

政策評価の概念および手法に関する一考察

—中小企業政策の評価を念頭において—

中 川 利 香

1. はじめに

政府によって実施されるさまざまな政策の原資は国民の税金である。よって、政府は政策の効果を特定し、計測し、それを評価して国民に説明する義務を負う。評価の結果、問題があるとされたなら、政策の再考が求められるだろう。問題を解決する策を講じて政策を継続させるのか、政策を中止するのかという決断しなければならない時もあるかもしれない。このように、政府はあらゆる政策の内容のみならず効果についても説明責任を果たすことが求められる。

ところが、政策の効果を計測して評価することは、それほど簡単ではない。その理由は第1に、政策を実施していなかったら何が起こっていたかという状況を正確に把握することができないからである。いちど政策を実施したら、実施していない状況を再現することはできない。よって、政策の実施前後を比較することは不可能である。第2に、政策実施後に何等かの変化が起こったとしても、それが純粋な政策の効果であると断言することができないからである。他の要因の影響を受けた可能性を排除することは極めて難しい。このように、政策評価の必要性が高まっているにもかかわらず、技術的な困難がともなう。

このような事情がある中、先進国ではさまざまな工夫を施しながら政策評価を行ってきた。以上のような背景を踏まえ、本稿は政策評価の手法に関する先行研究を整理することを目的とする。政策評価の歴史は1600年代まで遡ることができるといわれているが、さまざまな理由から現在に至るまで評価手法が確立されたとはいえない状況にあることを明らかにする。

本稿の構成は次の通りである。2では政策評価の議論を整理する。ここでは、政策評価の歴史および概念をまとめる。政策評価とは、プログラムの評価を意味することを明らかにする。3ではプログラム評価の手法について述べる。実務面では政策の目的によって評価の手法が異なっていることを説明する。また、定量的手法ではバイアスをいかに排除するかが重要であるが、その手法で把握できない部分については定性的手法で補うことが必要となる。さらに、評価にはトレードオフが

あることを指摘する。4では、評価対象の焦点を中小企業にあてた場合の政策評価に関する議論について触れる。最後に本稿をまとめ、今後の課題を述べ、結論とする。

2. 政策評価に関する議論

2-1. 政策評価の歴史

ロッシ・リップセイ・フリーマン [2005] によれば、評価の歴史は1600年代に始まり、体系的な仕組みとして取り入れられるようになったのは1900年以降のことであるとされている。最初に評価の対象となったのは、国民の識字率向上、職業訓練、保健衛生等の政府のプログラムであった。1920年代末まではこのような分野に対する政府の支出は相対的に大きくなかったため、プログラムの成果について注目されることはあまりなかった。ところが、世界恐慌を契機に政府支出は急増し、政府のプログラムの管理と効率性に関する意識が高まった（ロッシ・リップセイ・フリーマン [2005], p. 8）。

第二次世界大戦後、政府のプログラムの対象が教育や公衆衛生から都市計画や住宅等の社会開発にまで拡大し、1950年代末までには政策評価は一部の分野で一般的な現象となった。この背景には、政府のプログラムの規模が拡大したこと、それに関わる主体も政府から民間財団まで広がりを見せるようになったこと等が関係しているようである。政府支出の規模と範囲が拡大するほど、その成果に対する関心も高まったといえよう。

このように実務面での政策評価が活発になると、研究も進むようになった。ロッシ・リップセイ・フリーマン [2005] は、評価研究が1960年代末までに「ひとつの成長産業」となり、1970年代に入ると「社会科学におけるひとつの独立専門分野」になったと述べている（ロッシ・リップセイ・フリーマン [2005], p. 9）。ところが、当時の研究者と実務家の関心は若干異なっていたようである。研究者の関心が政府の意思決定や政策形成過程など、政府がどのように機能しているのかに向けられていたのに対し、政府内の実務家は、膨大な予算と人員を投入して実施するプログラムに対する正当性を証明する方法を模索していた（ロッシ・リップセイ・フリーマン [2005], p. 11）。

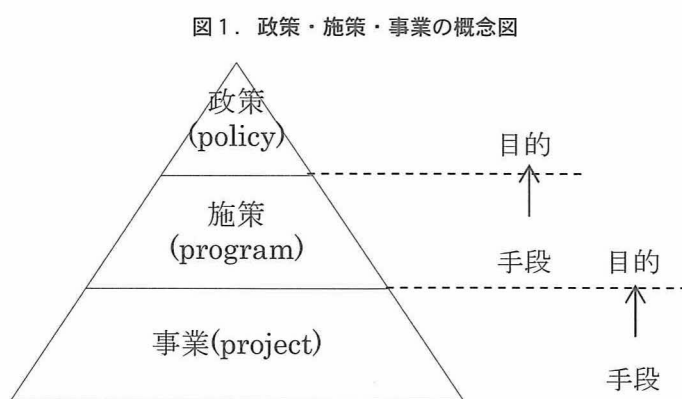
その後、政策評価に対する関心の高まりは、政府のプログラムに関与する主体や、プログラムによって、直接的にも間接的にも影響を受ける主体にまで広がっていった。ロッシ・リップセイ・フリーマン [2005] が、政策評価は「社会政策および行政活動において不可欠な一部としてみなければならない」（ロッシ・リップセイ・フリーマン [2005], p. 10）と表現していることから明らかなように、政府の活動における国民に対する説明責任を果たすための政策評価が求められるようになったのである。

2-2. 「政策評価」の概念整理

これまで、本稿では政策（policy）という用語を特に定義せずに使用してきた。しかし、政策の定義を明示しなければ評価できない。したがって、まず政策とは何かを明らかにし、その上で政策評価を論じる必要があるだろう。

山谷 [1997] によれば、政策は「政府の将来構想や基本姿勢を表現した理念の『器』」（山谷 [1997], p. 10）であると述べている。そして、政策（policy）の下位概念に施策（program）と事業（project）が存在し、「下位概念である施策や事業の達成すべき目標を示し、方向づけを行う指針としての役割を担う」（山谷 [1997], p. 11）ものであるとしている。図1は、政策・施策・事業の関係を表した概念図である¹⁾。政策（policy）は最上位目標であるのと同時に政府が達成しようとする目的を意味する。そのひとつ下の概念である施策（program）は、政策目的を達成するためにすべきことを具体化したものであると考えられる。さらにその下の概念である事業（project）は、施策で掲げられた目的を達成するための具体的な事業あるいは行動を意味し、施策（program）の目的を達成するための手段であると理解することができる。また、龍・佐々木 [2000] は、政策は「ある社会状況を改善するために、ひとつのあるいはいくつかの目的に向けて組織化された諸資源および行動」（龍・佐々木 [2000], p. 8）と定義している。このように、政策は複数の施策と事業から構成されており、ひとつの政策がひとつの政府の活動を意味するものではないと判断することができる。

以上の議論から、日本で一般的に用いられている「政策評価」は「プログラム評価」を意味することがわかる（山谷 [1997], pp. 14-15）²⁾。つまり、政策評価においてはより具体的にプログラム



（出所）山谷 [1997] 第1章および第4章を参考に筆者作成。

1) 政策・施策・事業のそれぞれの具体例は山谷 [1997] (p. 12) を参照されたい。

2) 政策評価に関する先行研究でよく引用されるロッシ・リブセイ・フリーマン [2005] (原書 [2004]) では、「プログラム評価」という用語を使用している。

の対象、目的、方法論を評価することになる。とりわけ、プログラムの対象が適切であったか否か、目的を達成したか否か、目的を達成するための方法論が適切であったか否か等が評価されることになる。

2-3. 何のためにプログラム評価を行うのか

そもそも、何のために政府はプログラム評価を行う必要があるのだろうか。それは、意思決定、資源配分、説明責任の改善のためである（OECD [1998]、ロッシ・リプセイ・フリーマン [2005]）。ロッシ・リプセイ・フリーマン [2005] によれば、プログラム評価を行う理由として、次の4つをあげている。第1に、プログラムのあり方を見直すためである。将来においてプログラムを継続するか否か、拡大するか否か、また改善すべき点はないか等を判断するために評価を行うとしている。第2に、プログラムの汎用性を判断するためである。プログラムの有用性が明らかになれば、他の地区や地域にも適用することができるかもしれない。第3に、プログラムの運営や管理面の効果ないし効率性を高めるためである。先の2つのポイントはプログラムの内容に関係していたのに対し、第3のポイントはプログラムの運用面に関係しており、評価を行うことによりプログラムの効率性を高めるための材料を提供することが可能としている。そして第4に、プログラムに必要な資金面をサポートする主体に説明責任を果たすためである。政府のプログラムであれば、その資金提供者は納税者ということになる。近年、税金が有効に使われているか否かは納税者の大きな関心事である。納税者に対して真摯に対応しているか否かは、政府の重要な責務である。このように、評価は政府のプログラムの運用や内容の改善に役立つ材料を提供するために行うものである。

プログラム評価の方法については、OECD [1998] が行っている「評価」の定義が参考になるだろう。それによると、評価とは「公共政策や組織、またはプログラムの結果（result）に着目し、それがプログラムの確実性（reliability）や有効性（usefulness）を説得的に示しているかを分析する」（OECD [1998], p. 3）ことであるとしている。なお、プログラムの結果（result）の捉え方は2通りあることに留意する必要があるだろう。ひとつは文字通り資源の投入に対する生産物（output）であり、もうひとつはプログラムの対象者に対する間接的な影響も含めた成果（outcome）である。プログラム評価ではこの2つを区別し、ロッシ・リプセイ・フリーマン [2005] はとりわけ後者に対する評価が重要であると指摘している。

3. プログラム評価の手法

政府が政策によって達成しようとする目標は実に多様なため、それを実現するためのプログラムも多種多様である。よって、政策評価研究会 [1999] が指摘するように、「万能な評価手法は存在しない」（政策評価研究会 [1999], p. 91）のが現状であり、後述するように評価の手法はさまざま

までである。したがって、それぞれの手法の特徴や利点と欠点を理解したうえで、プログラムの性質や評価の対象、目的を鑑みて適切な手法を選ぶ必要があるといえるだろう。

3-1. ロッシ・リプセイ・フリーマン [2005] による分類

ロッシ・リプセイ・フリーマン [2005] は、評価の方法として5つをあげている(表1)。ニーズアセスメントは、「社会問題の性質や大きさ、広がり、介入の必要性の程度、そしてその介入を設計するための社会状況の意味合いをアセスメントする」(ロッシ・リプセイ・フリーマン [2005], p. 53) ものである。プログラムの対象となる人に必要とされているかや、プログラムを提供する策の是非を判断することになる。プログラム理論のアセスメントは、プログラムが果たすべき役割について仮説を設定し、プログラムの目的を達成するまでのアプローチの合理性、実行可能性、倫理上の問題の有無などを判断する。プロセスアセスメントはプログラム理論の実施における運営面に着目し、理論の正確さの度合い (fidelity) や効果を評価するものである。インパクトアセスメントは、プログラム実施における成果 (outcome) がプログラムの意図する目的を達成したのか否か、想定しない副作用等の負の影響がないかを評価する。効率性アセスメントは、プログラムに必要な費用と便益を評価する。そのため、効率性アセスメントは費用便益分析 (cost-benefit analysis)、あるいは費用効果分析 (cost-effectiveness analysis) といわれることもある。

表1. ロッシ・リプセイ・フリーマン [2005] による評価の方法

評価の分類	内容
ニーズアセスメント (needs assessment)	必要とされるサービスとその提供の最善の策を判断する
プログラム理論のアセスメント (assessment of program theory)	目的を達成するためにプログラムが果たすべき役割を概念化し、それがどの程度適切かを判断する
プロセスのアセスメント (assessment of program process)	プログラムの運営面における正確さや効果を評価する
インパクトアセスメント (impact assessment)	プログラムが意図した目的の達成状況とその程度、副作用を評価する
効率性アセスメント (efficiency assessment)	プログラムの費用に対して十分な利益を生み出しているかを評価する

(出所) ロッシ・リプセイ・フリーマン [2005], pp. 51-60を参考に筆者作成。

3-2. 評価手法

(1) 定量的手法

評価の具体的な手法は、定量的手法と定性的手法に大別することができる。定量的手法は、統

計的にプログラムの実施による成果を観察し、実施前後を比較して評価を行う手法である。定量的手法には、無作為化フィールド実験法（randomized field experiment）と非無作為化準実験法（quasi-experiments）がある（表2）。前者は、プログラムを実施する介入群（intervention group）と実施しない（もしくは全く関係のない介入を受ける）対照群（control group）に分けてプログラムの成果を観察し、比較する方法である。ここで重要なことは、対象者をふたつのグループに無作為に割り付けることである³⁾。それに対して後者は、介入群のプログラムの成果を非参加者から成る対照群と比較するものである。この場合、対象者は無作為に介入群と対照群に割り付けられたとは限らない。また、後者においては介入群に対するプログラムの成果（観察値）と介入群に対してプログラムを実施しなかった場合の変化（推計値）を比較することでプログラム評価を行う場合もある。

ふたつの方法にはそれぞれ利点と欠点があり、また実用性の面でも特徴があるため両者の特徴を熟知したうえで実際の評価に使用する必要がある。無作為化フィールド実験法の利点は、プログラムの成果に関する予測妥当性について強い説得力を持つことである。ただし、介入群と対照群の質を均一にする必要がある。そのためには、介入群と対照群の構成を同一にする必要があり、またプログラムを実施しなくても両群は同じ成果（outcome）を示すように構成することが求められる。また、全ての観察期間においてプログラム以外の面を対象者を同じ環境におく必要もある（ロッシ・リプセイ・フリーマン [2005], p. 224]）。これは、プログラムの実施前から終了までの間の諸条件を等しくすることによってのみ、正確なプログラムの成果を観察することが可能となるからである。また、プログラムの途中で変更が加えられると、変更前と変更後のプログラムの成果の判別ができなくなるため、注意を要する。実際には実験の完全性を確保することが難しく、また時間と費用がかかることを理解する必要があるだろう。最も厄介なのは、実験的な介入と実際の介入の効果が異なる場合である。実験はあくまでも人為的にコントロールされた環境の中で行われるため、実際に介入が行われた場合には実験通りの成果が得られる保証はない。また、実験的手法そのものに批判が発生する場合もある。以上のように、無作為化フィールド実験法は、評価の正確性や説得力を追求するには望ましい手法であるが、実際に使用するには相当の準備と配慮が求められる⁴⁾。

3) 無作為化（randomization）とは、「対象者（あるいは対象単位）が、プログラムを受けるか、代わりに対照状況を受けるか、を確率的に決定する方法である」（ロッシ・リプセイ・フリーマン [2005], p. 225）。一般的には、乱数表を使って割り付けることが多い。

4) 龍・佐々木 [2000] では、無作為化フィールド実験法をランダム実験モデル（randomized experimental model）と表現している。この手法は、一時期あまり使用されなくなったようである。その理由は、本文でも述べた通り時間と費用がかかるうえ、評価結果を一般化することに疑問を呈する声が多く上がったことが関係している。実験的な手法は単なるふたつのグループの比較に過ぎず、それをプログラム成果の一般的な評価とみなすことには問題があるのではないかという批判である。しかしながら、龍・佐々木 [2000]

そこでより簡便な評価の手法として、非無作為化準実験法を使用することが可能である。この手法の利点は、無作為化フィールド実験法と異なり倫理面の批判が発生しにくい。また、十分なサンプル数を確保することが難しい場合は非無作為化準実験法を用いて評価することが可能である。しかしながら、無作為化フィールド実験法に比べて評価結果の説得性には欠けることは否めない。また、介入群と対照群の選別にバイアスがかかることが想定され、プログラムの成果を正確に計測することが難しい。そこで、統計的手法によりバイアスを排除する必要がある。

表2. プログラム評価の定量的手法に関する利点と欠点

<p>無作為化フィールド実験法 (randomized field experiment)</p> <p>無作為割付によるプログラムを実施する介入群 (intervention group) と実施しない (もしくは全く異なる介入を受ける) 対照群 (control group) に無作為に分けてプログラムの効果を観察し、比較する</p> <p>利点: プログラムの成果に関する予測妥当性について説得力がある</p> <p>欠点: 対照群と介入群の等質性を確保しなければ正確な評価結果が得られない プログラムの途中で変更が加えられると、結果に意味をもたなくなる 倫理面での批判が発生しやすい 実験的な介入と実際の介入の成果が異なる場合がある 時間と費用がかかる 実験の完全性を確保するのが難しい</p>
<p>非無作為化準実験法 (quasi-experiments)</p> <p>無作為割付によらないプログラム参加者から成る介入群 (intervention group) のプログラムの成果を非参加者から成る対照群 (control group) と比較する</p> <p>利点: 時間と費用を抑えることができる 倫理面での批判が発生しにくい</p> <p>欠点: プログラムの成果にバイアスが生じやすい</p>

(出所) ロッシ・リブセイ・フリーマン [2005] 第8章および第9章を参考に筆者作成。

(2) バイアスをどのように排除するか

非無作為化準実験法 (quasi-experiments) の場合、介入群と対照群は無作為に割り付けていないため、プログラム評価の結果にはさまざまなバイアスが発生することが想定される。つまり、評価結果の信頼性と妥当性を高めるには、バイアスをいかに最小限にとどめるかが重要となる。まず考慮すべきは、プログラムの成果に関する推計にバイアスがかかることである。プログラムを実施し

によれば、近年ではコンピュータの処理能力が格段に改善したことにより、実験的な手法が採用されやすくなったと述べている (龍・佐々木 [2000], p. 57)。実験的手法はサンプル調査であるから、情報技術の発達により膨大なデータの収集と処理が可能になったとはいえ、介入群と対照群のサンプルがプログラム実施対象者と非対象者を代表することができるか否かを常に考える必要があるだろう。その意味で、実験的手法の限界を常に意識しておかなければならない。

た介入群の変化は観察できるが、同じグループの介入しなかった場合の状態を観察することができない。したがって、非無作為化準実験法（quasi-experiments）を用いる場合は、介入群に対してプログラムを実施しなかった場合の変化を推計する必要がある。この推計にバイアスが発生し、プログラムの成果を過大ないし過少評価する可能性がある。これは選択バイアス（selection bias）とよばれる。例えば、対象者がプログラムの実施という介入を自発的に受け入れる場合、それが対象者の態度を変化させ、プログラムの成果（outcome）にバイアスがかかってしまうことがある。また、選択バイアスは評価者が対象者を選択する際にも発生する。評価者がプログラムの成果がより良好なグループを介入群として選択してしまう場合である。いずれの場合も、無作為割り付けによって形成されるグループを観察するのではなく、選択されたグループを観察することが原因でバイアスが発生するのである⁵⁾。

このようなバイアスを排除する方法として、第1に評価者が事後的に対象群を作成するマッチングという手法がある。評価者は介入群の特徴を熟知したうえで、それと同質の対照群をつくり、プログラムの成果を比較することができる。第2に、介入群と対象群の差異を統計的手法によってコントロールする方法がある。比較的よく使用されるのは、ロジスティック回帰分析や操作変数法などである。介入群と非介入群を区別する変数を回帰式に挿入することで、これらの差異を抽出する手法である。第3に、回帰・分断モデル（regression-discontinuity model）という方法がある。これは、成果の基準値以上と以下のグループに分け、一方にはプログラムを実施してもう一方には実施せずに各グループの変化を観察する方法である。第4に、異時点間のプログラムの参加者自身を比較する再帰的コントロール（reflexive controls）という方法である。例えば、before-after比較があげられる。これは、対象者のプログラム参加前と参加後の状態を比較する手法である。もうひとつの例として時系列分析がある。これはbefore-after比較の一種であるが、before-after分析は成果の観察がプログラムの実施前と後に限定されるのに対し、時系列分析はプログラムの実施中に何度か観察することで成立する手法である。世界銀行は、これらの手法は介入群、対照群の設定をする必要がない全国民を対象とするプログラムの成果を評価する際には有用であると言及している⁶⁾。しかし、プログラムの実施中にプログラム以外の要因の影響を排除することができないという欠点がある。青柳 [2007] は「プロジェクトの前後を通じて、プロジェクト以外の要因による変化は一切無いという仮定（条件）が必要」（青柳 [2007], p. 95）と述べている。確かに、そのような仮定を設定し

5) Storey [2006] は、選択バイアスが生じる例として自己選択（self-selection）によるものと、コミッティー選択（committee selection）によるものがあると解説している（Storey [2006], pp. 265-266）。

6) The World Bankウェブサイト

(<http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20188242~menuPK:412148~pagePK:148956~piPK:216618~theSitePK:384329~isCURL:Y,00.html>, 2012年3月21日アクセス)。

ない限り before-after 比較は難しいだろう。しかし、仮定そのものが現実的ではないという問題もある。

(3) 定性的手法

本来、定量的手法は数量処理をすることでプログラムの実施による変化を客観的に評価しようとするものである。しかし、既に述べたように、分析前の環境整備などにおいてバイアスが発生する余地を残していることが明らかになった。また、数量化できない事象については定量的手法で分析することはできない。そこで、用いられるのが定性的手法である。この手法は、プログラムによる評価対象の変化についてありのままを描写しようとする。具体的には、インタビュー調査やフィールド調査による観察などがあげられる。評価者による観察は、観察の方法や対象を操作せずによりのままを観察する場合もある⁷⁾。参与観察法は社会学等の分野で比較よく用いられる手法であり、評価者が対象者の集団あるいは地域に入り込んで観察する方法である。

定性的手法は、評価対象の変化について数量化できない部分を把握するうえでは有効であるが、評価者の主観に影響を受けやすいという欠点がある。したがって、定性的手法による調査結果には、常に客観性についての問題が発生することは否めない。

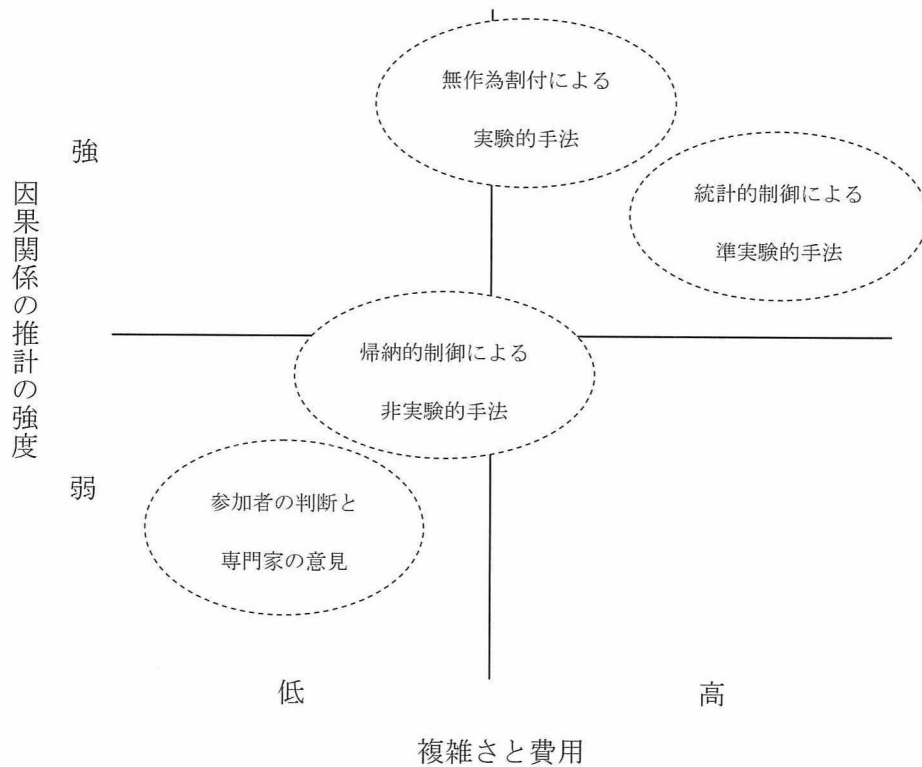
(4) 評価のトレードオフ

以上のように、評価手法について定量的手法と定性的手法についてそれぞれの特徴をまとめてきた。実際に評価手法を選択する際には、因果関係の頑強性や評価の煩雑さと費用の面においてトレードオフが発生することを理解しておく必要がある（図2）。プログラムとその成果の因果関係を最も頑強に示すことができる手法は、無作為割付による実験的手法である。統計的制御による準実験的手法は、バイアスを排除する手順を踏む必要があることから煩雑さと費用は無作為割付による実験的手法よりも高くなる。帰納的制御による非実験的手法は本稿で示すところの定性的手法であるが、これは煩雑さと費用の面で準実験的手法よりも低いものの因果関係の頑強性が劣る。

このように、プログラム評価においては、これらの点も考慮しつつ、定量的手法と定性的手法の両方を状況に組み合わせて進めていくことが望ましいのではないかと考えられる。

7) 山谷 [1997] は、評価の手法が適切か否か、プロセスや評価結果の妥当性を評価するメタ評価 (meta-evaluation) も定性分析値に分類している。参考までに、わが国の各省庁が行うプログラム評価のメタ評価は総務省が行っている。メタ評価に関する概念は源 [2009] で整理されているので参照されたい。

図2. 評価のトレードオフ



(出所) Storey [2006], p. 269より引用。ただし邦語訳は筆者による。

4. 中小企業政策の評価

本節では、政策の焦点が中小企業に絞られ、その育成を政策目標としている場合の評価をどのように考えたらよいか整理してみたい。石井 [2010] は、日本の中小企業・ベンチャー企業の公的支援策に焦点をあてて政策評価を論じている。石井 [2010] によれば、日本の中小企業・ベンチャー企業の政策評価は、その方法論が確立されておらず、各行政の自己評価が行われているという。これは次の2つの問題を引き起こす可能性を示唆している。第1に、各行政機関の目先の目標を達成したか否かに重点が置かれ、中小企業政策全体の評価を困難にさせる可能性である。第2に、政治の介入ないし影響が及ぶ可能性である。このような問題を発生させないためには、中立的かつ客観的視点を確保することが求められる。そこで、石井 [2010] は外部有識者や研究者が政策評価に関与できるような環境を整備することが重要であると提案している。また、欧米で進んでいる評価手法に学び、精度の高い政策評価の実施が急務となるとも述べている (石井 [2010], pp. 17-18)。

Storey [1998] は、中小企業支援策の評価として6つのステップを提示している。ステップ1では、特定の政策の対象となる企業の数、分野、規模、地域的な分散などの特性を把握する。ステッ

プ2では受益者の意見を聴取する。供与された支援策は有益だったか、支援申請および支援実施までに問題はなかったか、など政策の対象となった受益者の意見を収集する。ステップ3では、支援を受けた後に企業のパフォーマンスに違いが生じたか否かを調べる。ここでは、売上や利益などの数量的な部分も調査することが求められる。ステップ4では、支援を受けている企業と受けていない企業を比較する。ステップ3で収集したデータを比較することも必要であるが、生存率（survival rate）などを比較することも有効であろう。ステップ5では、支援を受けた企業間の比較である。ここでは、支援を受けているグループの中から操業年数、部門、所有者、立地などいくつかの要素で共通する企業を取り上げて比較する。ステップ6では、選択バイアスに配慮し、統計的手法を用いて評価を行う。これらのステップでは、監視（ステップ1～3）と評価（ステップ4～6）が分けられており、評価の実務に参考になる手順を示している⁸⁾。

中小企業政策の評価にあたっては、政策形成に関与する者が個々のプロジェクトの評価を行うことは重要であるが、それ以上にそのようなプロジェクトを実施して達成される最上位の政策目的を常に考慮する必要がある。一般的に、中小企業政策を講じる目的として、中小企業育成あるいは起業家育成が掲げられることが多い。したがって、各省庁の個別のプロジェクトが最上位の政策目的を達成しなければ、政策そのものの意味を問われることになる。とりわけ、中小企業政策はその内容が多様であることから、関係者による相互評価（peer review）の仕組みを導入することが有効であるとの提案もある（OECD [2007], p. 68）。相互評価を導入することにより、異なる省庁に所属する担当者間でさまざまな情報を共有することが可能となり、中小企業政策の全体の目標を再認識することにつながる。とりわけ、成功事例を共有することは政策形成に有効に作用すると考えられる。

5. むすびにかえて

以上、先行研究で展開されている政策評価に関する議論をまとめてきた。政策評価の歴史は古いが、実務での必要性が高まったことから1950年代から政策評価が一般的な現象となったことが明らかになった。

政策評価を行う場合、「何を（対象）」「何のために（目的）」「どのように（方法）」評価するのかを厳密に規定することが重要である。この点に関する議論はおおむね収束に向かっており、政策（policy）が最上位概念として存在し、下位概念として施策（program）と事業（project）があると

8) 6つのアプローチには、それぞれ問題点もある。例えば、ステップ2において支援策が有益であったか否かを中小企業に意見聴取しても、十分な意見が出てこない場合も考えられる。また、ステップ5では、支援を受けた企業の中からその属性が一致する企業を探すことが難しいかもしれない。このような問題があることはStoreyも認識しており、Storey [1998] の中で論じられている。

されている。日本で一般的に用いられている「政策評価」とは、「プログラム評価」であることが確認できた。すなわち、評価の対象はプログラムであり、最上位の政策ではないということが明らかになった。

プログラム評価の方法については、定量的手法と定性的手法があり、評価の対象、目的、内容などにより多様な手法があることも明らかになった。定量的手法ではそれぞれの手法に弱点があることを認識すること、バイアスをいかに排除するかを考慮することが必要である。また、評価のトレードオフが存在するため、定量的手法と定性的手法を併用しながら総合的に評価することが求められる。また、本稿では政策評価においては画一的な評価手法は存在しないことを指摘した。

以上の点は、中小企業に焦点を当てた政策を評価する場合においても、同様である。本稿ではモニタリングと評価を明確に分け、モニタリングでは情報およびデータの収集を行い、それを元にいくつかの視点から比較分析する手法を紹介した。しかしながら、中小企業政策の形成および実施が複数の省庁によって行われている場合、深刻な問題を引き起こすこともある。その問題を解決するには、省庁横断的な相互評価（peer review）の仕組みが有効であるという提案があることを述べた。

このように、政策評価には多くの困難が伴うが、政府の説明責任を果たすためには欠かせない作業である。諸外国の知見を集約して効果的な政策評価を行うことが求められよう。なお、本稿ではプロジェクト評価に関する実証研究については触れていない。筆者の関心は中小企業性政策の効果であるが、当該分野の実証研究に関する文献調査を行い、どのプロジェクトにおいてどの評価手法が多く用いられているのかを把握する必要があるだろう。この点については、今後の研究課題としたい。

【参考文献】

<日本語文献>

青柳恵太郎 [2007] 「インパクト評価を巡る国際的動向」、佐々木亮・湊直信・高木桂一・青柳恵太郎『国際開発における評価の課題と展望』、財団法人国際開発高等教育機構国際開発研究センター、pp. 87-153.

石井芳明 [2010] 「中小企業・ベンチャー企業の公的支援策の政策評価に関する考察」、早稲田大学産業経営研究所、『産業経営』、第46・47合併号、pp. 53-69.

政策評価研究会 [1999] 『政策評価の現状と課題—新たな行政システムを目指して』木鐸社.

ピーターHロッシ・マークWリップセイ・ハワードEフリーマン（大島巖、平岡公一、森俊夫、元永拓郎監訳） [2005] 『プログラム評価の理論と方法—システム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（Rossi, Peter H., Mark W. Lipsey and Howard E. Freeman [2004] *Evaluation: A Systematic Approach* (seventh edition), Thousand Oaks: Sage).

源由理子 [2009] 「評価の評価（メタ評価）—その概念整理」、財団法人行政管理研究センター「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報告書」、pp. 1-25.

(http://www.soumu.go.jp/main_sosiki/hyouka/seisaku_n/chousakenkyu/houkoku_2103_00.pdf).

2012年3月21日ダウンロード).

龍慶昭・佐々木亮 [2000] 『「政策評価」の理論と技法』 多賀出版.

山谷清志 [1997] 『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.

<外国語文献>

OECD [1998] “Best Practice Guidelines for Evaluation,” PUMA Policy Brief No.5, OECD Public Management Service (<http://www.oecd.org/dataoecd/11/56/1902965.pdf>, 2012年3月19日ダウンロード).

—— [2007] *OECD Framework for the Evaluation of SME and Entrepreneurship Policies and Programmes*, Paris: OECD.

Storey, David J. [1998] “Six Steps to Heaven: Evaluating the Impact of Public Policies to Support Small Businesses in Developed Economies,” CSME Working paper No. 59, Warwick Business School, University of Warwick (http://www2.warwick.ac.uk/fac/soc/wbs/research/csme/research/working_papers/wp59-six_steps.pdf, 2011年11月23日ダウンロード).

—— [2006] “Evaluating SME Policies and Programmes: Technical and Political Dimensions,” Mar Casson, Bernard Yeung, Anuradha Basu, and Nigel Wadeson (ed.), *The Oxford Handbook of Entrepreneurship*, New York: Oxford University Press, pp. 248-278.

<ウェブサイト>

World Bank, Evaluation Designs,

<http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20188242~menuPK:412148~pagePK:148956~piPK:216618~theSitePK:384329~isCURL:Y,00.html> (2012年9月20日アクセス)

(2012年9月30日脱稿)