

## 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。しかし現実はそのとほど遠い。

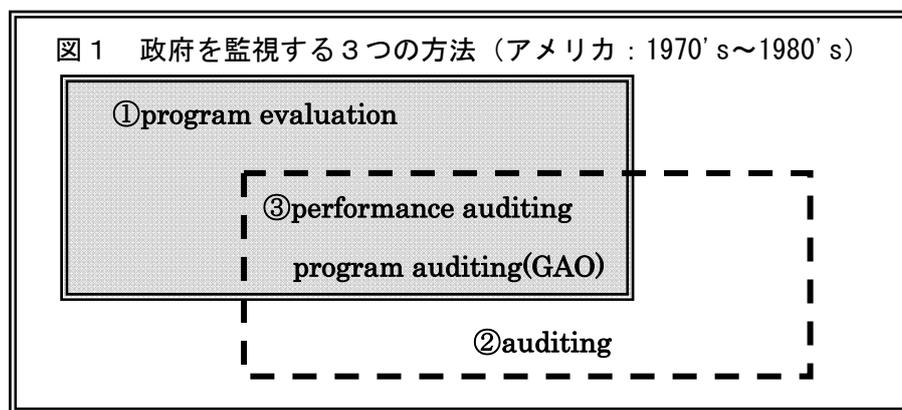
## 2 チェックシステムのプロトタイプ

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<sup>1</sup>。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

---

<sup>1</sup> 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

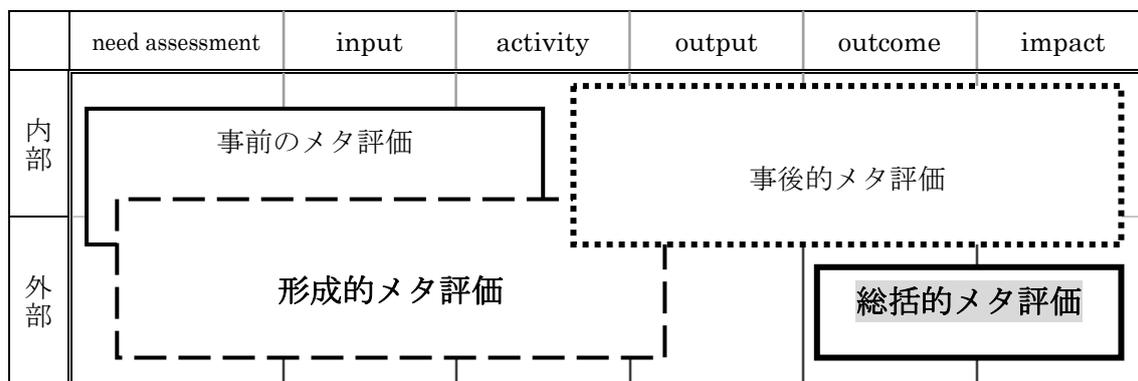
メタ評価についてはふたつの分類が可能である。その一つは事前・事後の時間的視点に立つ分類である (Hanssen et al. 2008)。組織の内部、外部にはそれほどこだわらない。

- 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><b>1. 案件実施の妥当性</b></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><b>2. 客観的検証の可能性</b></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><b>3. 有効性</b></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><b>4. 効率性</b></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><b>5. 総合評価</b></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"> <li>・基本計画・実施計画、ガイドライン、評価マニュアルの作成</li> <li>・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</li> <li>・評価業務の執行状況モニター。</li> <li>・評価サイクルのチェック</li> <li>・評価結果の予算・政策への反映支援</li> <li>・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）</li> </ul>
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"> <li>・基本計画、実施計画の事前チェック</li> <li>・評価可能性（evaluability）の確認</li> <li>・評価結果の妥当性の確認</li> <li>・評価結果の予算反映の可否チェック</li> </ul>
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"> <li>・各府省の基本計画、実施計画の確認</li> <li>・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</li> <li>・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</li> <li>・評価サイクルの確認</li> <li>・手続手順に関わる問題状況の指摘</li> </ul>

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"> <li>・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</li> <li>・各府省独立行政法人における事業評価との連携</li> </ul>
----------------------	---------	-----------------------------------------------------------------------------------------------------------------------------

(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。

