

プログラム評価の日本における理論と実際

－中央府省が実施する総合評価－

宗高 有吾

概要

日本の中央府省が行う政策評価制度における「総合評価」は、評価理論において「評価のプロトタイプ」とされるプログラム評価に該当し、政策の効果の把握、問題点の発見、情報提供といった機能が主に期待されている。しかしながら、2002年度からの総合評価の実施状況を見れば、その実施件数は低調であり、近年は減少傾向に転じている。また、実際の評価書を見れば、政策のアウトプットだけの記載にとどまる業績測定と変わらない内容であるため、政策効果の把握方法が不十分であり、それに伴い問題点の発見が不十分となっている例や、十分に情報提供機能が発揮されていない評価が一部で今なお存在する。

このような状態になっている背景・原因には、評価を実施する体制（人員、資金、時間）が不十分であることだけでなく、近年の「評価」の名を冠した業績測定や政策分析の普及や、プログラム評価の手法や概念の認識不足、プログラム評価は外部が行うという風潮などがあげられる。したがって、これらの課題を克服し、プログラム評価としての総合評価を浸透させるためには、プログラムの概念や総合評価の役割を再認識するとともに、業績測定との連携や役割分担について検討すべきである。また、評価を実施すること自体を法令にあらかじめ規定しておくなど、その実施を制度的に組み込んでおくといった方策も考えられる。

はじめに

評価の理論には、プログラム評価、業績測定、政策（プロジェクト）分析という評価の3類型がある。一方で、評価実務の世界においては、単に「評価」の名を冠した業績測定や政策分析の概念の普及が近年著しい。たとえば国の政策評価制度において、業績測定としての目標管理型の政策評価の導入、政策分析としての規制や租税特別措置等に対する評価の導入がこれに該当する。これら自体は間違っているわけではないが、プログラム評価を重視する評価理論の正統派にとっては、いささか違和感が残る普及状況である。それでは、日本国内におけるプログラム評価は今どのような状況に置かれているのであろうか。

国または都道府県あるいは市町村が実施するプログラム評価を扱った事例分析や研究は少ない。たとえば、政策評価制度を所管する評価専任組織¹としての総務省が行っている政策評価の点検においては、2002年度から2008年度までの総合評価（プログラム評価に相当する）の実施状況のうち、全体的な傾向や、参考となる優良事例が一部紹介されている。しかし、近年においては総合評価を対象とした点検自体が行われていないのが現状である。他方、研究面であれば、たとえば山谷はプログラムの概念に着目し、外務省、国土交通省及び防衛省における総合評価の枠組みについて事例を示しつつ論じている²。また、鎌田は2002年度から2007年

¹ 本稿では、政策評価制度を所管する立場としての総務省を政策評価に関する基本方針に則り「評価専任組織としての総務省」と表記する。一方で、一行政機関として内部評価（ないし自己評価）を実施する組織としての総務省は単に「総務省」と表記して区別する。

² 山谷, 2012年, 213, 223 ページ。

度に各行政機関³が実施した事後評価の政策への反映状況を分析し、そのなかで総合評価についても触れ、総合評価による政策の見直し率が比較的高いことを指摘している⁴。

このように国の行政機関における事例は優良事例を中心に一部整理され、いくつかの切り口で分析されているが、2009年度以降の総合評価の実施状況を中心として、プログラム評価の理論面や、総合評価としての制度導入時の意図をどれほど充足しているか、近年の評価を取り巻く状況等を考慮した分析はなされていない。一方で、都道府県あるいは市町村における評価については、『評価クォーター』といった評価の専門誌で事例が多く取り上げられてきたものの、たとえば小野が指摘するように、都道府県・市町村における「評価」の多くは業績測定 の範疇に含まれる⁵ため、事例分析の対象になりにくい。

したがって、本稿は国の行政機関においてプログラム評価に相当する総合評価を事例として取り上げ、その内容や手法に着目しつつ実態を整理し、プログラム評価の理論や制度導入時の想定と違った内容となっている事例があることを示すとともに、その背景や原因、今後の日本におけるプログラム評価を促進するための方策や検討課題について考察する。

1. プログラム評価と総合評価

本章では本稿で取り上げるプログラム評価とはどのような評価のことを指すのか、関連概念を整理しつつ定義する。また、日本においてプログラム評価に相当する総合評価はどのような議論を経て定められたかを整理し、総合評価の実際について考察する際の検討材料とする。

1.1 プログラム評価と関連概念

アメリカで単に「評価 (evaluation)」という場合、それはプログラム評価 (program evaluation) を指していることが多く、プログラム評価は評価のプロトタイプとされている⁶。プログラム評価のテキストとして代表的なロッシとリプセイとフリーマンらの著書、*Evaluation : A Systematic Approach* では、プログラム評価を「社会的介入プログラムの効果性をシステマティックに検討するために、プログラムを取り巻く政治的・組織的環境に適合し、かつ社会状況を改善するために社会活動に有益な知識を提供する方法で社会調査法を利用すること」と定義しており⁷、具体的な手法として、プログラムが改善しようとする社会状況及びプログラムのニーズに関する「ニーズ評価」、プログラムの概念化とデザインに関する「セオリー評価」、プログラムの運営、実施、サービスの提供に関する「プロセス評価」、プログラムのアウトカム (成果) やインパクト (効果) に関する「アウトカム・インパクト評価」、プログラムの費用や費用対効果に関する「効率性評価」の5つが示されている⁸。また、GAO (Government Accountability Office) もほぼ同様に、プロセス (実施) の評価、アウトカム評価、インパクト評価、費用便益・費用対効果分析の4つをプログラム評価として定義している⁹。さらに、山谷はプログラムの概念に着目し、「プログラムとは制度や仕組み、プロジェクトを動かすマニュアル、アイデア」であり、「プログラム評価とは、そうしたアイデアやノウハウを対象とする評価であり、官庁のジャーゴンでいえば政策をうまく『回す』方法の評価」と定義している¹⁰。

プログラム評価は広義の意味での評価の類型の1つであり、これに業績測定 (performance measurement) と政策分析 (policy analysis) を

³ 一般的に、「省庁」や「府省」などと呼ばれているものは、政策評価法において「行政機関」と表記されているため、本稿では「行政機関」に統一する。

⁴ 鎌田, 2008年。

⁵ 小野, 2002年, 30ページ。

⁶ 田中, 2010年, 1ページ。

⁷ Rossi Peter H, Lipsey Mark W, Freeman Howard E, (2004), p.16.

⁸ Rossi Peter H, Lipsey Mark W, Freeman Howard E, (2004), pp.52-61.

⁹ GAO, (2011).

¹⁰ 山谷, 2004年, 12ページ。

加えたものが評価の3類型であると指摘されている¹¹。業績測定は政策のアウトプットやアウトカムの測定・報告に重点が置いている。日本における実績評価がこの業績測定に該当し、アメリカにおけるGPRA (Government Performance and Result Act of 1993)や、NPM (New Public Management)の影響を受けている。また、政策分析は軍事上の作戦研究に起源を持つOR (Operations Research) やシステム分析が発展した政策科学 (Policy Sciences) の方法として成立したものであり、費用便益分析・費用対効果分析によって政策代替案の効果を数値化し、比較することで、政策選択に結びつけることを目指す。

これら3つの評価は、それぞれの学問的バックグラウンドが異なるために、これまで相互に連携が意識されていなかったが、互いに重なり合う部分があるため、相互に連携させて考える必要性についても指摘されている¹²。たとえば、プログラム評価と業績測定は互いに相反するものではなく補い合うものであり、業績測定における測定指標のモニタリングにおいて不十分な部分が発見されれば、プログラム評価においてその不十分な点を重点的に分析・検証することが想定されている¹³。一方で、プログラム評価と政策分析については、その役割分担について、政策代替案 (alternative) の提言を行うか否かに違いを見出して整理している場合もあり¹⁴、3つの評価の関係については議論の余地が残されている。

このようにプログラム評価には、業績測定や政策分析の手法が一部内在しているが、その役割や機能については業績測定や政策分析と異なる面があるとも分かる。そこで、本稿ではプログラム評価とは「方法論において業績測定と政策分析の要素を採用することはあるが、その主な機能・役割はプログラムの効果の検証と、それに伴う有用な情報提供である」と定義する。

1.2 プログラム評価としての総合評価

日本の国の行政機関におけるプログラム評価として位置づけられる評価は総合評価であり、これは「政策評価に関する基本方針 (2005年12月16日閣議決定)」で示される評価の3方式 (総合評価、実績評価、事業評価) の1つである。また、この基本方針は、総合評価 (方式) とは「政策の決定から一定期間を経過した後を中心に、問題点の解決に資する多様な情報を提供することにより政策の見直しや改善に資する見地から、特定のテーマについて、当該テーマに係る政策効果の発現状況を様々な角度から掘り下げて分析し、政策に係る問題点を把握するとともにその原因を分析するなど総合的に評価する方式」であると定義している。

総合評価の具体的な内容 (留意点) については、「政策評価の実施に関するガイドライン (2005年12月16日政策評価各府省連絡会議了承¹⁵)」 (以下、「実施ガイドライン」という) に記載されている (詳しくは表1を参照)。ただし、実施ガイドラインの前のガイドラインである「政策評価に関する標準的ガイドライン (2001年1月15日政策評価各府省連絡会議了承)」は、「評価のテーマや評価対象の性質等によって評価の内容は一定のものではなく、必ずしもすべての内容が該当するものではない」とも定めている。

なお、実施ガイドラインの定義とロッシとリブセイとフリーマンのプログラム評価の定義を比較すれば、実施ガイドラインは表1の第1の定義により、ニーズ評価、セオリー評価、プロセス評価、アウトカム・インパクト評価を、また、第4の定義により効率性評価の要素を含んでおり、とくに政策効果の解明という点に力点を置いていると分かる。また、実施ガイドラインの第2の定義においてはプログラムにおける問題点の発見にも重点を置いていると分かる。

¹¹ 梅田・小野・中泉, 2004年, 147-152ページ。

¹² 梅田・小野・中泉, 2004年, 150ページ。

¹³ GAO, (2011); 佐々木, 2010年, 93-94ページ。

¹⁴ May and Paul, (1999)。

¹⁵ 政策評価各府省連絡会議とは、総務省が開催する会議であり、「政策評価の質の向上を図る観点から、各行政機関間の連絡を密にし、政策評価制度の円滑かつ効率的な実施を図るとともに政策評価に関する取組を促進するため、各行政機関により構成される連絡会議」を指す (政策評価に関する基本方針Ⅲ-1)。

表 1 政策評価の実施に関するガイドラインが定める「総合評価」の具体的内容

1	評価対象政策の効果の発現状況を様々な角度から具体的に明らかにし、その際、政策の直接的効果や、因果関係、場合によっては、外部要因の影響についても掘り下げた分析を行い、さらに、必要に応じ波及効果（副次的効果）の発生状況及びその発生のプロセスなどについても分析する。
2	評価対象政策に係る問題点を把握し、その原因を分析する。
3	評価対象政策の目的が依然として妥当性を有しているかについて検討する。また、必要に応じて、行政関与の在り方からみて行政が担う必要があるかなどについて検討する。
4	必要に応じて、政策の効果とそのために必要な費用（マイナスの効果や間接費用を含む。）を比較・検討する。また、国民にとってより効率的で質の高い代替案はないかについて検討する。
5	関連する政策との間で整合性が確保されているかについて検討する。
6	場合によっては、他の政策よりも優先的に実施する必要があるかについても検討する。
7	評価の対象については、例えば、政策自体の在り方にもかかわるテーマを評価する場合には、政策（狭義）から施策、必要に応じて事務事業までを評価し、政策（狭義）を実現する具体的手段である施策に焦点を当てたテーマを設定して評価を行う場合には、施策を中心として必要に応じて事務事業までを評価するなど、設定した評価のテーマに応じて柔軟にとらえる。なお、分野横断的なテーマを設定して複数の施策を対象として評価を実施する場合も考えられる。
8	制度の改正など政策の大幅な見直しの検討を行う場合には総合評価方式を用いた評価を実施することを検討する。
9	総合評価方式を用いた評価と事業評価方式を用いた評価との関係については、例えば、総合評価方式を用いた評価において、事業評価方式を用いた事前の評価結果、事後の検証・評価結果を活用したり、総合評価方式を用いた評価の中で、事業評価方式を用いた評価の事前の評価、事後の検証・評価を実施することも考えられる。
10	総合評価方式を用いた評価は、政策の実施前の時点においても、必要に応じて評価を行うことがあり得る。その際、類似の総合評価方式を用いた評価の結果を参考にしたり、他の評価の方式の評価により蓄積された情報・データを用いる。
11	審議会等の答申や白書等による分析結果を評価に積極的に活用することも検討する。

出典：政策評価の実施に関するガイドラインから抜粋して筆者作成。

それでは、総合評価はどのような議論を経て導入に至ったのだろうか。1999年から2000年にかけて、政策評価に関する基本的な考え方の整理や政策評価の手法等の研究等を行うために立ち上げられた「政策評価の手法等に関する研究会¹⁶⁾」（以下、「研究会」という）において、「総合評価」という言葉は初めて登場した¹⁷⁾。なお、研究会ではアメリカのGAOによるプログラム評価を総合評価の手法として想定していた。また、現行の「行政機関が行う政策の評価に関する法律（2001年法律第86号）」（以下、「政策評価法」という）や政令、省令、ガイドライン等において、「プログラム評価」という言葉は直接登場しないものの、研究会の「政策評価制度の在り方に関する最終報告」（以下「最終報告」

という）においては、「総合評価の導入及び実施の在り方」に関する部分で用いられていたり、研究会の中間報告では総合評価方式の前の名称案であった「政策体系評価」の手法としてプログラム評価が例示されていたことから、プログラム評価として総合評価を位置づけていたことが分かる。また、最終報告によれば総合評価を導入した意図は次のとおりである（下線は筆者による）。

行政が国民のニーズや社会経済情勢に的確に対応するためには、政策・施策の効果を具体的に明らかにするとともに、行政として対応を求められる問題点やその原因などを分析し、その解決に資する情報を提供することにより、的確な改善・見

¹⁶⁾ 「中央省庁等改革の推進に関する方針」（1999年4月27日中央省庁等改革推進本部決定）等に基づき実施することとされている政策評価の具体化に向け、政策評価に関する基本的な考え方の整理や政策評価の手法等の研究等を行うために開催された。

¹⁷⁾ 当初は「政策体系評価」と呼ばれていたが、2000年6月27日の第15回会合で、総合評価に名前を変えた。「名称については、評価対象や評価の手法を説明するものではない、施策や事務事業に関する説明は理解が難しいなどさまざまな議論があり、その結果、最終的に名称については村松岐夫座長に一任し、変更が認められた経緯がある」（山谷、2012年、202-203ページ）。

直しにつなげていくことが必要である。特に、これまでの取組を見直し、新たな政策展開を行おうとする際には、このような評価が求められる。

総合評価は、このような要請にこたえるために必要な情報を提供するものであり、政策・施策の効果を明らかにしたり、問題点を解決することに資するものである。

このような記述を踏まえれば、当時から総合評価に求められていた性質は端的に、①：プログラムの効果の検証・分析、②：プログラムにおける問題点の発見、③：①または②によって得られた事実からの情報提供の3点に要約できる。なお、この趣旨は実施ガイドラインにも反映されている。

しかしながら、政策効果の検証については難しい領域であることが当時から指摘されていた。研究会の最終報告は当時のアメリカのGAOを引き合いに次のように示している。すなわち、

米国会計検査院(GAO:General Accounting Office)によるプログラム評価においては、行政活動の問題点について指摘する場合、評価のための限られた情報・データや資源の下で、比較的簡易な評価手法を用いて評価を行っている例や因果関係の分析にまで至らない例なども多いと指摘されている。

このようなことから、総合評価については、評価の目的、評価対象の性質等に応じ、適用可能で合理的な評価手法を選択し、組み合わせるとともに、信頼できる情報・データに基づいて評価を行うことが重要であり、必ずしも高度かつ厳密な評価手法を適用することを求めるものではない。また、性急に精緻な評価結果を求めるのではなく、上記のような方策を講じつつ、具体的な評価の実践の中で徐々にその質の向上を図っていくことが重要である。

このように、最終報告はGAOにおいてもプログラム評価が厳密に実施されてはいないことを指摘しつつ、必ずしも高度かつ厳密な評価手法を適用することを求めるものではないと示し

ている。

2. 総合評価の実際

国の行政機関が実施している総合評価の実施状況を把握するため、本章ではまず各行政機関が実施してきた総合評価の実施件数を整理し、その推移を確認する。次に、プログラム評価の理論や政策評価制度導入時の議論で考えられていた、政策効果の把握、問題点の発見、情報提供という総合評価に期待される3つの役割に着目しつつ、その実態を整理する。

2.1 総合評価の実施件数

国の行政機関においてどれほどの総合評価が実施されているのだろうか。総務省が毎年度公表している「政策評価等の実施状況及びこれらの結果の政策への反映状況に関する報告」によれば、2002年度から2013年度における各行政機関の総合評価の実施件数の推移は表2のとおりである。これを見れば、総合評価を実施している行政機関には偏りがあることが分かる。たとえば表2に計上していない宮内庁、特定個人情報保護委員会、消費者庁、復興庁、公害等調整委員会、経済産業省、環境省、原子力規制委員会の8行政機関は総合評価を実施していない。この理由については各行政機関の事情によるため、詳しくは不明であるが、近年に設置され比較的新しい行政機関や、総合評価以外の評価の実施件数も少ない機関が多い。また、総合評価を実施している機関は制度導入時の2002年から数年の間に総合評価に着手しているため、これらの行政機関を中心に総合評価が定着していると考えられる。

また、全総合評価の実施件数は減少傾向にある。件数の多くを占める外務省においては、2002年度の118件から2013年の19件まで減少しているが、外務省の場合は評価対象の重点化が実施された経緯があり、これが件数減少の主な理由となっている¹⁸。ただし、外務省によ

¹⁸ たとえば、評価専任組織としての総務省による政策評価の2004年度点検では、外務省の評価の枠組みや重点化について提言している。

る評価は、その内容から総合評価よりも実績評価としても分類されることもある¹⁹。このため、すべての総合評価の実施件数から外務省の総合評価を除けば、2008年度の件数をピークとし、2013年度にはその半数以下となっていることが分かる。この原因については後に考察するが、いずれにしても、総合評価の実施件数は制度導入時から低調な状態が続いている。

なお、行政機関ごとに評価1件の数え方や、評価1件当たりの分量が異なるため、各行政機関における取組み度合いについて、件数だけで単純に比較することは困難である。また、表2の最終行に記載した「全評価実施件数」は各年

度におけるすべての評価の実施件数の総和である。この数を総合評価の実施件数と比較すれば、各年度の全評価に占める総合評価の実施件数の割合は約0.3%～約3%であり、年度ごとにはばらつきがあることが分かる。ただし、全評価の実施件数の減少や増加の多くには、公共事業に係る評価が占めるため、この割合の変化を考慮することは参考にならない。たとえば、2006年度における件数の減少には農林水産省の公共事業評価の重点化が関係しており、2008年度における件数の増加は国土交通省において、再評価を行うべき公共事業評価の件数の増加したことによる。

表2 総合評価方式を用いた評価の実施件数

(単位: 件)

年度 行政機関名	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
内閣府		1	1	2			3					
公正取引委員会	1	2		4	2	9	3	4	7			
国家公安委員会 ・警察庁			1	1	1	1	1	1	1	1		1
金融庁				1			1					
総務省		1		1		1	9	8	7			
法務省		1			1			4	4	2	2	1
外務省	118		108	62	70	73	24	24	24	24	20	19
財務省		1		1		2		1				
文部科学省	2			1		1	2	1				
厚生労働省		3	5	7	7	10	3	1	3	9	6	7
農林水産省		1	5	2	1			1		1		2
国土交通省	11	8	5	9	3	5	4	5	9	9	8	3
防衛省	16	11	12	13	8	12	19	15	3		1	
合計	148	29	137	104	93	114	69	65	58	46	37	33
全評価の実施件数	10,930	11,177	9,428	9,796	3,940	3,709	7,088	2,645	2,922	2,748	2,631	2,559

出典:「政策評価等の実施状況及びこれらの結果の政策への反映状況に関する報告」の2002年度～2013年度分を基に筆者作成。

(注1) 総合評価方式を用いた評価を実施していない行政機関は表から除外している。

(注2) 行政機関名は2013年度時点のものである。

(注3) 「全評価の実施件数」とは、総合評価以外の評価も含めた当該年度全評価の実施件数である。

¹⁹ 外務省の総合評価は実績評価の側面も兼ね備えているものであるため、評価専担組織としての総務省による政策評価の点検においては、総合評価の点検対象からは除外されている。一方で、総務省が取りまとめている「政策評価等の実施状況等の国会報告、政策評価結果の予算要求等への反映状況」では、評価の実施件数は総合評価に分類されている。その理由について、総務省は報告書の中で次の段落のように説明している。なお、目標管理型の政策評価の導入後、外務省の総合評価は目標管理型の政策評価に分類されている。また、外務省においては他の行政機関において実績評価としているものが総合評価として計上されているため、他の行政機関と比べて総合評価の実施件数が多くなっている。

外務省の基本計画においては、「実績評価方式の手法を踏まえつつ、外交政策の特性を勘案し、総合評価方式の手法を取り入れた評価」としている。しかし、評価法第19条に基づく政策評価等の実施状況及びこれらの結果の政策への反映状況についての報告において、総合評価方式に分類されているため、「総合評価方式」として整理している。また、「総合評価方式」を用いて主要な行政目的に係る政策を毎年度網羅的に評価しようとしており、特定のテーマを設定して評価が行われている他の府省とは異なる枠組みの下で「総合評価方式」により評価が行われている。

2.2 プログラムの効果の把握と問題点の発見

前述のとおり、研究会の最終報告では、プログラム評価による厳密な効果の測定が難しいことを理由とし、簡素な評価手法を用いることも想定している。実際に、制度導入初期に評価専任組織としての総務省が実施した政策評価の点検結果では、政策効果の把握を行っており、なおかつ政策効果を定量化している場合は少ないとまとめている。しかしながら、近年の各行政機関の評価書を横並びで比較したところ、具体的なプログラムの効果ではなく、アウトプット・レベルでの記載による進捗状況の管理を行っていたり、業績測定の範囲からほとんど変わらないと思われる評価も依然として存在した。もちろん、業績測定という手法自体は否定されるべきではないが、プログラムの効果について記載されていないければ、プログラム評価としては不十分である。

たとえば法務省は2013年度の総合評価で、「社会経済情勢に対応した基本法制の整備」について評価している。この評価では、民事及び刑事に係る基本法制の整備について、それらもたらす「効果を分析して、必要かつ十分な法制の整備が行われているかを評価する」と説明している。しかし、この評価書は総合評価にもかかわらず、例年、法令の国会での成立状況を測定指標として示しており、いまだに成立していない法案がある点に言及しつつ、引き続き推進しなければならないと結んでいる。また、この評価書はプログラムの中身について具体的に触れておらず、さらに、審議会で検討中の内容に資するために評価を実施すると説明しているが、プログラム評価というよりも進捗状況の説明資料にとどまっている。

厚生労働省においても同様に、アウトプット・レベルでの記載が目立つ評価書が多い。また、単純に評価書のページ数において、その内容の充実さを判断することは不適切かもしれないが、1評価書当たり2～3ページ程度の記載となっているものもある。たとえば、厚生労働省は2013年度には「基本目標 XIII 国民に信

頼されるとともに、職員一人一人がやりがいをもって働くことができる体制を確立すること」という政策目標に関連して、6件の総合評価を実施しているが、このうち「施策大目標1 情報発信、情報公開と効率的・効果的な業務運営を図ること」のうち、「1-2 コスト意識・ムダ削減を徹底すること」の評価については、行政事業レビューを実施したことについてのみ記載しており、その評価書は2ページにとどまる。なお、この事例における評価対象はプログラムではなく、政策体系上に位置づけられる各項目（アウトプット・レベルの目標）であり、プログラムの概念を適切に捉えていなかったために、アウトプット・レベルの記載になったと考えられる。

なお、総合評価をプログラム評価の理論面と比較するのであれば、たとえばセオリー評価がプログラムの内容の記載に当たり、ロジック・モデルと呼ばれるフローチャート図によって、プログラムの効果の発現過程を「インプット→アウトプット→アウトカム→インパクト」の流れで外部要因も含めて示すことを求めているが²⁰、この例は多くない。また、アウトカム・インパクト評価や効率性の評価については、記載の程度に差はあるが、前述のとおりプログラムの効果に着目していないものもあるため、多くはない。

次に、プログラムにおける問題点の発見という視点で見れば、総合評価はどのように行われているのだろうか。前述のとおり、とくにアウトプットにのみ着目した評価書では、問題点の発見や、問題に対する具体的な方策まで指摘しているものは少ない²¹。一方で、問題点を発見することで、後の制度の見直しにつなげている例もある。たとえば、国土交通省が2003年度に実施した総合評価、「都市における緑地の保全・創出—都市緑地保全法等による施策展開の検証—」では、都市緑地保全法（1973年9月1日法律第72号）を根拠とする政策・施策について、制度導入時と現在の社会状況の比較や、制度の充実の余地などについて、市町村や民間企業に対してアンケートを実施し、具体的な問題点の把握を行い、その結果を列記している。

²⁰ Rossi Peter H, Lipsey Mark W, Freeman Howard E, (2004), pp.139-145; 佐々木, 2003年, 28-36 ページ。

²¹ この他、プログラムの問題点の発見や原因分析等の少なさについては、総務省行政評価局『政策評価の点検結果』（2008年度分）においても指摘されている。

そして、当該評価の実施後、緑の基本計画制度の拡充、緑地保全地域制度の創設、緑化地域制度の創設などを内容とする「都市緑地保全法等の一部を改正する法律案」を国会に提出している²²。

2.3 評価による情報の提供

総合評価の結果、有用な情報の提供はどの程度なされているのだろうか。この疑問に答えるには、評価結果を実際に利用する者や状況等を個々に想定する必要があるため、評価書に記載の内容だけでは一概に判断しがたい。たとえば、行政機関の担当者にとって有用な情報提供とは、実際の政策改善に資する具体的な情報である。一方で、一般国民のための評価であれば、それは情報公表や評価の内容の分かりやすさに

重点が置かれるべきであろう。この点は、評価は万能薬ではないことや、誰のための評価かといった論点にも通じる部分でもある。さらに、評価書に記載された情報が利用者にとって本当に有用か（有用だったか）否かは、その情報の内容や情報提供の時期、評価結果を利用する者の知識量にも左右されるだろう。そこで、本稿では実績評価で不十分とされた部分について、総合評価において詳しく調べているか、また、必要なタイミングで総合評価が適切に実施されているかという外形的な2点に絞って分析を行いたい。

前述のように総合評価が業績測定そのものであると考えられる例もあったが、日本で業績測定に相当する実績評価は総合評価においてどのように活用されているのだろうか。2010年度から2013年度の3年間に総合評価を実施した

表3 近年の各行政機関が実施した総合評価と実績評価との対応関係

行政機関名	総合評価	実績評価との対応
国家公安委員会・警察庁	・サイバー空間の脅威に対する総合対策の推進（2013）	あり（測定指標を設定）
	・振り込め詐欺対策の推進（2011）	あり（測定指標を設定）
法務省	・社会情勢に対応した基本法制の整備（2013）	なし（前年度も総合評価）
	・社会情勢に対応した基本法制の整備（2012）	／
	・入出国の適正な管理（2012）	あり（総合評価と実績評価のいずれかで毎年度実施）
外務省	（総合評価が実績評価を兼ねている）	—
厚生労働省	・施策目標1 今後の医療需要に見合った医療従事者の確保を図ること（2013）	あり（同一テーマ）
	・1-1 国民に伝わるように分かりやすく情報を発信するとともに、「国民の声」に耳を傾け、改善へ活かすこと（2013）	なし（前年度も総合評価）
	・1-2 コスト意識・ムダ削減を徹底すること（2013）	なし（前年度も総合評価）
農林水産省	・総合的な食料安全保障の確立（2013）	なし
	・農林水産分野の地球環境対策（2013）	なし
国土交通省	・不動産投資市場の条件整備（2013）	あり（測定指標を設定）
	・人口減少や少子高齢化の進展と乗合バスのネットワークやサービスの確保・維持・改善（2013）	あり（測定指標を設定）
	・地理空間情報の整備、提供、活用（2013）	あり（測定指標を設定）
防衛省	・東日本大震災への対応（2011）	なし

出典：2011年度～2013年度の3年間に実施された総合評価のうち、最大3件を抽出して筆者作成。

（注）総合評価の列のうち（ ）内に記載の数値は、評価が実施された年度である。

²² 総務省『政策評価等の実施状況及びこれらの結果の政策への反映状況に関する報告』（2003年度分）、431ページ。

行政機関を抽出し、総合評価と実績評価の実施状況と対応関係を整理した結果が表3である。

これを見れば、総合評価において取り上げられているテーマで、実績評価での業績測定結果と関連していると考えられる事例は、国家公安委員会・警察庁や国土交通省など一部の行政機関にとどまっていることが分かる。一方で、前述の法務省や厚生労働省の事例では、同一のテーマについて毎年度、総合評価が実施されている分野もあるが、これらの評価は業績測定に近い。また、農林水産省のように総合評価を実施するため、当該評価対象の分野については数年間評価を行わないと整理している場合もある。もちろん、総合評価に対応した実績評価が必ずしも実施されている必要はなく、防衛省の「東日本大震災への対応」のように急遽トピックとなるテーマもあるため、そのつど適切に評価対象分野の設定が必要である。

次に、総合評価が実施される時期や理由については、どのような傾向があるのだろうか。評価が実施される時期の1つは、各種計画の終了時期が迫っていたために、実施される場合である。たとえば、農林水産省が2014年度に実施した総合評価、「農林水産分野の研究開発（農林水産研究の重点目標の進捗状況の検証）」では、2010年3月に策定された「農林水産研究基本計画」の見直しに資するための実施された（当該計画は5年おきに見直しを行うものである）。

また、法令に見直しや評価を行うことがあらかじめ規定されていることを理由に、総合評価を実施している場合もある。たとえば、国土交通省が2012年度に実施した総合評価、「国土計画（全国計画）」では、国土形成計画法（1950年5月26日法律第205号）第7条が、計画の公表の日から2年を経過した日以後に評価を実施する旨を規定していたため、評価が実施された。

さらに、見直し時期ではないものの、必要に応じて評価が実施されている分野もある。たとえば、経済財政諮問会議が政策評価の重要対象分野として提示したものがこれに該当し、財務

省が2009年度に実施した「重要対象分野に関する評価書—地震保険—（総合評価書）」や、厚生労働省が2008年度に実施した「若年者雇用対策」、「少子化社会対策に関連する仕事と生活の調和の実現に向けた取組」、「子育て支援サービス」などに係る評価がある²³。このような事例はいずれも各行政機関の外部からの評価の要請によって行われたものである。

以上のような経緯で評価が実施された場合は、その評価手法にもよるが、とくに各種計画や法令の見直しのために実施される評価や、外部からの要請に基づいて行われる評価は、評価の目的や利用方法が比較的明確であるため、プログラム評価実施による情報提供が一定程度はなされていると考えられる。

一方で、評価対象とするプログラムの内容によって、毎年度総合評価を行ったり、総合評価と実績評価を交互に実施している「ルーチン型」も存在する。この場合、政策体系上の項目単位で実績評価か総合評価のいずれかが毎年度行われており、総務省²⁴、外務省、厚生労働省、法務省の評価がこれに該当する。この場合、業績測定としてのモニタリングや、定性的な記載という意味合いが強くなり、詳細なプログラムの分析がなければ、プログラム評価としての情報提供という意味は相対的に薄くなる。もちろん、毎年度、総合評価を実施すること自体はただちに否定されるべきではないが、総合評価には時間や労力がかかることが当初から想定されており、安易なルーチンワークとなってしまうと、プログラムの効果の把握、問題点の発見、情報提供といったプログラム評価が想定している内容を実施することは難しいかもしれない。

3. プログラム評価が十分に浸透しない背景・原因

以上のように、国における総合評価には、制度導入当初の意図やプログラム評価の理論面からみても不十分な点が残っている。また、日本における総合評価の実施件数も、近年減少の一

²³ なお、政策評価の重要対象分野となった厚生労働省の3つの評価は、先に指摘した総合評価の実例と比して、記載量が充実している。

²⁴ ただし、総務省が総合評価のルーチン型を採用していたのは、2008年度から2010年度に限り、2007年度までに総務省が実施した評価では、そのつど評価対象のテーマが設定されている。

途をたどっている。それでは、このような状況が生じている背景・原因として何が考えられるのだろうか。また、その背景・原因を踏まえた上で、国レベルでのプログラム評価の浸透に向けて、今後どのような方策をとり、また、どのような事項について検討していく必要があるのだろうか。

3.1 プログラムの概念とプログラム評価・業績測定・政策分析の関係

まず、プログラム評価の「プログラム」の概念については共通認識が十分にできているとは言いがたい。山谷は日本においてはプログラムの概念への理解が浸透しておらず、たとえば行政機関が道路課、住宅課、のようにプログラムよりも下位レベルのプロジェクト単位で縦割り化し、渋滞問題では道路課といったように、プロジェクトとプログラムが密接不可分に組織の中に組み込まれ、日本の行政関係者はそれに疑問をもってこなかったと指摘している²⁵。また、プログラムは「活動そのものではなく活動を導くアイデアやソフトウェアに近い」が、都道府県・市町村では「プログラム＝施策」と解釈したため、違った意味で使われる問題が生じたが、国においても同様の問題が発生しており、「政策」「施策」「事業」の意味や、プログラムの定義について、公式的な統一見解は存在しないことも指摘している。このような状況は、プログラム評価ひいては総合評価の位置づけを難しくしており、混乱が生じている可能性がある。

実際に、総務省（大臣官房政策評価広報課）が2007年度に実施した（内部評価としての）総合評価、「総務省が実施した政策評価の取組についての検証」では、総務省が実施した政策評価の実施状況自体を評価対象とし、省内部の職員に対してアンケートを実施し、評価に際しての問題点や職員の認識を調査している。この

結果、「総合評価方式をどういう政策に、どのような場合について採用すべきかの基準や考え方について共通理解がなかった」実態が明らかになっている。また、この評価書では、業績測定では測定指標の設定が難しい分野があることを理由とし、柔軟に総合評価を活用することが望ましいとしている。この評価結果を受けてか、2007年度までの総務省における総合評価の実施件数は3件であったのに対し、2008年度から2010年度の3年間で24件もの総合評価が実施されるに至った。しかしながら、以前に実績評価で行っていた内容を総合評価にかえて毎年度実施しているにとどまる²⁶。

次に、プログラムの概念への認識不足に加え、業績測定と政策分析の普及、プログラム評価に対するチェックが行われなくなったという背景も考えられる。たとえば、山谷は1996年から2002年までの国の政策評価のレビューにおいて、政策評価制度における「予期しなかった展開」として、政策評価を業績測定と同一視する傾向について既に指摘している²⁷。くわえて、近年に導入された目標管理型の政策評価²⁸に関連して、「政治家が熱心なマネジメント・サイクルでの活用方針が、いつの間にか評価をマネジメント・ツールと言い換えさせ、評価と測定が同義になっていった」と指摘している²⁹。さらに、目標管理型の政策評価では、施策レベルを評価対象と位置づけ、事前に設定した測定指標のモニタリングを毎年度行うものの、数年おきに測定値に関して「評価」を行い、測定値が基準に満たなければその原因分析を行うことを求めている。このような考え方は、業績測定とプログラム評価の関係性に類似する部分でもあるため、プログラム評価と業績測定の境界線がより曖昧になってきている。

また、2008年度以降の総合評価の実施件数が減少している時期の時代背景を見れば、たとえば民主党政権時の「予算編成の在り方」（2009

²⁵ 山谷, 2012年, 206-208ページ。

²⁶ なお、この評価の中で実施された「官房政策評価広報課の政策評価担当職員であった者からの意見聴取結果」では、実績評価において目標値の設定が困難な場合は、総合評価を5年に1度活用してはどうかという意見がある一方で、「【実績評価】が数値化を必須とするものであれば（やむをえず）総合評価方式を採用するだけのことにすぎないのではないか」という意見も出ており、意見も割れていた。

²⁷ 山谷, 2002年, 7-12ページ。

²⁸ 目標管理型の政策評価とは、「実績評価方式を用いた政策評価及びあらかじめ設定された目標の達成度合いについて評価する内容を含む、いわゆる『施策』レベルの政策の事後評価」とガイドラインは定義している。（「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日政策評価各府省連絡会議了承）」）

²⁹ 山谷, 2013年, 1-2ページ。

年10月23日閣議決定)で示された、政策の達成目標をあらかじめ示し、その成果(アウトカム)を評価する「政策目標達成明示制度」の導入への動きや、その後の目標管理型の政策評価の導入という、業績測定を強化する一連の動きがあった。なお、金本は「わが国では達成度の指標をどうやって客観的に作り出すことができるかという議論がなされることが多いが、GAOではこのような指標もほとんど用いられていない。GAOの基本的なアプローチは費用便益比や目的達成度などの指標を提示するのではなく、調査結果から客観的・論理的に導かれる結論や分析を報告し、政策決定者にとって有益な情報を提供することである³⁰⁾と1990年に主張していたが、ここで指摘されていた日本の状況は今もさほど変わっていないのではないだろうか。

ところで、業績測定だけでなく政策分析の拡大も、プログラム評価が浸透しない理由に、少なからず影響していると考えられる。日本では政策評価法施行令の改正によって、政策分析としての規制や租税特別措置等についての評価の実施が新たに義務付けられたが、これも総合評価の実施件数が減少し始めた2008年以降の出来事である。さらに、評価専門組織としての総務省による各行政機関に対する政策評価の点検についても、2010年度からは規制と租税特別措置等の評価と、従来から実施されてきた公共事業の評価に重点化されている。また、2014年度からは目標管理型の政策評価も加えて4分野が点検対象となったが、総合評価は2009年度から点検の対象から除外されているため、総合評価への注目度は相対的に低くなり、実施状況の把握もなされていない状況となっている。

このような流れは国の政策評価に関する各種ガイドラインの制定状況からも分かる。国の政策評価に関する制度の体系は、政策評価法、政令、省令に加え、基本方針、実施ガイドライン

等で構成されているが、このうち、情報の公表、規制の評価、租税特別措置等の評価、目標管理型の政策評価については、別途ガイドラインがそのつど制定され、また、当該分野が総務省による点検対象としても追加されるなどして、推進されてきた。しかしながら、総合評価については「政策評価に関する標準的ガイドライン」の内容をほぼ踏襲した実施ガイドラインが定められて以降、特段動きが見られない。さらに、最近の事例で言えば、政策評価・独立行政法人評価委員会³¹⁾が2014年度末をもって活動を終了することに伴い、今後取り組むべき課題について取りまとめた提言では総合評価について一切触れられていない³²⁾。

さらに、現在の政策評価法の制度上、評価の実施を義務付けられている分野は、公共事業、政府開発援助、規制、租税特別措置等、研究開発の5分野であるが、いずれも事業評価方式を用いた政策分析である。また、総合評価を実施するか否かは、各行政機関の裁量に委ねられており、義務付け対象の分野の評価が優先されることを考慮すれば、総合評価の優先順位は相対的に低い³³⁾。

以上のような背景・原因を踏まえると、プログラムの概念や、プログラム評価、業績測定、政策分析の関係、総合評価が制度導入時に意図していた効果の把握、問題点の発見、有用な情報の提供という機能を再確認する必要がある。

また、業績測定とプログラム評価は対立するものではなく、相互に補い合うものであるとされている。このため、業績測定において問題やその傾向がみられたものについて、優先的にプログラム評価の対象として選定し、詳細なプログラム評価を実施することが考えられる。近年、日本では施策レベルを見る目標管理型の政策評価と、事業レベルを見る行政事業レビュー³⁴⁾という整理によって、実績評価と行政事業レビューの連携が推進されているが、プログラム

³⁰⁾ 金本、1990年。

³¹⁾ 総務省行政評価局が行う政策評価に関する重要事項等について調査審議するとともに、各行政機関の独立行政法人評価委員会が行う評価の結果について当該委員会に意見表明等を行うため、総務省に設置された審議会。

³²⁾ 政策評価・独立行政法人評価委員会「政策評価・独立行政法人評価委員会の提言」(2015年3月9日)。

³³⁾ たとえば、政策ないし施策の改善情報を評価によって産出するための方策として、鎌田は現在の政策評価法体系における評価の義務付けの範囲を見直し、評価事務の総量についての相当の軽減化を図る必要性について主張している(鎌田、2008年、50-51ページ)。

³⁴⁾ 行政事業レビューとは、「各府省自らが、5,000を超える国の全ての事業について、概算要求前に、前年度の執行状況(支出先や使途)等の事後点検を行い、事業内容や目的、成果、資金の流れ、点検結果などを書いた各府省共通のレビューシートを作成、公表するもの」である(内閣官房ホームページ <http://www.cas.go.jp/jp/seisaku/gyoukaku/review.html> (2015年3月15日最終閲覧))。

評価については他方面と十分な連携が取られていないのではないだろうか。

ただし、プログラム評価と業績測定を連携させる手法は「健康診断」（定期検診と精密検診）にたとえられるが、これは業績測定が的確に行われているという大前提に基づくため、場合によっては不十分な点もあることに注意が必要である。たとえば、業績測定の方法や、測定値（目標）の設定自体が不適切であれば、実態として問題が表面化していても、業績測定に「引かかる」ことはない。したがって、業績測定が的確に行われているのかを探るためにも、業績測定に対するチェックや、プログラム評価において業績測定の情報把握したり、業績測定で得られた情報を吟味し直したり、場合によっては、新たな測定指標も加えて総合的に分析することが望ましい。

3.2 プログラム評価の実施体制

諸外国においては、このようなプログラム評価はどのように実施されているのであろうか。研究会の議論では GAO による評価がプログラム評価と位置付けられているが、各行政機関にとって GAO の評価は外部機関が行う評価に相当する。また、田辺はアメリカの労働省はプログラム評価と業績測定を関連付けた取組が先進的であると紹介しているが、労働省でのプログラム評価の多くは外部に委託されており、実質的に内部によって実施されている場合は少なく、日本でも評価の実施に際して一定程度の予算を配分する必要性を主張している³⁵。このように考えると、日本のように内部評価を中心とした総合評価の取組の可否についても疑問も出てくる。たとえば佐々木は、業績測定は内部機関が実施し、プログラム評価は外部機関に任せるという役割分担を主張している³⁶。したがって、プログラム評価を実施するための費用面（人員、資金、時間）への問題もあって、プログラム評価は外部機関が行うことが望ましいという風潮があるのではないかと考えられる。

しかしながら、国土交通省をはじめとして、

内部評価による総合評価を積極的に実施している事例もある。このため、組織内部におけるプログラム評価の実施は、専門的な調査・分析の問題を除けば、実際のところ不可能ではない。また、参考として、評価専担組織としての総務省による各行政機関に対する評価（総合性・統一性確保評価）についても、総合評価方式を用いた評価であると研究会の最終報告で整理されているが、実際に高度で厳密な分析技術は用いられていない。もちろん、評価専担組織としての総務省による総合評価は、調査の設計から勧告に至るまで約2年という長い期間が費やされているが、その多くは評価対象の機関との折衝や情報収集であることを踏まえれば、内部評価として実施すればより効率的になる可能性もある。

また、現在実施されている総合評価は、そのほとんどが原局・原課が中心となって実施している「自己評価」であるが、内部評価としての総合評価を考える場合、評価の取りまとめ部局（たとえば、大臣官房の政策評価の担当課）を積極的に活用することも考えられる。前述のとおり、原課においては縦割りでプログラムという考え方が浸透しにくいいため、内部評価としてプログラムに着目するのであれば、このような評価の取りまとめ部局が主体的に評価を主導することが望ましいといえる。

くわえて、評価の取りまとめ部局に属する者は、外部（たとえば、評価専担組織としての総務省）と内部（原局・原課）の中間的な立ち位置にいるという利点も考えられる。たとえば、評価の取りまとめ部局は、原課ほど評価に際しての情報量は持ち合わせていないが、外部よりは内部における細かい内情について理解を示せる余地があるため、外部ほど革新的すぎず、また、内部ほど保守的すぎない役割が期待できる。もちろん、ここで「中間的な立ち位置」といっても、評価の取りまとめ部局が1組織の内部組織に属することには変わりない。しかしながら、評価の取りまとめ部局は外部と内部のパイプ役であるため、たとえば評価専担組織としての総務省が行政機関に対して評価を行う場合、外部

³⁵ 田辺, 2014年, 13ページ。

³⁶ 佐々木, 2003年, 192-193ページ。

³⁷ たとえば、山谷, 2006年, 54ページ, 表2-2において、評価専担組織としての総務省による外務省の評価が行われた場合の照会先が整理されている。原課が評価に関係するか否かはその所管事項次第ではあるが、官房総務課、考査・政策評価官室といった評価の取りまとめ部局は必ず関係しており、外部（評価専担組織としての総務省）と内部（原局・原課）の間に入っていることが伺える。

と内部のやり取りは全て評価の取りまとめ部局を経由することになる³⁷。したがって、評価の取りまとめ部局は外部的な考え方と内部的な考え方の双方に触れる機会があるため、評価に際しては「中間的」な視点を備えているといえる。なお、これまで本稿で優良事例として紹介することが多かった国土交通省では、大臣直属の政策統括官という指定職クラスの担当が省内の評価を統括しているが、評価のために指定職を設置しているのは国土交通省のみである³⁸。このように評価実施のための体制が整備されていることも、比較的良く総合評価に取り組むことができている理由であると考えられる³⁹。

ところで、総合評価を内部評価で実施する場合、山谷が指摘するようにアカウントビリティの確保という面では外部評価よりも劣るが、組織学習という面では有効な手法となる⁴⁰。このため、総合評価を外部機関か内部機関かのいずれかで実施するかでいえば、どちらもあり得るのであり、それは評価結果をどのように活用したいかで変わる。プログラム評価でアカウントビリティの確保を重視するのであれば外部評価、情報提供による組織学習に重点を置く場合は内部評価、というように適宜使い分けられることが望ましい。いずれにしても、プログラム評価であることを理由にその多くを外部に委託するという事にはならない。

最後に、事例として先に取り上げたが、プログラム評価の実施を法令や各種計画の見直しと連動させることも重要である。研究会の議論では、法令等の見直し時に評価を行うことが期待されていたが、実際のところは事例としては少ない。このため、本稿で紹介した国土交通省の取組のように、計画策定後の一定期間経過後にプログラム評価の実施を法律で義務付けている場合は先進的な事例となる⁴¹。

おわりに

これまで日本における国の行政機関が実施している総合評価について、その実施件数の変化のほか、ガイドラインやプログラム評価の理論が意図するプログラムの効果の把握、問題点の発見、情報提供の有無といった点に着目しつつ、その実態を整理した。しかしながら、近年、総合評価の実施件数は減少しているだけでなく、プログラム評価としての精度についても依然として課題がみられる評価書もある。また、このような課題の背景・原因については、2008年以降の国における業績測定や政策分析の浸透やそれに伴うプログラム評価へのチェックの減少、プログラム評価を外部評価で行うことへの期待、日本においてプログラムの概念自体の理解が乏しいといった理由を取り上げた。そして、これらの課題を克服し、プログラム評価としての総合評価を浸透させるためには、プログラムの概念や総合評価の役割を再認識するとともに、業績測定との連携についても検討すべきである。また、評価を実施すること自体を法令にあらかじめ規定しておくなど、その実施を制度的に組み込んでおくといった方策も考えられる。

総合評価方式の利用が実態として減少していることは、ただちにその必要性を否定することにはつながらない。実際に、効果が不透明な政策・施策があり、測定指標が悪い結果を示しているも、必要な評価が実施されていない場合は依然として存在する。たとえば、評価専担組織としての総務省が実施した各行政機関に対する調査では、業績測定にも活用できるような基礎データの整備が不十分であるという指摘だけでなく、測定指標が不十分な傾向を示しているにもかかわらず、政策効果の検証（原因分析）が実施されていない旨を指摘した事例も存在する⁴²。このような状況を踏まえると、日本における総合評価の実施件数は近年減少してきてい

³⁸ 渋谷, 2001年, 86ページ。

³⁹ このほかに、たとえば鎌田は行政機関を取り巻く政策課題に対する認識や、その下で評価にどのように取り組んでいくかの戦略的な考え方を示し、それらと関連する形で評価テーマが提示するといった「評価政策」を各行政機関が作成する政策評価の基本計画に盛り込む必要性について指摘している（鎌田, 2008年, 50-51ページ）。

⁴⁰ 山谷, 2012年, 223-224ページ。

⁴¹ 国土形成計画法の他に政策評価法に基づく評価の実施を定めている法律としては、住生活基本法（2006年6月8日法律第61号）、社会資本整備重点計画法（2003年3月31日法律第20号）がある。

⁴² 総務省『医師等の確保対策に関する行政評価・監視結果に基づく勧告』（2015年1月27日）。

るものの、プログラム評価に対するニーズは各行政機関にとっても潜在的に高いのではないだろうか。たとえば政策効果の検証を十分に行うことで、政策がうまく進まない理由が解明され、評価結果が対策案・代替案を考える土台になれば、その評価による情報提供は政策の担当者や決定者にとって有用であったといえる。

ところで、プログラム評価はその実務と理論の双方において諸外国で展開されてきた歴史があるが、本稿では諸外国におけるプログラム評価の内容については十分なレビューを行わなかった。今後は、諸外国の事例を参考にしつつ、個別具体的なプログラム評価の事例研究の把握、蓄積が必要であり⁴³、とくに日本で応用することを考えるのであれば、内部評価の事例をより一層、調査する必要がある。

参考文献・ホームページ

- 梅田次郎, 小野達也, 中泉拓也『行政評価と統計』産業統計研究社, 2004年。
- 小野達也「地方自治体 行政評価 課題 展望」『日本評価研究』(日本評価学会) 第2巻第1号, 2002年, 29-37ページ。
- 金本良嗣「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』(会計検査院) 第2号, 1990年, <http://www.jbaudit.go.jp/effort/study/mag/2-2.html> (2015年3月15日最終閲覧)。
- 鎌田英幸「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』(行政管理研究センター) 第124号, 2008年, 39-54ページ。
- 佐々木亮『政策評価トレーニング・ブック』多賀出版, 2003年。
- 佐々木亮『評価論理』多賀出版, 2010年。
- 渋谷和久「国土交通省における政策評価の展開」『日本評価研究』(日本評価学会) 第1巻第2号, 2001年, 79-88ページ。
- 政策評価の手法等に関する研究会「政策評価制度の在り方に関する最終報告」(2000年12月11日) <http://warp.ndl.go.jp/info:ndljp/pid/283520/www.soumu.go.jp/hyouka/matome1.htm> (2015年3月15日最終閲覧)。
- 総務省『政策評価等の実施状況等の国会報告、政策評価結果の予算要求等への反映状況』(2002年度分～2013年度分)。
- 総務省行政評価局ホームページ「各府省の政策評価の点検(客観性担保評価活動) —これまでの取組結果—」。http://www.soumu.go.jp/main_sosiki/hyouka/seisaku_n/torikumi.html (2015年3月15日最終閲覧)。
- 田中啓「日本の自治体の行政評価」『分野別自治制度及びその運用に関する説明資料 No.14』(財団法人自治体国際化協会, 政策研究大学院大学比較地方自治研究センター) 2010年, 1-26ページ。
- 田辺智子「業績測定を補完するプログラム評価の役割—米国の GPRAMA の事例をもとに—」『日本評価研究』(日本評価学会) 第14号第2巻, 2014年, 1-16ページ。
- 山谷清志「わが国の政策評価—1996年から2002年までのレビュー—」『日本評価研究』(日本評価学会) 第2巻第2号, 2002年, 3-15ページ。

- 山谷清志「評価の理論と実践におけるプログラムの概念—政策評価と ODA 評価をめぐって—」『同志社政策科学研究』(同志社大学総合政策科学研究科) 2004年12月。
- 山谷清志『政策評価の実践とその課題』萌書房, 2006年。
- 山谷清志『政策評価』ミネルヴァ書房, 2012年。
- 山谷清志「政策評価制度 10年の経験—レビューと展望—」『日本評価研究』(日本評価学会) 第13巻第2号, 2013年, 1-2ページ。
- Government Accountability Office (2011), "Performance Measurement and Evaluation: Definitions and Relationships". <http://www.gao.gov/assets/80/77277.pdf> (2015年3月15日最終閲覧)。
- May, Iris Geva, and Pal, Leslie A. (1999), Good Fences Make Good Neighbors, *Evaluation*, Vol.5-3, pp.259-277.
- Rossi Peter H, Lipsey Mark W, Freeman Howard E, (2004), *Evaluation: A Systematic Approach*, 7th ed. Thousand Oaks, Calif.: Sage (大島巖, 平岡公一, 森俊夫, 元永拓郎 監訳『プログラム評価の理論と方法—システマティックな対人サービス・政策評価の実践ガイド』日本評論社, 2005年)。

⁴³ 海外におけるプログラム評価の事例研究の蓄積が少ない点については、田辺も指摘している(田辺, 2014年, 13-14ページ)。