

政策学における評価理論の貢献

—日本の評価システムから—

山谷 清志

概要

本稿では政策の実務と理論に評価がどのような形で貢献したのか、その検証をしたい。手かりは日本の評価システムの特徴である。日本の公共部門にはさまざまな評価が重疊的に混在し、しかし評価の理解が乏しい中で使われた結果として、日本の評価システムは「ガラパゴス化」（国際標準から逸脱）した。実務が評価学の知見を十分に生かしきれていないからである。そして、このガラパゴス化した評価システムが、政策学に悪影響を及ぼしてきたとの仮説も存在する。そこで本稿の前半はこの重疊化・混在化の背景、後半はガラパゴス化するに至った理由を政策学、行政学の知見から説明する。最後に、この説明をふまえ評価の大きな課題を明らかにする。

背景には政策学の「遠心力」問題がある。長年、政策学は学際的なマルチ・ディシプリンを自認し、政策学者は自分たちのマザー・ディシプリンの中で政策研究を進めてきた結果、政策学の「遠心力」が強まった（山谷 2019）。同志社大学・中央大学・慶応義塾大学・関西大学・関西学院大学・立命館大学・南山大学・津田塾大学の政策系学部学部長懇談会でも、この遠心力問題はたびたび指摘されてきた。政策学の研究者同士が共通理解をもたず、政策学以外のジャーゴンを使い、結果として政策学者相互の意思疎通が難しくなったこと、これが背景にある。

ここでは可能であれば、評価学の知見と政策現場の経験をふまえ、日本の評価システムを使った政策学の「求心力」を回復する方向を模索したい。

1. 行政学と評価学

政策学が政策の運用実態を知るために行政学の研究成果を吸収してきたことは、欧米の行政学、そして国際行政の実務、とくに政府開発援助（ODA）に関わる世界銀行や経済協力開発機構（OECD）では周知の事実である。

日本では 2001 年に政策評価が「行政機関が行う政策の評価に関する法律」（以下、政策評価法）によって制度化され、政策学は行政学と評価学の国際的な成果を取り入れ、ODA 評価の実践で得た教訓も吸収してきた。政策体系、政策デザイン、ロジック・モデルがその代表である。日本評価学会が 2000 年に設立されて以降この傾向は強まった。

ところで、過去半世紀以上におよぶ日本の政策研究の歴史は、行政学と評価学の視点から振り返ると、5つの時代に区分できる。

第1は 1960 年代の「計画の時代」である。核心は計画と予算の結合で、そのために事前の政策分析に期待した。それがアメリカ連邦政府で採用され、日本でも大蔵省、防衛庁、経済企画庁、建設省、運輸省で試行された PPBS (Planning Programming Budgeting System) である。ただし失敗した。理由は予算編成作業での使い勝手の悪さ、政策効果の予測困難、PPBS 推進者が行政と財政の実務知識を持っていなかった、データ処理作業の手間が膨大だったなどである（加藤 2008:147-152）。なお 2015 年頃から実務で議論されてきた Evidence Based Policy Making (EBPM) にも、似たところがある。

第2は 1970 年代から 1980 年代、本格的に政策の科学研究に着手した「政策科学」(Policy Sciences) の時代である。予測に基づき計画を策定して将来をコントロールしたいと考える未

来志向の時代精神は、システム分析を通じて実務に入り込み、科学（science）思考を政策研究に定着させた。当時、情報化という言葉が盛んに使われ、行政機関がコンピュータを活用する合理化が模索された時代でもある（OA化・電算化）。

なお、この時期欧米では後の政策評価につながる変化が生まれていた。事後評価（ex post evaluation）が福祉や教育、ODAの分野で多用され、一定の成果を出しており、それが政策研究に臨床の視点を導入して応用につながった。事前分析（ex ante analysis）重視の政策科学から、政策学（Policy Studies）への転換で、出ている結果を見てその原因を遡及的に調査する思考が登場したのである。

第3は、政策を対象にした応用研究が着実に進んだ1990年代後半から2000年代初頭である。この時期の象徴的な出来事は、日本公共政策学会の誕生である（1996年6月20日）。政治学、行政学、経済学、法学、社会学など多くの研究者、実務家、ジャーナリストが政策研究（Policy Studies）に取り組みはじめた。並行して日本各地で政策系学部・大学院のブームが起きたことも記憶される（同志社大学総合政策科学研究科の設立は1995年）。実務の経験を生かして、既存の学問分野を再編成しようとする試みであった。総合政策科学研究科の新川達郎教授は、2012年から2014年まで、この日本公共政策学会の会長であった。

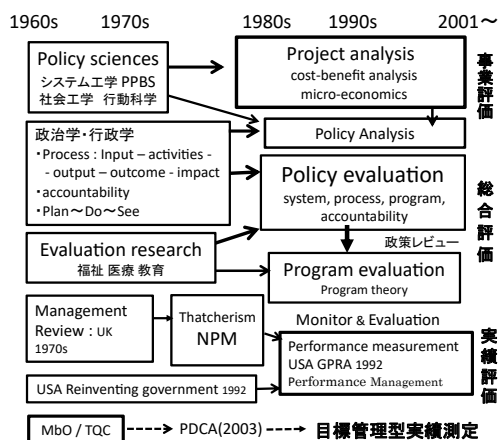
2001年のミレニアム後の第4の時期が、政府改革を経験した実践的政策研究、応用政策学が成果を出した時代である。この時期に影響を及ぼした重要なトピックは、行政改革会議（橋本行革、1996年～1997年）である。橋本行革は中央省庁の統合、再編成として注目されたが、その背景には政策の立案と実施とを分け、それぞれにアカウンタビリティを求め、そのアカウンタビリティ確保手段として評価（政策評価と独立行政法人評価）を置く新しいアイデアがあった。この改革のひとつ政策評価は、評価で得た情報を政策立案プロセスにフィードバックする可能性を開いた。2000年に設立された日本評価学会の学会活動は、この時代精神を反映した。

2010年以降の政策学の第5期には、社会にさまざまな非常事態が出現し、政策学はその研

究対象も内容も大きく変化させた。2011年の東日本大震災と福島原発事故、2014年の広島豪雨被害に代表される各地の豪雨災害、熊本地震（2016年）、そして2020年の新型コロナウイルス禍である。とくに新型コロナ禍の政策対応は、政策学の研究対象としては極めて重要である。コロナ禍で苦闘するエッセンシャル・ワーカーに、小泉内閣以来歴代内閣が続けた「新自由主義」マクロ政策が負の影響を及ぼしてきたが、その認識を市民も共有した時期でもある。

同時に、現場の活動を評価し、評価結果をフィードバックによって政策の立案、予算編成に反映させる実践が始まったのも、この時期である。政策評価と事業見なおしの連携で、行政刷新会議「新仕分け」2012年、各府省自らが行なう「行政事業レビュー」が2015年から実施された。

ところで、評価学はアメリカで始まった研究と実践の分野である（山谷1997、第2章）。いわゆる‘Evaluation research’であり、また1960年代にアメリカ連邦政府が展開した「偉大な社会プログラム」を評価したため、社会プログラム評価（Social Program Evaluation）と呼ばれた（山谷2012、第3章）。これが政策評価のプロトタイプである（図表1）。教育学、社会学、心理学、保健学、福祉学が中心で、それを統計学や社会調査、行政学（イギリスのSocial Administration）、そして政治学が支えた。紛争・貧困解決、人道支援、人権、ジェンダーの分野



図表1 さまざまな評価・分析・測定

出典：山谷清志作成

で実施され、国際協力では必須になる。日本にも、これらの研究分野と教育実践を通じて評価が浸透したことは知られている。

しかしこの評価の流れに水を差す動きがあった。1970年代から実践され、1980年代のイギリスの政府改革(当時の首相だった「サッチャー改革」で有名である)で使われた「Performance Measurement」が国際的に流行し、日本にも伝わったからである。その20世紀末、英語を使わない日本では、組織活動の業績測定が上記の評価と混同され、現在使われている政策評価にも実績評価として採用されてしまった。

ただし、そのまま模倣したわけではない。イギリスではアウトプットを指標にしたパフォーマンス測定であったが、日本ではアウトカム測定として使われる。ここに問題がある。アウトカムは外部の影響に左右され、この外部の影響を行政機関は完全にコントロールできないので、もし成果が出ないと行政機関を非難しても行政機関に責任が無いことが多い。このように、アカウントビリティ追及にアウトカム測定は向いていない事実も無視されている。

また、公共事業を代表とするプロジェクト評価の分野では、プロジェクトに投資するときに行われる事前審査において、費用便益分析に代表される事前分析(analysis)手法が、1960年代から経済学、システム工学、社会工学の知見を活かして採用された(たとえば代表は前述のPPBSシステム)。もちろん当時欧米で流行した行動科学の影響も大きい。この流れは現在、事業評価の事前分析として定着している。活躍したのは経済学、社会工学、経営学の分野で、政策学を構成するサブ・ディシプリンの政策科学になっている。

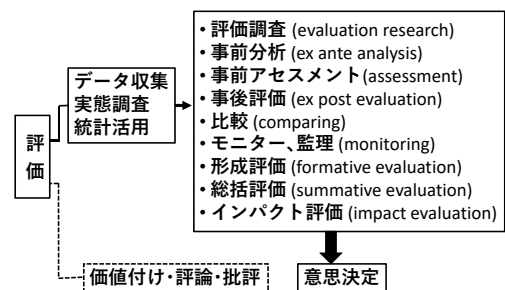
なお、2003年に経済財政諮問会議(2003年2月17日奥田碩議員の発言)で出されたPDCA(Plan Do Check Action; 下線は山谷)のアイデアは、発言者が自動車会社の経営者だったこともあって、政策や行政の議論よりは工場の生産現場の視点が強く出ていた。そもそもの発想がTotal Quality Controlにあり、生産現場の業務改善、生産性向上を目的としていた。和製英語で英語のネイティブに理解できなかったPDCAは、政策学と行政学、評価学とは全く違う背景をもつが、なぜか政策評価を説明する言葉として実務に定着してしまった。政策論を効率化や

節約、コストカットに矮小化する新自由主義的な政権の思惑とあいまって、全く異質な評価の議論が登場し、意図的な誤解がここで始まった。

こうして、図表1に示すように、日本ではさまざまな評価と測定の実践が溢れ、研究は錯綜し、市民の認識が歪められ、しかしこの事実が理解されないまま、第5期に至った。給与削減や人員カットを標榜する政治家が選挙戦を有利に闘う偏った世情もまた、こうした政策学と評価の理論研究と実践の教育に大きな混乱を招いていた。

2. 理論無き評価システム

誤解と混乱を招く理由は、評価の日本語と英語の両方にある。日本語で評価は特定の価値に基づき優劣を付ける価値判断のニュアンスを排除できない。また英語でも「e-valuation」と書くように、一定の価値判断だと考えられることが多い。しかし、こと公共部門の評価については、この考えは混乱の原因になる。価値観が違う人を説得できないからである。そこで評価学は、評価を冷静な議論の場で使うツールだと考えるべきだと主張してきた。その主張では、評価に使うデータ収集者と判断者は別であるとする(図表2)。そして、世間一般の日常会話の評価(評判や価値判断と同義)とは、点線の部分で、必ずしも客観的なデータに基づかない。



図表2 評価の概要

出典：山谷清志作成

つまり、現場で何が起きたのか、結果がどう出ているのか、この結果は予測と違ったのか、違った原因が何なのかを考える、冷静な議論のツールであると評価は定義される。図表2にあ

るように、実際の評価に着手する場合には、評価活動は前後の2つに分けられる。

前者は情報を収集して分析する活動で、客観的で正確な情報・データ収集を行う。エビデンス収集はその基本である。この場合、客観的とは情報収集の意図を明確にすること、集めた情報の性格を明らかにすることも意味する。ただ、データは政策目的ごとに意味が違ってくるので、客観的という場合には配慮が必要である。

2つめは判断、意思決定である。ただし、データを見れば機械的に判断できると主張するのは楽天的で、社会には優先順位、市民の選好、立場の違いなどがあって機械的に判断できる場面は多くない。評価学と政策学で政治信条、哲学、倫理学の議論が多いのはこのためである。

もちろん評価が政治化しないように、現実社会では用心する。この用心が評価の基本条件になる。評価が現状を知り、冷静な議論のツールとして機能すべきであれば、まず事後的な公開レビューを重視し、どうしてその政策判断をしたのか、政策判断の結果がどうなったのかを遡及的にレビューし、もし失敗していた時には判断者の結果責任を追及することになる。この事後レビュー・プロセスに不可欠なのは透明性（transparency）と追跡可能性（traceability）で、仕組みとしては情報公開と公文書管理が必要になる。

なお、政策判断はこのように結果責任になるため、アカウンタビリティは単なる説明責任ではなく、政治家や行政官、専門家たちにとっては結果について市民が納得できる説明をする能力、市民にとっては結果について公職者に説明を求める努力になる。評価がここで重要な役割を果たすのは言うまでもない。こうして、評価は市民にとって「民主主義のリテラシー」になり、専門的には応用社会科学（Applied Social Sciences）の学際（Multi-discipline）研究と呼ばれるようになる。

3. 評価の3目的

こうして考えると、評価の性格については再整理する必要がある。まずその目的である。

評価の目的は第1にアカウンタビリティの確保で、それは政策現場で何が行われているのか

実態の説明を求めることから始まる。すなわち政策の成功・失敗、政策効果の有無が関心事なので、有効性が評価規準になる。後述する政策とその手段を示すプログラムの評価（Program Evaluation）は、その主要な方法である。これが評価の第1目的である。このアカウンタビリティは長年、行政学が行政責任論の文脈で議論してきたテーマである。

他方、政策学が注目されてから、政策実施の現場ではさまざまな政策マネジメントに使う情報が求められる。これが評価の第2の目的である。一般にマネジメントで重視されるのは人事、予算、事業における合法性、合規性、コンプライアンス（準拠性・法令遵守）情報なので、監査（audit）と重なるが、実はこれは評価でない。非違や不正、過誤を追及する監査、監察（inspection）とは別の活動である。効率（行政学では能率）と節約をマネジメントは重視するので、この情報を集める方法として事業活動の進捗をモニターする目標管理型の‘performance measurement’が使われ、これを日本では評価に含めた。この部分で経営学や会計学の管理会計などに接近する。

行政経営型のマネジメントは総務と名前の付く組織が担当する内部管理を意味する一方で、インフラの整備プロジェクトや研究開発プロジェクトの現場にはプロジェクト・マネージャーがおり、別のマネジメントを行っている。その基本はプロジェクトの進捗管理、財務状況の把握、プロジェクト環境（ジェンダー、人権、先住民族、自然環境、動植物、気候、ワークライフ・バランス、環境汚染）への配慮である。評価がマネジメントの支援と言うときには、プロジェクト評価を通じた貢献、支援になるだろう。

評価の第3の目的は「専門分野に知的に貢献する」であるが、これは少し分かりにくい。実際の例で言えば児童虐待対策プログラムが成果を出していない、貧困対策プログラムが問題を解決しない、多くの予算が投じられても巧く行かない、このような場面を想定すればよいかも知れない。その原因を探るために福祉学や心理学の専門的知見からプログラム評価が行われる。このためプログラムをめぐるアカウンタビリティ追及と専門分野への知的貢献は、評価の重要な両輪になる。

前述のように、プログラム評価は1960年代に注目された手法で、教育、医療、社会福祉、研究開発のプログラムの分野で注目された。いずれも、従来の監査や会計検査によっては成果（アウトカム）の有無を確認するのが困難な分野で、数字やデータはあるがその解釈が難しい実践領域であった。また結果責任が重視されるので事後評価が重視された。なお、このプログラム評価を、アメリカのように連邦議会が行えば、アカウンタビリティの手段として政策実施機関（行政）の責任を議会が追及する手段として使える。他方、政策実施機関が自ら責任確認のために評価するのであれば自己評価、反省手段になり、英語の‘responsibility’である。外部評価が難しいプロフェッショナルの専門評価は、このレスポンスビリティになりがちで、ピア・レビューや参加型評価ワークショップなどの手法が開発されてきた。

このような3つの目的を持つ評価が日本では、実際にはどのように活用されるのか、以下で評価対象の視点から考えてみたい。

4. 7つの評価対象

何を評価するのかに関しては、実際の現場ではさまざまな対象を評価してきたため、その整理は難しく、対応に困る事態が多く見られる。

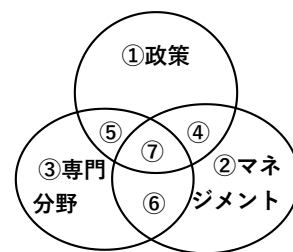
ただし、基本の了解はある。政策を対象にした評価（政策評価）と、組織活動を対象にした評価（行政評価・経営評価）とに別れるはずで、前者は政策学、後者は行政学や経営学の範疇にある。政策評価は、政策活動の結果として政策目標が達成されたかどうかを知りたいので、目標達成度すなわち有効性、そして効果が公平に出ているのかを調べる。他方、組織活動の評価は予算とスタッフのインプットを投入して活動（activity）を行い、その活動でアウトプットをどれほど生産できたかが重要である。この生産活動の効率（行政学では同じ efficiency を「能率」と呼ぶ）、節約、生産性が重要である。政策評価と独立行政法人評価を分けたのは、この政策評価と組織評価のアイデアが背景にある。

ただし、実態はこのように簡単ではない。とくに独立行政法人（国際協力機構 JICA が代表）と国立研究開発法人（宇宙航空研究開発機構

JAXA が代表）では、複雑な評価システムが形成されている（以下で法人と言う場合は独立行政法人と研究開発法人の両者をさす）。複雑な評価システムを解きほぐし、丁寧にたどると、公共部門には評価の諸制度を通して7つの評価対象活動が観察される。図表3で説明したい。

①政策の評価

政策活動は政府が策定する計画である。国会の議決や閣議を経た基本法、基本計画、大綱として公表され、その白書や外交青書が政策文書である。具体的なイメージとして宇宙関係の政策をあげると、政府が策定する2系統の法律とそれに基づく計画がある。科学技術・イノベーション基本法→科学技術・イノベーション基本計画→統合イノベーション戦略の系統、そして宇宙基本法→宇宙基本計画→宇宙基本計画工程表の系統である。それぞれ評価は大綱や基本計画の策定や改定の際に行われるが、それに加えて政策評価法に基づき定期的に政策評価を行っている。なお、地方自治体の場合、首長が選挙公約で提示したマニフェストがこの政策に該当する。



図表3 評価の7分野

出典：山谷清志作成

政策は政策、プログラム（施策）、プロジェクト（事業）の3層体系として認識される。政策の目標を達成する手段を説明するのがプログラムであり、そのプログラムには事業が複数含まれる。この体系を意識せずに政策を語ると目標・手段関係が分からず、事業だけが専ら注目される。地域経済の振興→交通網の整備→道路建設のようなトップダウン的な説明をする場合に政策体系は役に立つ。しかし、道路の整備以外にも地域経済の振興策はあり、さらに道路の補修や除雪の費用がかさみ、地方自治体の財政状態を圧迫する恐れもある。政策体系を使ってトッ

ブダウで行う説明には、難しいことが多い。

難しさを緩和する手がかりは、政策評価の評価方式にある。日本では政策評価が法制度化された2001年以来、3つの評価方式で実施されてきた。第1の方式は実績評価方式である。実績とはパフォーマンスのことで、目標の達成度合い、事業の進捗状況をモニターする。この手法は政策体系とは関係なく実施できるので汎用性が高く、また数字で示すので理解が簡単でもあり多用されている。実績評価方式はトップが全体の概略を把握するには適しているが、成果が出ていないときの原因を探るには不向きである。

第2の方式、事業評価は公共事業を代表とするプロジェクト評価、事前評価の場合にはプロジェクト分析である。政策評価法はこの事前評価を公共事業、政府開発援助（ODA）、研究開発に限って対象にしていた（第9条）。手法が確立しているとの理由だった。さらにその後、規制（2007年）、租税特別措置（2010年）が事前評価に加わって事前評価が5分野に拡大した。ただ、行政事業レビューを政策評価と併用する経済財政諮問会議提言（2013年3月8日）に従って、実績評価が事前評価の中心になった。

政策評価の方式の第3の「総合評価」が政策を対象にした評価にふさわしい。事後の総括的評価（summative evaluation）を行う‘comprehensive evaluation’がこの総合評価であり、政策評価をプログラム評価と考える諸外国の評価理論に一番近い。これであれば外部の影響を調査する評価も可能になる。

なお、政策評価導入時の所管官庁であった総務庁行政監察局が主催した「政策評価の手法等に関する研究会」で名称を検討していた段階では、「政策体系評価」の名称が候補に挙がっていたが、議論の結果、総合評価になった経緯がある（2000年6月19日第14回会合）。

② マネジメントと管理の評価

マネジメントを「管理」と考え、この管理の言葉が付く活動を対象にする。伝統的には定員管理、財務管理、人事管理、業務管理が伝統的であったが、事業を中心とする考えではプロジェクト・マネジメント、官公庁の現場ではプログラム・マネジメント、政策的発想では政策管理が使われてきている。他方、組織ではミドル・レベルでアドミニストレーション、現場で

はマネジメントがここで言う「管理」である。プリンシパル（P）とエージェンツ（A）の関係がこの背景にあり、プリンシパルに命じられたミドル・レベルの総務系組織は、予算管理や財務管理、財産管理、人事管理を行う。現場のAであれば庶務の内部管理とオペレーションが対象である。PがAの現状をコントロールするマネジメントでは、活動実績、アウトプットの出来高を見るために‘performance measurement’が必要になる。先の実績評価である。マネジメント・レビューと呼ぶマネジメント・コントロールで使う活動もこれに近く、地方自治体で言えば行政評価の名前で行う事務事業評価が似ている。

ここで留意すべき点は、政策実施組織がコントロールできるのはインプット、次いでインプットを投入して営まれる活動、その活動で得られる結果（アウトプット）までという点である。他方、成果（アウトカム）は組織の外部環境からの影響、例えば景気動向や災害、住民の選好、その時代の雰囲気や運によって左右されることが多いので、100%すべてAの責任だと言えないので、組織やメンバーの実績の評価としてアカウントビリティを追及するには留保が必要になる。成果主義は精神的スローガンとしては意味があるが、現場を無視した成果評価に拘泥するのは危険かも知れない。

なお、組織体系の上部を指すトップ・マネジメントという言葉もあるが、幹部の活動と現場監督とで混乱を生むので、上部は執政（executive）と呼んだ方が区別しやすい。

③ 専門分野の政策とその評価

この場合の専門分野とは、教育政策では教育学、社会福祉政策の福祉学、科学技術政策であれば研究開発政策である。そしてこの専門分野の研究と実務に知的に貢献する目的で、評価が行われる。この評価を政策評価やマネジメント評価と区別して、‘Professional Evaluation’の名称を使うこともあり、日本評価学会が認定する学校評価士はその代表である。したがって、専門学会の認証、認定もこの範疇に入る。

なお、前述のように政策評価制度は研究開発と公共事業、政府開発援助（ODA）の3分野に事前評価が義務づけたが、これらの分野では専門的な評価手法が確立していると考えたからである。確かに、たとえば2015年の経済財政諮問会議、2016年頃から政府与党、そして中

央府省で議論され始めた「エビデンス」に基づくPDCAサイクル（政策形成・評価）、そしてRCT（Randomized Controlled Trial）は、実は1970年代に専門評価の分野（とくに医療・福祉・保健）で流行した方法で、専門評価には長年の研究と実践の蓄積がある。

さて、現実にはこれら①②③の基本そのものが明確に意識されないまま、政府や法人の評価、さまざまな要請から派生した評価や測定が重層的に実施されている。以下それについて説明する。

④政策実施活動

政策とマネジメントの両方に関わる実施活動（implementation）は、組織のエグゼクティブから指示をうけたミドル・レベル管理者や現場マネージャーが営む活動である。橋本行革の最終報告（1997年12月）で政策の企画立案部門から切り離れた実施部門がイメージされる。もちろん企画立案が①で、この④では実施が行われるが、その評価の実態はモニターと呼ばれる監視、業務実績の測定が中心である。

この政策実施活動測定の代表は独立行政法人や研究開発法人の評価対象業務であり、この場合の評価とは業務を委託された法人が本省（大臣）に報告するもので、本省が指示する方向で現場の仕事をマネジメントした結果の業務実績報告になる。実施庁（気象庁、特許庁、海上保安庁）の実績評価も同じカテゴリーである。考え方としては、政策を国の府省が作るが、政策実施を任される法人、実施庁、地方自治体、民間（企業、病院、私立大学）が評価や測定を求められ、それを前提に補助金や交付金が渡される方法である。

法人の場合、各府省の長である大臣が5～7年間の中期目標を提示し、それによって法人が中期計画を策定し、活動を行う。前者の中期目標は政策評価の対象になり、後者の中期計画は業務実績評価の対象である。年度ごとの業務実績評価もある。

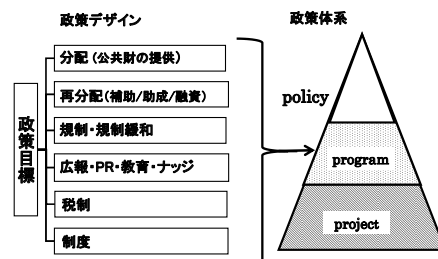
⑤プログラムとその評価

プログラム活動とは政策目的を達成するためのアイデアとスケジュールを考え、またその成功や失敗の評価を政策活動に組み込む活動である。1960年代から70年代にかけて積極的にプログラム評価に取り組んだアメリカの連邦政府の例で言えば、かつて連邦政府が決めた社会プログラムが基礎自治体に補助金を交付、交

付を受けた基礎自治体はこのプログラムが規定した方法でプロジェクトを実施する（Urban Institute, 1970）。似た例は日本で2014年度から始まった地方創生関係交付金における内閣府と地方自治体の関係にもある。

また、このプログラム評価は、政策目標を達成する手段として作成されたプログラムに従って専門分野が活動しているかどうかを確認する作業になる（プロトコルの意味）。日本の政策医療、政策金融、政策年金がその代表である。

この確認作業が‘evaluation’であるが、確認ポイントは第1に政策体系、第2に政策デザイン、第3に政策の形成から終了に至るまでの流れの論理的展開を説明するロジック（logic-model）、この3つである。政策の責任者はプログラムの結果について市民が納得できる説明をする責務を担うが、このアカウントビリティが‘program accountability’である。その意味でプログラムとは、さまざまな専門分野で行われている活動を玄人が素人に説明する時に使う共通言語で、プログラム評価は政府が市民に対して説明する基本である。プログラム・アカウントビリティ、プログラム評価は政策医療、政策金融、ODA政策、科学技術政策の分野で有効である。もし政策にプログラムが無いならば、それはそのデザインが杜撰だからである。杜撰さを指摘された場合、④や⑥で代用され、効率や目標数値に拘泥し、あるいはKPI基準を多用する。



図表4 政策体系と政策デザイン

出典：山谷清志作成

ところで、宇宙開発をはじめとする科学技術政策の分野でプログラムが使われるのは、アカウントビリティのためもあるが、構造としての政策体系の中心をプログラムが占めているからでもある。このプログラムは政策目標とそれを

達成する政策手段の組み合わせのデザインで（図表4）、プログラム評価とはそのデザインの適切さのチェックだと言うこともできる（山谷2020）。

プログラムにはそれを構成するプロジェクトの他に、関係者、受益者、地域、必要とする専門などが記述されている。もちろんスケジュールや工程表、言いかえると成果に至る「道筋」も書き込まれている。道筋とは、評価理論の中で使われてきたロジック・モデルを日本語で説明する言葉である。

このロジック・モデルとはプログラムの流れを論理的に説明する考え方で、プログラム作成時に組み立てられるが、他方でプログラムが終了した後から逆算的に遡及して追跡する（trace）時にも使われる。プログラムの道筋を構成する段階は5つある。

- ・input：資源の投入
- ・activity：投入された資源を使う活動
- ・output（結果）：活動が産出した結果
- ・outcome（成果）：結果による成果
- ・impact：このプログラム活動の影響

これらの流れを論理的に説明できるプログラムが「良い」プログラムと考えられる。もちろん想定した結果が出ないとき、あるいはこの結果が成果につながらないときには事後評価の登場になる。

⑥プロジェクト・マネジメント評価

プロジェクト・レベルでのマネジメント活動が基本になる。研究開発部門での代表はプロジェクト・マネジメントで（予算の執行管理は②）、これを統括するポストを特に設けている組織が多く、⑥を専門に行うスタッフも増えている。なお、JAXAのような巨大なプロジェクトを抱えている法人は、市民から見るとどうしてもこのプロジェクトが目立ち、プロジェクトだけが独り歩きする「木を見て森を見ず」状態になる。

公共事業をはじめとするプロジェクトの失敗研究は多い。失敗原因は過去の経緯へのこだわり、コストの過少計算、社会環境変化の読み間違い、歴史の教訓無視、リスクの過少評価、技術力の不足、科学的合理性を欠く政治判断、少数者の苦悩に鈍感、陳腐化した前提への拘泥、「空気の読み過ぎ」、「同調圧力」、特定利益の付度など多様である。これらはプログラムがその

作成者が意図した動きと違う動作をする原因、いわゆる「プログラム・バグ」であるが、プログラム評価やプロジェクト評価の対象にならず、公表されないことが多い。

なお、⑥の実例は非常に多い。たとえば義務教育は国の政策であるが、それを地域で実施する役割を担っている教育委員会については、事務局における教員出身の教育職職員（指導主事）以外の、一般行政職員の専門性をどのように高めるべきかという議論があった（村上2015:71）。ここまでであれば⑥の議論になるが、実際の2014年の教育委員会制度改革は首長の意向をより反映する改革（図表3では①に接近する方向）になった。

また、宇宙航空開発研究機構JAXAでは文科系の学部卒業者をこのポストに当てているし（宇宙探査イノベーション・ハブがその例）、同じく地方独立行政法人化した公立病院でも病院事務担当者として文科系学部卒業者が採用されている。

⑦全体評価

全体評価は、すべての活動を統括する一段上位の活動である。Executiveと呼ばれる人の活動である。その中心は判断とその説明（結果責任の意味のアカウンタビリティ）が中心で、科学的な分野であればサイエンス・コミュニケーション、防災や原発のようにリスクが伴う課題についてはリスク・コミュニケーションの能力が求められる。

ただし、この活動を担当する人が①～⑥全部について熟知している例は少ないし、現場の状況をトップがすべて把握できるわけでもない。したがって、たとえば政府の司令塔の内閣府・内閣官房が何らかの指標やKey Performance Indicator（KPI）に頼るのはやむをえない。政府全体の政策活動を統括する（知る）ために、トップダウン型の指標が最適だからである。KPIは政策内容に詳しくないトップのエグゼクティブが使い易いが、同じ理由で評定やレーティングも専門知識がない政治家や市民にとっては意味がある。

なお、この⑦に評価を入れて、そのために外部有識者委員会を設置し、政策内容に関わる専門の研究者を任命するのは適切ではないかもしれない。図表3の③の活動を、同じ③の専門家に見せるだけに終わり、専門家同士の納得と確

認に終わる恐れがある。あるいは①～⑥に関わる専門家をすべて招聘し、それぞれの意見を求めるのがバランスをとる方法として良いと考える委員会もあるが、実際にはそれぞれ専門が違うため会議では議論がすれ違い、意見の言いっ放し、招聘した行政機関は「専門の先生に見せた」とアリバイづくりに使う。審議会政治問題の復活である。また、委員選任が恣意的だと批判されても反論できない難もある。

これら7種類の活動には、それぞれふさわしい評価がある。もし不適切な評価を選択すると誤った政策情報を提供し、政策判断ミスにつながる。喫緊の課題に対応すべき専門家が評価のペーパーワークに動員され、課題対応に使うべき時間を浪費する恐れもある。間違った評価を命じられ、専門家が提供したデータが政策全体の判断を誤誘導する危険もある。また20年、30年たってはじめて成果が分かるような研究については、判断ミスの影響が分からないという課題がある。

5. 評価が導く論点

評価の目的が3つあり、評価が対象にする領域は7種類あった。以下でこうして出来上がった評価システムから浮上した論点を指摘したい。いずれも評価に際して登場した課題で、政策学にも重要検討事項である。

論点1：政治と政策の関係

言葉がよく似ているため政治（politics）と政策（policy）の関係は分かりにくいことが多い。地方自治体の行政改革を目的とする行政評価が自治体の立法機関である議員定数の削減を提案した誤謬、閣議決定の対象になる大綱そのものを「行政機関が行う政策の評価」の対象にして何を評価するのかという戸惑いなど、政治と政策の関係に混乱は多い。

ただし、21世紀になってから整理は進み、政策評価では政治と政策のデマケーションが明記される。たとえば府省の政策評価書は標準化が進み、ここには施策名、施策の概要、達成すべき目標、そして施策に関係する内閣の重要政策が記述されている。この重要政策の部分の決定が「政治」で、それ以外が行政機関が行う政

策である。

なお、専門化が進んでいる政策領域で発生した問題に対しては政治的決着が求められるが、新型コロナ禍問題に代表されるように、専門家でない政治家の政治的判断が必ずしも正しいとは限らない。また、政治的配慮から政策目的が二つになってしまい、相互がトレードオフ状態に陥ることもある。新型コロナ禍対策の経済対策と感染者拡大阻止の例である。

論点2：政策と行政

政策と行政（administration）の区別も難しい。行政は制度としての理解（憲法・国家行政組織法・地方自治法）、機能としての理解（いわゆる行政活動）の2つに分けられる。ここで問題になるのが機能としての行政と政策との区別である。

政策が政策→施策（プログラム）→事業（プロジェクト）と上からピラミッドの政策体系を形成しているように、行政機能も行政体系、すなわち執政（executive）→中間管理（administration）→現場の業務（operation）の3層構造の体系で説明できる。それぞれ政策体系に対応するが、行政体系のうち中間管理と業務が世間がイメージする行政である。

評価の視点を入れた政策評価と行政評価の違いを使って整理すると、政策と行政の機能の違いがさらに明確化する。政策評価は政策の有効性、政策の成果が第1の関心事である。これに対して行政評価は、行政という組織活動の評価なので合法性、合規性、コンプライアンス、効率、経済性（節約）が中心的関心事である。もちろん政策の成功・失敗の議論と行政機能の成功・失敗の話は次元が違う。行政機能を人、予算、法令などの管理情報を使ってコントロールする手段がマネジメントである。

論点3：output と outcome

アウトプットと、アウトカムを使用してくるアウトカムとは意味が違う。政策学は評価学の影響を受けて、以下の考えを使うようになった。「プロジェクトがインプットを使った活動によってアウトプットを生産し、このアウトプットが社会に貢献して出てくる成果をアウトカムと呼ぶ」。ここが評価学の貢献である。

アウトプットとアウトカムの関係で問題にな

るのは3つある。1つはアウトプットとアウトカムの区別をつけ難い場合（外交や防衛のような場合）、2つめはアウトカムが発現するまで長時間（20年）必要で担当組織の責任を問えない場合である。

3つめがアウトカム主義の限界である。実施担当組織の努力だけではアウトカムを達成できない場合で、アウトカムでアカウンタビリティを問われるのは心外だと実施担当組織が反発するのは当然であろう。先の「政策評価の手法等に関する研究会」（4. ①）が行った各府省ヒアリングでも警察庁、労働省、外務省、厚生省、法務省、通商産業省から反発があった（2000年5月11日第12回会合の資料）。評価でアウトカムを明らかにし、プログラムのアカウンタビリティを問う場面に使いたいときの困難がここにある。

論点4：アカウンタビリティの混乱

アウトカム問題が責任概念に影響して、日本ではアカウンタビリティとレスポンスビリティを混同、あるいはすり替えている状況が頻発している。アメリカでは1970年代に連邦議会が行政機関のアカウンタビリティを向上させるために自らを含めて改革を行ったが、日本ではアカウンタビリティ主体をはじめとする制度設計が弱い。国会が直接政策評価に関わることもない。政策評価法に見られるように、政策担当行政組織が「自ら責任を負う」レスポンスビリティの考えで、アカウンタビリティを達成すると言う混乱が見られる。

日本の政府内で評価によってアカウンタビリティを確実に達成するシステムを作るためには、地方議会や国会の改革、高校での政策評価教育をはじめとするシチズンシップ教育（山谷2017）から着手する必要があるが、それをしないのであればレスポンスビリティの考えで評価システムを運用する方法が实际的である。

論点5：事前評価と事後評価

評価は基本、事後評価である。何が起こったのか、どんな理由で失敗したのか、なぜ失敗するような政策手段を採用したのか、誰がその判断を下したのか、事後評価を使って行う原因究明はデモクラシーの大前提であり、また現実に出ている結果を見ながら行うので技術的には容

易である。しかし日本の行政は事後評価に熱心でない。その理由は多い。1つは責任を追及されるのを厭う人間心理があり、これ乗り越えてアカウンタビリティを追求する制度設計が甘い現実である。過去のことをすぐ忘れる国民性もある。

他方、政策評価に限って言えば、政策評価結果を政策の企画立案にフィードバックする考えが、政策ではなく「予算に反映する」にすり替わり、それを2003年から繰り返し奨励した経済財政諮問会議の行動も、国民と国会の関心を事前評価に誘導している。

ただし事前評価は難しい。技術的に難しいのは前述のPPBSの時からすでに明らかであった。また、将来発生するであろう政策の成果を予測しても、災害の発生や突然の政権交代で、その通りになるかどうかかわからない。将来成果は出ないと誰の目にも明らかでも、政策原案を作った担当者は人事異動や選挙で交替してなくなっている。また市民や有権者も忘れていて。事前に行なった評価と現実との齟齬を放置する悪条件が事前評価には多い。

政策評価を導入するとき基本的視点は、「反省」であった。21世紀を目前にして「この国のかたち」を見直す主張は明らかに事後評価であったが、いつの間にか政策評価は事前評価主体にすり替わってきた。アカウンタビリティ追及を免れたい側と、アカウンタビリティ追求意識が弱い側の双方の思惑が、アカウンタビリティを矮小化し、説明責任に言い換え（国語審議会1998年～2000年）、それが「説明すれば良い責任」になった。原因の一つは事前評価の流行で、結果責任が政策現場で忘れられたからである。

論点6：評価と監査の区別

監査（audit）のマインドで評価（evaluation）を行う弊害は至る所で指摘され、図表3の②④⑥に多く見られる。その遠因は政策評価や自治体評価を導入する際、日本の官公庁で採った行動にある。政策評価の制度を所管する官庁であった総務庁行政監察局「政策評価の手法等に関する委員会」（1999年8月～2000年12月）をはじめとする各省庁の検討委員会や地方自治体の委員会では、公認会計士や経営コンサルタントを委員に加えたからである。また評価

(evaluation) を日常会話の「評価」と誤解した地方自治体が、無駄や冗費の削減だけでなく、違法な公金使用の追究、浪費・不正・非違の摘発、不適正な業務処理のチェック手段として考え、評価システム導入業務を経営コンサルタントに業務委託したところにも原因がある。独立行政法人や研究開発法人では現在も、評価監査部の名称をつけた組織が存在するのは、その名残である。

評価が政策のプログラムとプロジェクトのアウトカムを確認するのに対し、監査の中心的事は組織の適正な業務執行にある。また評価が政策学や評価学、行政学を理論背景にしているのに対し、監査は会計学と法律学である。このように異なる評価と監査は、アカウンタビリティをめぐる分業体制を組んでいるが、しかし上記の理解を持たない評価活動が評価と監査を混同した結果、勘違いの無駄な作業が「評価疲れ」を招き、間違った理論は「評価の経年劣化」を重ねた。

なお、評価に詳しい担当者は政策デザイン、ロジック・モデル、政策システムを使った評価を実践するが、手間と時間がかかる難点と、上司が理解しない問題がある。

論点7：評価の規準と基準

評価基準 (standards) と評価規準 (criteria) の問題である。客観的な評価を標榜しているが、特定の価値を反映した規準を選択する場合の問題である。特定の規範、価値観に導かれたものさしを規準とよび、解釈の幅がない機械的に計算できるものさしを基準と呼ぶ。基準に該当するのは人数、回数、金額であるが、それが客観的かどうかは状況次第である。公務員の数、人件費は新自由主義思想の下では少ない方がよい。そうすると一見客観的な議論をしているような効率基準も、特定の価値観に導かれている規準になる。

ポリシーの目的が評価のものさしに適合しているかどうかはこの議論に重要である。地方県立高校や予備校は国公立大学への進学率を高校の評価で重視するが、甲子園や大相撲を目指す中学生には無意味である。こうした、一般化できないものさしは規準のカテゴリーに含めて語る必要があるが、女性の社会参画、障がい者雇用のように、昔は特定の運動の価値規準として

判断されていたが、いまや社会常識となって基準として語るようになることもある。

自治体改革や行政改革がらみの政策評価においては基準と規準の区別が重要な第一歩になる。これはEBPMの議論に重なり、注意が必要だったが、ここでも勘違いが多かった。

論点8：外部評価と内部評価

外部評価と内部評価の論議は、内部評価がお手盛り評価になる、内部評価は客観性を欠く、だから外部の有識者に評価してもらう、この20数年前の主張が前提であった。もちろん、外部にそうした専門家が少ない、内部者が持つ程度のデータや情報、政策情報を集めて処理するノウハウを外部者は持たないという現実認識が、評価の外部・内部の議論を終わらせた。日本の評価現場ではこの論争は、もう少し実務と学問の協力を待つ必要がある。

こうした中で、評価の客観性を担保するのは難しい。そのため2016年頃から政権与党内で、評価に際して科学的根拠を示し、エビデンスを使うよう求めた意図は理解できる。しかし、社会には立場の違う人が複数存在し、絶対真理はない。科学的エビデンス、同じ数字でも解釈が違うことはよくあり、それを想起すると科学的エビデンスの議論は医療や技術の世界に限定され、政策領域一般に拡大するのは牧歌的すぎる。

現実の政策では、いろいろな意見を忖度しながら政策を決め、実施せざるを得ない。また、その時は正しくても、10年後、20年後に大きな間違いであったことが分かることがある。市町村合併に対する「3.11」後の評価、公務員の人員削減策の新型コロナ後の評価が典型である。この場面では、政策の現実が、評価学に貢献する。

6. 評価専門家の役割

これらの論点を前提に考えると、アカウンタビリティのさらなる再検討が必要になる。

日本では評価を説明ツールとして使い、説明すれば責任を果たしたと見なすため、多くの分野でさまざまな評価が重複して行われるようになった。1人の担当者や1つの組織にあらゆる評価業務が集中することもある（図表5）。評

価値の重複の回避、交通整理をする必要があるが、現場には評価知識と時間的余裕が無い。ここに評価専門家の出番がある。

このような反省すべき事態にもかかわらず「評価は民主主義のリテラシー」と評価の専門家は考えるが、それは評価が2つの長所を持つからである。第1は、評価とは知ること、つまり政策現場で起きている実情を把握し、冗長な説明を整理し、ポイントを絞った重要な情報を提供するからである。整理のポイントは、政策体系と政策デザインの確認、ロジック・モデルによるプロセス追跡で、じつはこうした確認と追跡こそ、アカウンタビリティ追及に評価を使うメリットである。

評価が巧く運用されているODAの分野では、外務省の『開発協力白書』や『ODA評価年次報告書』が毎年刊行されており、また『国際協力機構 年次報告書』、『国際協力機構 事業評価年次報告書』がある。これらの評価文書を通じて外務省やJICAはODAについて情報発信している。

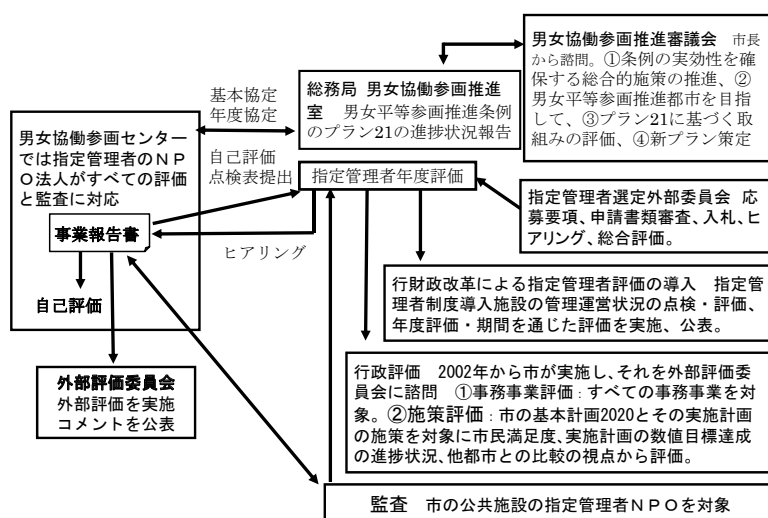
評価の第2の長所は、公共部門のさまざまな「制度」とそれらに付随する仕組みの理解に、評価が貢献することである。たとえば国の政策評価制度と独立行政法人制度、実施庁評価制度を熟読すると、中央官庁の複雑な所掌事務の理解が進む。また地方自治体の行政評価を見ると、

国の府省と地方自治体の財政関係、とくに補助金を通じた中央地方関係が理解できるようになる。そうすると、「地方創生」政策は1995年から続けられてきた地方分権政策に逆行していることが分かる。

もちろん、宇宙政策で言えば国際機関と関係各府省、研究開発法人との複雑な関係は評価を通じて理解できるだけでなく、産業政策、外交政策、防衛政策、研究開発政策、高等教育政策を所管する官庁の権限が交錯し、その要にある内閣府ではゼネラリストが担当している姿も見ることでもある（南島 2020）。行政学の古典的テーマ、スペシャリストとゼネラリストの関係の生きた教材である。

つまり評価を通じて公的部門の実像が把握できるのである。それは研究者にとって絶好の臨床研究になる。こうした評価の長所を生かすのも、評価専門家の仕事である。

この評価専門家の定義は、評価の専門ポストに就けばプロフェッショナル、プロがさらに研鑽すればエキスパートである。プロやエキスパートに必要な「知」は理論研究から得られる専門知（評価に関わる修士や博士の学位）と、実務から得られる経験知（評価担当ポストでの勤務経験）である（山谷 2018）。日本評価学会が認定する上級評価士は、まさにこの専門知と経験知を持つ人を対象に認定する。



図表5 評価マップ Q市の男女協働参画センターの例

出典：内藤・山谷, 2015, P.216 を修正して引用

ただし、この評価専門家の育成について難問がある。日本では体系的な評価の高等教育（大学院レベル）を受けた人は少ない。政策評価と法人評価の実務に関しては総務省が専門的な研修を行っているが、対象は官公庁の評価担当者なので、プロを育てる研修ではない。また、各府省の政策評価担当者に評価の専門的な基礎教育（主に大学院レベルでの行政学と評価学）を受けた人は少なく、On the Job Trainingや独学で必要な知識を身につけた人が多い。この人びとも繰り返される人事異動の中にいるので、「通過者」でしかない。つまり、官公庁では最新の状況を知り得ても、過去の経緯や日本の特殊事情に詳しくないので、その時に流行している評価の意味が分からない（EBPMやPDCAがその代表）。

そうした状況を予期して2000年に日本評価学会が設立され、「評価士養成講座」制度を設け、評価専門家の養成に努めてきた。この講座は大学院の修士レベルを想定し、実務経験を持つ社会人を対象に、データ収集と分析のスキル、評価デザインの方法、中間評価、終了時評価、インパクト評価などの方法論、法人評価などで見られるパフォーマンス測定の考え方、政府全体における評価の位置づけ、国際機関の評価の現状などを講義した後、最終試験を経て評価士の資格を付与している。また各地の大学院でも、このような評価専門家養成の動きを見せている。

ところで評価専門家に期待される役割は何であろう。教育や医療、ODA、科学技術など専門分野に対しては政策学と評価学の知見を持つ‘friendly advisor’、一般市民には‘communicator’や‘facilitator’、社会に啓蒙を行う‘Empower-Mentor’、利害対立の場では仲裁者・調停者（mediator）である。また、政策やプログラムが沈滞した場面では触媒（catalyst）としての役割も期待できる。世間で誤解されるような公開処刑人、告発者、事業のターミネーターではない。

ただし、科学技術の世界、とくに宇宙政策の分野ではこうした評価専門家が少なく、活躍していない。原因の一つは日本評価学会側にある。評価学会では科学技術政策、宇宙政策に対して関心を持つ会員が少なく、学会の報告や機関誌『日本評価研究』でほとんど取り上げられてい

ないからである。こうして科学技術の専門家と評価専門家との意思疎通が少なかったために、評価作業を分野違いの技術者・研究者に任せ、貴重な時間を浪費する事態になる。

7. おわりに

本稿は、評価学の知見と政策現場の経験をふまえて、政策学の「求心力」を回復する方向を模索したいと考えた。具体的な作業は評価課題の論点整理だったが、この整理は評価学の理論研究だけでなく、政策の実践課題を考える参照枠組み（frame of reference）づくりになった。しかし、政策学と評価学には、この枠組み作りを困難にする課題がある。はじめに述べた「評価の大きな課題」はこれである。それを説明し、今後の研究展望にしたい。

大きな課題とは、マクロ政策とミクロ政策、それぞれの評価の不整合である。たとえばマクロ政策として安倍政権は、2014年12月「まち・ひと・しごと創生総合戦略」を掲げ、内閣府に地方創生推進事務局を置き、2020年度まで数多くの施策を進めてきた。他方、ミクロの各施策については、内閣府の自己評価「令和元年度内閣府実施施策に係る評価書」の政策評価シートを見ると、各施策評価結果においては「進捗あり」「目標達成」の文言がある。その実際はどうだったのか。地方都市の人口減少は止まらないどころか、加速化し、10万人規模で住民が減っている県が複数存在する。内閣府各課が自己評価したミクロの施策評価では目標を達していることになるが、マクロでは全く達成できていない。

この政策体系（図表4）に存在する不整合の課題はあまり指摘されない。理由は、政策体系の知識が普及していない（政策学）、政策体系を専門的に見る組織や研究機関が無い（行政学）、政策体系評価が断念されたこと（4.の①で言及）が考えられる。つまり、政策学とその実務では、そもそも「求心力」を考えていないのかも知れない。

また、マクロの視点は存在しても、それを忘れさせる現実がある。たとえば、経済財政諮問会議「財政の質の改善に向けて～実効性のあるPDCAサイクルの構築に向けて～」（2013年3

月8日)がある。この提言の柱は、PDCAサイクルの実効性向上、行政事業レビューの活用(政策評価、会計検査報告、予算執行調査などの事務負担の軽減と基礎情報の統一)、司令塔機能の強化、エビデンスにもとづく政策評価であった。現場の議論がミクロ視点に傾斜する恐れは指摘されたが、それは共有されず、人事異動を繰り返す官僚組織の宿命として、前任者の置き土産としてミクロ視点傾斜は安易に受容され、継続した。

組織上の不整合もある。エグゼクティブ・レベルでの評価が、一般市民を意識しているとは思えない難解なレベルになっていることである。具体例は総合科学・イノベーション会議の「国の研究開発評価に関する大綱的指針」で、専門家が集まって作成した難解なこのマクロ指針は、これに従ってデザインされた研究開発評価を理解困難にする。その一方で、現場のミクロ作業では年度ごとの研究計画策定や研究予算要求の時に大量の文書を作るが、これらを分かりやすくするため、そして内部の予算査定を通りやすくするため、「Excel職人」や「パワポ職人」が登場する。分かりやすいキレイな文書をつくるために本来の研究時間を削り、膨大な時間をかけて資料を作成する。ここに几帳面な管理者が異動してくると、さらに文書が複雑化、緻密化する。こうして評価とその文書はどんどん精緻化し、ボリュームも信じられないほど増加し、一般市民の理解を超える。もちろん資料を作成するのは研究者や技術者なので、本来研究に向けるべき時間が大きく削減される。「アカウンタビリティのジレンマ」状態である。

政策の大きな課題の解決を模索するなら、評価専門家の育成が喫緊の課題になる。そのために行政学・行政の経験知と、評価学・評価の実践知を活用した教育が必要であろう。ただし、それは大学院レベルの教育になるはずで、総合的な学際研究(multi-discipline)、知の交流と共有を目指す学修(inter-discipline)、諸学横断型アプローチ(cross-disciplinary)、現実世界の問題を学問分野を超越して考察する‘trans-discipline’の姿勢が必要になる。

遠心力を求心力に変える研究教育には時間がかかるが、これ自体がまた大きな課題になる。

参考文献

- 加藤芳太郎(2008)『予算論研究の歩み』自治総研叢書25(戦後地方自治の証言Ⅲ)、敬文堂。
- 内藤和美・山谷清志・編著(2015)『男女共同参画政策 行政評価と施設評価』見洋書房。
- 南島和久・編(2020)『JAXAの研究開発と評価』見洋書房。
- 村上祐介(2015)「教育委員会制度改革と教育行政の専門性」『政治主導改革と教育の専門性』日本教育行政学会年報第41号。
- 山谷清志(2012)『政策評価』ミネルヴァ書房。
- 山谷清志(2017)「政策評価とアカウンタビリティ再考―18歳選挙権のインパクト―」、『日本評価研究』第17巻2号。
- 山谷清志(2018)「政策評価における『官』・『学』協働の可能性」『評価クォーターリー』、行政管理研究センター、No.47、1-20。
- 山谷清志(2019)「巻頭言・平成時代と公共政策研究の30年―『遠心力』―」日本公共政策学会『公共政策研究』第19号。
- 山谷清志・監修、日本評価学会(2020)『プログラム評価ハンドブック』見洋書房。
- The Urban Institute(1970), *Federal Evaluation Policy: Analyzing the Effects of Public Programs*, the Urban Institute: Washington DC..

- * 本稿は日本航空宇宙学会の第64回宇宙科学技術連合講演会(2020年10月29日)での報告原稿に加筆修正したものである。
- * ここでの記述は「国立研究開発法人における組織マネジメントと評価のあり方に関する研究」(基盤研究C、2018-2020年度、研究代表者・南島和久)の研究成果から得られた。