

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。

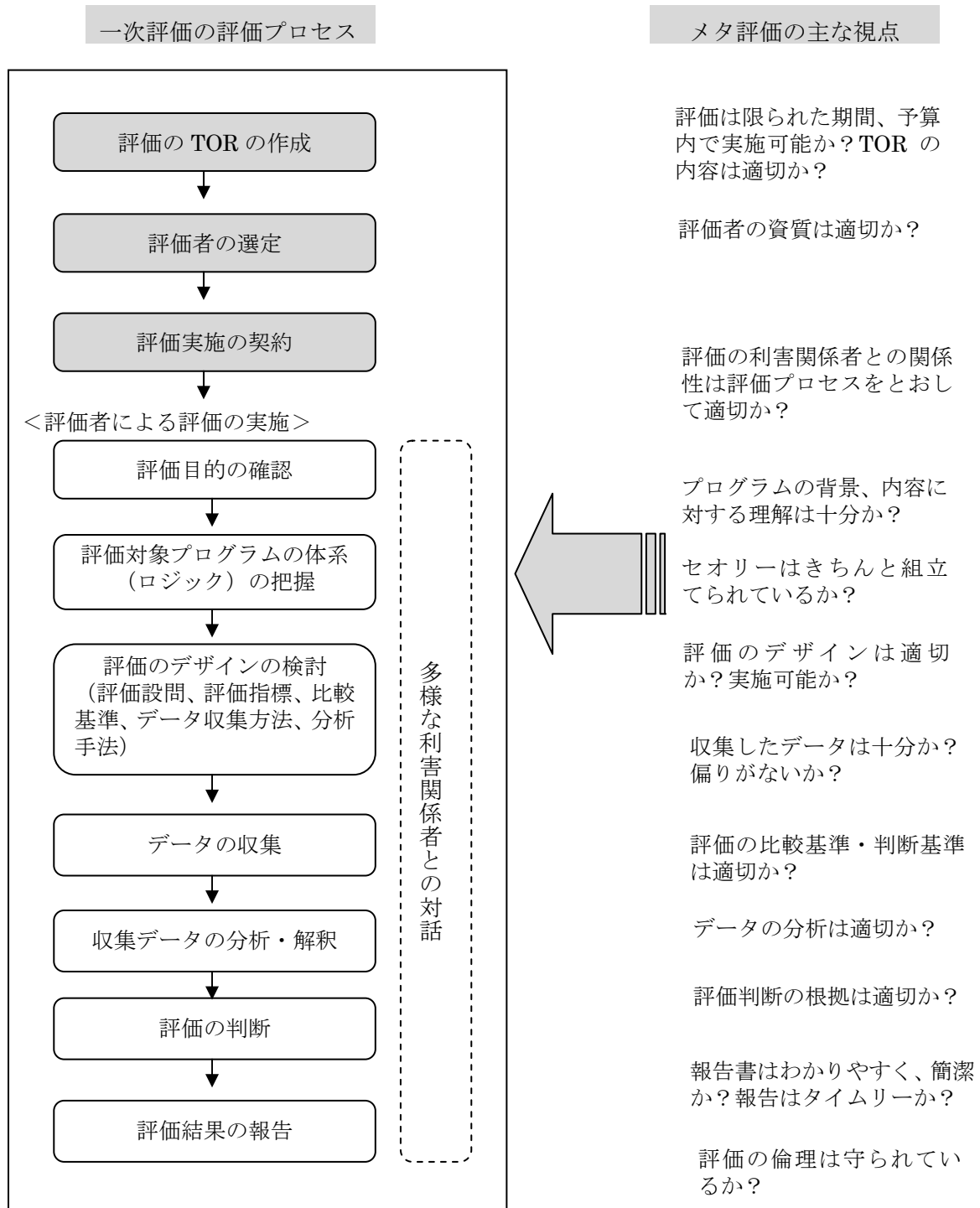
(2) プログラム評価の実施過程とメタ評価の視点

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、あるいは的を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、効率的にかつ遅れが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 of 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、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 「形成的メタ評価」と「総括的メタ評価」

	形成的メタ評価	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。省エネをとおり、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、チャータースクール学生、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)

