

政策評価のメタ評価システム

－客観性と評価の質－

山谷 清志

概要

政策評価は行政改革会議(1996年～1997年)、いわゆる橋本行革が、日本の行政を「21世紀型行政システム」へと転換する手段として導入した(行政改革会議『最終報告』1997年12月3日、p.6)。21世紀型行政システムとは、国民に対する説明責任を充実させ、事前の手續チェックから事後評価へ行政運営の力点を移し、政策の企画立案と実施は分離、その政策実施手段を持つ独立行政法人制度を創設、そして縦割り行政の視野狭窄を超越するために内閣官房の政策調整機能を強化したシステムであり、システムの実現には政策評価が重要な役割を担うと考えられていた。その後政策評価は、1997年夏頃からの試行、「行政機関が行う政策の評価に関する法律」(2001年、以下「政策評価法」)による制度導入を経て、2003年頃から実用段階にある。

しかし政策評価にはいくつか問題が浮上した。評価方法に対する理解不足、評価(evaluation)と測定(measurement)との混同、評価担当者の異動による経年劣化、評価に使用するデータの不足や改変問題、政策形成・政策実施に反映されない評価結果、評価者の倫理問題などである。

その中で一番深刻なのは客観性問題であった。内部評価であるが故の「お手盛り評価」批判への対応は、政策評価制度導入以前の検討段階から議論になり、いまだに解決されていない困難な問題である。しかし、この批判に適切に応えられない限り、政策評価導入を決定した橋本行革の理想は矮小化される。もちろんこうした事態を実務担当者が看過してきたわけではない。たとえば、総務省行政評価局は『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』(2009年)においてチェックシステムの可能性を模索していたが、アカデミズムにおいて、客観性問題をめぐる議論は低調であった。

本稿では政策評価に当初から付随した客観性確保問題を、メタ評価によって解決すべきではないかと考える。しかし、客観性問題とメタ評価のそれぞれにもいろいろな議論が存在し、あるいは混乱が見られる。これまでの議論を含めて、再度その整理を試みるのが本稿の目的である。

1. 政策評価の客観性とそのチェック

1.1 アカウンタビリティと客観性

政策評価システムを欠いていた日本の公共部門は長年、政策に関するアカウンタビリティ欠如の状態にあり、しかもその問題点に気づき議論することも少なかった。国民に対してどのように説明し、いかに納得してもらうか、その視点がない政策運営の慣行が続いてきたのである。政策評価法が問題視したのはこの点にあり、したがって法律のはじめに、政策評価を運営する基本的方針を具体的に定めている(下線は筆者)。

第1条 この法律は、行政機関が行う政策の評価に関する基本的事項等を定めることにより、政策の評価の客観的かつ厳格な実施を推進しその結果の政策への適切な反映を図るとともに、政策の評価に関する情報を公表し、もって効果的かつ効率的な行政の推進に資するとともに、政府の有する

その諸活動について国民に説明する責務が全うされるようにすることを目的とする。

第3条 行政機関は、その所掌に係る政策について、適時に、その政策効果（当該政策に基づき実施し、又は実施しようとしている行政上の一連の行為が国民生活及び社会経済に及ぼし、又は及ぼすことが見込まれる影響をいう。以下同じ。）を把握し、これを基礎として、必要性、効率性又は有効性の観点その他当該政策の特性に応じて必要な観点から、自ら評価するとともに、その評価の結果を当該政策に適切に反映させなければならない。

2 前項の規定に基づく評価（以下「政策評価」という。）は、その客観的かつ厳格な実施の確保を図るため、次に掲げるところにより、行われなければならない。

- 一 政策効果は、政策の特性に応じた合理的な手法を用い、できる限り定量的に把握すること。
- 二 政策の特性に応じて学識経験を有する者の知見の活用を図ること。

しかし、そもそもこの法律の文言には四つの無理がある。第1に、政策評価の客観的で厳格な実施を求めながら、評価担当者は政策を所管する行政機関自身である。ここに根本的な無理がある。第2に、行政機関自らが実施するにもかかわらず、国民に説明する責任（アカウンタビリティ）を全うする、と求めている。しかし、自ら全うする責任はアカウンタビリティとは言わない。レスポンシビリティである（山谷2012, p. 5）。

第3の問題は、この説明の手段として合理的な手法を求めているが、それは定量的な手法であると、政策評価法第3条2項を評価担当者は読む。自民政権時代の2003年経済財政諮問会議「骨太方針」、そして民主党政権2009年の菅副総理の提言は、ともに数字をもとに政策を目標管理する方式を提唱し（山谷2012, pp. 144-145）、定量分析は事業の「数字による管理」に拡大した。しかし、政策担当者が付けた目標値、指標の数字だけでは政策課題の真の動きを追うことはできない。本来、統計を使った定量分析に加え、定性分析も同時に行って現場の事実に向ける評価が「良質の評価」であるが、時間

と能力に限界がある政策所管課の多くは、政策課題の現状認識を数字の管理に代えてしまう。結果として、政策現場で何がどうなったのかは見過ごされ、あるいは放置されかねない。

最後の第4の無理は、客観的かつ厳格な実施の確保を求めるために政策の特性に応じて学識経験を有する者の知見の活用を図るのであるが、この学識経験者の意味が混乱している。ときには政策内容にも、評価にも関わりのない人が任用される。素人の任用であるが、素人に事前説明するのは政策所管課なので情報の非対称性問題が生じ、言葉は悪いが「洗脳」と同じになることも少なくない。その結果として政策評価の外部有識者会議は役所の政策主張の追認機関になる。

総務省で長年政策評価の実務、とくに府省の政策評価の客観性をチェックするセクション（プロジェクト・チームという）で仕事に携わった人びとが、頻繁に口にする警句が「二次評価は、二度評価することではない」である。つまり、二次評価は一次評価とは別の視点でチェックするべきであると言っている。この二次評価で考慮する重要なポイントは6点ある。

- ①チェック機関の組織・体制
- ②チェックの実手順・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事例指摘
- ⑥チェックをめぐる理論

これらに着目したときに客観性の議論を始めることができるのである。ただし、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。

①に関しては広義で言えば比較政策学、狭義では比較政治制度論・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（その代表は連邦議会補佐機関であるアメリカ会計検査院 GAO: General Accounting Office, 2004年7月から Government Accountability Office）の統治構造上の配置、その統治構造を導いた理念・理論に関する理解があれば、チェック体制そのものの有効性が判明する。日本では府省の官房に一次チェック機関として政策評価課が置かれ、府省評価外部有識者はこの課に意見を言う。総務省行政評価局は二次チェック機関として存在し、政策評価・独立行政法人評価委員会がそのご意見番になる。日本では、このようにチェッ

ク機能を重層化することによってチェックの有効性を担保させているが、後述するように外部有識者の資質に問題があり、実際はそうっていない。

他方、②・③・④・⑤はマイクロ・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」を導く現場の経験と、この経験をもとに発展させた評価戦略、評価ポリシーが必要になってくる。この評価戦略、評価ポリシーをオープンにすると評価の透明性が高まり、国民の信頼は増す。②～⑤を間違えると組織活動のアウトプット測定・活動報告に終始する似非評価と、長期の経済政策や対外政策のマクロ動向のレビューとが混在することになる。

ところで「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だろうか。言い換えると‘policy evaluation’と呼ばれる活動をチェックすることによって、何をしようとしているのか。これを明確にするのが上記⑥の議論であるが、そこにはどのような理論があるのか、あらかじめ認識しておく必要がある。

1.2 チェックシステムとその課題

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、まず法令や標準、基準、ガイドラインを設定し、それらにしたがって活動が行われているかどうか確認する点検作業であろう。それは監査に似たコンプライアンス・チェックの作業になる。

日本の政策評価システムは政策評価法と、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を3つ置いている。政府全体としての統一性を確保する見地から総務省行政評価局が行う統一性確保評価、複数の府省にまたがっている政策課題を総合的な視点から行政評価局が実施する総合性確保評価、そして客観的かつ厳格な実施を府省に求めるために行政評価局がチェックする客観性担保評価の3つである。

統一性確保評価も総合性確保評価も政府全体、府省横断の視点から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がもつ専門的知識、あるいは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、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担保評価については、悩ましい事情がある。それが外部者の外部評価を意味するのか、それとも第三者評価なのかははっきりしないからである。

たとえば、日本の政策評価制度における外部委員会は、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見る限り、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、府省が知りうる範囲での専門家を府省自ら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を併任することもあり、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、日本の政策評価制度は無関係の第三者による「独立性」を想定していないことになる。外部有識者、外部評価委員に求めるのは評価の指南役、政策のご意見番の役割であることが、府省が公表する政策評価会議の議事録から伺うことができる。

そもそも、評価理論において外部評価は、評価の任務にたえたと信用(accreditation)され、鑑識眼(connoisseurship)があると認められた人が行うとき機能すると考えられている(*Encyclopedia of Evaluation*, p.150)。逆に、信用できない人や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、府省が鑑識眼を持つと信用した専門家を連れてくるのは最低限の条件である。「外部委員を選任しても、人選の原案を作るのが当該行政機関であれば、そして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」と言うシニカルな批判は、この意味での外れなのかも知れない。また、公募方式を採用した場合、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応募して来て会議が混乱を極めるリスクがある。

ただ、「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チェックしてもらおうと考えるのは、間違いではない。ただし、熱心な

表1 政策評価のチェックシステムの類型

	客観性を確保する制度と方法	客観性を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。結果重視。手続チェックが中心。	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、日本評価学会の上級評価士を活用。
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、入札条件・TORの事前チェック・事後検査が必要。手続でチェック。	D：研修。職員を大学院に派遣して学位取得。資格（評価士）取得。有識者の叡智を借用。他の官庁の相互交流、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。

(筆者作成)

素人が過剰に意識して、重箱の隅を突つつくような的外れで細過ぎる注文を付け、会議が混乱、遅延することがある。そこで、あらかじめ設定した項目について事後チェックを行う方法が選択されるが、設定した項目の妥当性、適切性が疑われると政策評価外部有識者会議は府省が自己正当化するツールに堕落していると非難されても、反論しにくい状況にある。

そこで、チェックシステムを公式の制度にして、非難に対応する方法が考えられる。たとえばアメリカ連邦政府の会計検査院（GAO）や、イギリスの会計検査院（NAO:National Audit Office）の例が考えられる。いずれも議会との関係が密接であり、議会のイニシアチブを受けて行動する場面も多いので、議会制民主主義の枠内で説明でき、正統性がある。日本でもこの方向での改革案があったが（1996年日本版GAO創設の民主党提案）、当時の総務庁行政監察局の参議院移動や会計検査院法の改正を前提としており、実現性に乏しく頓挫した。制度改革には著しくエネルギーが必要であり、むしろ現行制度を前提とするチェックシステムを、表1を参考に考える方が現実的であろう。

2. チェックシステムの現実的アプローチ

2.1 チェック担当組織

日本の政策評価においてチェックシステムを考える際に考慮すべき点が3点ある。その第1はチェックを担当する制度・組織、そしてスタッフである。日本では前述の「第三者評価＝独立機関」制度の構築や、行政評価局・会計検査院

をGAOのような議会付属組織に制度変更するのは今のところ非現実的である。そこで「外部評価」を専門性で代用する方途を考えるべきなのである（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。この場合の評価専門家とは政策評価の著書を持ち政策評価を講義している大学教員（スペシャリスト）、府省の政策評価の担当部局に任期付きで採用され実務経験が多いコンサルタントのスタッフ（エキスパート）、日本政府、諸外国や国際機関の政策評価、ODA評価を担当する評価専門家（プロフェッショナル）である。日本評価学会が認定する上級評価士とはこれらの二つ以上の役割を担っている人たちである。

シンクタンクや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、日本ではそこまで考えない行政実務者が多い。理由は、行政側にコンサルタントを「下請け業者」として低く見る気風が残っている、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、良質なコンサルタントを雇うには費用がかかるので「安からう・悪からう」会社に発注する、また社会全体で合意された資格確認手続がないので評価能力を確認する手続がうまく機能しないからである。理想論としては、評価業務に詳しい府省スタッフが、質的に高レベルのTOR（terms of reference：委託のための条件を定めた文書で、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書）を作成・公表することによって内部で客観性を確保する方法がある（表1のC）。もっとも、それはゼネラリスト育成型の行政組織では難しく、府省が委託時のTORをきちんと提示しなかったので、

表2 評価手法ごとのチェック・ポイント例

評価方式	評価の方法	チェックすべきポイント
総合評価	プログラム評価	プログラムは政策目標と政策手段について、ロジカルな思考に基づき、適正に構築されているか確認。
	セオリー評価	評価手段を運用するプログラム、目標と手段を組み合わせたデザインの理屈（セオリー）確認が正しくできているかどうかの確認。
	プロセス評価	評価プロセスは、政策プロセスと並んで、正しいセオリーに基づいて実施されているかをチェック、モニターする。
実績評価	業績測定	設定された目標数値あるいは指標の適切さをチェック。無理に定量評価にしていないかも確認。
	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。
	アウトカム評価	評価対象はアウトカムなのかアウトプットなのか。
事業評価	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、評価可能性（evaluability assessment）、形成評価（formative evaluation）、総括評価（summative evaluation）で終わる評価サイクルの適切な実施状況のチェック。

(筆者作成)

後にコンサルタントが提出した評価報告書が「ゴミ」になった例は少なくない。

また、表1のB欄の場合、難しい課題がある。高度な評価技術を持ち定性分析や数量的政策分析を駆使できるコンサルタントを活用した時、分析内容がかえってブラックボックス化してしまうという指摘もある（宗高2013）。たとえば、道路建設事業の経済効果を事前評価する場合、コンサルタントは数万人の人口を前提条件に考え、評価方法の政策分析モデルを現実に合わせてカスタマイズ（悪い言葉で言えば操作）することが多いが、このカスタマイズ・プロセスが表に出ないと、府省評価担当者は分析結果の正誤が判断できず、事後評価に対するチェックは機能しなくなる。結果として、コンサルタントの分析は正しいのか間違っているのか分からないだけでなく、コンサルタントがデータを捏造するリスクを排除できない。

なお、表1のA欄において無能な素人を第三者として連れてきた場合、最悪のチェックシステムができあがる。

ところで政策評価において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、必ずしも専門家（エキスパート、スペシャリスト、プロフェッショナル）とは言えないからである。専門家と言われるようになるのは、政策内容関連分野あるいは評価関係分野で学位・資格を取得する、評価の実務経験を

複数年積む、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、事業評価（プロジェクト分析）・実績評価（パフォーマンス測定）・総合評価（プログラム評価とレビュー）などの方式からいずれを選択するか、それらにどのようなデータが必要か、評価を行う時にどのようなポイントを押さえていなければならないか、評価のタイミングをはかり評価結果を上司や国民に提示する場合にどんな見せ方がよいかなどを助言できる人であろう。なお評価方式と評価方法は複数存在するので、そのそれぞれをチェックするときは当然視点が違ってくるので、注意が必要になる（表2を参照）。

ちなみに、評価の専門家（エキスパート）のイメージに近い存在は、日本評価学会が認定する上級評価士である。そしてこの学会が上級評価士と認定する場合には、以下の5項目の審査項目基準をクリアしていることが条件になる。

①研究実績

- ・学会誌への投稿実績（評価関連のテーマに限る）
- ・評価関連著書の出版実績

②政策評価委員会等への外部有識者としての参加実績あるいはそれに相当する実績

- ・府省、地方自治体、公益・非営利組織、市民組織などにおける評価委員会への外部有識者としての参加実績
- ・府省、地方自治体、公益・非営利組織、市民組織などにおける評価委員会への組織代表者としての参加実績（評価委員会を主催する組織の代表者としての参加を含む）
- ③評価実施の実績
 - ・自身が総括等を務めた外部評価の評価報告書
 - ・組織の評価専担部署の職員（社員）として実施した評価の評価報告書
 - ・組織において評価関連の管理職位に複数年（1年以上）従事した実績
- ④評価教育に関する実績
 - ・大学等の教育機関における研修講師の実績
 - ・その他の研修機関における研修講師の実績
- ⑤その他特筆すべき実績
 - ・評価関連の国際会議における活動実績
 - ・評価文化の醸成・普及に貢献する講演会等の実績
 - ・評価関連テーマによる取得学位

ただし、日本ではこのように研究業績、教育実績、実務経験、実務能力、学位（博士）などにわたって厳密な資格要件を課す「士」「師」は他に見られないので、議論の対象になりにくい。また上級評価士の数も、極めて少ない。つまり、上級評価士とは極めて特殊な事例なのである。期待は大きいですが、メタ評価の担当者として上級評価士を期待するのはかなり難しく、非現実的かも知れない。

2.2 プロセス・チェック

さて、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、評価の専門家がいなくて有効である。ただし、これは監査に似てくる。なぜなら以下のような形式的要件を強調することになるからである。

- ・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）に込めているか
- ・政策評価の基本計画（3～5年）、それ

に合わせて毎年策定される実施計画、実施要領（評価担当課が政策評価を実施する原課に配布する要領）にもとづき評価が行われているか

- ・評価に使用したデータは信頼できるか（とくに費用と便益のデータ）
- ・評価結果、報告書の提出はタイムリーか
- ・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）

この2番目の現実的思考は、多くの政策評価の現場で見られる。理由は、外部有識者のみならず政策評価担当課のスタッフまでも、原課が行っている政策の実質的内容に詳しくないことが多いからで、消去法的な選択がこのプロセス・チェック方式の採用につながっている。そしてこのチェックの問題点は、これらの要件を満たしても、政策が良い政策だったかどうか、有効性があったのかが分からないことである。

2.3 チェック項目の構築

第3の考慮すべき点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く、この決定から逆算してチェック項目を配置することである。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だが、その場合でもまた空理空論では困るので、実際のアプローチが必要である。その一例は、2009年度3月末に府省に通知された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、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、表3のように分類される（下線は筆者）。

この事例で見える限り、公共事業に関しては事後的チェックではあるにしても、方法や考え方は政策形成の場で重視されるポイントを対象にする「形成的メタ評価（formative meta-evaluation）」であり、一段高い処から（meta-）評価の実施状況を見て、評価のデータの集め方やデータそのもの、分析方法、マニュアルを対象にしている「評価の指南役」である。ま

表3 総務省の内容点検事例 2008年度

I 公共事業 (24件)
1. 費用算定の前提となる需要予測の妥当性に疑義があるもの 2. 便益算定に際しての評価方法の妥当性に疑義があるもの 3. 便益算定に用いられているデータ等の信頼性に疑義があるもの 4. マニュアルの適用の妥当性に疑義があるもの
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、その原因分析が行われていないと考えられるもの 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、政策効果に着目した指標の設定が必要と考えられるもの 3. あらかじめ設定した指標による効果の測定が行われていないもの 4. 判断基準・指標等の設定について改善が必要と考えられるもの 5. 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と考えられるもの

(筆者作成)

た、一般政策については結果や成果を事後的に総括する総括的メタ評価 (summative meta-evaluation)、つまり「評価のご意見番」に近い使い方をしている。

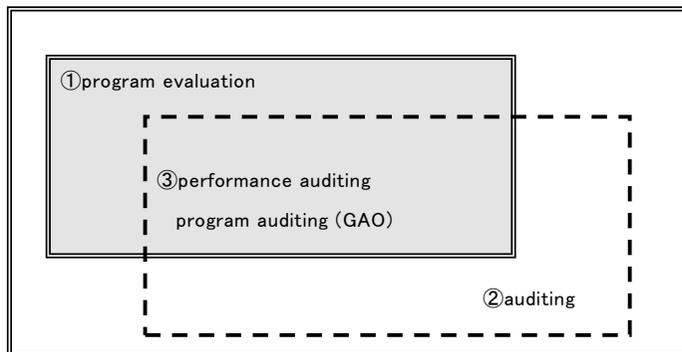
政策評価に関する方法論について認識が深まり、専門家が育ち、アメリカの大学院教育のように評価の専門家を目指す人びとがほぼ同種のテキストを使い、類似のカリキュラムでリテラシーの標準化と知的レベルアップが可能になれば、表3の公共事業に対して事後的に行われる形式的メタ評価は必要なくなる。そのときには政策に対する総括的メタ評価、われわれが「政策評価」と一般的に考える活動の本格的議論が可能になり、国民的レベルでの政策リテラシーも向上するはずである。

3. 「メタ評価」のプロトタイプ

政策評価の先進国のアメリカではかつて、日本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、政策評価チェックシステムについて多くの議論が見られた。この時期には予算編成と評価とは別立てのプロセス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、1990年代はじめに一定の決着を見ている。

3.1 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



(筆者作成)

図1 政府を監視する3つの方法 (1970's~1980's)

‘evaluation’ と ‘audit’ であり、両者の関係についてはアメリカ連邦政府における実務においてさまざまな議論が存在した。

すなわち、GAO のプログラム評価が 1980 年代にある程度定着した後、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 GPRA (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の部分で理論的にどのように考えるべきかが論争になっていた。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ているのか、また監査を主たる業務としてきた GAO (会計検査院) は、監査と評価をどのように使い分けているのかをめぐる論争である。

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’ (会計検査院主導なので業績「検査」と日本では呼んだ) 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 (Kloman1979)。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, p. 36)。なお、‘performance auditing’ の実務は後に ‘performance measurement’ に接近し、‘program auditing’ 活動は ‘evaluation research’ の方法をより多く反映させたため、これもまた事実上 ‘program evaluation’ に吸収された形になる。政策評価の実際の現場で観察される監査と評価の違いは、おおよそ以下の通りである。

- 目的：評価は有効性の確認、監査は法規・手続き・規定・指示のコンプライアンス・チェック。
- 関心事：実質的合理性の確保を考える評価に対して、監査は形式合理性に関心。
- 対象：評価はプログラムの対象・対象者の実際を見るが、監査は書面審査が多い。
- 担当機関の独立性を強く求め、強調する監査に対し、評価はそれほど強く求めず、むしろ現場を知りたいためにプログラム実施者と協調的な場合もある。

このように、評価と監査は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無視して実施してきたのであろう。

これにならえば、評価をチェックする方法にもふたつあると考えられる。1つは「評価の評価 (meta-evaluation)」であり、他は「評価の監査 (auditing evaluations)」である。

3.2 日本のチェックシステム

「評価の監査」とは評価の実施プロセスに対する形式審査であり、評価の運用手続き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。評価そのものの実質的内容には踏み込まない。逆に、評価の実質的内容を議論するのが「評価の評価」(以下「メタ評価」)であり、これは一段高いところから (meta-)、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。

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) 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はない。日本の運用実態では公認会計士や、行政の実務を知る人が担当する場面が多い。ただし、「外部」評価と言える体制を作らなければ、客観性を疑われ「お手盛り」の批判を受ける。実は、政策評価の客観性をめぐる論争は、まずこの点で出てきたのである。

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。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が必要になる (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、社会調査や統計学、経済学、財政学、評価学、そして政策学などの専門的知見を持つことが重要な前提である。したがって日常的に評価を担当していても、

これらの専門能力・知識に欠ける人には不向きである。

なお、図1の③に似た活動は日本にも存在する。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)、府省評価実施課への問い合わせや確認(事業評価ではなく実績評価を選択した理由、評価マインドを欠く記述と「作文」の文書をなぜ「総合評価」と呼ぶのかなど)、定量分析の強調、予算への反映件数の確認などである。日本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、メタ評価を担当する行政評価局がなじんだかつての行政監察(administrative inspection)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で形式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「重箱の隅を突つく」ことが多い。

3.3 メタ評価の類型

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, Lawrenz, Dunet, 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)。終わった評価の適否を、外部者や他機関が判断するのに役立てようとデータをそろえる内部作業。

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。

「PME・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。府省の官房で政策評価の取りまとめを行っている課の役割に近い。なお、2011年4月から、全府省では政策評価(実績評価)の実施にあたって政策目的と政策手段の対応関係を明示しつつ、あらかじめ政策効果に着目した達成すべき目標を設定する「事前分析表」を作成することになった。各政策担当課は予算額、業績指標、実績値とあわせて事前分析表を作成し、政策担当課に提出するが、その事前分析表を府省の評価担当課がチェックするとこのPMEになる。

また「RME・事後のメタ評価」とは評価の妥当性、有用性を「事前のメタ評価」と同じ項目を見て事後に判断するのであり、もし妥当性を欠き役に立たない(有用でない)との結論が出たときはその原因を評価プロセスに求める。

表4 4つの評価の位置関係概念図

	need assessment	input	activity	output	outcome	impact
内部	事前のメタ評価 PME			事後のメタ評価 RME		
外部	形成的メタ評価 FME				総括的メタ評価 SME	

(筆者作成)

そのため、評価手法やデータの内容にまで踏み込んだ調査をする。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。

「FME・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、一段高所の外部から見ているため評価実施担当者とは距離があり、評価活動の適正かつ円滑な運用を重視するプロセス・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の5つを事前にチェックする。詳細に行うにはコストと時間がかかるので評価ガイドラインや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。なお、この形成的メタ評価を事後的に行えば、総務省行政評価局が行う統一性確保評価、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「SME・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、まさに評価のオーソリティに「総括」を求める場合が想定される。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。もっとも、日本では総括的メタ評価を行うレベルまで実務が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、次の架空の話である。

X省がQ県で行っている干潟埋め立て事業は、住民の反対運動を引き起こしたため知事は工事続行に躊躇、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」判決が出たにもかかわらず、その事業の再評価がX省の政策評価対象に上ってこないとき、これに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」「状況が変わったら再評価すべきである」というプロフェッショナルリズムから発する意識

であり、それは評価専門職の責任感である。

4. おわりにー政策評価の変質

政策評価は政策評価法の施行後3年を経て政策評価法附則第3条に従って見直され、それをきっかけに2005年に大きな変更が見られた。ただ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。見直し作業の前提になっていたのは、①2005年6月「骨太方針」指示の予算との連携強化、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。そして2013年12月「目標管理型の政策評価の実施に関するガイドライン」(2013年12月20日政策評価府省連絡会議了承)が提示され、全府省共通で「施策レベルの実績評価」を行うことにしたのである。日本の政策評価は、諸外国で業績測定、パフォーマンス測定と呼ばれる手法への集中を意識させられた。

予算との連携強化が求めた政策評価の変質は、政策評価とは何かを説明するときに難しい問題を提起する。そもそも政策評価を所管する官庁が財務省ではなく総務省だったため、当初はそれほど強く予算を意識しない制度設計で、事業評価と総合評価(プログラム評価)、そして実績評価とで構成されていた。しかし予算との連携の新しいミッションは現実問題として政策評価の変化を求め、2012年に政権交代があった後、行政事業レビューと政策評価との役割分担関係の構築、両者の関係をつなぐ事前分析シートの作成という形でこの変化は現れた。もちろん、予算は数字であるため、それになじむ実績評価方式(業績測定・パフォーマンス測定)が政策評価の中心になってくる。この活動は節約と無駄削減には有効で、また評価の素人、政策内容をよく理解しない人には全府省同じ様式で書くので表面的には分かった気になるが、政策目標達成の現実を知ることは難しい。府省政策評価担当課と外部有識者会議のメンバーは、この事前分析表のチェック作業に時間をとられるようになったので、ますます本来の政策評価から遠ざかる。いずれにしても政策評価とその

チェックシステムを取り巻く体制には、後付けでさまざまな仕事が追加され、政策評価とパフォーマンス測定、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かったのである。

一般に、日本ではどこでも評価に対してはそれほどのエネルギーを使わず、実際に使う費用も用意していない。したがって人も費用も限られ、時間的余裕もないまま行われる。まして、メタ評価においてはさらにこの状況は深刻である。残された方法としては、選択と集中である。日本の政策評価の実務経験から考えると、選択と集中の提案は3つ考えられる。

その第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、府省の政策評価担当課とその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。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があり、

その課には「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った。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。整理するために使う枠組みは、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、以下の表5のようになるであろう。

選択集中策の第2は、アカデミズムとの連携をさらに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、「共同調査」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

表5 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
府省・政策評価担当課	事前メタ評価 形成的メタ評価	府省評価計画、ガイドライン、評価マニュアルの作成、そ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。年度実施施策に係る事前分析表のチェック。可能であれば、会計課と連携して「行政事業レビュー」に活用できるようにする。
府省・政策評価外部委員会	形成的メタ評価 事後メタ評価	評価基本計画、評価実施計画の事前チェック、評価可能性 (evaluability) の確認、年度実施施策に係る事前分析表の総合レビュー、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。人数は3～5人。
総務省行政評価局	形成的メタ評価 総括的メタ評価	府省の評価基本計画、評価実施計画のチェック。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	府省が選定した評価対象・評価手法、評価結果の妥当性確認。府省独立行政法人における事業評価との連携。

(筆者作成)

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いる。「御用学者」批判も変化しているなかで、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にはかなり有望であろう。

第3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分科会と独立行政法人評価分科会のふたつの分科会に分かれているが、一種形式的な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公共事業を実施していたばあい、その事業が10億円以上の公共事業・ODA事業であれば、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。さらに、国際協力機構や農畜産業振興機構をはじめとして、府省の政策手段を専門に担当している独立行政法人も少なくない。このように独立行政法人評価と政策評価は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。

結論として言えるのは、政策評価の客観性確保とは評価の質の確保であり、それに必要なのは評価能力涵養のことであって、しかし政策の中身の話とは別のことであった。したがって、上級評価士の資格を持ち政策評価を専門とする大学教員が、複数の府省の外部有識者を兼務しても問題はない。ただ、各政策領域の専門家のアドバイス（地域事情・平和構築・中小企業政策・防衛装備など）は不可欠である。政策評価の形式と政策の実質の狭間で知的基盤を涵養することが、生産的な「政策評価の客観性の議論」を導くはずである。

参考文献

1. 伊多波良雄・編著『公共政策のための政策評価手法』、中央経済社、2009年。
2. 宗高有吾「評価の客観性の現状とその課題」、(日本公共政策学会関西支部大会、2013年7月28日、セッションDにおける報告。)
3. 山谷清志『政策評価の理論とその展開－政府のアカウントビリティ－』、晃洋書房、1997年。
4. 山谷清志『政策評価の実践とその課題－アカウントビリティ

のジレンマ－』、萌書房、2006年。

5. 山谷清志『政策評価』、ミネルヴァ書房、2012年。
6. Hanssen, Carl E., Lawrenz, Frances, and Dunet, Diane, "Concurrent Meta-Evaluation: A Critique," *American Journal of Evaluation*, Vol. 29, No. 4, 2008, pp.572-582.
7. Kloman, Erasmus H., ed., *Cases in Accountability: The Work of the GAO*, Westview Press, 1979.
8. Mathison, Sandra ed., *Encyclopedia of Evaluation*, Sage, 2005.
9. Stufflebeam, Daniel L., "The Metaevaluation Imperative," *American Journal of Evaluation*, Vol. 22, No. 2, 2001, pp.183-209.