 	GAOが行う年間約1,000件の評価のうち、90%は議会からの要請で実施している。【p33】
	政府業績成果法 (GPRA) により各省庁が行うプログラム評価	行政管理予算庁 (OMB)	予算査定ツール「施策の査定と格付けツール」(PART)による。各省庁はPARTを適用して評価を行い、評価書をOMBに提出。OMBの予算審議官がPARTを用いてそれらの審査及び格付けを行う。【pp40-41】	PARTのチェック項目の中にエビデンス/データの項目があり、それによりエビデンスの提出は義務とされている。【pp44-45】	<ul style="list-style-type: none"> PARTによれば以下のとおり。 質、範囲、独立性 (何者にも従属せず自らの意思で評価)、頻度 プログラムが有効であり目的を達成しているか。【pp44-45】 	大統領予算に添付して議会に提出され、予算査定に活用することを前提としている。【p41】
	教育省が行うプログラム評価	教育省教育科学研究所 (教育学の専門家を多数雇用)	教育省教育科学研究所が運営する“What Works Clearing House (WWC)”による。特定の施策に関する複数の報告書を選定し、当該施策の一般的効果を6段階で評価。その評価の過程を公開し教育関係者に情報提供。【pp48-49】	収集した評価報告書を基に評価。【P49】	<ul style="list-style-type: none"> WWCの「効果レーティング」によれば以下のとおり。 複数のエビデンスにより正の有意性が示されるか。 エビデンスがWWCエビデンス基準を満たしているか。 負の効果を示すエビデンスがあるかどうか。【p51】 	<ul style="list-style-type: none"> 本メタ評価は教育省独自の取組である。【p48】 WWCによる評価結果は、省内の次年度の予算配分に使用 NCLB法においては、エビデンスに基づく有効性が証明、確認される政策のみに連邦予算が配分されるように求められている。【pp47-48】
英国	各省庁と大蔵省との間で締結されるPSA (公共サービス協定) (業績測定)	大蔵省 (HM Treasury)	PSAの作成に当たり、①各省の担当者と大蔵省歳出担当各課で固めた草案を、②大蔵省一般歳出政策課でチェック。その後同省主計大臣から、公共サービス出関係会議(PSX)に報告され、重要な点について各省大臣と協議。【p70】	(PSA作成の主導権は大蔵省にある。【p63】)	<ul style="list-style-type: none"> PSAに関するガイドラインによれば以下のとおり。 組織目標と業績目標との妥当性、誤ったインセンティブの回避、帰属性 (外部要因の強さや行政責任の所在)、明確な定義、タイムリー性、信頼性、比較可能性、検証可能性【p71】 	予算案とともに議会に提出され、予算審議のための参考資料とされる。【p63】
		会計検査院 (NAO)	<ul style="list-style-type: none"> 優良事例の確定・普及【p72】 PSAに対する外部評価【pp72-73】 	検査に必要な文書を入手することができる。(会計検査院法第8条)	PSAのデータに対する質的保証が中心。【p73】	2001年の「シャーマン報告」においてはPSAの外部評価に際して統計局等との連携も示唆されている。【p72】
NZ	各省庁が作成する年次報告書のサービス業績表 (業績測定)	会計検査院 (Office of the Auditor-General)	年次監査の一環として、各省庁の年次報告書のサービス業績表 (アウトプット群とその指標・達成度) を、SOI (業務計画書) と併せてチェックし、監査意見を表明。【pp85-87】	会計検査担当部署としての権限に基づき、検査に必要な書類には提出の義務がある。【p87】 (公監査法第24条)	<ul style="list-style-type: none"> 「会計検査院監査基準」の中の「サービス業績報告書の監査基準」に以下のとおり示されている。 SOIの法令遵守性、目的・アウトプット群・指標・目標値の適正性、その設定プロセスの適切性【pp87-89】 	
		財務省 (Treasury)	予算書及び予算に関するSOIを予算作成の一環としてチェックする中で、各省庁からの要望に応じて年次報告書のチェックを行う。【p90】	予算調整担当省庁としての権限に基づき、予算調製のデータとして入手する。【p90】	(担当する財務省国家部門業績局は、省庁の効率性、有効性の改善に関する提言を行う部署である。【p90】)	各省庁の指標の達成度を用いて予算査定をすることはない。【p91】
		政府サービス委員会 (State Services Commission)	同委員会の行政システム改善局が、省庁の行う業績測定について必要に応じて相談にのりアドバイスを行う。実施手順が定式化されたものではない。【p92】	(相談にのる中で、アドバイスを行うために必要な情報を入手する。)	「各省庁の業績を向上させる」という視点から、委員会の公表する数々のガイドラインに依拠し、各省庁のアウトプット・アウトカムの設定および指標の適切性に対しアドバイスを行う。【p92】	
	各省庁が重要な分野に絞って任意に行う評価 (Evaluation) (通常、事後評価)	(社会開発省の場合) ニューゼaland経済調査研究所 (NZ最大級の民間非営利の研究所)	<ul style="list-style-type: none"> 各省庁では評価担当部署を設置して評価を行うことが一般的。行った評価のチェックの方法は各省庁によってさまざまである。【pp94-95】 社会開発省では、省内の社会調査評価センターが同省の実施した政策について評価を実施。センターの行う評価は、ニューゼaland経済調査研究所による評価を年1回受けている。【pp95-99】 	評価の委託契約に基づいて入手する。	評価の手法・プロセスが適切かどうか。例えば定量的分析を行っている場合には、定量的分析を採用することが適切なケースであるか、適切な場合には、当該定量的分析が適切に行われているかどうか。【p99】	労働庁や教育省でも外部の専門家による評価は行われている。【p99】

参考文献一覧

(第1章)

- 石橋順三 (2007) 「政策評価制度～制度運用の課題と展望～」『立法と調査』No.269
- 鎌田英幸 (2008) 「政策評価の本来的機能の発揮に向けて」『季刊行政管理研究』No.124
- 国際協力機構 (2008) 『2007年度事業評価年次報告書』2008年2月
- 総務省行政評価局 (2008) 「政策評価の点検結果—評価の実効性の向上に向けて—」2008年3月
- 南島和久 (2007) 「府省における政策評価の中立性および客観性：グリッド／グループ文化理論に基づく考察」『法学志林』104巻4号
- 山谷清志 (2006) 「政策評価の実践とその課題～アカウンタビリティのディレンマ」萌書房
- ALNAP, (2005) *Assessing the Quality of Humanitarian Evaluations, The ALNAP Quality Proforma 2005 (v. 02/03/05)* (<http://www.alnap.org/pdfs/QualityProforma05.pdf>)
- Cooksy, J.L. & Garacelli, V.J., (2005) “Quality, Context, and Use: Issues in Achieving the Goals of Metaevaluation”, *The American Journal of Evaluation*, No.1, Vol. 26
- Office of Management and Budget, (2008) *Program Assessment Rating Tool Guidance No. 2008-01, January 29, 2008* (http://georgewbush-whitehouse.archives.gov/omb/part/fy2008/part_guid_2008.pdf)
- Patton M.Q., (1997) *Utilization-Focused Evaluation*, The New Century Text, 3rd edition, Sage Publications
- Robson, C., (2000) *Small-Scale Evaluation*, Sage Publications
- Rossi, Peter H., Reeman, Howard E., and Lipsey, Mark W., (1999) *Evaluation: a systematic approach*, 6th ed., Sage Publications
- Scriven, M., (1991) *The Evaluation Thesaurus*, 4th ed., Sage Publications
- _____ (2009) “Meta-Evaluation Revisited”, *Journal of Multi Disciplinary Evaluation*, No.11, Vol.6
- Stufflebeam, D., (1999) *Program Evaluation Models Metaevaluation Checklist*, Based on the Program Evaluation Standards (http://www.wmich.edu/evalctr/checklists/eval_model_metaeval.pdf)

_____ (2001) "Meta-evaluation Imperative", *The American Journal of Evaluation*,
No.2, Vol. 22

Stufflebeam & Shinkfield, (2007) *Evaluation Theory, Models, and Applications*, Jossey Bass

Weiss, C.H., (1998) *Evaluation: Methods for Studying Programs and Policies*, 2nd ed.,
Prentice-Hall,

(第2章)

伊藤勲 (2006) 「ワシントンD.C.調査記録 (1) ~ (6)」『行政評価月報』

金本良嗣 (1990) 「会計検査院によるプログラム評価—アメリカ GAO から何を学ぶか—」『会
計検査研究』第 10 号第 2 号 (<http://www.jbaudit.go.jp/effort/study/mag/2-2.html>)

黒田忠司 (2003) 「米国会計検査院における評価活動プロセス」『季刊行政管理研究』(2003.6,
No.102)

左近靖博 (2008) 「米国連邦政府ブッシュ政権における「予算と業績の統合 (BPI)」の結果考
察—予算編成プロセスにおける業績情報活用の成果総括—」『季刊政策・経営研究』(三
菱UFJリサーチ&コンサルティング) 2008.Vol.4

田辺智子 (2006) 「エビデンスに基づく教育—アメリカの教育改革と What Works Clearinghouse
の動向—」『日本評価研究』(日本評価学会) 第6巻第1号、

廣瀬淳子 (2006) 「アメリカにおける行政評価と行政監視の現状と課題GAOとCIAを巡る最近の
状況から」『レファレンス』(2006.5)

龍慶昭・佐々木亮 (2004) 『政策評価の理論と技法』(増補改訂版) 多賀出版

Bustelo, M. (2002). *Metaevaluation as a Tool for the Improvement and Development of the
Evaluation Function in Public Administrations*. Retrieved on March 9, 2009.

http://www.evaluationcanada.ca/distribution/20021010_bustelo_maria.pdf

Cooksy, L. & Caracelli, V. J. (2009). *Metaevaluation in Practice: Selection and Application
of Criteria*. Journal of MultiDisciplinary Evaluation, Volume 6, Number 11.

<http://www.jmde.com/>

Davidson, J. (2005). *Evaluation methods basics*. Thousand Oaks, CA: Sage.

Donaldson, S. I., & Christie C. A. (2005) . *The 2004 Claremont Debate: Lipsey vs. Scriven*.
Journal of MultiDisciplinary Evaluation. <http://www.jmde.com/>

Institute of Education Sciences, Department of Education.(2008). *What Works*

- Clearinghouse Procedures and Standards Handbook (Version 2.0)*. MOE.
- Mathison, S. (2008). *Encyclopedia of Evaluation*. Sage Publications
- National Association of Schools of Public Affairs and Administration (NASPAA). (2008). *Commission on Peer Review and Accreditation*. NASPAA.
- Patrick, R.M. (2006). Congressional Reporting: A Management Process to Build a Legislative-Centered Public Administration. Retrieved on March, 9, 2009. <http://scholar.lib.vt.edu/theses/available/etd-04202006-104259/unrestricted/Chapter3MANAGEMENT4-19.pdf>.
- Rossi, P. H., Lipsey, M., & Freeman, H. E. (2004). *Evaluation: A systematic approach (7th ed.)*. Sage Publications.
- Sanders, J. (Ed.). (1994). *The program evaluation standards: How to assess evaluations of educational programs*. Thousand Oaks, CA: Sage.
- Sanders, J.R. & Joint Committee on Standards for Educational Evaluation (2008). *The Program Evaluation Standards*. Sage Publications.
- Smith, R (2005). *Peer Review Report: International Peer Review of the Performance Audit Practice of the United States Government Accounting Office*. (Submitted to The Comptroller General of the United States).
- Sherman, J. (2009). *Survey of the United States Government's arrangements for monitoring and evaluating support to security reform*. Retrieved on March 9, 2009. www.cic.nyu.edu/Lead%20Page%20PDF/USDonorSurvey.pdf
- Scriven, M. (1969). *An introduction to meta-evaluation*. Educational Product Report, 2.
- Scriven, M. (1991). *The Evaluation Thesaurus* (4th ed.). Sage Publications.
- Scriven, M. (2005). *The logic and methodology of checklists*. Kalamazoo: The Evaluation Center, Western Michigan University.
- Scriven, M. (2006). *Key evaluation checklist*. Kalamazoo: The Evaluation Center, Western Michigan University.
- Scriven, M. (2007). *Key evaluation checklist*. Kalamazoo: The Evaluation Center, Western Michigan University.
- Scriven, M. (2009). *Meta-Evaluation Revisited*. Journal of MultiDisciplinary Evaluation, Volume 6, Number 11. <http://www.jmde.com/>

- Stufflebeam, D. (1974). *Meta-evaluation*. The evaluation center, Occasional paper series #3.
- Stufflebeam, D. (1999). *Program Evaluation Metaevaluation Checklist*. Retrieved on March 9, 2009. http://www.wmich.edu/evalctr/checklists/program_metaeval.pdf
- Stufflebeam, D. (2000). *The Methodology of Metaevaluation as Reflected in Metaevaluations by the Western Michigan University Evaluation Center*. *Journal of Personnel evaluation in education* 14:1.
- Stufflebeam, D. (2002). *CIPP Evaluation Model Checklist*. Retrieved on March 9, 2009. <http://www.wmich.edu/evalctr/checklists/cippchecklist.htm>
- U.S. Government Auditing Office (GAO) (2006). *Generally Acceptable Government Accountability Standards (GAGAS)*. GAO.
- U.S. Government Auditing Office (GAO) (1992). *Evaluation Synthesis*. Program Evaluation and Methodology Division, GAO.
- U.S. Government Auditing Office (GAO) (2004). *Results-oriented Government: GPRA Has Established a Solid Foundation for Achieving Greater Results*. GAO.
- U.S. Government Accountability Office (GAO) (2006). *Government Auditing Standards 2006 Revision*. GAO. <http://www.gao.gov/govaud/ybk01.htm>
- U.S. Office of Management and Budgeting (OMB) (2008). *Guide to the Program Assessment Rating tool (PART)*. Office of Management and Budget.
- U.S. Office of Management and Budgeting (OMB) (2004). *What Constitutes Strong Evidence of a Program's Effectiveness?*. Retrieved on March 9, 2009 http://www.whitehouse.gov/omb/part/2004_program_eval.pdf
- U.S. Government Auditing Office (GAO). (2005). *Program Evaluation: OMB's PART Reviews Increased Agencies' Attention to Improving Evidence of Program Results*. GAO.

(第3章)

- 稲継裕昭 (2001) 「英国ブレア政権下での新たな政策評価制度：包括的歳出レビュー (CSR) ・公共サービス合意 (PSAs)」 (『季刊行政管理研究』 (93))
- 君村昌編著 (2001) 『行政改革の影響分析』 ((財) 行政管理研究センター)
- 国土交通省国土交通政策研究所 (2002) 『New Public Management：歴史的展開と基礎理論』

- 児玉博昭 (2001) 「英国における政策評価システム：公共サービス合意及びサービス提供合意を中心として」(『白鷗法学』(18))
- 財政制度等審議会 (2007) 『財政制度分科会 海外調査報告書』
- 総務省行政評価局政策評価官室 (2001a) 『諸外国における政策評価の制度の概要 (政策評価関係資料 I)』
- (2001b) 『諸外国の政策評価の実際例 (政策評価関係資料 II)』
- (2002) 『諸外国における政策評価の制度と運営 (第4巻)』
- (2003) 『諸外国における政策評価の制度と運営 (第6巻)』
- 谷藤悦史 (2001) 「英国における公共サービス管理と政策評価制度の展開と特性」(総務省行政評価局政策評価官室 (2001a) 所収)
- 南島和久 (2007) 「府省における政策評価の中立性および客観性」(『法学志林』104 (4))
- (2009) 「信頼確保をめぐるアカウンタビリティの課題：政策評価と内部統制をめぐる」(行政管理研究センター『行政の信頼性確保、向上方策に関する調査研究報告書』)
- 農林水産政策情報センター (2007) 『英国における包括的歳出見直し等の状況及び影響評価の制定等政策評価の現状』((財)農林水産奨励会)
- 東 信男 (2006) 「イギリスにおける公会計制度改革の検証：効果と課題」(『会計検査研究』(34))
- (2009) 「イギリス中央政府における国際会計基準 (IAS/IFRS) の導入：公会計の目的に対応させながら」(『会計検査研究』(39))
- 平松英哉 (2001) 「議院内閣制下における会計検査院による『政策評価』：英国会計検査院の動向に着目して」(『同志社法学』53 (2))
- (2002) 「イギリス中央政府におけるアカウンタビリティと会計検査院：チャーマン報告の概要とその意義を中心として」(『同志社法学』54 (4))
- (2004) 「イギリス会計検査院による評価事例の研究：有効な外部評価の条件を求めて」(『日本評価研究』4 (2))
- 山谷清志 (1990) 「行政管理におけるサッチャーの『革命』：『エージェンシー』と業績評価」(『國學院大學紀要』(28))
- 和田明子 (2007) 『ニュージーランドの公的部門改革：New Public Management の検証』(第一法規)
- HM Government (2007) Service Transformation Agreement
- HM Treasury (1998) *Public Services for the Future: Modernisation, Reform, Accountability:*

- Comprehensive Spending Review: Public Service Agreements 1999-2002*, Cm4181
- (2001a) *Managing Resources: Full Implementation of Resource Accounting and Budgeting*
- (2001b) *Holding to Account: The Review of Audit and Accountability for Central Government*
- (2002a) *Audit and Accountability in Central Government: The Government's response to lord Sharman's report "Holding to Account"*, Cm5456
- (2002b) *Opportunity and Security for all: Investing in an enterprising, fairer Britain: New Public Spending Plans 2003-2006*, Cm5570
- (2003) *The Green Book: Appraisal and Evaluation in Central Government* (総務省行政評価局政策評価官室 (2003) に訳文所収)
- (2004) *Stability, security and opportunity for all: Investing for Britain's long-term future: 2004 Spending Review: New Public Spending Plans 2005-2008*, Cm6237
- (2006a) *Long-term challenges and opportunities for the UK: Analysis for the 2007 Comprehensive Spending Review*
- (2006b) *Releasing the resources to meet the challenges ahead: Value for money in the 2007 Comprehensive Spending Review*
- (2007a) *Audit Committee Handbook*
- (2007b) *Meeting the aspirations of the British People*, Cm7227
- , Cabinet Office, National Audit Office, Audit Commission, Office for National Statistics, *Choosing the Right: A Framework for Performance Information*. (総務省行政評価局政策評価官室 (2002) に訳文所収)
- Hood, C. (1991) "A Public Management for all Seasons?," *Public Administration*.(69),spring
- *et al.* (1999) *Regulation Inside Government: Waste-Watchers, Quality Police, and Sleaze-Busters*, Oxford
- National Audit Office (2005) *Public Service Agreement: Managing Data Quality- Compendium Report*
- (2006a) *Second Validation Compendium Report: 2003-06 PSA data system*

————— (2006b) *Third Validation Compendium Report : Volume 1&2*

————— (2007) *Third Validation Compendium Report : Volume 1&2*

————— (2008) *Annual Report 2008*

(第4章)

和田明子 (2007) 『ニュージーランドの公的部門改革—New Public Management の検証—』 第一法規。

Boston, J., *et al.*, (1996) *Public Management: The New Zealand Model*, Oxford University Press

Gill, D. (2008) “Managing for Performance in New Zealand: The Search for the “Holy Grail”?” in *Holy Grail or Achievable Quest?: International Perspectives on Public Sector Performance Management*, KPMG International.

Gregory, R. (1994) “Consolidating the Reforms: Can the Production Model be Counter-Productive?”, *Public Sector* 17(2)

Hood, C., (1990) “De-Sir Humphreyfying the Westminster Model of Bureaucracy: A New Style of Governance?”, *Governance*, Vol.3

Office of the Auditor-General, (2002) *Report of the Controller and Auditor-General: Reporting Public Sector Performance 2nd edition*

Office of the Auditor-General, (2008) *Annual Report 2007/08*

Office of the Auditor-General, (2008) *The Auditor-General's Observations on the Quality of Performance Reporting*

(<http://www.oag.govt.nz/2008/performance-reporting/>)

State Services Commission & Treasury, (2003) *Doing the Right Things and Doing Them Right: Improving Evaluative Activity in the New Zealand State Sector*

State Services Commission & Treasury, (2008) *Performance Measurement: Advice and Examples on How to Develop Effective Frameworks*

Statistics New Zealand, (2008) *New Zealand Official Yearbook 2008*

Steering Group Management for Outcomes, (2002) *Managing for Outcomes: Output Plans Guidance for Departments*

Treasury, *Budget Process Guide 2009*

<http://www.treasury.govt.nz/publications/guidance/budget/process/>
Treasury, (2005) *Cost Benefit Analysis Primer*

<http://www.treasury.govt.nz/publications/guidance/costbenefitanalysis/primer>
Treasury, (2008) *Guidelines on the Regulatory Impact Analysis Requirements*

<http://www.treasury.govt.nz/publications/guidance/regulatory/impactanalysis>
Treasury, (2008) *Preparing the Annual Report: Guidance and Requirements for Departments*

Treasury & State Services Commission, (2009) *Statement of Intent: Process and Production Guidance for Departments*

Treasury & State Services Commission, (2008) *Statement of Intent: Technical Guide for Departments*

Treasury & State Services Commission, *Performance Information Measures and Standards in the SOI and Annual Report*

Turner, D. (2003) “New Zealand’s Core Public Sector Response to Evaluation: The Case of the Department of Labour”. in Lunt, N., Davidson, C. and McKegg, K. (eds) *Evaluating Policy and Practice: A New Zealand Reader*, Auckland, Pearson Education.

(ニュージーランド省庁等ホームページ)

New Zealand Institute of Economic Research (ニュージーランド経済調査研究所)
<http://www.nzier.org.nz>

Office of the Auditor-General (会計検査院) <http://www.oag.govt.nz>

State Services Commission (政府サービス委員会) <http://www.ssc.govt.nz>

Treasury (財務省) <http://www.treasury.govt.nz>

The Auditor-General’s Auditing Standards (会計検査院監査基準)
<http://www.oag.govt.nz/2008/auditing-standards/index.htm>

Development Goals for the State Services
<http://www.ssc.govt.nz/display/document.asp?DocID=6545>

Performance and Accountability
<http://www.ssc.govt.nz/display/document.asp?navid=339>

(第5章)

伊多波良雄・編著 (2009) 『公共政策のための政策評価手法』 中央経済社。

山谷清志 (1997) 『政策評価の理論とその展開－政府のアカウンタビリティー』 晃洋書房。

山谷清志 (2006) 『政策評価の実践とその課題－アカウンタビリティーのジレンマー』 萌書房。

Carl E. Hanssen, Frances Lawrenz, Diane O. Dunet, (2008) “Concurrent Meta-Evaluation:

A Critique,” *American Journal of Evaluation*, Vol. 29, No. 4.

Sandra Mathison ed., (2005) *Encyclopedia of Evaluation*, Sage.

Daniel L. Stufflebeam, (2001) “The Metaevaluation Imperative,” *American Journal of*

Evaluation, Vol. 22, No. 2.