

日本評価研究

Japanese Journal of Evaluation Studies

Vol. 14, No. 2, December 2014

ミニ特集：プログラム評価の現状と可能性

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—
田辺 智子

科学的根拠に基づく実践とその形成評価アプローチが日本社会に定着しない現状と
要因～改善への示唆～

大島 巖

研究論文

証券アナリストと評価者による倫理向上への取り組み—職業倫理の比較研究—
小林 信行

研究ノート

「業績情報提供型」業績予算：その実践概要と使用される情報及び手法
荒川 潤

実践・調査報告

参加型モニタリング・評価手法MSC (Most Significant Change)
—バングラデシュNGOでの実践から4つの特色を考察する—

田中 博

『日本評価研究』編集委員会

Editorial Board

編集委員長
Editor-in-chief

山谷 清志(同志社大学)
Kiyoshi YAMAYA

副委員長
Vice-Editor-in-chief

西野 桂子(関西学院大学)
Keiko NISHINO

常任編集委員
Standing Editors

三好 皓一(立命館アジア太平洋大学)
Koichi MIYOSHI

牟田 博光(東京工業大学)
Hiromitsu MUTA

編集委員
Editors

青山 温子(名古屋大学)
Atsuko AOYAMA

岩淵 公二(NPO政策21)
Koji IWABUCHI

大島 巖(日本社会事業大学)
Iwao OSHIMA

岡本 義朗(新日本有限責任監査法人)
Yoshiaki OKAMOTO

小野 達也(鳥取大学)
Tatsuya ONO

窪田 好男(京都府立大学)
Yoshio KUBOTA

佐々木 亮(国際開発センター)
Ryo SASAKI

渋谷 和久(内閣府)
Kazuhisa SHIBUYA

田中 弥生(大学評価・学位授与機構)
Yayoi TANAKA

西出 順郎(岩手県立大学)
Junro NISHIDE

林 薫(文教大学)
Kaoru HAYASHI

松岡 俊二(早稲田大学)
Shunji MATSUOKA

源 由理子(明治大学)
Yuriko MINAMOTO

事務局
Office

〒108-0075 東京都品川区港南1-6-41 品川クリスタルスクエア 12階
一般財団法人国際開発センター内
特定非営利活動法人日本評価学会
TEL: 03-6718-5931, FAX: 03-6718-1651
E-mail: jes.info@evaluationjp.org

『日本評価研究』は、和文年2回、英文年1回を刊行します。

日本評価研究

第14巻 第2号 2014年12月

目次

ミニ特集：プログラム評価の現状と可能性

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－……………1

大島 巖

科学的根拠に基づく実践とその形成評価アプローチが日本社会に定着しない現状と要因
～改善への示唆～ ……………17

研究論文

小林 信行

証券アナリストと評価者による倫理向上への取り組み－職業倫理の比較研究－ ……………29

研究ノート

荒川 潤

「業績情報提供型」業績予算：その実践概要と使用される情報及び手法 ……………43

実践・調査報告

田中 博

参加型モニタリング・評価手法MSC (Most Significant Change)

－バングラデシュNGOでの実践から4つの特色を考察する－ ……………61

日本評価研究刊行規定 ……………79

日本評価研究投稿規定 ……………81

日本評価研究執筆要領 ……………83

日本評価研究査読要領 ……………86

【研究論文】

業績測定を補完するプログラム評価の役割 —米国GPRAMAの事例をもとに—

田辺 智子

国立国会図書館

stanabe0509@gmail.com

要 約

日本の政策評価では、業績測定（評価方式としては実績評価）が中心的な手法となっており、プログラム評価（総合評価）の実施は少ない。本稿では米国の事例をもとに、業績測定を改善するためにプログラム評価をどのように併用できるか検討を行った。米国では、政府業績結果法近代化法（GPRAMA）に基づき、業績測定だけでは業績を正しく把握できないという認識のもと、プログラム評価が積極的に併用されている。具体的には、妥当性の高い指標設定、業績の分析、施策改善のための意思決定などにプログラム評価が利用されている。

日本の政策評価では、指標設定の困難性や評価結果が利用されていない等の問題が指摘されるが、プログラム評価によってそうした点の改善が期待できる。

キーワード

業績測定、プログラム評価、インパクト評価、政府業績結果法（GPRA）、
政府業績結果法近代化法（GPRAMA）

はじめに

わが国の政策評価をめぐるのは、評価の質が不十分である、評価がマネジメントに使われていないといった問題がしばしば指摘される。本稿は、こうした問題の一因は、業績測定が中心でプログラム評価を適切に併用していないことにあるのではないかという問題意識から出発している。

わが国の政策評価で中心となっているのは、指標の実績値を継続的に測定する業績測定の手法である。政策評価制度では「実績評価」として位置付けられ、最も広い範囲の施策を対象に行われて

いる。これに対し、特定の施策を対象に掘り下げた分析を行うプログラム評価は、政策評価制度では「総合評価」として位置付けられているものの、その実施は必ずしも活発ではない。しかし、もともと業績測定とプログラム評価は補完的な関係にあり、近年、諸外国では両者の違いを生かして併用すべきという見方があらためて強調されている。

業績測定とプログラム評価が異なる種類の情報を生み出すのは周知のことであり、これまでも両者の補完性や日本でのプログラム評価の必要性は議論されてきた。本稿ではこれをさらに一歩進め、

業績測定を改善するために、プログラム評価をどのような形で併用できるかに焦点を当てて議論を行う。

1. 課題と方法

(1) 用語の定義

本稿では、プログラム評価と業績測定については、米国の政府監査院（Government Accountability Office: GAO）の定義を採用する。それによれば、業績測定（Performance Measurement）とは、施策の実績、特に事前に設定した目標や基準の達成度合いについての体系的で継続的なモニタリングである（GAO 2012, p.3）。このために、インプット、アウトプット、アウトカム等に関わる指標を設定し、定期的の実績値の測定を行う。これに対しプログラム評価（Program Evaluation）は、施策がどの程度、またどのように効果をあげているかを検証するために、研究的手法を用いてデータを収集・分析する体系的な研究である。

後述するように、業績測定はプログラム評価というところのモニタリングの手法を切り出して独立した評価手法として体系化したものであるため、業績測定をプログラム評価の一部として位置付ける考え方もありうる。しかし、本稿では両者の併用について議論するため、両者が重複していると捉えると混乱が生じる可能性がある。このため、本稿で議論の対象とするプログラム評価は、プログラム評価全体のうち、業績測定に当たるモニタリングを除いた部分とする。

なお、日本では、「評価」という用語が実績測定にもプログラム評価にも使用されるが、米国では“evaluation”という用語はプログラム評価を指す用語であり、評価専門家は通常、業績測定のことを“evaluation”とは呼ばない（Hatry 2013, p.24）。つまり、米国で“evaluation”と呼ばれているもの（＝プログラム評価）と、日本で「評価」と呼ばれているもの（＝プログラム評価＋業績測定）にはずれがあることに注意が必要である。本稿では、米国の用法にならない、プログラム評価のことは「プログラム評価」または「評価」、業

績測定のことは「業績測定」と呼ぶ。わが国の政策評価制度では業績測定とプログラム評価がどちらも含まれるため、政策評価制度について言及する場合は「政策評価」という用語を用いる。

本稿では、プログラム評価と業績測定の併用について議論するが、ここでの「併用」とは、業績測定を実施する過程でプログラム評価を併せて利用することを指す。その逆の方向、つまり、プログラム評価を実施する過程で業績測定を利用することもありうるが、本稿ではその方向については検討の対象外とする。「評価の利用」という概念については、田中（2014）が整理を行っている。田中は、「利用」は多面的な概念であるとして、①用途・目的、②利用対象、③利用者、の3要素に分けて概念整理を行い、実務や実証的研究の上では目的に応じてその都度「評価の利用」を具体的に定義する必要があるとしている（田中 2014, pp.142-146）。本稿では、「用途・目的」については、業績測定の目的の達成のためと定義する。業績測定の目的は実施する機関によって異なりうるが、一般的には、業績の改善とアカウンタビリティの確保が代表的な目的である。「利用対象」については、評価のプロセスではなく、評価の結果とする。評価の「利用者」については、業績測定を実施する主体、つまり行政内部の関係者とする。以上をまとめると、本稿でいうプログラム評価と業績測定の併用とは、「行政内部の主体が、業績測定の目的がよりよく達成されるようにするために、業績測定の実施過程でプログラム評価結果を用いること」と定義する。そしてその具体的な用い方を分析するのが本稿の目的である。なお、本稿では、「利用」、「併用」という用語を特に区別せずに用いる。

議論の対象としては国レベルの評価を想定する。評価の単位については施策（program）を念頭に置くが、議論の多くは事務事業レベルにも適用可能と考える。

(2) 課題と方法

本稿の課題は、事例分析をもとに、業績測定の実施過程で、プログラム評価をどのように併用できるかを明らかにすることである。そのために、まず文献調査によって、併用の前提となる業績測

定とプログラム評価の理論的關係を整理する。そこで焦点となるのは両者の性格の違いと補完性であるが、補完性については、論者によって想定する補完性の内容が異なることに留意して整理を行い、事例分析のベースとしたい。

その上で、事例として米国の政府業績結果法 (Government Performance and Results Act: GPRA) の改正法である政府業績結果法近代化法 (GPRA Modernization Act: GPRAMA) を取り上げる。GPRAMAは業績測定とプログラム評価の併用を想定しており、業績測定とプログラム評価の關係性を分析するという本稿の課題に合致する。また、米国は理論・実践の両面で世界の評価・業績測定をリードしてきており、政策評価制度の導入に当たってもGPRAがモデルとされていることから、わが国にとっても参考となる知見が得られるものと考えられる。業績測定の構成要素は、「指標の設定」、「目標水準の設定」、「実績値の測定」、「分析」、「報告」、「業績に関わる意思決定」の6要素に整理できる¹⁾。本稿では、米国の事例をもとに、この構成要素のうち、どの要素について、どのような形でプログラム評価が併用されているかを抽出する。政府全体の状況についてはGAOの報告書を素材とし、具体的な事例として労働省の取組みを取り上げる。

最後に、以上の分析を踏まえ、わが国の政策評価を改善する上でのプログラム評価の必要性について考察を行う。

2. 業績測定とプログラム評価の理論的關係

(1) 両者の性格の違い

現在のアウトカム重視の業績測定手法の生みの親であるハトリーは、業績測定はプログラム評価の一部と見なすことができると述べている (Hatry 1997, p.41)。では、業績測定はプログラム評価の体系のどこに位置付けられるのだろうか。

プログラム評価は、実施される文脈に応じて課題を設定し、それに合致した分析手法を採用して行われるものである。代表的なテキストである Rossi et. al (2004) にまとめられている評価の体系を表1の左側に示す。評価の観点としては、必要性評価、セオリー評価、プロセス評価、インパクト評価、効率性評価がある。一方、業績測定は、指標を設定して定期的・継続的にその実績値を測定するものであり、プログラム評価でいえば「モニタリング」という手法に当たる。インプットやアウトプットなど、施策の実施プロセスに関する指標を測定する場合は特に「プロセス・モニタリング」、アウトカム指標を測定する場合は「アウトカム・モニタリング」と呼ばれる。つまり業績測定で採用している手法は、表1の右側に示したように、プログラム評価の体系ではプロセス評価、インパクト評価の一部に位置付けることができる。

近年主流となっている業績測定は、このモニタリングの手法を特に切り出し、独立した評価手法として体系化したものである。そこでは、アウトカムの測定が特に重視されており、アウトカム・モニタリングの要素が強い。このアウトカム・

表1 業績測定とプログラム評価の守備範囲

評価の観点	プログラム評価	業績測定
	性格	
必要性評価	問題の性質と程度を分析し、施策の必要性を評価する。	
セオリー評価	施策の論理的な構造をロジックモデル等で明らかにし、その質や内容を評価する。	
プロセス評価	施策が意図したとおり実施されているかを評価する。	↑ ↓ プロセス・モニタリング (指標：インプット、活動、アウトプット) アウトカム・モニタリング (指標：アウトカム)
インパクト評価	施策による社会状況の改善の有無と程度を評価する。	
効率性評価	施策の費用対効果を評価する。	

(出所) 筆者作成。

表2 インパクト評価の手法

手法	施策実施前後のデータ計測	対照群の設定	性格
ランダム化比較試験 (RCT)	あり	あり	施策の実施前に、対象をランダムに介入群と対照群に分け、施策実施の前後の変化を比較する。
疑似実験デザイン	あり	あり	何らかの方法で介入群とできるだけ条件の似た対照群を設定し、施策実施の前後の変化を比較する。(回帰・分断モデル、マッチングモデル等)
統計解析法 (回帰分析)	あり	なし	サンプル中のデータのばらつきを利用し、指標の変化のうち施策の有無によって説明できる部分がどれだけあるかを統計的に推計する。(クロスセクション・モデル、時系列モデル、パネル・モデル等)
反復測定モデル	あり	なし	施策実施の前、途中、後の指標を測定する。
単純事前・事後比較	あり	なし	施策実施の前後の指標を比較する。
人為的判断による評価	なし	なし	専門家や行政官などが経験に基づいて評価を行う。(専門家評価、受益者評価、行政官評価等)

(出所) 筆者作成。

ニタリングをプログラム評価でいうところのインパクト評価の一種と考える場合、他のインパクト評価手法と比較して、どのような特徴を持つのだろうか。

インパクト評価には表2に示すような分析手法が用いられる。最も厳密な手法は、施策の対象者をランダムに介入群と対照群に分け、施策実施の前後の変化を比較するランダム化比較試験 (Randomized Controlled Trial: RCT) である。これにより、施策と成果との因果関係の有無や、施策以外の外部要因の影響を取り除いた施策の純粋な貢献度を測定することができる。表2では、概ね下に行くほど科学的厳密性が低い手法となる。アウトカム・モニタリングは、表2でいえば反復測定モデルに近いものである。このモデルは、施策実施に伴う指標の変化を測定するが、施策の非対象者との比較を行わないため、アウトカム指標が改善したとしても、それが施策がもたらした効果なのか、偶然や施策以外の外部要因によるものなのかを区別することができない。つまり、アウトカムに焦点を当てた業績測定は、インパクト評価の一種と考えることができるが、精度としてはかなり低いということになる。

ではなぜ、プログラム評価の一部ともいえる業績測定が、独立した業績検証手法として発展し、普及したのだろうか。米国における政策評価の歴史をたどると、プログラム評価の発展が業績測定に先行している³。プログラム評価が今ある形で

行われるようになったのは1960年代からであるが、初期のプログラム評価では、施策が意図した効果をあげているかを厳密に検証するインパクト評価に主眼があった。当時は、低所得者への所得移転が労働供給に与える影響についての評価などで、RCTが数多く実施された。しかし、RCTには長い期間を要するため意思決定に必要なタイミングで評価結果が得られないことが多く、分厚く詳細な報告書は必ずしも現場で利用しやすいものではなかった。このため、評価は科学的に厳密であるよりも、現実の政策改善に役立つものであるべきという考え方が生まれ、評価の利用者、つまり評価を意思決定に利用する議会や行政府、現場の管理職者等のニーズに合った情報を産出すべきという、実用重視の評価が生まれた。

今ある形の業績測定は、こうした実用重視の方向性の中で生まれたものである。業績測定自体はそれ以前からあったものではあるが、アウトカム重視の今ある形の業績測定は、地方政府が現場で使いやすい簡便なデータ収集法として実用重視評価の流れの中で開発された。この経緯からわかるように、まず厳密なプログラム評価が先にあり、その問題点を解決するための一方策として業績測定が開発されている。つまり、業績測定はもともと、プログラム評価を補完するものとして発展したことになる。

(2) 不足補完性

続いて、業績測定を生みの親であるハトリーの最近の著作をもとに、両者の補完性の中身を詳しく見てみたい (Hatry 2013)。ハトリーによれば、業績測定では、多くの施策を対象に、タイムリーな情報を定期的に提供することができる。実施に当たってはプログラム評価ほど特別なスキルが必要とされず、施策1件当たりで見れば比較的、費用がかからない。そのかわり、得られる情報は浅く、科学的厳密性には限界がある。一方、プログラム評価では、信頼性の高い詳細な情報が得られるものの、実施には年単位の期間と多額の費用がかかるため、少数の限られた施策に不定期にしか行うことができない (Hatry 2013, p.27)。

つまり業績測定とプログラム評価は、それぞれ異なった種類の情報、それももう一方が提供できない性質の情報を提供できるという意味で、補完的な関係にある。この関係は、定期健診と精密検査の関係にたとえると理解しやすいだろう。定期健診では健康状態を見る上でポイントとなる多数の項目を毎年度測定し、定期健診の結果、異常な数値が見つかったら精密検査を行って異常の原因や病気の有無を検査する。これと同様に、業績測定ではポイントとなる指標を幅広く測定し、必要に応じて詳細なプログラム評価を行って指標の背景を分析することが考えられる。

本稿では、このハトリーの指摘する補完性を、互いに足りない点を補い合うという意味で「不足補完性」と呼ぶことにする。

(3) 相乗補完性

以上で見た業績測定とプログラム評価の不足補完関係を前提とすれば、両者の性格の違いを生かして併用するのが合理的ということになる。しかし、両者を併用すべきという議論は、必ずしも評価専門家の間で共有されていたわけではなかった。

米国評価学会における議論を見ると、評価専門家の間には、業績測定に対する根強い懐疑が存在してきたとされる (Nielsen and Hunter 2013, p.117)。RCTのような厳密な手法でプログラム評価を行ってきた専門家の立場からすれば、業績測定から得られる情報は粗いものであり、それを政府のマネジメントに利用することに懸念が示されたのは、ある意味、自然な反応だったともいえる。

しかし、業績測定の実施経験が蓄積された現在、業績測定とプログラム評価の関係をあらためて問い直す動きが出ている。米国評価学会の発行する雑誌 *New Directions for Evaluation* の2013年春号では、「業績マネジメントと評価」という特集が生まれ、業績測定とプログラム評価の関係について様々な国やセッティングでの事例が報告されてい

表3 業績測定とプログラム評価の補完性

名 称		内 容
不足補完性		業績測定は、情報としては浅いが、プログラム評価がカバーできない幅広い施策について高頻度の情報を提供する。プログラム評価は、少数の施策にしか実施できないが、業績測定が産出できない掘り下げた情報を提供する。
相乗補完性	連続補完性 (Sequential complementarity)	業績測定の結果から、プログラム評価で解明すべき問いが明らかになる。あるいは逆に、プログラム評価によって、継続的なモニタリングが必要な点が明らかになる。
	情報補完性 (Information complementarity)	業績測定とプログラム評価は同じデータソースを共有できるため、情報収集ルートを並立して作成する必要がない。
	組織補完性 (Organizational complementarity)	業績測定とプログラム評価は、しばしば組織の同じユニットで行われる。
	手法補完性 (Methodical complementarity)	データ収集、分析等の面で、業績測定とプログラム評価には共通する部分がある。
	階層補完性 (Hierarchical complementarity)	国レベル、政策レベルの業績測定で収集された情報は、地方レベルのプログラム評価で比較対象やベンチマークとして利用できる。

(出所) 筆者作成。連続補完性から階層補完性については、Nielsen and Hunter (2013), p.120による。

る。

この特集を編集したニールセンとハンターは、業績測定と評価については、補完関係が主張されてきたわりにはその理論化が進んでいなかったとして、5種類の補完性を提示している。業績測定から評価で明らかにすべき問が明らかになる、あるいは逆に評価からモニタリングが必要な点が明らかになるという「連続補完性」、両者が情報収集ルートを共有できることによる「情報補完性」、両者を組織の同ユニットで実施することによる「組織補完性」、データ収集や分析面での共通性による「手法補完性」、国レベルの業績測定データが下位の地方レベルの評価で利用できる「階層補完性」の5種である。前節で見た不足補完性は、互いに足りない点を補い合うという性格の補完性であったが、ここでニールセンらが提示している補完性は、それとは性格が異なり、両者を併用することで双方がより効率的・効果的に行え、相乗効果が上がるという性格の補完性となっている。本稿では、このタイプの補完性を前述の不足補完性と区別するために、「相乗補完性」と呼ぶ。

以上で見た業績測定とプログラム評価の補完性は、表3のようにまとめることができる。

3. 米国GPRAMAに見るプログラム評価の併用

以下では、まずGPRAMAの概要とプログラム評価の位置付けを確認した上で、GPRAMAの中でプログラム評価がどのように併用されているかを分析する。業績測定の構成要素は、前述のように、「指標の設定」、「目標水準の設定」、「実績値の測定」、「分析」、「報告」、「業績に関わる意思決定」の6要素に整理できる（表4左側）。これらは概ねこの順番に実施されるが、報告を行った後でさらに詳細な分析を行う、といったように前後する場合もありうる。これらの構成要素ごとに、具体的な評価の併用の例を抽出する。

(1) GPRAMAにおけるプログラム評価の位置付け

1993年に成立したGPRAは、連邦政府に業績測定を導入するものだった。政府機関は中期的な戦略計画を定め、毎年度、年次業績計画で目標と指標を設定し、その結果を年次業績報告で報告することが求められている。それから18年後の2011年に成立した改正法であるGPRAMAでは、政府全体の優先目標の設定、その四半期毎のレビュー、目標達成に責任を持つ目標責任者の任命等を含む

表4 業績測定の構成要素とプログラム評価併用の類型

	業績測定の構成要素		プログラム評価併用の類型
	構成要素	内容	
1	指標の設定	施策等の評価対象ごとにアウトカムを特定し、指標を設定する。	・インパクト評価で施策実施とアウトカムの因果関係を確認した上で、妥当性の高い指標を設定する。 ・アウトカム指標の設定が難しい場合に、評価によって中間アウトカムを特定して指標とする。
2	目標水準の設定	個々の指標について目標値やベンチマーク対象を設定する。	—
3	実績値の測定	定められた頻度で実績値を測定する。	—
4	分析	実績値を目標水準と比較し、問題がある場合は原因等を分析する。	・インパクト評価によって他施策の影響や外部要因を取り除いた施策の効果を測定する。 ・既存のプログラム評価をレビューし、提言を抽出する。
5	報告	測定結果を報告する。	・指標の推移だけではわからない詳細な業績情報を評価で明らかにして提供する。 ・施策の有効性について自治体等の関係者に情報を提供する。
6	業績に関わる意思決定	必要に応じて業績向上のための意思決定を行う。	・施策のデザインや実施を見直すための判断材料・裏付けとする。 ・予算要求、資源配分の裏付けとする。

(出所) 筆者作成。

改定が行われ、業績情報の利用をさらに促進することが目指されている（GAO 2013a, pp.5-7）。

プログラム評価については、GPRAの時代から、戦略計画には目標設定に利用したプログラム評価と将来的なプログラム評価実施スケジュール、年次業績報告には対象期間中に完了したプログラム評価結果の要約を含めることが求められていた（同法Sec.306.(a)(6)；Sec.1116.(d)(5)）。GPRAMAでも、これらの規定に変更は見られない（同法Sec.306.(a)(8)；Sec.1116.(c)(7)）。

現オバマ政権では、プログラム評価の役割について、行政予算管理局（Office of Management and Budget: OMB）のメモランダムという形で方針が示されている。2009年の最初のメモランダムでは、連邦政府は長い間プログラム評価に投資してきたが、多くの重要な施策が一度も評価されたことがなく、有効性が確認されないまま毎年存続しているという認識が示され、新規の評価提案の募集が行われた（OMB 2009, pp.1-2）。2012年と2013年のメモランダムでは、予算要求でのプログラム評価とエビデンスの利用について記述している（OMB 2012; OMB 2013）。ここでいうエビデンスに含まれるのは、プログラム評価、業績指標、その他のデータ分析や研究結果などであり、中でも質の高い実験的・疑似実験的手法がのぞましいとされている（OMB 2013, p.2）。つまり、エビデンスには幅広い業績情報が含まれるが、最良のエビデンスとなるのはRCT等による厳密なインパクト評価だということになる。

2013年のメモランダムによれば、省庁は有効性についての強固なエビデンスに裏付けられた施策に予算を振り向けることが推奨されている。新規の政策提案や年次業績計画には、関連するエビデンスについての詳解を添付することとされた。また、既存施策の改善や、新規施策の意思決定に利用しうるプログラム評価の実施が推奨されている。厳しい予算制約があるため、重要な政策課題の解決に資する評価や、RCTや疑似実験などによる少数の質の高い評価に特に注力することが推奨されている（OMB 2013, p.2）。

OMBは、省庁のエビデンスの産出や利用を後押しするため、省庁職員への助言や技術的支援を行うとともに、エビデンス産出・利用の先進的な

事例を紹介している。たとえば、複数省庁の行政データを統合したプログラム評価、低コストで迅速なインパクト評価方法、補助金支給にあわせて評価を義務付ける方式等である（OMB 2013, p.3, 6-11）。

(2) 政府全体におけるプログラム評価の併用

GAOは、GPRAMAに基づいて同法の執行状況を調査しているが、その一環としてプログラム評価の利用についての調査を行った（GAO 2013b）。14万8,300人の連邦政府の部局長・監督者クラスの行政官から4,391人を無作為抽出しアンケートを行うとともに、プログラム評価経験の豊富な機関及びOMBを対象にインタビューを行っている。

①プログラム評価の利用状況

アンケートでは2,762人から有効な回答が得られたが、そのうち、自らが関わる施策や事業で過去5年間に完了したプログラム評価があると回答した行政官は37%であった。この37%の行政官に対し、当該プログラム評価を施策マネジメントや政策立案活動にどの程度、利用したかを質問した結果が図1である⁴。各選択肢が表4でまとめた業績測定の構成要素のどこに関係するかを見ると、「指標の設定」、「分析」、「報告」、「業績に関わる意思決定」に該当していることがわかる。

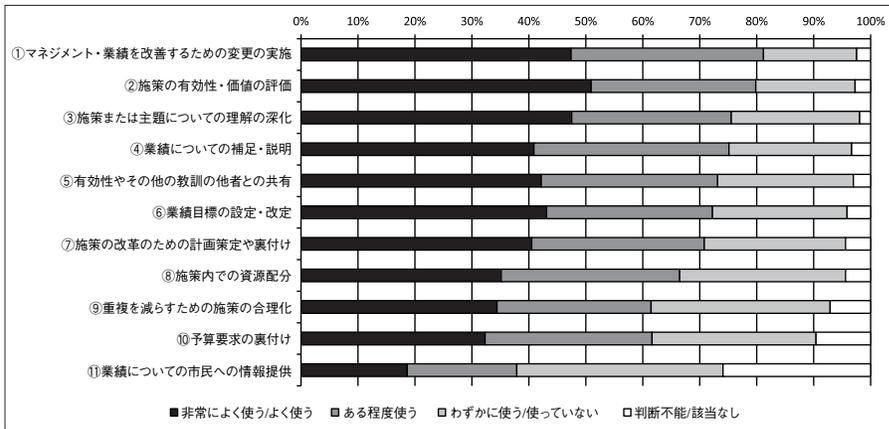
<指標の設定>

選択肢には直接的に指標の設定に言及したものはないが、「⑥業績目標の設定・改定」は、施策の目指す目標やアウトカムの特定を意味していると考えられ、指標の設定に関係すると見てよいだろう⁵。この利用目的については、7割以上の回答者がプログラム評価を「非常によく使う/よく使う」または「ある程度使う」と回答している。

<分析>

選択肢のうち、分析のための利用に当たるものは、「②施策の有効性・価値の評価」、「③施策または主題についての理解の深化」、「④業績についての補足・説明」である。これらは業績測定だけではわからない詳細な業績情報を把握するための利用に当たり、7~8割の回答者が「非常によく使

図1 プログラム評価の目的別の利用状況



(注) 過去5年以内に完了した評価があると回答した人数を母数として割合を算出。無回答を除く。

(出所) GAO 2013b, p.9. (原データはGAO. *Managing for Results: 2013 Federal Managers Survey on Organizational Performance and Management Issues*, GAO-13-519SP, an E-supplement to GAO-13-518.)

う/よく使う」または「ある程度使う」と回答している。

<報告>

「⑪業績についての市民への情報提供」は、他の選択肢に比べ利用割合は低いが、アカウントビリティ目的での利用に当たり、表4でいえば「報告」に関わる利用といえるだろう。また、「⑤有効性やその他の教訓の他者との共有」も、報告に関係する選択肢といえる。GAOの説明によれば、連邦政府の施策では具体的な運用が州・地方政府に委ねられ連邦が直接関与していない場合も多く、そういったケースでは評価の結果明らかになったエビデンス等の情報を州・地方政府に提供することが、連邦政府にできる最も直接的な関与となるとのことである (GAO 2013b, p.10)。

<業績に関わる意思決定>

施策の業績改善のための意思決定に関係する選択肢としては、「①マネジメント・業績を改善するための変更の実施」、「⑦施策の改革のための計画策定や裏付け」、「⑨重複を減らすための施策の合理化」がある。このうち、①は実施プロセス等の比較的小さい見直し、⑦と⑨は施策全体の方向性に関わるような、より大きな見直しに相当すると考えられる。①の小規模な見直しのほうが⑦や

⑨の大規模な見直しよりも利用が多く、①については回答者の約8割が評価を「非常によく使う/よく使う」または「ある程度使う」と回答している。OMB担当者へのインタビューでは、プログラム評価利用の一般的な形として、ある施策を存続するか否かといった重要な利害に関わる問題に評価が答えを与えるというよりは、施策の業績をどうしたら改善できるかといった利害に関わらない意思決定に利用される傾向があると述べられており (GAO 2013b, p.14)、このアンケートでも整合的な結果となっている。

資源配分への意思決定に関わる選択肢としては、「⑧施策内での資源配分」、「⑩予算要求の裏付け」があるが、これらの目的に利用するとした回答者は相対的には多くない。

②プログラム評価利用の促進策

この報告書では、プログラム評価の利用状況に加え、利用の促進策が分析されている。インタビュー対象機関の評価担当者が強調した利用促進の基本的な戦略としては、評価を後押しするリーダーシップ、強固なエビデンスの確立、評価過程へのステークホルダーの関与の3点があげられている (GAO 2013b, p.24)。二点目の強固なエビデンスについては、多くの場合、1件のプログラム評価のみに基づいて意思決定をすることはなく、施

策実施の様々な段階での評価や、異なるサンプルを用いた評価が蓄積されることによって、より理解が深まり、確信を持って行動に移せると述べられている（GAO 2013b, p.27）。評価手法については、必ずしもRCTでなければいけないわけではなく、どの手法を用いる場合であっても厳格な実施と評価の質が重要であると指摘されている（GAO 2013b, p.26）。

(3) 省庁におけるプログラム評価併用の取組み

業績測定とプログラム評価を積極的に併用している例として、労働省を取り上げる。労働省は、GPRAMA関係文書の中でプログラム評価について詳細に言及しており、前項で見たGAOの報告書でも省内の2局が評価経験の豊富な機関としてインタビュー対象となるなど、従来からの評価の蓄積も有している。

① 労働省のプログラム評価

労働省のプログラム評価についての方針は、最新の戦略計画（2011-2016年度）にまとめられている。そこでは、個々の施策が目標を達成しているか、また業績を改善するためにどのような見直しが必要かを理解する上ではプログラム評価が最善の方法であり、業績指標から得られる情報だけでは不十分であるという認識が示されている（Department of Labor 2011, p.85）。同省では、政府全体の方針に沿い、プログラム評価のための施策を強化し、特にインパクト評価に力を入れている。また同省は、全省のプログラム評価を統括するため、2010年に最高評価責任者（Chief Evaluation Officer: CEO）という役職を新設している（Department of Labor 2013, p. 84, 87）。

労働省の評価予算については、同省の予算概要書にまとめられている（Department of Labor 2013）。もともと、雇用・職業訓練の施策については、労働力投資法（Workforce Investment Act of 1998）でプログラム評価を実施するよう規定されており、そのための予算は2012年度で956万ドルであった。さらに労働省は、2010年度から、それ以外の施策の評価を拡充しており、2009年度には350万ドルだった予算を2012年度には848万ドルまで拡大している。また、2014年度の予算要求の中

では、主要施策の予算の1%までを評価に割り当てることを要求している（Department of Labor 2013, p.16）。件数としては、2006年度以降で40件以上のプログラム評価が実施されており、ウェブサイト分野別、手法別に閲覧可能となっている⁶。

② 業績測定とプログラム評価の関係

労働省では、前述のようにインパクト評価に力を入れており、特に施策と目指す成果との因果関係の確認が重視されている。たとえば、査察の実施が法令遵守状況を改善することが評価によって確認されれば、査察を増やすのが合理的な選択ということになる。評価のねらいは、こうした因果関係を明らかにし、業績指標の精査や、戦略・施策の改善につなげることだとされている（Department of Labor 2013, p.84）。また、労働省の評価担当者は、業績測定は、着目すべき点を明らかにすることで将来的なプログラム評価のアイデアの源泉となるという見解を示している（GAO 2013b, p.13）。

プログラム評価が特に大きな影響を与えた例として、労働者保護行政がある（Department of Labor 2011, p.85; Weil 2010）。労働省は、賃金や労働時間等の労働基準の遵守状況について、様々な業界を対象にした評価を実施した。その結果を受け、2011-2016年度の戦略計画では、従来の監督業務を大きく見直し、各業界の状況に応じた監督方法を採用する戦略策定が行われた。そうした新しい監督方法については、今後プログラム評価を実施し、戦略策定にフィードバックするとしている。同時に、業績指標についても改定が行われた。（表5項番2, 7を参照。）ここでの評価の利用は、表4でいえば「指標の設定」、「業績に関わる意思決定」での利用に当たるだろう。

(4) プログラム評価併用の具体例

ここまで用いてきた資料をもとに、業績測定に関連してプログラム評価が利用された具体例をまとめたのが表5である。労働省の例が中心であり網羅的なものではないが、プログラム評価の併用にどのような形がありうるか考える上で参考になる。

表5では、表4で示した業績測定の構成要素のう

表5 業績測定におけるプログラム評価併用の具体例

業績測定の構成要素	項番	プログラム評価併用の例
指標の設定	1	労働省では、雇用者の権利保護のため、雇用者が自らの権利について理解を深め違反に対し仕返しを恐れず行動できるようにするためのアウトリーチ活動を行っている。これにより、職場での雇用者の「声」を強くし、最終的にはより安全で健康的な職場を作るのがねらいである。しかし、これまでの「声」の強さを測る適切な指標がなかったため、一連のプログラム評価を実施し、アウトリーチ戦略の成果を測る指標の設定を目指している。(Department of Labor 2011, p.88)
	2	労働省では、プログラム評価をもとに労働者保護行政の全面的な見直しを行ったが(項番7参照)、それに合わせて指標の改定を行った。従来は査察や取り調べの成果を指標としていたが、それでは対象となった一部業者の遵守状況しかわからないため、新たにサンプル調査による対象業者全体の法令遵守状況をアウトカム指標として採用した。(Department of Labor 2011, p.13)
分析	3	労働省では、失業保険の効果を、他の低所得者向け施策の効果と切り離して把握するため、州レベルのデータを用いて過去の景気悪化時の分析を行った。その結果、失業保険のスタビライザーとしての有効性が確認された。(GAO 2013b, p.12)
	4	財務省では、人々の行動には社会規範が大きな影響を与えるという社会心理学の知見に基づき、滞納者への督促状の内容について、簡易な表現を用い社会規範に言及する等の変更を行った。その変更の効果はRCTによって検証され、従来の督促状より支払率が高いことが確認された。(OMB 2013, p.7)
	5	早期教育の評価・研究に関する諮問委員会が2012年にまとめたレポートは、厳密なプログラム評価を含む大量の早期教育についての知見をレビューし、将来の研究、政策、実践のための多数の提言を行った。(GAO 2013b, p.12)
意思決定への利用	6	低所得者の食料購入を補助する農務省の補助的栄養支援プログラムでは、かつてはフードスタンプという金券を支給していたが、一連の費用対効果分析の結果、スーパー等のレジの支払に使用可能なデビットカード形式で支給する方式に変更された。(GAO 2013b, p.10)
	7	労働省は、賃金や労働時間等の労働基準の遵守状況について様々な業界を対象にした評価を実施した。その結果、近年の経済構造の変化、労働組合の機能低下、フランチャイズや業務委託等による雇用形態の多様化等のため、査察や取り調べを中心とした従来型の監督業務では、たとえ投入資源を増やしたとしても大きな効果は望めないこと、業界によっては現場レベルではなく業界上層部への働きかけが有効であること等の知見が得られた。労働省ではこれを受け、2011-2016年度の戦略計画では、各業界の状況に応じた監督方法を採用した。(Department of Labor 2011, p.30, 85)
	8	保健社会福祉省では、糖尿病にかかるリスクが高い人々を対象とした予防プログラムを策定する際、過去のプログラム評価等のエビデンスのレビューを行った。糖尿病予防と強い関連のある要素として減量、運動習慣、ストレス管理等をエビデンスに基づいて抽出し、予防プログラムに反映した。このエビデンスの裏付けが、予防プログラムが関係者の支持を得る上で役立ったとされている。(GAO 2013b, p.10)
	9	米国の失業者支援策では、各々の州や地方政府が、職業訓練のための給付金の金額や、カウンセリング・ガイダンス等のサービス実施内容を自ら決定する。労働省では、給付金額やサービス内容の違いが、どのように職業訓練への参加や参加者の所得に影響しているかを比較するプログラム評価を行った。その結果、より高額で柔軟な支給方法の費用対効果が高く、求職者の長期的な所得に正の効果があることが明らかになった。労働省では、評価結果を施策の意思決定に役立ててもらうため、州や地方の施策実施部門に情報提供を行った。(GAO 2013b, p.11)

(出所) GAO (2013b); OMB (2013); Department of Labor (2011) をもとに作成。

ち、「指標の設定」、「分析」、「業績に関わる意思決定」での具体例が含まれる。「指標の設定」では、適切なアウトカム指標がない場合に評価を用いて指標を設定しようとしている例(項番1)などがある。「分析」では、インパクト評価によって他施策の影響を取り除いた当該施策の純粋な効果を明らかにした例(項番3)、既存のプログラム

評価をレビューすることによって提言をまとめた例(項番5)などがある。「業績に関わる意思決定」では、施策の基本的な枠組みはそのまま実施方法を変更した例(項番6)、施策のそもそものデザインを見直した例(項番7~9)がある。こうした場合、それに伴って指標が見直される場合がある(項番2)。

ここでの例は、政府文書で特に言及されていることから、比較的目立つ大きな事例が多いと考えよう。これら以外にも、より小規模で日常的な利用がされている可能性がある。

(5) 米国の事例を理解する上での留意点

米国の事例を理解する上で重要だがここまででカバーできなかった点を補足すると、まず第一に、プログラム評価の浸透という点では、米国の政府内部でも省庁によって格差があるということである（GAO 2013b, p.25）。このため、GAOの報告書は、評価を従来から特に積極的に実施・利用してきた省庁のグッド・プラクティスを政府全体に紹介するものとなっている。

第二に、評価を実施する側も、意識して利用されやすい評価をめざしていることである。前述のように、米国では、かつて、せっかく実施されたプログラム評価が利用されないことが問題視され、その反省から評価を利用する人々のニーズに沿った評価をするという実用重視評価が生まれた。GAOの報告書では、評価の利用を促進するための方策として、ステークホルダーとの連携重視等があげられている。

第三に、プログラム評価を施策の変更につなげるためには、ある程度の評価の蓄積が必要となることである。GAOの報告書でインタビュー対象となった評価担当者達は、単一のプログラム評価で施策を変更することは通常なく、複数の評価・研究の結果が時間をかけて施策に影響するのだと述べている（GAO 2013b, p.10）。

4. 結論

(1) 米国の事例に見る業績測定とプログラム評価の補完性

以上の事例分析から、米国GPRAMAでは、3(3)①で見た労働省の戦略計画の記述に端的に見られるように、業績指標だけでは業績を正しく把握できないという認識のもと、プログラム評価の併用が進められていることがわかる。これは、表3でいうところの不足補完性の認識がベースになっているといえる。これに加え、相乗補完性のうちの

連続補完性も確認できる。3(3)②で見たように、労働省の評価担当者は、業績測定によって将来的なプログラム評価のアイデアが明らかになると述べている。

その他の種類の相乗補完性は、今回の事例分析では確認できなかった。情報収集ルートの共有による情報補完性、データ収集・分析面の共通性に基づく手法補完性については、米国のケースではどちらかといえば逆で、業績測定では得られないデータをプログラム評価で入手し、業績測定とは異なる手法を用いて詳細な分析を行うという形が多いものと見られる。業績測定とプログラム評価が組織の同じユニットで行われる場合に見られるとされる組織補完性についても、今回の米国の分析では確認できなかった。また、国レベルの業績測定で得られた情報を地方レベルのプログラム評価で利用するという階層補完性については、今回は連邦政府のみを分析対象としたために確認できなかった。これらの相乗補完性の分析は今後の課題である。

(2) 業績測定とプログラム評価の併用の形

具体的にどのような形での併用が可能か、事例分析から明らかになった類型をまとめたのが表4の右側である。業績測定の構成要素でいえば、「指標の設定」、「分析」、「報告」、「業績に関わる意思決定」において、評価が併用できるという結果となった。

<指標の設定>

まず、プログラム評価は、妥当性の高い指標の設定に利用できる。労働省では、施策とアウトカムの因果関係をインパクト評価で確認することが重視されており、それを施策の見直しにつなげるとともに、指標に反映させてマネジメントに組み込むことを目指している。わが国でも、セオリー評価のツールであるロジックモデルを用いることで、インプットからアウトプット、アウトカムまでの流れを明らかにし、指標の設定が容易になることは評価専門家によってしばしば指摘されている。ただし、その際は、単にロジックモデルを作成するだけでなく、既存研究との照合や論理的妥当性の吟味によるセオリーの評価を行うのが本来

の姿である(田辺 2012)。このロジックの確認をインパクト評価をもとに行えば、最も妥当性の高い指標の設定が可能となるだろう。

また、表5の項番1のように、適当なアウトカム指標がない場合に、プログラム評価によって中間アウトカムを特定し、指標に利用することも考えられる。項番1の例では、最終アウトカムは安全で健康な職場の実現だが、その中間アウトカムである職場での雇用者の「声」を測る指標を設定しようとしている。

<分析>

プログラム評価は、業績についての詳細を分析し、施策の有効性について理解を深めるためにも利用できる。表5の項番3のように、異なる省庁や地方レベルで競合する事業がある場合に、他の事業の効果を除外して特定の事業の真の効果を見極めるという利用も可能である。

業績測定に伴うデータの分析には内訳数の計算やアンケート調査など様々な方法があり、必ずしもプログラム評価を行わなければならないわけではないが、国レベルの施策のように施策以外の外部要因の影響が大きい場合には、プログラム評価の有用性は特に大きいと考えられる。

<報告>

業績測定は、指標という形で端的にわかりやすい業績情報を提供できるものではあるが、アウトカム指標の推移は施策の効果そのものではないため、そのままでは誤解を与える場合もありうる。米国の例では、プログラム評価による詳細な情報が、市民へのアカウントビリティ確保の目的でも利用しうることが示された。

また、国レベルの政策を受けて施策や事業を実施する地方政府などに対し、プログラム評価結果について情報提供し、有効性の高い施策を普及するためにも利用されている。

<業績に関わる意思決定>

指標が目標水準を達成しなかった場合、原因を分析して、施策の見直しを行うのは業績測定に想定される機能の一つである。労働省やその他の省庁の例からは、プログラム評価が様々な形の施策

の改善に利用できることが明らかになった。施策のデザインの根本的な見直しや、新たな施策の立案に利用する場合もあれば、施策のデザインはそのままでも実施方法に改善を加える場合もある。また、予算要求や資源配分の裏付けとしても評価が利用できる。

5. 日本への示唆

(1) 日本の政策評価の課題とプログラム評価の必要性

以上、本稿では米国の事例を分析した。米国一國のみの事例であるため、そこから得られた知見が、制度や文脈の異なる他の国にどこまで適用できるかは議論の余地があろう。しかし、米国はプログラム評価においても業績測定においても世界で最も長い経験を持ち、理論面でも世界をリードしてきた国である。業績測定とプログラム評価の併用についても20年の経験を持ち、そこから導かれる併用のあり方は、他国にとっても一つのモデルを提示するものといえる。本稿の最後に、米国の事例分析から得られた知見を踏まえ、日本の政策評価を改善する上でのプログラム評価併用の必要性について考察を行う。

日本の政策評価制度で定められた評価方式でいえば、業績測定は「実績評価」、プログラム評価は「総合評価」にそれぞれ対応する。そして、わが国の場合、評価方式は行政府省の選択に委ねられており、実態としては事後評価については業績測定がその中心となっている。総合評価の実施件数は全省庁合計で年間20~40件台で推移しており、活発に実施されているとはいえない⁸。また、総合評価では分析方法の自由度が高く、本稿で見たようなプログラム評価の手法は必ずしも採用されていない(田中 2012)。

日本の「実績評価」、すなわち業績測定の課題として、多くの論者が共通して指摘しているのは指標設定の困難性である。指標としてはアウトカム指標が望ましいとされながらも必ずしも設定できておらず(原田 2011, p.5)、指標の妥当性や測定の品質という点でも問題を抱えている(小野 2013)。このため、目指す方向としては、アウト

カムと施策実施の因果関係を分析し、それに基づいて指標を設定するのが望ましいと指摘されている（林 2004, pp.193-194）。つまり、施策の実施からアウトカム発現までの因果関係を確認することで、アウトカムに焦点を当てた妥当性の高い指標の設定が可能になるということである。本稿で見たように、これを実現するためには、プログラム評価の実施が有効な選択肢となる。

もう一つ指摘されている重要な課題は、分析の欠如である。測定結果の分析は、業績測定が真価を発揮するために極めて重要であるにもかかわらず、わが国では軽視されていると指摘される（田中 2014, p.232）。指標の実績値を測定するのみならず、分析を行わなければ、判断や意思決定に有用な知見は得られない。分析の方法が必ずしもプログラム評価でなければならないわけではないが、国レベルの施策における因果関係の確認では、プログラム評価を視野に入れる必要があるだろう。

この指標設定と分析の欠如が相まって、業績測定が意思決定やマネジメントに利用されないという現状につながっているものと考えられる。本稿で見てきたように、業績測定における指標設定、分析、意思決定への利用という面で、プログラム評価に期待できる効果は大きい。たとえ時間がかかっても、米国の事例を参考に、プログラム評価の実施を進めるべきではないだろうか。

(2) 米国の事例からの示唆

これまでプログラム評価の蓄積がほとんどない日本で、今後、業績測定とプログラム評価の併用を目指す場合、米国の事例から以下のような点が示唆される。

第一に、業績測定を「使える」ものにしていくためには、ある程度のコストがかかることを覚悟する必要がある。もともと、業績測定の仕組みを作るには相応の時間と資源がかかることが指摘されている（Newcomer 1997, p.11）。わが国でも、コストをかけずに実施する政策評価では、それなりの結論しか出ないのは当然だといわれる（山谷 2012, p.23；西出 2011, p.20）。このコストには、業績の測定そのもののコストだけではなく、必要に応じてプログラム評価を実施するコストを含め

て考えるべきだろう。諸外国では、財政状況が厳しいからこそ、有効な施策に予算を振り向ける手段としてエビデンスに基づく施策が志向されている。プログラム評価にかかる経費は、政府全体の施策を改善し、業績を高めるために必要なものと理解すべきだろう。

第二に、プログラム評価の併用を目指すためには、プログラム評価の実施を後押しする仕組みが必要である。米国の労働省の例のように、重要な政策については根拠法でプログラム評価の実施を義務付ける、予算の一定割合を評価に割り当てる等の制度を検討すべきだろう。また、米国で行われているような、評価の義務付けを伴う補助金支給方式や、低コストで評価を行えるような手法も研究の余地がある（OMB 2013, p.3, 6-11）。

第三に、プログラム評価の実施に必要な人材育成の必要性である。米国では厳密なインパクト評価は基本的に外部委託で実施されており、日本でも内部評価というよりはコンサルタントや大学、学術研究機関等への委託が選択肢となるだろう。ただし、米国では政府内部にも評価の専門知識を持った担当者があり、テーマや手法の選定、発注管理に携わっている。そうした人材の確保や、府省の評価担当者を技術的に支援する仕組みも必要となる。プログラム評価の手法自体はすでに確立されたものがあるため、それを吸収し、個々の施策に応じて実施していく能力が求められる。

第四に、プログラム評価が蓄積されるまでの間は、諸外国で行われたプログラム評価を翻訳、紹介していくことも意味があると考えられる。社会経済状況が異なるわが国でその結果が一般化できるかは検証の必要があるが、その点を明らかにするために、諸外国で行われた評価と同じ手法を用いた評価を計画することもできる。また、プログラム評価の実施例が少ないわが国では、どのようなプログラム評価が可能かについて、評価専門家の間にも必ずしもイメージが共有されていない。諸外国で採用されている手法を吸収する上でも、評価事例の紹介が必要といえるだろう。

おわりに

本稿では、業績測定を改善する上でのプログラム評価の役割に焦点を当てて議論を行った。業績測定を施策の見直しにつなげていくためには、プログラム評価併用を避けて通ることはできないのではないだろうか。

プログラム評価は決して万能薬というわけではなく、たとえば、非常に大きい制度改正や国民の間に論争があるような問題について、単一のプログラム評価が回答を与えるといったような過剰な期待をするのは適切ではないだろう。米国の事例が示しているのは、むしろ、具体的で限定的なテーマでの評価の積み重ねが、施策への理解を深め、実施プロセスや施策のデザインを次第に変えていくという姿である。日本でそのようなプログラム評価の蓄積を目指すのであれば、相当に息の長い取り組みが求められる。日本評価学会としても、プログラム評価の普及に向けて、議論を続けていく必要があるだろう。

注記

- 1 Hatry 1999, 田中 2014 (pp.176-179) を参考に整理した。
- 2 本節は、Hatry 1999、Rossi et. al. 2004、龍・佐々木 2004、田辺 2002等をもとにまとめた。
- 3 業績測定、プログラム評価の発展過程については、田辺 2002、田辺 2001を参照のこと。
- 4 各項目とも6割程度の無回答があるが、このグラフでは無回答を除いている。
- 5 ここでの「業績目標」は“performance goals”と表現されており、個々の指標の具体的な目標値(“targets”と表現されることが多い)ではなく、施策が最終的に目指す目標を意味していると考えられる。
- 6 <<http://www.dol.gov/asp/programs/REIDL/index.htm>>.
- 7 Hatry 1999 (Part III)、田中 2014 (p.232-)、西出 2011 (p22) 等。
- 8 田中2013 (p.41) による。外務省の総合評価は内容的に実績評価に近いので、この件数から除いている。

参考文献

- 小野達也 (2013) 「政策評価と実績測定—府省の実績測定における計量・係数の現状—」、『日本評価研究』、13 (2) :21-36
- 田中啓 (2012) 「国の機関の総合評価の現状と課題—期待と現実のギャップ—」、『日本評価学会 第13回全国大会要旨集録』、157-158
- 田中啓 (2013) 「府省における政策評価の所見—現状の課題と解決の方向性—」、『評価クォーターリー』、25
- 田中啓 (2014) 『自治体評価の戦略』、東洋経済新報社
- 田辺智子 (2001) 「米国における政策分析と評価」、『レファレンス』、610
- 田辺智子 (2002) 「政策評価の手法—米国の評価理論と実践をもとに—」、『季刊行政管理研究』、97
- 田辺智子 (2012) 「セオリー評価実施における留意点」、『日本評価学会 第13回全国大会要旨集録』、159-162
- 西出順郎 (2011) 「行政評価の使い道を探る」、『評価クォーターリー』、16
- 林和喜 (2004) 「実績評価の現状と有効性検査」、『会計検査研究』、29
- 原田久 (2011) 「各府省における実績評価—「政治主導」は10年の制度運用を乗り越えうるか?—」、『評価クォーターリー』、18
- 茂木康俊 (2013) 「政府業績成果近代化法 (GPRAMA) 制定後の米連邦政府における戦略計画・目標設定・評価結果の活用」、『評価クォーターリー』、24
- 山谷清志 (2012) 「政策の評価の制度化をふまえた理論の再構成—政策評価「学」に向けて—」、『評価クォーターリー』、22
- 龍慶昭・佐々木亮 (2004) 『「政策評価」の理論と技法 (増補改訂版)』、多賀出版
- Department of Labor. (2011). *Strategic Plan: Fiscal Years 2011-2016*.
- Department of Labor. (2013). *FY2014 Department of Labor Budget in Brief*.
- Government Accountability Office. (2012). *Designing Evaluations: 2012 Revision*, GAO-12-208G.
- Government Accountability Office. (2013a). *Managing for Results: Executive Branch Should More Fully Implement the GPRA Modernization Act to Address Pressing Governance Challenges*. GAO-13-518.
- Government Accountability Office. (2013b). *Program*

- Evaluation: Strategies to Facilitate Agencies' Use of Evaluation in Program Management and Policy Making*, GAO-13-570.
- Hatry, H. P. (1997). Where the rubber meets the road: Performance measurement for state and local public agencies. *New Directions for Evaluations*, 75.
- Hatry, H. P. (1999). *Performance Measurement: Getting Results*, The Urban Institute Press. (上野宏他訳『政策評価入門』東洋経済新報社 2004)
- Hatry, H. P. (2013). Sorting the relationships among performance measurement, program evaluation, and performance management. *New Directions for Evaluations*, 137.
- Hunter, D. E., and Nielsen, S. B. (2013). Performance management and evaluation: exploring complementarities. *New Directions for Evaluations*, 137.
- Newcomer, K. E. (1997). Using performance measurement to improve programs. *New Directions for Evaluations*, 75.
- Nielsen, S. B., and Hunter, D. E. (2013) Challenges to and forms of complementarity between performance management and evaluation. *New Directions for Evaluations*, 137.
- Office of Management and Budget. (2009). Increased emphasis on program evaluations. *Memorandum to the Heads of Executive Departments and Agencies*, M-10-01, Oct. 7, 2009.
- Office of Management and Budget. (2011). *Preparation and Submission of Strategic Plans, Annual Performance Plans, and Annual Performance Reports*. Circular No. A-11, Part6.
- Office of Management and Budget. (2012). Use of Evidence and Evaluation in the 2014 Budget. *Memorandum to the Heads of Executive Departments and Agencies*, M-12-14, May 18, 2012.
- Office of Management and Budget. (2013). Next Steps in the Evidence and Innovation Agenda. *Memorandum to the Heads of Executive Departments and Agencies*, M-13-17, Jul. 26, 2013.
- Rossi et. al. (2004). *Evaluation: A Systematic Approach*, 7th ed., Sage Publications. (大島巖他訳『プログラム評価の理論と方法』日本評論社 2005)
- U.S. Government. (1993). Government Performance and Results Act of 1993.
<<http://www.whitehouse.gov/omb/mgmt-gpra/gplaw2m>>
- U.S. Government. (2011). Government Performance and Results Act Modernization Act of 2010.
<<http://www.gpo.gov/fdsys/pkg/BILLS-111hr2142enr/pdf/BILLS-111hr2142enr.pdf>>
- Weil, D. (2010). *Improving Workplace Conditions through Strategic Enforcement: A Report to the Wage and Hour Division*.

(2014.9.29受理)

The Role of Program Evaluation as Complement to Performance Measurement: The Case of GPRAMA in the United States

Satoko Tanabe

National Diet Library
stanabe0509@gmail.com

Abstract

In the Policy Evaluation system in Japan, performance measurement (called “jisseki hyoka” in the system) is most frequently used and the use of program evaluations (“sogo hyoka”) is limited. In this article, the author discusses how we can utilize program evaluations as a means of supplementing performance measurement, by analyzing the case of the United States. Under the framework of GPRA Modernization Act (GPRAMA), program evaluations are being utilized for developing and revising performance goals, for setting performance indicators, and for supporting program reforms.

Program evaluations could solve part of problems in the Policy Evaluation system in Japan, including the difficulty in setting indicators and non-use of evaluation results.

Keywords

Performance measurement, Program evaluation, Impact evaluation,
Government Performance and Results Act (GPRA), GPRA Modernization Act (GPRAMA)

【実践・調査報告】

科学的根拠に基づく実践とその形成評価アプローチが 日本社会に定着しない現状と要因～改善への示唆～

大島 巖

日本社会事業大学

iwao_oshima@yahoo.co.jp

要 約

保健・福祉領域など対人サービス領域において効果的プログラムモデル形成と評価の経験に基づいて、科学的根拠に基づく実践（EBP）および効果的プログラムモデル形成評価アプローチに対して社会的な関心が乏しい日本の現状とその要因を国際的な視点と対比させて明らかにする。その上で、そのアプローチを社会に定着させるための社会的条件と課題を検討した。

日本社会においては、まず社会プログラムが使命として達成すべきプログラムゴールの実現を、利害関係者、特にプログラムに深く関わる施策立案者や実践家・管理者が強く共通認識することがその立脚点になる。その上で、成果中心の科学的プログラム評価の実施、EBP等効果的プログラムモデル（EBP等効果モデル）の重視・尊重、EBP等効果モデルの情報集積とデータベース構築、関係者間での共有化、実践家参画型アプローチの尊重、その実現のための体制整備が図られる必要がある。

キーワード

科学的根拠に基づく実践（EBP）、成果中心のプログラム評価アプローチ、
実践家参画型形成評価、評価統合、評価文化

1. はじめに

本稿では、プログラム評価の理論と方法論（Rossiら、2004）を用いて、主に保健・福祉領域における効果的プログラムモデル形成評価と、科学的根拠に基づく実践（EBP; evidence-based practices）の構築に関わってきた著者らのこれまでの経験に基づいて、まずはプログラム評価、中でもインパクト評価（定義は後述）を用いた効果的プログラムモデルの形成評価アプローチ（同）、

その成果として生み出されるEBPプログラムを日本社会に定着させることが困難な現状とその要因を明らかにする。その上で、そのようなアプローチを社会に実施・普及するための社会的条件と課題を検討する。

保健・福祉領域を含む対人サービス領域では、近年、プログラムの成果（アウトカム）に焦点を当てて、ランダム化比較試験（RCT; randomized control trial）等のインパクト評価の知見を主に用いたEBM（evidence-based medicine）やEBPなど

成果志向アプローチが重視されるようになった(Drakeら,2003; 2005; Corriganら,2008)。このうちEBPは、改善すべき社会問題や社会状況をプログラムゴールとする「社会的介入プログラム(社会プログラム)」が対象になる。EBPでは、EBMと同様にRCTなどのインパクト評価を主に用いるが、一方で、プログラム評価のプロセス評価や理論評価をも体系的に活用する(Rossiら,2004)。すなわちEBPでは、対象となる社会プログラムをプログラム評価によって、エビデンスレベルが高く、より効果的なEBP等の効果的プログラムモデル(以下、EBP等効果モデル、あるいは単に効果モデルと呼ぶ;後述)を形成し、さらには認定されたEBPプログラムへと発展させること、その上で制度化して社会の中に実施・普及させることを目指している。

世界的には、限られたサービス資源の配分に当たり、より良い成果(アウトカム)を生み出すEBP等効果モデルに予算配分上の優先順位を与える国際的潮流が生じている。社会的合意形成の上で優先順位を設定するために、EBP等効果モデルに関する情報を社会的に広く集積し、関係者が参照可能にするためデータベースを構築するなど情報共有の基盤も形成されるようになった(佐々木,2010)。社会プログラム実施に関わる実践家や利用者もEBP等効果モデル形成に関心を持ち、関与することが期待されている(Chen, 2010; Chinmanら,2004)。

しかしながら、日本の現状はこれら世界の潮流とは大きく異なる。本稿では、まずEBP等効果モデルの形成評価アプローチに対して社会的な関心が乏しい日本の現状とその要因を、著者が関わったプログラム事例に基づいて、国際的な視点と対比させて検討し、その上で今後の課題と示唆を提示する。

2. 本研究の視点と分析方法

2.1. 用語の定義

- ・EBP等効果モデルの形成評価アプローチ：エビデンスレベル(正木ら,2006)が高く、より効果的なEBP等効果モデルを形成するために行う

プログラムモデル改善のために行う評価アプローチ。世界的な専門団体によって認定される、エビデンスレベルの高いEBPプログラムになることを目指す。

- ・効果的：エビデンスレベル(正木ら,2006)が高く、エフェクトサイズ(Rossiら,2004)が大きいこと。
- ・効果モデルがより効果的に発展することを「形成」とし、そのための評価を「形成評価」とする。なお効果モデルの「構築」とは、効果モデルが社会的に認定され、制度化されることを示す。

2.2. 本研究の視点

EBP等効果モデルの形成評価アプローチに関する国際的な視点として、本研究では次の5点を用いる。すなわち、①成果志向アプローチ、②効果モデル形成に向けたプログラム評価の活用、③EBP等効果モデルの重視・尊重、④EBP等効果モデルの構築に向けた評価統合(Evaluation Synthesis)への取り組み、⑤実践家参画型アプローチの重視である。以下にそれぞれの背景と根拠を提示する。

視点①：成果志向アプローチ

本稿で取り上げるのは、社会問題や社会状況を改善するために設計され、組織的で計画された、通常は継続的な取り組みである「社会プログラム」である。「社会プログラム」は社会的に解決を目指すべき明確なプログラムゴールを持つことが前提となる。このため自ずから成果志向のアプローチになる(Rossiら,2004; 大島ら,2012c)。ゴールを達成するために、最も合目的で有効な、そして組織的に計画された実現可能な取り組みの単位(構造・機能・プロセス)、すなわち「プログラム単位」を同定し、さまざまな社会条件の下でも有効なプログラムモデル(効果モデル)を、プログラム評価の方法論を用いて形成することが社会的に求められる(Rossiら,2004; 大島ら,2012c)。EBP等効果モデルは、さまざまな社会条件の下での検証を経て、EBPプログラムとして世界的な認定を受けることになる。

視点②：効果モデル形成に向けたプログラム評価の活用

プログラム評価は、ある社会的な問題状況を改善

するために導入された社会プログラムの有効性を、ニーズへの適合性、プログラムの設計や概念の妥当性、介入プロセスの適切性、プログラムの効果と効率性という諸側面（評価階層）から、総合的・体系的に査定・検討し、その改善を援助して社会システムの中に位置づけるための方法である（Rossiら、2004）。導入された社会プログラムに対してプログラム評価の方法を用いて科学的な検証を行い、より効果的なプログラムモデル、EBPプログラムへと改善するために不可欠な方法論である。EBPプログラムを形成するには、インパクト評価の実施のみならず、ゴール達成に有効な効果的プログラム要素の検討（プロセス評価）、プログラム理論の構築（理論評価）が同時に重視されなければならない（大島ら、2012c; 大島、2010）。

視点③：EBP等効果モデルの重視・尊重

近年世界的には、プログラム評価の方法論を用いて生み出されたEBP等効果モデルが社会的に重視され、政策的な優先順位が与えられるようになった。ランダム化比較試験（RCT）が社会プログラムに対しても実施され、その成果が公表されている。RCTの結果に対してシステマティックレビュー（Systematic Review; SR）が行われ、メタ分析のエビデンスが蓄積されている（大島、2010）。その結果、多くのEBPが効果モデルとして世界的な認知を得ている。その知見に基づいて、EBPプログラムが、世界的な治療ガイドラインや支援ガイドラインなどに盛り込まれるようになった（Drakeら、2005）。

限られた社会的資源を活用して、社会的問題を解決するためには、科学的な裏付けのある、EBP等効果モデルに優先的に資源配分することが求められる。このような社会的合意形成が、世界的に進んでいる。

このため、さまざまな社会プログラムは、プログラムゴールの達成により効果的なプログラムモデル、EBPプログラムとなり、社会の中での実施・普及が進むことが求められている。効果的なプログラムモデルに改善・発展するための形成評価のアプローチ、実施・普及の方法論が重視されるようになった（大島ら、2012c; Palinkasら、2012）。

視点④：EBP等効果モデルの構築に向けた評価統

合（Evaluation Synthesis）への取り組み

アメリカでは1993年に政府業績結果法（GPRA: Government Performance and Result Act）が導入され、2011年にはその改正法である政府業績結果法近代化法（GPRA Modernization Act; GPRAMA）が制定された。各省庁には戦略計画の策定・実施・モニタリングだけではなく、プログラム評価の実施を義務付けた。それに伴いアメリカ議会予算局（OMB）や会計検査院（GAO）には、政策評価、制度・施策評価にEBP等効果モデルの評価統合（Evaluation Synthesis）とそれを進めるための評価統合システムが体系的に導入される。EBP等効果モデルを実施することが不可欠の情勢になった（佐々木、2010）。

保健・医療・福祉・教育など対人サービスに関わる連邦政府各部署には、EBPを含む効果的な各種対人サービスプログラムの情報を評価統合して、蓄積された情報を関係者に体系的に整理し、提供する仕組みが構築されつつある（佐々木、2010）。たとえばアメリカ連邦教育省はWWC（What Works Clearing House）というデータベースを構築した。WWCはアメリカ教育省教育研究所が国際的EBPのデータベースであるキャンベル共同計画と共同で開発・運営を行う。またアメリカ連邦保健社会福祉省では、N-REPP（National Registry of Evidence-based Programs and Practices）というEBPデータベースを構築し、情報提供を行うと共に補助金支給の根拠にも活用する（大島、2012b）。

視点⑤：実践家参画型効果モデル形成評価アプローチの重視

保健・福祉領域の社会プログラムを含めて、対人サービス領域のプログラムでは、EBP等効果モデルを開発、形成・発展、実施・普及するために、実践家参画型アプローチが不可欠と考えられている（大島ら、2012a; 2013）。地域を基盤とした実践家参画型協働型研究（CBPR）が注目されている。このほか、EBP等効果モデルを形成・発展させ、そのプログラムを実施・普及させるために、実践家参画型のボトムアップアプローチが重要と指摘される（Chen, 2010）。

日本では、「措置行政」と呼ばれるトップダウンでプログラム実施を進め、実践現場からの「創

意・工夫」を反映することが困難な文化が存在する。これに対して大島ら（2012a）はCD-TEP評価アプローチ法という新しい評価方法を提案し、成果中心の評価アプローチの観点から実践家参画型でEBP等効果モデルを発展させる可能性を検討している。

2.3. 分析方法

著者が主に関与して来た保健・福祉領域について、社会的課題の解決のために近年導入された、プログラムゴールの設定が比較的容易に考慮される社会プログラムを事例として取り上げる。

事例検討を行うのは、そのうち著者らが何らかの形で関わりを持ったものである。具体的には、①精神障害者退院促進支援事業（2003-2012年）、その後継の②精神障害者地域移行・地域定着支援事業（2012年から）、③知的障害者通勤寮事業の廃止（2005-2012年）、④障害者就労移行支援事業（2005年から）、⑤自殺総合対策大綱に基づく自殺予防対策プログラム（2008年から）である。

それぞれについて、本研究の5視点から現状と課題を明らかにする。

2.4. 日本の保健・福祉領域における社会プログラムの全体状況

本研究は、著者が主に関与して来た保健・福祉領域の社会プログラムについて、さらには著者が関与したことのある5プログラム事例を主に検討するものである。紙幅の制約から、保健・福祉領域の社会プログラムの全体状況を以下に示す。

近年の社会福祉制度改革・医療改革の中で、保健・福祉領域の社会プログラムに関して達成目標を明示した新しい実践プログラムが導入されるようになった（大島ら、2010b）。介護保険法の介護予防プログラム、障害者自立支援法の就労移行支援プログラム、退院促進支援プログラムなどである。これは、成果志向アプローチが日本社会でも進展して来たことを示している【視点①】；上記「視点」に対応、以下同様）。

しかしながら一方で、日本の政策立案者や実践家の間ではプログラム評価【視点②】や、EBPへの関心と必要性の認識【視点③】があまり認められない。実践と実証の積み重ねがないままに、行

政主導のプログラム導入がされ【視点④】、それを受け入れる実践現場では導入上の課題を認識したとしても、要綱に沿ったプログラムの導入と実施を比較的粛々と進めている【視点⑤】。

3. 結果とその考察：日本の社会プログラムに対する事例分析

3.1. 精神障害者退院促進支援事業

a. プログラムの概況

本プログラムは、人口当り精神病床数が世界最多である日本の精神科病院長期入院者の退院を進めるために、障害者基本法に基づく障害者基本計画の重点施策実施5か年計画である障害者プラン（2002年）に基づいて導入された。同プランでは約72,000人の入院患者を10年のうちに退院・社会復帰を目指すとして明記した。

しかしこの事業の事業目的は、「症状が安定しており受入条件が整えば退院可能である者に対し、活動の場を与え退院のための訓練を行うことにより精神障害者の社会的自立を促進」と曖昧に設定されている。そのため多くの自治体は、本来この事業に期待される退院実現やその後の地域定着、地域における生活の質向上よりも、入院患者や医療機関関係者との関係作りや、啓発活動や退院する患者を地域で支えるための地域ケアネットワーク作りを重視した（大島ら、2011）。実施事業所に自治体から委託される事業費は少なく（1事業当り500万円程度）、常勤専門職職員を雇用する補助金は用意されていない。保健福祉圏域内の長期入院者の退院・地域移行を本格的かつ体系的に進める体制は構築できない。

この事業により退院したのは6年間で約2000人程度の実績（2008年）に留まる。また2002年の精神科病院入院患者数は330,050人だったが、2010年には308,615人で2万人程度の削減に留まり、障害者プランの目標は達成されていない。また事業実施主体である自治体レベルで、退院・地域移行をアウトカム指標に設定して事業評価することはほとんどない（大島ら、2011）。

b. 本研究視点からの分析（考察）

この事業の問題は、障害者プラン（2002年）の

入院患者約72,000人の退院・社会復帰という目標に対応するほぼ唯一のプログラムだったにも関わらず、退院・地域移行、地域定着というプログラムゴールを明確に設定し【視点①】、それを体系的に実現できる体制を国が用意しなかったことにある。また関係者が、プログラム評価を用いてより効果的なプログラムモデル、EBPプログラムを目指すことも志向しなかった【視点②③】。

3.2. 障害者地域移行・地域定着支援事業

a. プログラムの概況

事例1の後継事業（2012年-）である。精神科病院からの脱施設化にとって不可欠な要素として知られる地域定着・地域継続支援の要素を、制度的に「地域定着支援事業」に位置づけた。また補助金の支給方法を前制度の包括払い方式から個別給付による出来高払い制度へと変更した。

これら事業は、前事業（退院促進支援事業）を直接的に引き継ぐ「地域移行支援事業」と、「地域定着支援事業」に分かれている。二つの事業は本来は有機的に連携して、全体として一つのゴールを目指す必要がある。しかし制度上において、これらが有機的に運用されることは考慮されていない。またプログラムゴールは、2つのプログラム（移行事業・定着事業）ごとに別個に設定された。

さらに事例1で事業所の努力で経験を蓄積した「入院患者・医療機関関係者との関係作りや啓発活動」はこの事業から除かれ、補助対象外とされた。また前制度ではこの活動は包括払いで対応したが、新制度においては出来高払いのため、この活動が財政的に実施困難となった。

なお、2012年度以降これら事業の利用者数は減少している。

他方、制度変更に当り事業実施関係者との意見調整や十分な合意形成がなされないまま、制度改正が行われた。

b. 本研究視点からの分析（考察）

これらの事業は、事例1の後継事業であるが、プログラムゴールが事業ごとに設定され、それが分散化しており全体のゴールが明確になっていない【視点①】。これに関連して、プログラム要素として「地域移行」「地域定着」「継続支援」に加

えて「入院患者・医療機関関係者との関係作りや啓発活動」という効果的援助要素が有機的に連携することで、全体のプログラムゴールが実現できると、大島ら（2011）はプログラム理論の検討から明らかにした。特に長期入院を続けている人たちは、退院への動機付けを見失った人たちである。「入院患者・医療機関関係者との関係作りや啓発活動」の活動は、退院促進支援事業の要になる最重要の援助要素と考えられる。これら援助要素がどのように有機的に関連し全体ゴールを実現するのか、プログラム理論を設定して理論評価を十分に行う必要がある【視点②】。

なお制度変更に当り、事業実施関係者との意見調整・合意形成がない点は、9年間の試行的な実践家たちの取り組みの反映という観点から大きな課題である【視点⑤】。

3.3. 知的障害者通勤寮事業の廃止

a. プログラムの概況

知的障害者通勤寮事業は、知的障害者福祉法に基づく事業であり、大規模施設から地域生活への移行を段階的に支援する、ハーフウェイハウス事業（中間住居事業）であった。2年間の訓練期間に地域生活の訓練を行い、地域移行をスムーズに行うことを目的とした。

障害者自立支援法（2005）の導入により、この事業は「生活の場」としてのケアホーム・グループホーム等と、「日中活動の場」の生活訓練等に支援要素に分離され継続された。しかし通勤寮という事業名は十分な議論もなく消滅した。そのため、これら支援要素（「生活の場」「日中活動の場」「身近な対人サービス」）を有機的に結びつけて、プログラムゴール（段階的に地域移行と地域定着を進める）を明確に再設定し、ハーフウェイハウス事業としての実施要綱を作成することはなかった。さらに前プログラム事例（事例2）と同様に、制度変更に当り事業実施関係者との意見調整や十分な合意形成がなされないまま、制度の改正が行われた。

b. 本研究視点からの分析（考察）

厚生労働省の基礎調査（2005）によれば、全国の知的障害のある人たち547,000人のうち、施設入所している人は128,000（23.4%）を占めており、

国は脱施設化・地域生活移行の方針を出している。施設入所する知的障害のある人たちが地域生活に移行するためのハーブウェイハウス事業は一定の有効性を有している。しかし地域移行へのニーズは未だ大きいにも関わらず、このプログラムが十分な議論なく消滅した点は、【視点⑤】から大きな課題と言える。またこれも前事例と同様に「生活の場」と「日中活動の場」、そして「身近な対人サービス」という援助要素が分散化された。ある特定のプログラムゴールを達成する取り組みとして、プログラム理論を設定し理論評価を行う必要があったが、実施されていない【視点②】。

3.4. 障害者就労移行支援事業

a. プログラムの概況

障害者就労移行支援事業は、障害者自立支援法(2005)の法目的を実現する中核的な施策として導入された。この事業は、一般就労を希望する障害のある人たちが2年間の利用期限(訓練期間)のうちに、一般就労に移行することを支援する。施設内外の活動に参加することで、就労に必要な知識及び能力の向上のために必要な訓練、求職活動に関する支援、その他適性に応じた職場の開拓の支援を受け、2年間で一般就労を目指す。就職後には職場定着のために必要な相談等の支援を提供することになっている。しかし段階的に就労移行支援を目指すプログラムであるため、世界的エビデンスがあるOJT(On the Job Training)という援助要素(Beckerら、2003)は、この事業に含まれていない。

事業費の補助額が比較的高額のため取り組む事業所が多い。しかし「就労移行数=なし」が36.2%を占める現状がある(2011年4月、厚生労働省障害福祉課調)。

一方で、就労移行を積極的に進め、速やかに一般就労を実現すると事業所の定員確保が難しくなる。さらに補助事業の支援期間は受け入れ後から一般就労移行までが主であるため、一般就労するまでに一定期間在籍することを方向付ける財政的な圧力が働く。他方で、一般就労後に就労先でOJT支援を継続しようとしても、財政面で制度的に対応が難しい(小佐々ら、2011)。また利用者の動機付け支援や、導入までの支援に対応すること

も財政的な担保はない。このように、支援開始と継続支援という支援の入口機能、出口機能に、この事業の制度設計上の課題がある。

b. 本研究視点からの分析(考察)

この事業には一般就労移行というゴールが明示されているが、障害のある人たちの地域生活にとって大切な就労移行後の就労定着、就労継続による生活の質向上というプログラムゴールが明確には位置づけられていない【視点①】。そのため、特に就労定着に必要なOJTなどの支援要素が含まれていない【視点②】。取り組みの熱心な事業所はこれらの成果を上げているが、就労移行実績を納めていない事業所が4割近くを占めるなど事業所が2極分化する状況がある。成果を上げる事業所の取り組みが、他と共有されたり、より良い取り組み要素を制度に反映する試みも、未だ行われていない【視点③④⑤】。

3.5. 自殺総合対策大綱に基づく自殺予防対策プログラム

a. プログラムの概況

自殺総合対策大綱(2007年)に基づいて多額の予算(287億円、2013年度)が投入されるが、自殺予防成果は現時点では十分には認められない。大綱に基づき多種多様なプログラム要素が用意されるも、それらが有機的・戦略的に運用されない。実践現場では自殺率を真に低減しようとする動機付けを持たせる成果と結びつけた取り組みは未だ行われていない。精神保健福祉一般の啓発事業にふだん予算が付かないので、それを活用するという意識すらある。事業成果によりプログラム内容を見直すことや、予算の増減をすることはない。

b. 本研究視点からの分析(考察)

自殺予防というプログラムゴールは明確に設定されている【視点①】。しかし自殺率低減という成果との関係から、自殺予防の多種多様なプログラム要素が検証されることはほとんどない。多種多様な取り組みについて、どのように有機的・戦略的に組み合わせるのかに関するプログラム理論が形成されていない【視点②】。

4. 総合考察

4.1. EBP重視、効果モデル形成に向けて社会的 取り組みが進まない要因

前節では、効果モデル形成に向けて社会的取り組みが進まない日本の現状を示すと共に、本来であれば取り組みが促進されなければならない社会問題への対応において、EBP等効果モデルの構築が進まない厳しい状況を事例的に明らかにした。

いずれの事例とも共通して、①達成を目指す具体的で明確なプログラムゴールが欠如しており、②ゴール実現に対する合理的で戦略的なプログラム設計（プログラム理論の構築）に欠けている。また③実施後の有効性に関する科学的検証と改善方針の作成がされず、④施策立案者からトップダウンで実践現場に実施内容が通告される問題があると思われる。

a. 確実に達成を目指す具体的で明確なプログラム ゴール設定の欠如

事例1～事例5のいずれにおいても、導入された社会プログラムが確実に達成を目指す、具体的で明確なゴールが示されていない。また関係者間で合意形成の上でゴールを設定しそれを共有化していない。そのため、ゴール達成を事業所単位、地域単位で検証しようとする取り組みがほぼ行われていない。これは前項で示した【視点①】成果志向アプローチに関わる課題である。

それぞれの事例においては、一見すると「退院促進」「地域移行」「地域定着」「一般就労移行」「自殺率低減」というプログラムゴールが設定されているように見える。しかし、たとえば「退院促進」「地域移行」では、その実現に関わる入院患者や医療機関関係者との関係作りや、啓発活動や退院する患者を地域で支えるための地域ケアネットワーク作りを重視する自治体が多くを占めていた（瀧本ら、2011）。事例3では「地域定着」のゴールが関係者間の十分な議論がなく消滅しても、特段問題になっていない。このように明確なプログラムゴールの設定が欠如していると考えられる。

b. ゴール実現に対する合理的で組織的・戦略的な プログラム設計および効果的プログラム要素 検討の不在

社会プログラムが「成果中心」に焦点化され、達成すべきプログラムゴールが明確になれば、それを最も合目的的に実現する社会プログラムの輪郭や枠組みを設定することが可能になる。そのため合理的で戦略的なプログラム設計を行い、成果に結びつく効果的プログラム要素を検討して、効果モデル形成のためにプログラム評価を行うことになる。これは【視点2】プログラム評価の重視に関わる課題である。

しかしながら前項で示したように、日本の社会プログラムは、確実に達成を目指すプログラムゴールが不明確である場合が多い。また「漠然としたゴール」であっても、それを実現するためのプログラム要素（効果的援助要素）が有機的・体系的に機能すれば良いが、要素分解的にプログラムが位置づけられている。たとえば事例2の障害者地域移行・地域定着支援事業は、事例1の退院促進支援事業を引き継いだ事業であったが、「地域移行」と「地域定着」は連続的に達成すべきゴールである。また2要素が断片化することにより、最も重要なプログラム要素と考えられていた、長期入院している入院患者との関係づくり・動機付けの機能が欠落することになった。

事例1～5にすべてに共通して、ゴール実現に対する合理的で組織的・戦略的なプログラム設計、プログラム理論構築への視点が乏しい。プログラム要素を分解しても、それぞれのゴールを実現できると考慮しているようにみられる。その一方で、ゴール実現に有効なプログラム要素・効果的援助要素の検証が、実証的に行われることはほとんどない。

c. 有効性に関する科学的検証と改善方針作成の 欠如

4.1項でアウトカム指標値の変化がモニターされることが少ないことに触れた。行政による社会プログラムの導入・実施後に、その有効性を科学的に検証する取り組みはほとんどない。さらに効果的援助要素を尺度化したものとして、近年世界的にはフィデリティ尺度の作成が行われ、プログラム実績が評価される（Bondら、2000）。しかし、そのような取り組みは日本では一般化していない。

世界的には社会プログラムの制度化に当たっ

て、EBPプログラムが重視される。しかし日本の社会プログラムから、新たなEBPプログラムに発展させる取り組みはほとんど行われていない。導入された社会プログラムをより効果的なプログラムモデル、さらにはEBPプログラムに改善・発展させる取り組みが日本社会に定着していない。

これらは、【視点②】プログラム評価の重視、【視点③】EBP等効果モデルの尊重・重視に関わる課題である。【視点④】EBP等効果モデル構築に向けた評価統合への取り組みも関連する。これまで述べて来たように、EBP等効果モデルの形成という志向性が、日本社会の中には未だ定着していない。

d. 施策立案者からトップダウンで実践現場に実施内容が通告される点

以上検討した社会プログラムは、いずれも行政主導で施策立案者からトップダウンで実践現場に降ろされてくる。実践現場の関係者に対して事前提示があり、十分な合意形成を行った上で導入されることはほとんどない。また提示された社会プログラムの内容は、実践と実証の積み重ねがないままに、行政主導で導入が決められる（大島ら、2011）。

一方、プログラムゴールが明確な社会プログラムについては、国際的にはEBP等効果モデルが構築され、世界の関係者間で共有されるようになった。たとえば、事例4の障害者就労移行支援事業では、IPS援助付き雇用は、利用者の動機付けとOJTを重視するEBPプログラムであり、十分な科学的根拠の蓄積がされている（Beckerら、2003）。しかしながら、日本で就労移行支援事業を設計する際には、IPS援助付き雇用の効果的援助要素はほとんど盛り込まれていない。事例1・事例2の事業についても、脱施設化を進めた先進諸国の実践から集中型・包括型ケアマネジメントACTの有効性が明らかになっている（大島、2004）。しかしACTなどの効果的援助要素は、障害者地域移行・地域定着支援事業にはほとんど盛り込まれていない。これは【視点④】EBP等効果モデル構築に向けた評価統合に関わる課題である。

さらに日本では、【視点⑤】実践家参画型アプローチが注目されることはほとんどない。新しく導入された社会プログラムは、もとより「完全」

なプログラム、EBPプログラムではない。ところが施策立案者が、その新規導入プログラムに改善の必要があると考えることは稀である。より効果的なプログラムモデルに発展させるために、実践現場からの創意・工夫、アイデアの反映がプログラム改善にとって不可欠であるが、そのような発想と意識が、施策立案者にも実践家にも欠けている。

これらの点に関して、特に社会福祉領域においては、「措置行政」という国の福祉事業を民間事業所が担って来た長い歴史を振り返る必要がある。「措置行政」の下では、要綱の記載事項は「お上」が示した「完全」なものであり、実践現場からの「創意・工夫」はまったく期待されない。したがって、要綱上記載される以外のことを「創意・工夫」で行おうとすると、それは「コンプライアンス違反」と警告され、補助金の打ち切り対象になる虞が生じる。

4.2. EBPおよび効果モデル形成評価アプローチを日本社会に根付かせるための条件と課題

前項で述べた日本の社会プログラムを取りまく状況は、さまざまな観点から抜本的な改善を必要とする深刻な事態である。まず最も重要なことは、限られた社会的資源、国民から負託された公的財源を有効に活用していない「資源の無駄遣い」の問題がある。またエビデンスが不明確なまま、細かく実施内容を規定して実践現場を拘束する煩雑な行政上の要綱の存在がある。しかもそれが朝令暮改で改正されることにより（たとえば事例1・2）、実践現場は振り回され徒労感に陥る。そのため実践家は専門性の高い科学的な実践を行うモチベーションを低下させる。それにより世界的な潮流であるEBP等効果モデルを国際的水準で検討し、より良いものに発展させることが妨げられる。

これらに対応するには、当面、特に事前に設定した5つの視点に対応する5つの取り組みが重要になろう。

これらは世界的にも新しい潮流ではあるが、関係者間に徐々に定着してきた。しかし日本の取り組みは十分ではなく、実施には困難が伴う。

a. 成果中心のプログラム評価アプローチの定着 社会プログラムが使命として達成すべきプログ

ラムゴールの実現を、利害関係者、特にプログラムに深く関わる施策立案者や実践家・管理者が強く意識する必要がある。営利企業であれば、利益や売上の追求は至極当然である。社会プログラムにおいては社会から付託された目標達成・実現を、それぞれの立場で使命として取り組む必要がある。また社会プログラムに関わる関係者は、その取り組みをモニターする環境を整備することが求められる。

b. 効果モデル形成に向けたプログラム評価の活用

社会プログラムが「成果中心」に焦点化され、達成すべきプログラムゴールが明確になれば、それを最も合目的的に実現する社会プログラムの輪郭や枠組みを設定することが可能になる。そのために合理的で戦略的なプログラム設計を行い、効果モデル形成のためにプログラム評価を行うことになる。

国際的には多くランダム化比較試験（RCT）が行われ、RCTの結果に対してシステマティックレビュー（SR）が行われてエビデンスが蓄積されている。それに基づき、多くのEBPが効果モデルとして世界的な認知を得ている。これに対して、日本では社会プログラムに対するRCT評価研究は限られている。保健福祉領域、中でも精神保健福祉領域ではEBPプログラムのRCT研究が行われるようになった（大島、2010a）。日本でもEBP等効果モデルのエビデンスを積み重ねる時期に差し掛かっている言えよう。

c. EBP等効果モデルの重視・尊重

限られた社会的資源を活用して、社会的問題を解決するためには、科学的な裏付けのある、効果的な社会プログラムに優先的に資源配分することが求められる。このような社会的合意形成が世界的に進んでいることは既に述べたとおりである。

日本のこのような現状を乗り越えるため、新しく導入された社会プログラムに科学的な方法でプログラム評価を積極的に導入し、EBP等効果モデルを形成する取り組みが重視される必要がある。また世界各国で蓄積されている国際的なEBP等効果モデルを参照することが、社会的・行政的にも進められる必要がある。

d. EBP等効果モデルの情報集積とデータベース構築、関係者間での共有化

保健・医療・福祉・教育など対人サービスに関わる連邦政府各部署には、EBPを含む効果的な各種対人サービスプログラムの情報を評価統合して、蓄積された情報を関係者に体系的に整理し、提供する仕組みが構築されつつある（佐々木、2010）ことについては既に述べた。キャンベル共同計画、アメリカ連邦教育省のWWC、保健社会福祉省のN-REPPなどのEBPデータベースである。

日本には、キャンベル共同計画の日本語版が開設されている。しかし、政策的な優先順位決定のためにEBPのデータベースが活用されることはない。これに対して日本でも、社会プログラムの効果モデルに関する情報蓄積のためのデータベース構築と、関係者間での共有化を進める必要がある。

e. 実践家参画型アプローチの重視・尊重、その実現のための体制整備

社会プログラムに関わる実践家が、EBP等効果モデル形成のために積極的に関与するためには、社会プログラムの設計図に当たるプログラム理論や、効果的プログラム要素を可視化して、プログラムに関わる関係者とともに共有し、実践現場の創意・工夫を「カイゼン活動」の一環として社会プログラム形成に反映させる仕組みを整備する必要がある。そのために、EBP等効果モデルを、プログラム理論、効果的援助要素、実施マニュアルなどによって可視化する方法論の確立が必要であり、同時に実践現場を実践的・教育的にサポートする体制の整備も求められる（大島ら、2013）。

5. 結論

本稿では、主に保健・福祉領域におけるEBP等効果モデル形成と評価の経験に基づいて、インパクト評価を用いたEBP等効果モデル形成評価アプローチ及びEBPプログラムを、日本社会に定着させることが困難な現状とその要因を明らかにし、そのアプローチを日本社会に根づかせるための社会的条件と課題を検討した。日本社会においては、まず社会プログラムが使命として達成すべきプログラムゴールの実現を、利害関係者、特にプログラムに深く関わる施策立案者や実践家・管理者が強く意識することがスタートラインになる。その

上で、科学的な方法で行われる成果中心のプログラム評価の実施、EBP等効果モデルの重視・尊重、効果モデルの情報集積とデータベース構築、関係者間での共有化、実践家参画型アプローチの重視・尊重、その実現のための体制整備が図られる必要がある。

以上のことは、EBP等効果モデルの形成のために、施策立案者、実践現場、そしてその他の利害関係者が共通してプログラム評価の理論と方法論を有効なツールとして共通認識し、そのアプローチ法を共有する評価文化を形成することが前提になる。そのための関係各層への教育システムや、実施支援システムの構築が必要となるであろう。

参考文献

- 大島巖編 (2004) 『ACT・ケアマネジメント・ホームヘルプサービス～精神障害者地域生活支援の新デザイン』、精神看護出版
- 大島巖 (2010a) 「精神保健福祉領域における科学的根拠にもとづく実践 (EBP) の発展からみたプログラム評価方法論への貢献」、『日本評価研究』、10 (1) : 31-41
- 大島巖、小佐々典靖、贄川信幸、道明章乃 (2010b) 「科学的な実践家参画型プログラム評価の必要性和実践的評価者・評価研究者育成の課題」、『リハビリテーション研究』、(145) : 32-37
- 大島巖 (主任研究者)・古屋龍太 (分担研究責任者) 他 (2011) 「精神障害者退院促進・地域定着支援プログラムを効果的プログラムモデルに再構築し発展させるための方策～プログラム理論・エビデンス・実践間の円環的対話による、効果的福祉実践プログラムモデル形成のための評価アプローチ法 (CD-TEP法) からの示唆」平成22年度文部科学省・科学研究費補助金基盤研究 (A) 「プログラム評価理論・方法論を用いた効果的な福祉実践モデル構築へのアプローチ法開発」 (主任研究者：大島巖) グループ分担研究報告書
- 大島巖他 (2012a) 「CD-TEP：円環的対話型評価アプローチ法実施ガイド」平成22年度文部科学省・科学研究費補助金基盤研究 (A) 「プログラム評価理論・方法論を用いた効果的な福祉実践モデル構築へのアプローチ法開発」報告書 (主任研究者：大島巖)、
<http://cd-tep.com/> (2014.12.8取得)
- 大島巖 (2012b) 「制度・施策評価 (プログラム評価) の課題と展望」、『社会福祉学』、53 (3) : 92-95
- 大島巖他 (2012c) 「福祉系大学におけるプログラム評価教育ガイドライン」文部科学省 (平成21年度採択事業) 組織的な大学院教育改革推進プログラム (2009-2011年度) 福祉サービスのプログラム評価研究者育成報告書、日本社会事業大学
- 大島巖・贄川信幸・中越章乃他 (2013) 「効果的プログラムモデル形成のための実践家参画型評価アプローチ法の開発：その方法と現状の到達点、課題」、『日本評価学会春季第10回全国大会要旨集録』、143-150
- 小佐々典靖・大島巖・高原優美子他 (2011) 「障害者就労移行支援事業の現状と課題」平成22年度文部科学省・科学研究費補助金基盤研究 (A) 「プログラム評価理論・方法論を用いた効果的な福祉実践モデル構築へのアプローチ法開発」 (主任研究者：大島巖) グループ分担研究報告書、「障害者就労移行支援プログラムを効果的プログラムモデルに再構築し発展させるための方策～プログラム理論・エビデンス・実践間の円環的対話による効果的福祉実践プログラムモデル形成のための評価アプローチ法 (CD-TEP法) からの示唆」 (分担研究責任者：植村英晴)、7-10
- 佐々木亮 (2010) 「アメリカの政策評価におけるメタ評価の現状」、『日本評価学会第11回全国大会要旨集録』、93-100
- 瀧本里香・大島巖・古屋龍太他 (2011) 「効果的あがる精神障害者退院促進支援プログラムモデル構築に向けた実証的アプローチ (その4) : 全国都道府県事業モデルの実情把握調査からみた効果モデルの検討」、『病院・地域精神医学』、54: 202-205
- 竹島正・稲垣正俊・高橋祥友他 (2012) 「自殺総合対策大綱見直しの提言について」、『精神神経学雑誌』、114: 1428-1431
- 正木朋也・津谷喜一郎 (2006) 「エビデンスに基づく医療 (EBM) の系譜と方向性：保健医療評価に果たすコクラン共同計画の役割と未来」、『日本評価研究』、6 (1) : 3-20
- Becker DR, Drake RE (2003) . *A working life for people with severe mental illness*. Oxford University Press, New York (大島巖、松為信雄、伊藤順一郎監訳) : 精

- 神障害をもつ人たちのワーキングライフ～IPS：チームアプローチに基づく援助付き雇用ガイド. 金剛出版、2004)
- Bond GR et al (2000) . Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research*, 2, 75-87.
- Chen HT (2010) . The bottom-up approach to integrative validity: A new perspective for program evaluation. *Evaluation and Program Planning*, 22, 205-214.
- Chinman M, et al (2004) . *Getting to outcomes 2004: Promoting accountability through methods and tools for planning, implementation, and evaluation*. RAND Corporation.
- Corrigan PW, Mueser KT, Bond GR, Drake RE, Solomon P (2008) . *Principles and practice of psychiatric rehabilitation: An empirical approach*. Guilford Press.
- Drake RE, Goldman HH (2003) . *Evidence-Based Practices in mental health care*. American Psychiatric Association.
- Drake RE, Merrens MR, Lynde DW (2005) . *Evidence-Based Mental Health Practice*. WW Norton & Company.
- Palinkas LA, Soydan H (2012) . *Translation and implementation of Evidence-Based Practice*. Oxford University Press.
- Rossi PH, et al. (2004) *Evaluation: A systematic approach (7th edition)* , Sage (大島巖他監訳 (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)
- (2014.11.20受理)

**Present situation and its social conditions that inhibit
evidenced-based practices and their formative approaches
from established in the Japanese society:
Implications for improvement**

Iwao Oshima, PhD

Japan College of Social Work

iwao_oshima@yahoo.co.jp

Abstract

Aims of this article are firstly to describe difficult but necessary situations to place program evaluation and methods for constructing effective program models using scientific impact evaluation in the fields of human services including health and social welfare arena solidly in the society of Japan, and secondly to discuss issues and social conditions to promote the approaches.

We found that the first step and the fundamental point to start formative evaluation approaches for building effective program models is for principal stakeholders, especially practitioners and administrators associated with the program to be sharply conscious of the goal attainment with which the social program fulfills its mission in Japan. We also found it important in the development of evaluation approaches for building effective program models in Japan to conduct result-centered scientific program evaluation, to place emphasis on effective program models including EBPs, and to accumulate evidence of the program models building their database, and to share the information with stakeholders in the society with improving systems for achieving them.n.

Keywords

Evidence-Based Practices (EBP), result-oriented program evaluation approach,
practitioner participatory formative evaluation, evaluation synthesis, evaluation culture

【研究論文】

証券アナリストと評価者による倫理向上への取り組み —職業倫理の比較研究—

小林 信行

マーストリヒト大学

n.kobayashi@student.maastrichtuniversity.nl

要 約

プロフェッションとは自らを律するものであり、評価倫理ガイドラインの策定は、評価者がプロフェッションとして一層の成長を遂げたことを意味する。職業倫理の向上を図る上で、その目標に長年取り組んできた他プロフェッションの動向は参考となるだろう。本稿は、職業倫理における「理論」、「制度」、「実践」の観点から証券アナリストと評価者の比較分析を行った。比較の結果、評価倫理ガイドラインの利用価値を高めるには、倫理上の問題がおきやすい局面の記述拡充が一案と考えられ、事例を継続的に汲み上げるプラットフォーム作りが課題として浮かび上がった。また、倫理規程の解釈と適用にコンセンサスが見出し難いという特質から、評価分野では個々人に判断を委ねる部分が多い。そのため、ケース・ディスカッションによる訓練は有意義であり、ディスカッション用の教材作成や演習も価値のある取組みとなる。

キーワード

職業倫理、評価倫理ガイドライン、証券アナリスト、評価者、エージェンシー理論

1. はじめに

2012年12月に日本評価学会総会にて評価倫理ガイドライン（以下「ガイドライン」）が承認され、プロフェッションとしての評価者に行動の規範が示されることになった。応用倫理学の研究者である Boatright（2014）が整理したプロフェッションの定義では、特殊な知識の体系、高度の組織化と自主規制、公共の利益への寄与、の3つが挙げられている（p.44）。自主規制の核となるガイドラインの成立により、評価者はプロフェッションとして一層の成熟を迎えた。本稿は、倫理向上に長年取り組んだ他プロフェッションを比較対象と

し、日本評価学会のガイドラインや取組みを精査し、評価者という新しいプロフェッションの職業倫理の高揚に向けた具体的な示唆を導出する。

評価者と他プロフェッションの職業倫理を比較対象とする研究では、Fitzpatrick（1999）が米国評価学会（AEA）と米国の各種専門職団体との比較分析を行っている。また、小林・一寸木（2010）では、ガイドライン策定作業の一環として、評価学会、学術団体、専門職団体の倫理関連文書における基本原則項目の比較を行い、それぞれの特徴を導き出した（pp.148-149）。同稿では、日本の専門職団体として、日本公認会計士協会、日本弁護士連合会、日本内部監査協会を採りあげた。そ

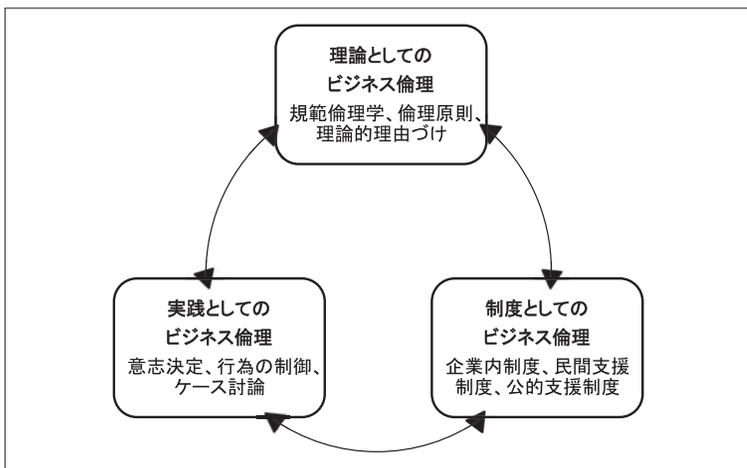
ここで、本稿ではガイドライン策定において参考とされなかった証券アナリストを採りあげ、その団体の倫理綱領や取組みを比較対象に採りあげる。評価分野における倫理研究では、これまで証券アナリストは比較対象とはなっておらず、同プロフェッションとの比較分析は新たな試みとなる。

証券アナリストの専門職団体である日本証券アナリスト協会は、1962年に設立された東京証券アナリスト協会をその前身としており、1969年に現在の団体名に改称した。また、同協会は1979年に証券アナリスト試験制度を導入しており、一定の技能を有する専門職を認定する役割を担っている。2013年12月時点で、個人会員は25,782名（うち25,702名が証券アナリスト試験に合格し、3年以上の証券分析経験を有する検定会員）、法人会員は235社となっている。法人会員は金融機関（銀行、証券会社、生損保、投資顧問等）、シンクタンク、格付会社となっている。日本証券アナリスト協会は1976年に倫理委員会（1979年に規律委員会に改称）を設置し、職業倫理の核となる倫理綱領については1987年に証券アナリスト職業行為基準（以下、「職業行為基準」）を策定し、これまで四度にわたる改定を行った（2014年2月時点）。なお、「職業行為基準」では、証券アナリストが従事する証券分析業務の分野として、「投資情報提供」、「投資推奨」（有価証券の売買推奨）、「投資管理」（資産運用）に言及している。

2. 職業倫理分野の先行研究レビュー

証券アナリストと評価者の職業倫理を比較するにあたり、分析の視点となる3つの論点を提示したい。第1の論点として、一見すると捉えどころのない職業倫理をより具体的に理解するためのフレームワークに言及したい。職業倫理を包括的にとらえる概念的枠組みとして、梅津（2002）はビジネス倫理を3つの側面からとらえるフレームワークを提示している（p.7）。このフレームワークでは、倫理は倫理の「実践」とそれを反省し、行動の理由づけを考察する「理論」から成り立つ。それぞれが異なった社会的機能を持つため、この2つの側面を分けるべきとし、また原則という全体と個々のケースの個別性を統合する「制度」が「実践」より外化してきたと説明している。つまり、「理論」は職業上重視される価値の根拠を提供し、「実践」は個々人が実際に職業倫理をどう行使するかを扱う。「制度」はプロフェッションの行動指針を明文化し、モニタリングする仕組みを構築する。そしてこの3つの側面が相互に関連し補いあうことで、職業倫理は発展し続ける（図1を参照）。具体的には、「理論」は「実践」が必要とする判断根拠を示し、「実践」は時に「理論」に対し修正を要求する。「制度」は「理論」と「実践」を踏まえ形成されるが、「制度」が現実と合致しなければ、「理論」や「実践」に新たな視

図1 ビジネス倫理における「理論」「制度」「実践」の関係



（出所）梅津（2002）より抜粋

点がフィードバックされる。八田・町田（2003）は会計プロフェッションの職業倫理に梅津（2002）のフレームワークを適用し、「理論としての職業倫理」として規範倫理学、倫理原則、倫理的理由づけ、「制度としての職業倫理」として倫理基準、倫理規則、関連法規等、「実践としての職業倫理」として、会計事務所内の研修、専門職団体の継続教育、ケース・メソッドを挙げている（pp.42-43）。

第2の論点として、クライアントとの関わりから倫理上のジレンマが生じやすいことを踏まえ、クライアントの特徴を明確にする切り口を示したい。プロフェッションの職業倫理を比較する研究には、クライアントの持つ側面に着目する分析がある。クライアントにはサービスの選任者、費用負担者、受益者という三側面があることが指摘され、クライアントの持つ3つの側面が一致しない現象はインセンティブの捻じれと呼ばれている（小俣 2012、p.20）。インセンティブの捻じれはプロフェッションの倫理的ジレンマを理解する上で無視できない視点であり、三大プロフェッションと呼ばれる医師、弁護士、公認会計士のうち、医師と弁護士の倫理と、公認会計士の倫理とを分ける特徴となっている（表1を参照）。インセンティブの捻じれは医師、弁護士には見られず、その2つのプロフェッションでは、クライアントは自己にとって最善のサービスを提供する専門職を選び、サービスの提供を受け、その費用を負担する。両者の関係は双務的であり、利害関係は二者間のものとなる。一方、公認会計士はインセンティブの捻じれを抱えており、公認会計士の提供する主たるサービスである財務監査においては、受益者は不特定多数の市場参加者であるものの、選任者は監査対象となる企業の取締役会である¹。そのため、選任者からの圧力が受益者への適切なサービス提供を妨げる可能性がでてくる。この捻じれ

により、公認会計士は利益相反に陥りやすく、不正会計問題時においては適切な監査の実施につき疑義が生じるため、公認会計士には独立性が強く求められる（桑野 2003、pp.158-159）。

第3の論点として、倫理面での問題を誘発する要因にも触れたい。金融サービス業における倫理研究では、エージェンシー理論を踏まえ、金融分野のプロフェッションと顧客間との情報格差が倫理面での問題の一要因と考えられている。エージェンシー理論では、情報の非対称性があり、プリンシパルがエージェントの業績を判断できない場合、適切なインセンティブの設定が困難で、エージェントが自己の利益を追求し、利益相反が起こりやすいとの結論を導きだしている（Milgrom and Roberts 1992,邦訳版 p.35）。首藤（2011）は、実物市場と比較した金融市場の特徴として、交渉力の格差に加えて、情報の非対称性が大きく、金融プロフェッションを利益相反に誘う状況が生じやすい点を指摘し、この問題への対処を金融市場の機能に関わる公共政策的課題と位置づけている（pp.4-5）。

3. 証券アナリストと評価者の類似点

証券アナリストと評価者の比較分析の前に、両者の共通点をまず詳述したい。

(1) 資格の任意性

日本証券アナリスト協会と日本評価学会ともに自らの定めた基準により専門職資格の認定を任意で行っている。評価業務も証券分析業務も特定の有資格者のみによる独占業務ではなく、専門職資格が確立される以前から実践されていた業務である。

表1 三大プロフェッションのインセンティブの捻じれ

	公認会計士	医師	弁護士
1. 選任権者	取締役会	患者	依頼者
2. 費用負担者	クライアント会社	患者	依頼者
3. 受益者	主に市場参加者	患者	依頼者
インセンティブの捻じれ現象	あり	なし	なし

(出所) 小俣（2012）より一部抜粋

金融機関や調査会社においても検定会員以外が証券分析業務を担当することがある。同様に評価士資格を持たない担当者が政府職員やコンサルタントとして評価業務を実施している現状がある。そのため、両団体が職業倫理の普及を進める上で、有資格者以外に対し直接影響を及ぼすことには制約がある。

(2) 利害関係の多様性

証券アナリスト、評価者ともに担う役割は多様であり、顧客との利害関係も一様ではない。そのため、プロフェッション全体を網羅する単一の利害関係モデルに沿って、倫理上のジレンマが生じる局面を想定し、行動指針を策定することは困難である。

日本証券アナリスト協会の検定会員の所属先は、引受金融機関（銀行、証券会社等）、機関投資家（生損保、投資顧問会社等）、調査会社（格付会社等）と幅広い。同様に、日本評価学会の認定する評価士の所属先は政府機関、大学、コンサルティング会社等と多岐にわたる。表2に例示するように、クライアントの三側面の一致という視点でみると、三側面が一致するケースとして、バイサイド・アナリストと改善目的の評価を実施する外部評価者が挙げられる²。一致しない例としては、セルサイド・アナリストと説明責任目的の評価を実施する外部評価者が挙げられる³。

証券アナリストと比した評価者の特徴として、評価者は担当業務の目的毎に担う役割が異なり、利害関係も変わる上、担当業務の多くは説明責任と政策改善の双方が目的となる点が挙げられる。つまり、評価者においては、業務毎ばかりでなく、

業務の局面毎に利害関係が変化しうる。一方、証券アナリストにおいては、利害関係が比較的固定されている点が指摘できる。具体的には、セルサイド・アナリストはクライアントの三側面が一致しない利害関係、バイサイド・アナリストはクライアントの三側面が一致する利害関係の基にある。

(3) インセンティブの捻じれ

上述の通り、証券アナリストと評価者はクライアントの三側面が一致せず、インセンティブの捻じれを抱える層（説明責任目的の評価を実施する外部評価者、セルサイド・アナリスト）を抱えている。

日本証券アナリスト協会の「職業行為基準」は会員に対し「公正かつ客観的な判断を下す」ことを求めている。セルサイド・アナリストは証券会社等に所属し、アナリスト・レポートを通じて中立的な立場から投資家に投資情報提供、投資推奨を行う役割を担う。アナリスト・レポートは調査対象企業に関する事実関係（証券分析業務のうち、投資情報提供）に加えて、投資レーティング、つまり「投資する価値がある/ない」という価値判断（証券分析業務のうち、投資推奨）も含むことが一般的である。

セルサイド・アナリストのサービス利用者は本来、機関投資家が想定されていたが、証券会社のウェブサイトを通じて個人投資家がアナリスト・レポートを直接入手し、また証券会社自身がアナリスト・レポートの要約を個人投資家向けに作成することも増えてきた。アナリスト・レポートの利用者が十分な金融リテラシーのない層に広がる

表2 証券アナリスト、外部評価者のインセンティブの捻じれ

	アナリスト		外部評価者	
	セルサイド・アナリスト	バイサイド・アナリスト	外部評価者（説明責任）	外部評価者（改善）
1. 選任権者	引受金融機関	所属企業	評価委託者	評価委託者
2. 費用負担者	投資家	所属企業	評価委託者	評価委託者
3. 受益者	投資家	所属企業	主に一般国民	評価委託者
インセンティブの捻じれ現象	あり	なし	あり	なし

(出所) 著者作成

一方で、1999年の株式委託手数料の完全自由化後、手数料は低下傾向にある。証券会社は証券の引受業務により注力する中、セルサイド・アナリストが社内の引受部門や発行会社からレポート内容、特に投資レーティングに圧力を受けるケースが生じている。2000年以降のITバブルの崩壊時に、米国のセルサイド・アナリストの利益相反（「売り」推奨に比して相対的に多い「買い」推奨、証券会社における引受部門への協力、セルサイド・アナリストの個人的な株式売買等）が顕在化し、わが国でも同様の利益相反が生じるリスクが強く認識された（日本証券アナリスト協会 2002、pp.12-17）。

図2 セルサイド・アナリストの利害関係

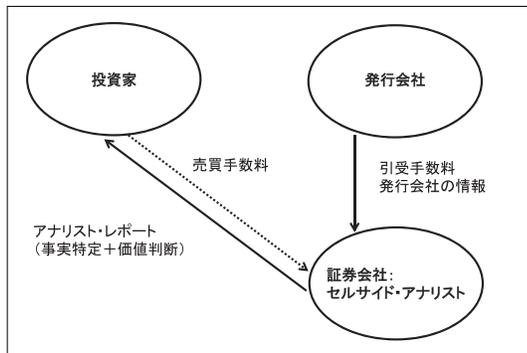
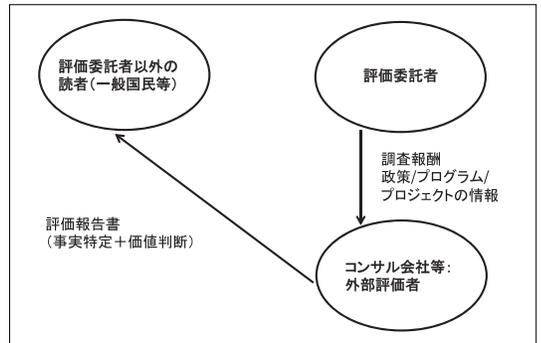


図2に示すように、投資家は売買手数料を支払うことでサービスの対価を間接的に負担しているものの、証券会社全体では発行会社の利害を重視する方向にある。2002年の発行会社へのヒヤリングでは、発行会社がセルサイド・アナリストや証券会社の引受部門に投資レーティングへのクレームをつけることがかつては横行し、ヒヤリング当ても投資レーティングが引受証券会社を選ぶ際に参考とされ、発行会社にとってネガティブな投資推奨を行うセルサイド・アナリストが情報提供等で冷遇されることもあるとのコメントがあった（日本証券アナリスト協会 2003a、p.33）。

外部評価者が説明責任を目的に含む評価を受託した場合、インセンティブの捻じれが生じる（図3を参照）。説明責任を求められる局面においては、外部評価者は偏りのない価値判断に基づく評価報告書の作成が必要となる反面、外部評価者を選任し、調査費用を支払う評価委託者の意見は無視で

図3 外部評価者（説明責任）の利害関係



きない。また、調査費用に加えて、評価委託者は評価対象となる政策、プログラム、事業に関する主要な情報提供者でもあるケースも多く、この点からも評価委託者の意向を踏まえて、協力的な関係を築くことは、外部評価者が業務を円滑に遂行する上で非常に重要である。しかしながら、評価委託者が評価報告書の記載内容に関して圧力を加える状況では、良心的な外部評価者は公益と私益の間で板挟みとなる。このジレンマは、インセンティブの捻じれがある評価の業務そのものに根差すものであり、洋の東西を問わず発生している。服部（2013）は日本評価学会会員への質問紙調査を精査した結果、評価結果の変更を評価委託者より依頼されるケースがあることを明らかにしている（p.25）。同様に、米国でも、AEA会員への質問紙調査に基づき評価委託者から評価結果を変更するよう圧力を受けるケースがあることが指摘されている（Morris and Cohn 1993, pp.628-632, Morris and Clark 2012, pp.60-64）。

エージェンシー理論の枠組みを適用すると、受益者（プリンシパル）以外がサービス提供者（エージェント）を選任している点で、説明責任を目的に含む評価を実施する外部評価者とセルサイド・アナリストは共に利益相反に陥りやすい環境にある。また、情報提供の面でも、プリンシパル以外の利害を優先しやすい。加えて、公正な価値判断がそれぞれのプロフェッションの提供するサービスの核となるが、プリンシパルがエージェントの作成する報告書から価値判断の適切さを判断しにくいという課題（情報の非対称性）を抱えている。

証券アナリストと評価者は、資格の任意性、利害関係の多様性といった類似する環境にあり、インセンティブの捻じれが発生するケースがある点で共通する課題を持つ。インセンティブの捻じれは、利益相反の温床となるものであり、職業倫理のあり方を考える上で、看過できない特質である。

4. 証券アナリストと評価者の比較分析

本稿では、証券アナリストと評価者の職業倫理の比較に際し、「2. 職業倫理分野の先行研究レビュー」で言及した梅津（2002）のフレームワークを適用する。

具体的には、「理論」としての職業倫理では職業倫理研究、倫理綱領及び綱領内の基本原則を、「制度」としての職業倫理では倫理綱領内の行動指針、罰則規定、その他の規則、「実践」としての職業倫理では専門職資格における倫理研修、職業倫理関連の各種資料、ケースの活用を採りあげる。

(1) 「理論」としての職業倫理

①職業倫理研究

日本証券アナリスト協会は月刊機関誌として「証券アナリストジャーナル」誌を発行しており、検定会員に配布されている。同誌は特集論文、査読付き投稿論文、識者のオピニオンページ、書評、活動報告等を掲載している。職業倫理関連では、上述の米国における証券アナリストの不祥事を契機に2002年から2004年に集中して合計20本以上の論文や関連記事が掲載された。2005年以降に関しては、職業倫理そのものをテーマとした論文の掲載はない。但し、証券アナリストの職業倫理は金融分野における職業倫理に内包される論題であるため、関連論文の掲載先が分散する傾向にある。例を挙げると、「証券経済研究」誌や「資本市場クォーターリー」誌は米国の証券アナリスト不祥事についても関連論文や記事を掲載しており、近年ではサブプライムローン危機における格付会社の利益相反に関連し、格付アナリストの職業倫理にも言及する論文が見られる⁴。日本における証券アナリストの職業倫理研究は、金融不祥事と関連

して増減する傾向がある。

日本評価学会は年3回（うち和文2回、英文1回）学会誌を発行し、その内容は査読付き論文が中心となっている。評価倫理や倫理上の課題を主たる研究テーマとした論文は少ないが、「ガイドライン」の理事会承認を受けて2013年の第13巻第1号でミニ特集が生まれ、関連する2本の論文が掲載された。また、特定の評価手法に関連し手法適用上の課題として、倫理に一部言及する論文がある。一例を挙げると、佐々木（2006）ではインパクト評価における倫理上の課題を論考している（p.47）。

証券アナリストの職業倫理研究の動向をみると、職業倫理は不祥事といったプロフェッション全体の危機感がない限り取り組まれることが少ない研究テーマである。評価分野においても職業倫理を主題とする論文は少なく、研究の蓄積を図るには定期的な特集といった取り組みが必要だろう。一方、評価手法研究は関心の高い研究テーマであり、かつ手法実践上の課題が扱われることも多い。そのため、執筆者、査読者共に、意識して倫理面の課題を評価手法関連の論文に反映することで、より多くの読者が評価倫理を考える機会を持つものと思料される。

②倫理綱領

分析する倫理綱領として、日本証券アナリスト協会は「職業行為基準」（2012年3月23日改定）、日本評価学会は「ガイドライン」を採りあげる。「職業行為基準」は、前文、定義、総則（基本原則に相当する）、行動指針から構成され、「ガイドライン」も同じ構成となっている。倫理綱領の内容のうち、倫理的行動の根拠を示す前文や基本原則は「(1) 理論としての職業倫理」で、具体性のある行動指針は「(2) 制度としての職業倫理」で採り上げる。

「職業行為基準」及び「ガイドライン」共に、倫理綱領の策定に至った経緯や倫理綱領の意義を明示しており、概略的に職業倫理全体を順守すべき根拠を述べている。一方、規範倫理学といった理論の源泉まで遡って基本原則個々の根拠を示してはならず、基本原則間のコンフリクトに対処する目安は提示されていないため、コンフリクトへ

の対処は個々の専門職に委ねられている。

「職業行為基準」では会員の権威と信用向上に向けた自主的ルール確立を制定の理由に掲げており、プロフェッションの自己規制を強く意識した内容となっている。「職業行為基準」は日本証券アナリスト協会の所属会員（個人会員、法人会員の双方）を対象とし、会員外への職業倫理普及に関して明示的な言及はない。但し、「職業行為基準」は同協会のホームページを通じて会員以外も入手ができ、関心のある層がその内容を参考とすることは可能である。

「ガイドライン」は、評価文化の醸成に向けた評価に携わる者すべての倫理・行動規範の提示、評価に関する知識や技能の向上のための標準的な手順の整理、の双方を制定の理由として言及し、自己規制ばかりではなく、プロフェッションに特有の知識体系を整理することも考慮されている。「ガイドライン」では、評価文化の醸成の観点から対象者を広く設定し、所属会員のみならず、「評価に携わる者」を対象としている。この相違点は、倫理綱領の目的の違いを反映している。「ガイドライン」は日本評価学会のホームページを通じて学会員以外も入手が可能である。

③基本原則

基本原則はプロフェッションが重視する価値を明示したものであり、行動指針の根拠となるものである。「職業行為基準」は基本原則を五項目、「ガイドライン」は七項目を設定している（表3を参照）。

公共の利益への配慮、誠実さは両者に共通するが、「ガイドライン」は公共の利益への責任を独立した項目として扱い、よりその価値を重視している。専門的な能力を保持し、向上させる点でも両者は共通する。評価者の独立性は客観性を損なう圧力の排除を意図しており、証券アナリストが重視する「公正かつ客観的な判断」と一致する価値である。

「職業行為基準」の特徴としては、法人会員を対象とした基本原則項目があることが挙げられ、法人会員には所属する個人会員の指導と証券分析業務の独立性・客観性を確保することが求められている。上述の通り、セルサイド・アナリストの投資推奨には同じ組織内の引受部門から圧力がかかりやすいため、個人会員ばかりでなく、法人会員も倫理面の課題に対処することを基本原則の段階で規定している。また、業務実施において抵触すべきでない事項を定めた法令や規則があるため、それらの順守が明示的に示されている点も「ガイドライン」とは異なる点である。

「ガイドライン」はデータに基づく体系的調査の実施や有用性の確保を含むが、「職業行為基準」は、それに該当する項目を含んでいない。また、「ガイドライン」上では、有用性の項目で利用者の意思決定への情報提供に寄与すべき点が明記されている。「職業行為基準」と「ガイドライン」の差異からは、社会科学分野の調査手法を体系的に活用し、公共の利益に配慮しつつも、クライアントに意思決定に資する情報を提供するという評価者のサービスの特徴が浮かび上がる。

表3 基本原則項目の比較

	日本証券アナリスト協会 職業行為基準 総則	日本評価学会 評価倫理ガイドライン 基本原則
共通項目	<ul style="list-style-type: none"> ・社会的な役割を踏まえた、誠実な職務 ・専門能力の維持・向上 ・公正かつ客観的な判断 	<ul style="list-style-type: none"> ・公共の利益への責任
		<ul style="list-style-type: none"> ・誠実
		<ul style="list-style-type: none"> ・専門的能力の保持・研鑽
独自項目	<ul style="list-style-type: none"> ・法令・規則の順守 ・(法人会員) 会員の指導及び所属会員の独立性の確保 	<ul style="list-style-type: none"> ・人々への敬意 ・体系的調査 ・有用性

(出所) 著者作成

Fitzpatrick (1999) が指摘した点—評価者の職業倫理は研究専門職 (Scholarly Professional)、コンサルティング専門職 (Consulting Professional) の双方の性格を有する—は証券アナリストと評価者の比較でも明らかである⁵ (p.8)。

「ガイドライン」は「評価に携わる者」という個人を対象者とするため、個人ばかりでなく、所属組織にも規範を求める点はそのまま導入できない。しかし、外部評価者の所属組織が評価委託者との間に有する利害関係が、評価判断の客観性に影響する可能性がありうる点は考慮すべきだろう。

(2) 「制度」としての職業倫理

①行動指針

「職業行為基準」、「ガイドライン」の行動指針は行為者を特定し、行為者の望ましい行動を提示している点では共通する。「職業行為基準」であれば、会員の行動を、「ガイドライン」であれば、評価を行う者、評価を依頼する者、評価に携わる者の行動を、それぞれ規定している。

他方、プロフェッション全体で共有できる単一の利害関係モデルがない点への対応は異なる。「職業行為基準」の行動指針は、特定層の利害関係をモデルとして想定せず、かわりに倫理上の問題がおきやすい状況を特定し、その状況下での適切な行動を提示する。また、「職業行為基準」では、ほとんどの行動指針において「～しなければならない」や「～してはならない」といった強制力のある表記が見られる点が特徴的である。強制力のある行動指針の表記は、「職業行為基準」の目的であるプロフェッションの自己規制を色濃く反映している。

「ガイドライン」は、外部評価者の利害関係をモデルに設定し、利用者の利便性を考えて適切な行動を評価のフェーズ毎に網羅的に提示している。行動指針の文末は、「～しなければならない」や「～してはならない」という表記は非常に少なく、「～する」といった行動を推奨する記述となっている。行動指針に強制力を持たせない記述は、(1)の②で言及した両者の目的の違いに由来し、また (2)の③で述べる倫理規程の解釈・適用の困難さも考慮すべきため、「職業行為基準」を単純

に参考にはできない。しかし、倫理面でのジレンマが起きやすい状況に関する規程を重点的に拡充することは「ガイドライン」でも取組み可能であり、判断の誤りが深刻な問題に繋がる局面において利便性を高める。但し、そのような局面を特定するためには、ステークホルダー (学会関係者、評価士、評価委託者等) から事例を収集する仕組みの構築が重要となる。

②報告書作成時の留意事項

日本証券アナリスト協会の規律委員会は、2003年5月に「アナリスト・レポート作成の際の留意事項」という公式文書を発表している。投資情報及び投資推奨の主な伝達手段であるアナリスト・レポートの表記について、この文書はA4で6枚にわたって注意点を列挙している。この文書は投資推奨を行うセルサイド・アナリストのレポートを特に念頭に置き、その内容は「職業行為基準」を改めて強調した事項 (事実と意見の明確な区分、重要な事実の遺漏のない正確な記載等) に加え、投資レーティングの定義の記載、ネガティブな情報を記載する必要性、論理的で冷静な表現、比較基準の具体的な提示等が含まれ、総じて投資家が価値判断の質を確認するための情報提供に主眼が置かれている (日本証券アナリスト協会 2003b)。エージェンシー理論の枠組みからこの措置を捉えなおすと、セルサイド・アナリストと投資家間にあるサービスの質をめぐる情報の非対称性を緩和し、利益相反が起りにくい環境を作り出すことを狙ったと言えるだろう。

「ガイドライン」の行動指針でもレポートティングを1つの章に位置付け、「ガイドライン」は、委託者を始めとする関係者の利害を踏まえつつも、評価結果を客観的に提示し、結果の解釈や価値判断の根拠となる価値観、前提、理論に関しても記載することを推奨しているが、報告書表記に関する部分は上記規則ほど具体的で微細に及ぶ内容ではない。報告書はいずれのプロフェッションについても、サービスを提供する重要なツールであり、関連する規則が策定されているが、すでに問題が明らかとなっているセルサイド・アナリストに関してはより詳細な規程が定められている。評価分野においても、外部評価者が一般に公開する報告

書を作成する場合には、情報の格差を埋める上で、報告書の表記や評価判断に関する規程の拡充は検討に値するだろう。

③懲戒及び注意

日本証券アナリスト協会は、2002年以降、懲戒事由には該当しないが、職業倫理に照らし問題がある行為には、規律委員会の議決により、書面もしくは口頭で会員に注意を与える制度を導入している。2002年以前においては「職業行為基準」に違反した場合、会員への懲戒処分を行うことを同協会の定款に規定していた。この制度改定は倫理面の問題が深刻になる前に倫理的な問題に対処する手段であり、対処手段を増やすことでより柔軟に制度運営を行う仕組みと言える。「職業行為基準」が目標とするプロフェッションの自己規制とも懲戒・注意制度は整合的である。但し、懲戒・注意制度の導入が可能な理由として、「職業行為基準」や「アナリスト・レポート作成の際の留意事項」にて倫理面で問題の生じやすい状況を特定し、強制力のある行動指針が設けられていることに留意すべきである。

一方、「ガイドライン」では懲戒や注意は規定しないことが明記されている。AEAの倫理綱領である「評価者の基準原則」(Guiding Principles for Evaluators)でも、注意・罰則といった強制メカニズムには言及がない。その理由として、評価における方法論の多様さ、プロフェッションとしての新しさにより倫理規程の解釈や適用に明確なコンセンサスがなない点が指摘されている(Fitzpatrick 1999, p.12)。プロフェッション内において、倫理規程の解釈や適用を共有しにくく、懲戒・注意制度の運用が容易でない点は、評価者の「制度」の特質と指摘できるだろう。

(3)「実践」としての職業倫理

①職業倫理に関するハンドブック

日本証券アナリスト協会は1995年に会員向けに「職業行為基準」を解説した「証券アナリスト職業行為基準実務ハンドブック」(以下、「実務ハンドブック」)を刊行した。2002年には実際に発生している問題をふまえて事例を拡充し、適切な行動について具体的な理解を深める方針を打ち出し

た。実務ハンドブックは「職業行為基準」のすべての基準項目に解説を加え、併せて事例を使って行為を記述し、その行為が当該基準に違反しないかを明示的に述べている。このような解説書は行動指針の実務への適用に困難を感じる初学者には特に有意義と考えられ、証券アナリスト資格取得時のテキストとしても活用されている。

他方、「ガイドライン」は策定されて間もないため、解説資料は作成されていない。実践的な解説資料の作成には、幅広く実務者の経験を吸い上げ、継続的に事例をあつめるプラットフォーム作りが必須であり、実務者への資料作成への協力、質問紙調査、ステークホルダーへのヒヤリングといった取組みがタスクとして視野に入ってくる。

②職業倫理に関する資料集

日本証券アナリスト協会は資料集「職業倫理を考えるために ルールと参考資料」を2004年に発表した。2000年代初頭に米国のセルサイド・アナリストの利益相反が問題となった際に、倫理綱領の改定が内外で進められ、それに関連する資料が多数作成された。この資料集は、その時期に同協会が発表した各種倫理規程やポジションペーパー、発行会社からのヒヤリング結果、国際的な職業倫理への取組みに関する論文等を取り纏めている。倫理綱領や制度の改定の背景を理解し、個人が職業倫理を再考する上で有用な資料であり、本稿で言及している文献も多数この資料集に含まれる。この資料集は一般公開され、現在も同協会のホームページから入手可能である。

日本評価学会では、職業倫理関連の資料集は作成されていない。「ガイドライン」の策定から日が浅く、評価者による深刻な不祥事もないため、日本評価学会が職業倫理に関するポジションペーパーを作成する必要性は薄い。また、発表された倫理関連の論文も多くはないため、資料集の作成に十分な研究の蓄積がないことも指摘できる。資料集は職業倫理を考える貴重な資料であり、「ガイドライン」を実践する上でも参考となる。評価分野においても、倫理研究の蓄積が進むに伴い可能となる取組みであり、長期的には導入の検討が望ましい。

③資格取得時の倫理教育

日本証券アナリスト協会の検定会員となる資格試験を受験するにあたり、受験者は通信教育講座に参加する制度となっている。職業倫理はこの通信教育講座で取り上げる独立した科目の一つとなっており、上述の通り「実務ハンドブック」が通信教育講座のテキストとして使われている。証券アナリスト試験でも職業倫理に関連する設問が含まれるため、検定会員になるには職業倫理の知識が必須となっている。また、講座を通じて、初学者の意見を取り入れ、内容を理解しやすいよう「職業行為基準」のうち投資情報の提供に関する項目を整理する変更が2012年に行われた。この変更は、実務者の意見を吸い上げ、倫理の「実践」と「制度」の間でフィードバック関係が成立していることの証左と見なされるだろう。

評価分野においても、評価士資格を取得する上で、職業倫理の学習は必要となっている。評価士養成講座においても評者者倫理を扱う講義が設けられ、「ガイドライン」内容や評価者の社会的な役割に関して初学者への教育が実施されている。評価士認定試験に関しても評価倫理はその出題範囲に含まれている。職業倫理はプロフェッションとしての判断の根幹を扱うものであるだけに、証券アナリスト、評価士共に初学者向けの講義にも含まれている。さらに、証券アナリストでは、初学者向け講座から行動指針へのフィードバックを得ている。この点を踏まえると、より理解しやすい行動指針を作る上で、評価士養成講座から意見を吸い上げる取組みは今後、検討に値するだろう。

④資格取得後の継続教育

「職業行為基準」は職務にふさわしい専門能力の維持・向上を基本原則に含め、日本証券アナリスト協会は、2015年1月より検定会員向けの継続教育制度の導入を予定している。継続教育自体は検定会員に対して推奨するものであり、義務ではないが、倫理に関しては例外的に年間3クレジット（約90分相当）の活動（講演会やセミナー参加、講座受講等）が義務付けられる。

「ガイドライン」でも「専門的能力の保持・研鑽」は基本原則に含められており、評価士につい

ては資格取得者を対象にすでにフォローアップ講座が実施されている。しかしながら、フォローアップ講座ではこれまで倫理を扱った講義は実施されていない。評価者の模範となるべき評価士にはより高いレベルでの職業倫理の実践が求められるため、職業倫理は継続教育の一部に位置付けられることが望ましい。

⑤ケース・メソッド

ケース・メソッドは具体的な事例を討論し、その討論を通じて意思決定やその理由を考えるものであり、事例の当事者の立場に立ち、実践的な意思決定の訓練を行うものである。ケース・メソッド用の教材では、問題点や結論は明示されず、利用者は自ら状況を読み解く。明確な結論を得ることは困難で、オープンエンドな議論を想定した題材も多い。ケース・メソッドは、研究手法として具体事例に基づき問題や結論を提示するケース・スタディとは性格を異にするものである（梅津2002, pp.75-77）。

日本においては、証券アナリストを対象としたケース集は準備されていない。証券アナリストの実務ハンドブックでは、事例を利用しているが、これは教育目的の議論を期するものではない。事例は短い記述にとどまり、行為の適否はわかりやすく、あくまでも行動指針の理解を助けるためのものである。ケース・メソッドの導入は、証券アナリストにおいて取組みが十分ではない分野と言えるだろう。

評価分野では、米国においては Morris (2008) が編集したケース集があるものの、同ケース集は和文には翻訳されておらず、わが国の評価者教育において幅広く利用されるものではない。同ケース集は主要な評価プロセス毎（例えば、契約時、レポート作成時等）に1章を設け、章毎に2ケースが用意されている。各章の冒頭でケースを1つ提示して評価経験のある識者がコメントを述べ、もう1つのケースに関しては識者のコメントはなく、議論を導くための質問のみを提示している。利用にあたっては、最初のケースとコメントを見て、倫理的な行動を考える手掛かりを得た上で、各自がケースを検討する順序となっている。評価分野では、倫理規程の解釈や適応にコンセンサス

が少なく、個々人に判断を委ねる部分が大きくなる。また、評価者の利害関係は業務毎、さらには業務の局面にて異なってくるため、自らの置かれた状況を的確に把握する能力も培う必要がでてくる。これらの観点から、ケース・ディスカッションによるトレーニングは評価者にとって有意義であり、ディスカッション用の教材作成と演習は取り組む価値があると言える。

5. おわりに

本稿は、職業倫理における「理論」、「制度」、「実践」の観点から包括的に証券アナリストと評価者の比較分析を行い、倫理上の共通課題であるインセンティブの捻じれについては特に論考を深めた。証券アナリストの取組みとして特筆される事項は、(1) 行動指針や追加のガイダンスにおいて倫理上の問題が生じやすい局面とその対応を詳述する、(2) 詳述された規程に基づき罰則を設けて抑止のインセンティブを作る、(3) 各種資料提供や有資格者への継続教育通じ個々人に職業倫理を意識させている、の3点であろう。

評価者においては、倫理綱領の目的や倫理規程の解釈と適用の難しさから、罰則を伴うアプローチをそのまま取り入れることはできない。しかし、深刻な倫理上の問題が起こりうる局面につき記述を拡充することは、「ガイドライン」の利用価値を高めるであろう。特に評価報告書において価値判断の適切さを確保し、利用者がその質を確認できるような取組みは、インセンティブの捻じれへの対策として検討に値する。

職業倫理の要となる「ガイドライン」はすでに策定され、今後は「実践」から「制度」に対して、行動指針の修正や追加に向けたガイダンスをフィードバックする段階にある。そのため、「制度」においては、実務者から事例を継続的に汲み上げるプラットフォーム作りが課題となる。また、職業倫理に関する各種資料の作成には事例の蓄積が必要であり、この観点からもプラットフォーム作りは重要な役割を果たす。

評価分野において倫理規程の解釈や適用が共有されにくいことは、他のプロフェッション以上に、

個々人が状況を読み解き、職業行為に反映させる必要性を示唆する。この観点から、ケース・ディスカッションによる訓練は有用性が高いと考えられる。「実践」においては、評価倫理に関するケース・ディスカッション用の教材作成や演習が今後、取り組むべきタスクであろう。

最後に、今後、研究を掘り下げるべき事項に言及し、本稿のまとめとしたい。まず、本稿は専門職団体の取組みを中心に扱ったものであり、所属組織や業界団体からも職業倫理向上への示唆を導出できるだろう。本稿で取り扱った証券アナリストを例にとると、引受金融機関、格付会社、日本証券業協会は別途倫理規程を有している。また、インセンティブの捻じれを持つ他のプロフェッション（一例として、公認会計士）との比較も、利益相反という深刻な問題を防止する上で、有意義な示唆があるだろう。

注記

- 1 会社法及び金融商品取引法では、株主総会の決議で会計監査人を選任するが、取締役会は議案を株主総会に提出する機関であり、実質的な選任は取締役会が行う。
- 2 バイサイド・アナリストは、機関投資家に所属し、社内向けの投資情報提供や投資推奨に従事する。
- 3 セルサイド・アナリストは、証券会社等の引受金融機関に所属し、投資家向けの投資情報提供や投資推奨に従事する。
- 4 格付アナリストは格付会社に所属し、債券元利払いの確実性をレーティングする。
- 5 コンサルティング専門職の例として、弁護士、医師、建築家、会計士、心理療法士等が挙げられる。

参考文献

- 梅津光広 (2002) 『ビジネスの倫理学』、丸善出版
小侯光文 (2012) 「第2章 倫理研究の基本的視角」、藤沼亜起『会計プロフェッションの職業倫理』、同文館書店、13-16
桑野忠雄 (2011) 「第9章 会計プロフェッションの職業倫理と行動規範」、首藤恵『金融サービスのイノベーションと倫理』、中央経済社、155-176

- 小林信行・一寸木英多良 (2010) 「『評価者倫理・スタンダード策定分科会』中間報告-基本原則 (Principles) 案について」、『日本評価学会第11回全国大会発表要旨集録』、147-154
- 佐々木亮 (2006) 「ODA分野における『エビデンスに基づく評価』の試み：「貧困アクションラボ」の動向」、『日本評価研究』、6 (1) : 43-54
- 首藤恵 (2011) 「第1章 金融サービス業の職業倫理」、首藤恵『金融サービスのイノベーションと倫理』、中央経済社、2-24
- 日本証券アナリスト協会 (2002) 「証券アナリストの職業倫理を高めるために」、『職業倫理を考えるために』、9-22
- 日本証券アナリスト協会 (2003a) 「発行会社の証券アナリストに対する見方・意見-発行会社からのヒヤリングの概略-」、『職業倫理を考えるために』、31-35
- 日本証券アナリスト協会 (2003b) 「アナリスト・レポート作成の際の留意事項」、『職業倫理を考えるために』、25-30
- 八田進二・町田祥弘 (2003) 「会計プロフェッションの自主規制における職業倫理の位置づけ」、『企業会計』、55 (3) : 39-48
- 服部浩昌 (2013) 「日本の評価者が抱える倫理的な問題 「評価者倫理・スタンダード」策定に向けた質問紙調査の分析結果」、『日本評価研究』、13 (1) : 19-31
- Boatright, J.R. (2014). *Ethics in Finance 3rd edition*. Chichester, West Sussex: John Wiley & Sons.
- Fitzpatrick, J. L. (1999). Ethics in Discipline and Professions Related to Evaluation. In Fitzpatrick, J. L. and Morris, M. (Eds.) , *Current and Emerging Ethical Challenges in Evaluation*. San Francisco, CA: Jossey-Bass, 5-14.
- Milgrom, P. and Roberts J. (1992). *Economics, Organization and Management*. Prentice Hall (奥野正寛等訳 (1997) 『組織の経済学』、NTT出版)
- Morris, M. (2008). *Evaluation Ethics for Best Practice*, New York: Guilford Press.
- Morris, M. and Cohn R. (1993). Program Evaluators and Ethical Challenges, *Evaluation Review*, December 1993, 621-642.
- Morris, M. and Clark B. (2012). You want Me to Do What?: Evaluators and the Pressure to Misrepresent Findings, *American Journal of Evaluation*, 34 (1) , February 2013, 57-70.

(2014.11.7受理)

Financial Analysts' and Evaluators' Efforts to Improve Ethics: Comparative Study of Professional Ethics

Nobuyuki Kobayashi

Maastricht University

n.kobayashi@student.maastrichtuniversity.nl

Abstract

The Japan Evaluation Society approved the Guidelines for the Ethical Conduct of Evaluations in December 2012. As an important feature of professionals is self-governance, this means that evaluators are more matured as a profession. Since other professions have tackled this task for long periods, their efforts can provide useful implications to evaluators. This paper compares financial analysts and evaluators in light of three ethical aspects: theory, institution, and practice.

Detailed description on situations which may cause ethical problems can improve practical value of the Guidelines but this requires a platform to continuously collect actual cases. Each evaluator has more room for judgment because consensus is hardly made on interpretation and application of ethical codes in this field. As case discussion is a meaningful training, preparation and use of teaching materials for case discussions is an effort worth making.

Keywords

Professional Ethics, Guidelines for the Ethical Conduct of Evaluations,
Financial Analyst, Evaluator, Agency Theory

【研究ノート】

「業績情報提供型」業績予算： その実践概要と使用される情報及び手法

荒川 潤

愛知県庁¹

arakawa525jun@gmail.com

要 約

本稿では、政策評価の結果が政府の資源配分の決定にどのように活用されるのかを、「業績予算」(Performance Budgeting)を例にして議論する。具体的には、「業績情報提供型」業績予算(Performance-informed Budgeting: PIB)の実践概要を整理するとともに、どのような情報がどのような手法により活用されているのかを、国内外の事例により分析して議論する。

PIBの実務を見ると、個別案件の業績情報は成果重視の経営の下で最も重要かつ有益な情報であるものの、それだけでは予算の増減を含めた今後の方向性を議論できず、政府や省における優先事項についての情報や、個別案件についての(業績以外の)情報、他案件についての情報なども提供されて初めてそのような議論が可能となる。言い換えれば、PIBでは省や政府全体の成果がまず重視され、その予算配分の効率性を高めるための個別案件の在り方が、業績情報や他の多様な情報を活用しながら判断されている。

このような多様な視点からの情報を効果的に分析するツールとして、レーティング手法が用いられている。レーティングには手法に内在する多くの制約があるものの、実務的に有用なツールであるため、多くの政府によって使用されている。業績に加えてそれ以外の情報も含めて構築されたレーティングの結果は、予算配分の決定にやや直接的に反映されている。

キーワード

業績予算(Performance Budgeting)、「業績情報提供型」業績予算(Performance-informed Budgeting: PIB)、業績情報、レーティング、成果重視の経営

1. はじめに

本稿では、政策評価²の結果がどのように資源配分に活用されるのかについて論じる。具体的には、国や自治体の「業績予算」(Performance Budgeting、Performance-based Budgeting、Results-based Budgetingなど)の実務にて、どのような情

報がどのような手法により活用されて判断が下されているのかを、国内外の事例により分析する。

政策評価の結果を用いてより適切な資源配分を実現することは、政策評価を実施する目的の一つとして位置づけられることが多い。わが国の「行政機関が行う政策の評価に関する法律」でも、政策評価の結果は、当該政策に適切に反映させるこ

とに加えて、予算の作成に適切に活用すべき旨が規定されている。

しかし政策評価の結果を具体的にどのように資源配分に反映させるかについては、国内外にて様々な取組みがなされてはいるが、その手法の確立にはまだ至っていない状況にあると思われる。そのような中、本稿は今後の手法の確立に向けた検討の一端となることを試みるものである。

なお本稿では「業績情報」を、成果（アウトカム）についての情報（例えば目標達成度など）を指し示すものとして用いている。

1. 「業績情報提供型」の業績予算

まず、先行研究及び実践事例などにより、業績予算、特に本稿での議論の中心となる「業績情報提供型」業績予算（Performance-informed budgeting : PIB）の概要について整理する。

(1) 業績予算とは

業績予算は、政策評価の結果を活用して予算を編成することを意図した公共経営の枠組みである。OECD（2007）では業績予算を、「配分される資金を測定可能な成果に関連づける予算方式の1つである」と定義づけている。（なおわが国には業績予算との言葉は用いていないものの、取組みの実態からそのように判断しても差し支えないと思われるものがあり、本稿ではそれらを含めて議論している。）

(2) 業績予算の類型化

OECD（2007）は業績予算を、業績情報と予算との連携の程度により、以下の3類型に分類して

いる（表1参照）。まず両者の連携の程度が最も低いのが「提示型業績予算（Presentational performance budgeting）」である。ここでは、業績情報は単に予算文書もしくは他の行政文書にて提示されるのみであり、業績情報は資源配分の意思決定にて役割を果たしていない。

逆に、最も連携の程度が強いのが、「直接/計算式型業績予算（Direct/formula performance budgeting）」であり、業績情報と資金配分とは直接的にリンクしている。ここで直接的なリンクとは、業績を資源配分に直接反映させることを意味する。業績予算として一般的に想定されやすいのがこの分類であろうと推察されるが、OECD（2007）は、この類型は限定的な国の限定的な分野（教育、保健など）のみにて使用されているとしている。

そして両者の中間に位置するのが「PIB」であり、ここで業績情報と資金配分とは間接的にリンクしている。間接的なリンクとは、業績情報がそれ以外の情報と共に予算の意思決定に体系的に使用されていることを意味する。業績情報は意思決定のプロセスにて重要な役割を果たすが、しかし配分される資金の量を必ずしも直接的に決定づけている訳ではない（OECD 2007）。

World Bank（2010）は、OECDによる業績予算の分類を更に進めて、業績予算を、予算プロセスにて、「成果情報を制度に基づいて体系的に使用するかどうか」、また「成果と資源配分とを厳密にリンクさせるかどうか」の2軸から、4つの類型に分類している（表2参照）。そして、予算プロセスにて、制度に基づいて体系的に業績情報を使用するが、但しそれだけで自動的に予算額を決定するわけではないという意味で「緩やかなリンク」である業績予算の類型を、PIBと定義づけている。

表1 業績予算の3類型（OECDの分類）

分類	業績情報と資金配分とのリンク	計画上の、又は実際の業績	予算プロセスにおける主要な目的
提示型業績予算	リンクなし	業績目標、及び/又は業績(成果)	説明責任
業績情報提供型業績予算(PIB)	緩い/間接的なリンク	業績目標、及び/又は業績(成果)	計画、及び/又は説明責任
直接/計算式型業績予算	強固な/直接的なリンク	業績(成果)	資源配分及び説明責任

(出所) OECD（2007）

表2 業績予算の4類型（World Bankの分類）

		予算配分とのリンク	
		厳密なリンク	緩やかなリンク
業績情報の利用	制度的な利用	直接／計算式型業績予算 ■教育、保健分野などに事例がある。	業績情報提供型業績予算（PIB） ■実例が最も多い類型。OECD諸国でも多くの事例がある。
	アドホックな利用	日和見主義型業績予算 ■特定の施策に対して、日和見主義的に、業績情報が厳格に分析される。	提示型業績予算 ■高度の社会責任が課せられており、政府が自主的に社会的な要請を追求するケース。例として北欧諸国。

（出所）World Bank（2010）を一部修正

しかしここで「緩やかなリンク」とは、業績情報を、単に予算プロセスにおける分析ツールとして体系的に用いるのではなく、予算額の決定に少なくとも何らかのインパクトを及ぼすことを意図して、例え緩やかであっても、成果情報と予算配分との間にある程度のリンクがあることを指す、としている（World Bank 2010）。

PIBはこのように定義されるため、業績目標を達成しても予算が減額される（もしくは案件が廃止される）場合もあれば、業績目標が未達成でも予算が維持されたり増額されたりする場合もある。国内外にて政策評価の結果を資源配分に結びつけようとする実務的な試みの多くは、このPIBに該当するものだと考えられる。

（3）PIBの導入状況

表2にも示すように、World Bank（2010）によるとPIBはすでに多くのOECD加盟国にて実施されている。OECD（2007）では、加盟8カ国の事例が紹介されている³。

米国では、州政府や地方政府でのPIBの取組みも確認されている。例えばMelkers & Willoughby（2007）は、全米50州の州政府への調査結果（2004年）を基に、50州の全てにおいて、州法（33州）もしくは行政規則（17州）に基づく何らかのPIBが実施されていると報告している。

わが国（国レベル）の場合も、例えば各府省により「行政事業レビュー」などを活かした予算要求がなされ、財務省も自ら実施する「予算執行調査」などを活かしながら予算編成している実態を

踏まえると、PIBが実施されているとみて差支えないと考える。また地方自治体の場合にも、政策評価の結果情報を予算編成プロセスにて体系的に活用している例は、本稿で取り上げる自治体を含めて少なからず存在していると考えられる。

PIBはまた、先進国に限らず開発途上国でも導入が進められている。World BankやIMFも近年、PIBについての技術協力、情報提供（概念整理、開発途上国での導入事例紹介、マニュアル開発、等）を行うなど、様々な支援を行っている⁴。

（4）PIBに使用される業績情報の種類

表1が示すように、PIBにて使用される業績情報には、「計画上の業績目標」と、「実際の業績（成果）」の双方がある。これは、PIBを更に分類する軸となっている。

この内、計画上の業績目標を用いるのは、ある目標水準のアウトカムをある水準の予算によって購入するという契約に近いアプローチにおいてであり、例えば英国のブレア、ブラウン両政権により実施されていた公共サービス合意（Public Service Agreement : PSA）やオランダ等にその例が見られる（World Bank 2010）。また、オズボーンなどが提唱する「アウトカム予算（Budgeting for Outcomes）」もこの範疇に含まれよう（Osborne & Hutchinson2004）。このアプローチの下では、設定された計画上の業績目標を如何に達成するか、という業績の進捗管理にマネジメントの主眼が置かれる。

逆に、実際の業績（成果）を用いるのは、業績

を基に次期（次々期）の予算を検討するアプローチであり、米国、カナダ、韓国、日本（国、地方自治体）などにその例が見られる。ここでは、次期（次々期）の予算編成に対して、より合理的な根拠を如何に提供するかという業績を基にした方向づけにマネジメントの主眼が置かれる。

(5) 伝統的な予算編成とPIBとの相違

Joyce (2007) によると、伝統的な予算編成とは個々の案件における増分主義に基づく限界的な予算配分である。これに対して、PIBでは、成果重視の経営の下で、政府や省庁の戦略計画や優先事項を踏まえた上で、予算額（インプット）と成果との関係に着目して、政府や省庁全体として「予算配分によってどれだけ多くの成果が生み出されるようになるか」などを分析した上で予算編成がなされるとしている（表3参照）。

(6) PIBの目的（成果重視の経営への貢献）

①業績マネジメントへの貢献

PIBは、予算と成果とを結びつけて議論することによって、「成果重視の経営」（Results-based management, Managing for results）の下での業績マネジメントに直接的に貢献する。PIBは、政府機

関がより良い成果をあげるような圧力として機能すると共に、政府職員の成果を上げようとするインセンティブを高めることにもつながる。

またPIBは、それ単独で導入されるのではなく、その「成果重視の経営」の実現に向けた一連の改革の一環として実施される（Robinson 2007、Robinson and Last 2009）。例えば、政策評価制度改革、公務員制度改革、組織改革、分権化などの制度改革（現場への権限移譲など）などが想定される。この点について、Schick (2007) は、予算と業績情報とのリンクが如何に行政職員の成果をあげるインセンティブにつながるとは言え、経営や組織文化の支えがないとPIBは成功せず、政府や組織内で「成果重視の経営」が実施されていて初めて、成果のための予算編成もなされうと指摘する。

②公共支出管理への貢献

上記のような「成果重視の経営」の下での業績マネジメントに対するPIBの貢献は、視点を変えて公共支出管理（Public Expenditure Management: PEM）の観点からは、PEMの目的である、ア）予算配分の効率性（allocative efficiency）、イ）予算執行の効率性（operational efficiency）、ウ）総

表3 伝統的な予算編成とPIBとの相違

	伝統的な予算	PIB
着目点	■予算額そのもの	■予算額と成果の関係
予算額決定の視点	■個々の施策での増分主義による限界的な予算額の変化(例えば、昨年度に比べていくら増額するか)	■全施策における予算額と成果の変化(例えば、どの程度の増額により、どの程度多くの成果が得られるか)
予算と計画・経営との関係	■予算と、省庁の計画・経営との乖離	■予算と、計画・経営の統合
「予算」の意味	■配分された資源	■成果を実現する費用

(出所) Joyce (2007) を一部修正

表4 公共支出管理（PEM）の目的

PEMの目的	概要
予算配分の効率性 (allocative efficiency)	<ul style="list-style-type: none"> ■より良い成果が得られる分野・施策・事業への予算配分(生産性の低い資源活用から高い活用への転換) ■限られた財源を最大限に活用しようとする、ニーズや優先事項に応じた予算配分 ■政府の目標に照らした、古い優先から新しい優先への資源配分の転換
予算執行の効率性 (operational efficiency)	<ul style="list-style-type: none"> ■執行段階における、より少ない費用によるより多くの成果の追求 ■資源が配分された財・サービスのコストの削減
総合的な財政規律 (aggregated fiscal discipline)	<ul style="list-style-type: none"> ■上記の効率性が追求された結果としての、予算総額の抑制（コントロール）

(出所) Schick (2001)、Robinson and Last (2009)、World Bank (2010) を基に筆者作成

体的な財政規律 (aggregated fiscal discipline) の実現に、PIBが貢献すると整理される (Robinson and Last 2009、World Bank 2010) (表4参照)⁵。

社会のニーズや優先順位に応じて、それまでの予算の構成 (分野間の配分比率等) を変えていくことは、伝統的な予算制度の下での増分主義的な予算配分においては難しいが、PIBにより、配分される資源と業績目標 (計画・実績) と結びつけることにより、政府の優先順位に沿った予算編成が可能になる。そのことで、「予算配分の効率性」を高めることが可能となる。

またPIBのプロセスを通じた、より良い成果を求める圧力により、またその成果をできる限り少ない費用にて実現するような圧力により、「予算執行の効率性」を高めることが可能となる。

これらの効率性の追求の結果として、効率性の低い事業 (意図した成果を上げていないのに高額の予算は費やされている事業) に対する見直しの圧力が高まり、成果に結びつかない無駄な予算配分が避けられるようになる。ひいては、予算総額を抑制して、「総合的な財政規律」の維持に貢献

することとなる。予算総額の抑制は、「財政余力 (fiscal space)」を生み出すことにつながり、財政赤字の削減や、より優先度合の高い分野への再配分を可能にする。

(7) 予算プロセスの各段階におけるPIB

PIBは、予算編成時の予算査定のためにだけ実施されるのではなく、予算プロセスの各段階にて、様々なアクターにより実施されうると指摘されている (Joyce 2007、Joyce 2008、World Bank 2010)。予算プロセスの各段階とは、米国連邦政府を例にすると、①各省の予算要求、②大統領府行政管理予算局 (Office of Management and Budget : OMB) による「大統領予算」編成、③連邦議会による予算承認、④大統領による予算承認、⑤予算の執行、⑥監査及び評価、の6段階である。

Joyce (2007) は、それぞれにおけるPIBの概要を以下のように説明している (図1参照)。なおこれら全てが既に米国連邦政府にて実施されているという説明ではなく、この内容に沿った良い実践例が既に数多くみられる段階 (特に①②⑤) と、

図1 予算プロセスの各段階におけるPIBの展開 (米国連邦政府の例)

段階	準備段階	承認段階	執行段階	監査・評価段階		
	① 予算要求 <各省>	②「大統領予算」 <OMB/大統領>	③立法措置 <連邦議会>	④署名・否認 <大統領>	⑤施策の実施 <各省/OMB>	⑥監査・評価 <監査人/評価者>
実施・検討する事項	<ul style="list-style-type: none"> 予算要求の正当化 (justification) トレードオフ (省間での戦略的配分) の検討・実施 (施策の) 生産性の分析 (省内) 重複施策の分析 実施主体の決定 (内製・外部委託) 	<ul style="list-style-type: none"> トレードオフ (省間の戦略的配分) の検討・実施 予算案の正当化 (justification) (省間) 重複施策の分析 実施主体の決定 (内製・外部委託) 	<ul style="list-style-type: none"> 「業績決議」採択 (予算決議の一部として) 具体的な業績期待値の設定 (歳出権限決定プロセスの一部として) 明かな業績期待値の設定 (歳出法の一部として) 限界効用を得るための費用の比較 (立法措置のプロセスにて) 	<ul style="list-style-type: none"> 署名が否認かの決定 (歳出権限を決める根拠法案、歳出法案) 業績への影響を踏まえた政府見解 (Statements of Administration Policy) の表明 	<ul style="list-style-type: none"> 省の業績目標達成の観点からの法制上の制約の理解 業績目標の共有 内部・下部組織への予算配分 第三者への予算配分 予算執行期間中の費用と業績のモニタリング 業績向上策の検討・実施 適切な公共調達 	<ul style="list-style-type: none"> 財政面での規則遵守に加えた業績面での監査・評価の重視への転換 成功・失敗の判断 規則遵守状況の把握 経営改革の重点分野の特定 経営上・業績上の課題の抽出
使用する情報等	<ul style="list-style-type: none"> 省の戦略計画・業績計画 施策体系 (日々の業務と省の目標をつなぐもの) 費用/アウトプット/アウトカム 有効性・生産性 効率性及びその比較 (官民比較を含む) 人員体制 	<ul style="list-style-type: none"> 政府全体の戦略計画・実績計画等 各省の戦略計画・業績計画 国家目標達成に要する全体費用 (国の) 限界費用と限界効用 費用/アウトプット/アウトカム 有効性 効率性及びその比較 (施策間・官民間を含む) 	<ul style="list-style-type: none"> 大統領予算 業績情報 正確な費用見積 戦略計画・業績計画 費用 (プログラム毎) 「社会指標」 財政見直し 異なる水準の配分をした際の費用/アウトプット/アウトカム (限界費用・効用の比較) 	<ul style="list-style-type: none"> 立法府が承認した予算に対する、各省庁による政府の戦略目標達成の観点からの評価 承認された施策を通じて期待される成果 法における費用/アウトプット/アウトカムの関係 (「大統領予算」との比較) 	<ul style="list-style-type: none"> 政府全体及び省の戦略計画 承認された予算額の詳細 費用/アウトプット/アウトカム (省内外、目標と実績) 	<ul style="list-style-type: none"> 省の戦略目的 費用 業績報告書 ロジックモデル データの質と信頼性

(出所) Joyce (2007) の記述 (表及び本文) を基に作成

今後の取組みを推進していく必要がある段階（特に③）の双方が混在しているという前提での説明である。

最初のプロセスは予算準備（①②）であり、まずは、各省がOMBに提出する予算要求を準備する段階（①）である。省の戦略計画・業績計画を基に、省の重点事項に整合するように、施策体系に沿って、個々の活動・事業・施策の業績や費用の情報を踏まえて、予算案を編成する。省の戦略目的をより高い水準にて実現しうるように、様々な活動・事業・施策の「取捨選択や重視度合の相対的な変更」（trade-off）、類似・重複案件との比較を踏まえた実施判断、外部委託・内製化の判断などを戦略的に行う。

次は、行政府として議会に提出する「大統領予算」を編成する段階（②）である。政府全体の目標を実現する観点から、各省からの予算要求を査定して、予算案を編成する。各省が行った「取捨選択や重視度合の相対的な変更」（trade-off）や類似・重複案件との比較を踏まえた実施判断などを、省を超えて全政府的な観点から実施する⁶。現実的には、想定される予算総額の規模に合うように各省の予算要求額をカットしていくプロセスでもある。

そして次は予算を承認するプロセス（③④）であり、まずは行政府が提出した大統領予算を踏まえて、連邦議会が予算を承認する段階（③）である。連邦議会による予算承認は、ア）予算規模や大まかな方向性を定める「予算決議」（Budget Resolution）の採択、イ）個々の施策・事業の実施根拠であり歳出権限・規模等を定める「権限法」（Authorization Acts）の制定、ウ）各省の具体的な予算額を予算費目ごとに決定する「歳出法」（Appropriation Acts。通常は13本の分野別の法で構成）の制定、にて構成される。Joyceは、現時点でこの段階ではまだPIBが実施されているとは言えず従来からの方法による予算承認がなされているが、例えば、「予算決議」を業績面から方向づける「業績決議」（Performance Resolution）の採択、歳出権限に加えて議会としての期待成果の表明、及びそれを踏まえた「歳出法」の制定などのPIBが実施されうるとしている。またその前提として、「歳出法」の単位を施策単位に揃えるこ

とを提案している。

予算承認プロセスの次の段階は、大統領による承認である（④）。議会が採択した歳出法案に、大統領が署名する（もしくは否認する）。この段階では、政権の成果目標を達成させる観点から議会が承認した予算内容を分析した上で、政府見解を表明したり、歳出法案への署名を判断するなどのPIBが実施されうるとしている。

予算承認の次のプロセスは、各省による予算執行である（⑤）。大統領によって最終的に承認された予算額を踏まえて、省が現実的に目指しうる成果目標を設定し直すと共に、予算を実際に執行して活動・事業を実施する。歳出権限や規模を規定する権限法と年度の個別の歳出額を決める歳出法とは互いに密接には連携しておらず、個別の活動・事業・施策ごとの年度予算額とその費目は、予算承認後も必ずしも明確になっていない。そのため、それらは各省が予算の執行段階にて各省の裁量により決定することとなる。業績情報や他の情報を基に、省内各部門や省外部（外部委託や補助金等）への予算配分がなされる。単に過年度の業績を踏まえた個別案件の増減ではなく、省の戦略目標・業績目標を踏まえて、それを確実にかつより高い水準にて、しかも効率的に実現する観点から、配分額が決定される。

そして最後のプロセスが、省内外の監査人・評価者による監査・評価の段階である（⑥）。従来からの予算執行をめぐる規則への遵守状況の監査・評価に加えて、成果目標の達成状況の視点も踏まえた監査や評価（プログラム評価など）が実施される。

全体として、何れの予算プロセス（段階）でも、個々の事業・施策の予算額の過年度からの単なる増減に着目するのではなく、個々の活動・事業・施策における次年度以降の予算の増減が省もしくは政府全体の目標達成にどのような影響を及ぼすのか、との観点からの分析・検討がなされて判断が下されることがPIBの特徴となる。

（8）PIBの実施に必要な情報

PIBの実施に業績情報が必要なことは言を待たないが、同時にPIBの定義は、それ以外の情報も使用されることを明示している。ここで「それ以

外の情報」には、大別して2種類あると思われる。

①政府全体や各省の方向性や方針に関する情報、及び②個別の施策・事業に関する（業績以外の）情報である。

予算編成の際には、個別の施策・事業を超えて、政府全体や省としての政治的・政策的な優先事項、戦略計画や施策体系、財政政策、予算編成方針などの情報が活用されることとなる。

また個別の施策・事業については、例えば、政府の重点・優先事項への整合性、引続き施策・事業を行う必要性（新規であれば新たに実施する必要性）、効率性、実施プロセスや案件マネジメントの適切性、他の類似施策・事業との関係、など

の情報が必要となる（表5参照）。

業績情報により業績の良し悪しは分かるが、それを今後の方向性の判断に結び付けるためには、他の情報が必要になるとの趣旨のHatry（2006）（表6参照）の指摘、及び、成果に基づいて予算編成をするには、結果・成果情報によりもたらされる以上の施策（及び代替策）への深い理解が求められる、とするSchick（2007）の指摘は、これらの双方の情報の必要性を指しているものと思われる。また、上記（7）にて概観した米国連邦政府におけるPIBも、個別の業績情報に加えて、これら①②の情報が使用される（及び使用されうる）ことを示している。

表5 PIBの実施に際して必要となる情報

①政府全体・各省の情報／外部環境の情報	②個別の予算対象(施策・事業)の情報
<ul style="list-style-type: none"> ■政治・経済・社会状況 ■短期・中長期の財政見通し ■財政政策、予算編成方針 ■重点・優先事項（政治上・政策上） ■政府全体の戦略計画・業績計画・業績報告 ■各省の戦略計画・業績計画・業績報告 ■施策体系（個々の活動・事業と組織・政府全体の目標とをつなぐもの） <p style="text-align: right;">／等</p>	<ul style="list-style-type: none"> ■重点・優先事項との整合性 ■戦略目標・業績目標との整合性 ■必要性・緊急性 ■成果（有効性） ■費用、効率性 ■実施プロセスの適切性 ■マネジメントの適切性 ■他の類似施策・事業との比較、代替策 <p style="text-align: right;">／等</p> <p>※ 全ての施策・事業について、これらと同様の情報が<u>必要になる</u>。</p>

（出所）Robinson（2007）、Joyce（2008）、Robinson and Last（2009）、World Bank（2010）などを参考に筆者作成

表6 業績情報の体系的な活用がもたらすこと

項目	概要
業績の悪い施策の特定	■変化の必要性、予算の増減の必要性を示唆（どのような変化をすべきかの判断には、他の情報も必要）
業績の良い施策の特定	■大きな変化は必要ないと想定（しかし変化の必要性の判断には、他の情報も必要）
新規施策の評価	■単に費用や期待価値についての一般的説明ではなく、新規施策は費用に照らして価値があるか
代替案との比較	■「費用」と「期待する成果」の双方の比較
同様の成果指標を持つ活動の特定	■連携や予算配分の仕方の変更
予算選択の正当化	■省庁、政治家、市民に対する効果的な正当化
より良い説明責任の基盤構築	■適切な業績目標が設定されている際には実績と比較

（出所）Hatry（2006）より作成

PIBの実務では、多岐にわたる施策・事業についての多岐にわたる情報を、短期間にかつ的確に処理する必要性から、個別の施策・事業についての業績情報及び上記②の情報を総合的かつ効果的・効率的に把握する手法として、レーティングが用いられるケースがある。

2. レーティング（格付け）

(1) レーティングの効用と制約

PIBにてレーティングを用いることには、多くの実務的な効用があると同時に、構造的な制約や懸念も多く存在する（表7参照）。

具体的に、個別案件の有効性の情報のみならず、予算額決定の判断に必要な情報を網羅的に、しかも政策分野を超えて比較可能な形で把握できるようになることなどの効用が認められる。

これらの効用によって、PIBをより効果的・効率的に実施しようとの判断から、多くの政府にて採用されているものと考えられる。予算担当者は、必ずしも各施策分野や個別事業の専門家ではなく、その意味で原課担当者との情報の非対称性を有しており、そのような予算担当者にとって、全施策・事業についての情報を横断的にも比較可能な形で容易に入手しうるレーティングは、便利な

ツールとなりうる。

このような効用の一方で、レーティングには、多くの制約や懸念も存在する。例えば、妥当性・有効性・効率性・プロセスなどの全く異なる要素を、一つのレーティングの枠組みで取扱ってしまうことや、その結果として、合計値の多寡や総合判定の高低が「何を示すのか」が必ずしも明らかではないこと、などである。

この点に関連してHatry（2006）は、本稿でも下記で議論する米国ブッシュ政権の「プログラムの評価及びレーティングのツール（Program Assessment and Rating Tool: PART）」について、それを構成する4要素それぞれの分析は有益であるとしながらも、意味の異なるそれら要素を単純合計して得点化して、それを基に5段階にレーティングしている点を批判している。また50%のウェイトが設定されている「成果」もその詳細を見れば、成果そのものについての基準はその極一部に過ぎず、プロセスに関する基準も多く含んでいることにも懸念を表明している（(2)①を参照）。

実務上の有用性から、PIBにレーティングを用いる傾向は今後も続くものと思われる。その際、上記のような制約・懸念への対応としては、Hatry（2006）が指摘するように、合計値や総合判定の結果のみならず、各構成要素のスコアも公表して使用者がアクセスできるようにして、使用

表7 PIBにてレーティングを用いる効用と制約・懸念の例

効用	制約・懸念
<ul style="list-style-type: none"> ■結果、成果（アウトカム）などの有効性の情報のみならず、判断に必要な情報を全て把握できること。 ■これら情報が、施策分野や実施手法、施策をめぐる環境等の相違を超えて、政府全体として比較可能になること。 ■数多くの行政施策の状況（全体傾向・個別状況）が短時間で把握できるようになること。 ■適切なウェイトの設定により、（単なる合計ではなく）自らの価値観を踏まえた結果を導き出せること。 ■これらの結果として、金額の多寡を何も示さない評価結果と予算額との間にブリッジをかけうること。 	<ul style="list-style-type: none"> ■妥当性・有効性・効率性・プロセス等の異なる要素を1つの枠組みにて扱ってしまうこと。 ■そのため、合計値の多寡や最終の総合的な判定結果の高低が「何を示すのか」が明らかではないこと。 ■また構成要素間の比較が可能なのか、明らかでないこと。 ■施策分野や実施手法の相違を超えた比較が、本当に意味ある比較なのか、明らかでないこと。 ■ウェイトの根拠を客観的に説明しにくいこと。（例え各要素を同じ比重にて扱う場合も、同様の困難に直面する。） ■各要素の判定や総合判定が、主観的・判断的な傾向を帯びやすいこと。 ■このような制約をもつ手法が意思決定の重要なツールとして用いられること。

（出所）Hatry（2006）などを参考に作成

者が各要素がどのように判定されているのかを確認できるようにすることや、使用者が（例えば異なるウェイトによる）異なるレーティングを構築することを可能にすること、などが考えられる。併せて、レーティングを使用する際に、その結果を用いて今後の方向性を機械的に判断しないこと、判断に当たっては個別案件の個別状況を勘案すること、なども重要となろう。

(2) レーティングの実施例

PIBの実務にて、どのようなレーティングが導入されているのか、ごく一部ではあるが国内外PIBにおける先進的な取り組み事例を以下に示す。ここでは、業績目標ではなく「実際の業績（成果）」を用いるアプローチの事例を示している。

①アメリカ

ブッシュ政権にて「大統領の経営課題（President's Management Agenda）」として示された5課題の1つが、「予算と業績の統合」（Budget Performance Integration: BPI）イニシアティブによる業績予算の推進であった。BPIイニシアティブは、政府予算が最大の成果をあげよう、意思決定者や市民が、良い施策・欠陥のある施策を特定できるようにすると共に、業績改善の方策を検討できるようにすることをその目的とする（Redburn, Shea, and Buss 2008）。そのツールとして導入されたのが、PARTである。このPARTにより、連邦政府の全プログラム（約1,000）が評価され、レーティングされている。

PARTは、「プログラムの目的とデザイン（20%、5問）」「戦略計画（10%、8問）」「プログラムのマネジメント（20%、7問）」「プログラムの成果と説明責任（50%、5問）」という4視点から構成されている。これらについての25問を標準として、更に7種類の施策タイプ（連邦直轄、規制、補助金、研究開発など）についてはそれらに固有の質問も追加される。そして、各視点の得点の合計により、「有効（Effective）」「ある程度有効（Moderately effective）」「普通（Adequate）」「有効ではない（Ineffective）」「成果は示されず（Results not demonstrated）」の5段階に格付けされる。（「成果は示されず」は、評価手法（業績指標など）が適切

ではないものなどが該当する。）

各省のプログラム担当者がPARTの質問への回答及びその根拠を準備して、OMBの各省担当の予算査定官との協議（対話）を経て、最終的には当該査定官がレーティング結果を確定させる。なお、各省によるBPI及びPARTへの取組み状況（進捗状況）も、「省庁スコアカード」という別のレーティング手法によって評価され管理される。

②韓国

アジア金融危機に端を発する財政状況の悪化及び急速に進展する高齢化などを背景として、公共支出管理（PEM）の改革に迫られていた韓国では、当時の計画予算省（Ministry of Planning and Budget: MPB）が、2003-07年に一連の財政改革を実施した。その一つが、業績予算を支える制度として2003年に導入された「予算施策の業績マネジメント・システム（Performance Management System of Budgetary Programs: PMSBP）」である。PMSBPは、「モニタリング（年次業績計画・報告書）」「レビュー（チェックリスト）」「評価（詳細な評価）」の枠組みにより構成されている。この内、「レビュー」のツールとして2005年に導入されたのが「予算施策の自己評価（Self-Assessment of Budgetary Programs: SABP）」である。SABPは、業績評価と予算配分とのリンクを強化するために導入されたものであり（Park, 2010）、また米国のPARTを参考に開発されたツールであることから、「K-PART」という俗称を持つ（韓国版PARTという意味）。

SABPでは、「施策デザインと業績計画（30%）」「マネジメント（20%）」「成果と説明責任（50%）」という3つの視点にて、標準化されたチェックリストに基づく評価が行われる⁷。各省がチェックリストの質問に回答する形で評価を行って、各省の評価結果をMPBがレビューする。各視点の得点を合計して、その合計得点を基に、「有効（Effective）」「ある程度有効（Moderately Effective）」「普通（Adequate）」「有効ではない（Ineffective）」の4段階にレーティングされる。

毎年1/3のプログラムが評価され、2005-07年の3か年で、約1,700のプログラム／サブ・プログラムが評価された。

③カナダ（プログラム・レビュー・テスト）

カナダのプログラム・レビュー・テスト（Program Review Test）は、カナダ連邦政府の財政難を背景に、財政支出の削減を目的として1994年に実施されたものである。

目標水準を設定した全政府一律の支出削減には限界があったことから、各省の個別の施策を対象とした評価が実施されることになった。考え方として、「何をカットするかよりも、何を残すか」（Bourgon 2009）が重視された。プログラム・レビュー・テストは、全体で6つのテスト（表8参照）から構成されており、各省が自らの歳出プログラムに対してこの6つのテストを上から順番に適用して、6つのテスト全てをクリアしたプログラムに対してのみ、今後の連邦政府による支出を継続する仕組みになっている。クリアできないテストがあった場合には、その段階で、そのプログラムの廃止もしくは州政府・民間への移管が決定する。

6つのテストではまず、政府が自ら役割を果たすべき公共の利益があるのか、その役割は連邦政府ではなく州政府や民間企業・非営利機関などによって果たせないのか、が問われている。そしてその後、つまり連邦政府が自ら予算を投じて関与すべきプログラムであることが確認された後に、その効率性を高める可能性（支出に対してより多くの結果・成果を産み出す可能性、若しくは同水準の結果・成果をより少ない支出にて実現する可能性）が問われている。その上で最後に、現在の厳しい財政制約の下で本当にそのプログラムを実施しうるのか、が問われている。

プログラム・レビュー・テストは、6つのテストを適用した結果として、合計得点が算定されたり、ABCなどの順位が付されたりする訳ではないことから、厳密にはレーティングの事例とは言にくい側面も存在する。しかし、6つのテストによる評価を通じて、政府がこれまで予算を投じてきた全プログラムについて、今後の予算付与についての位置づけ（分類）が6段階にて明確にされていること及び、米国のPARTやカナダの「戦略的見直し」（下記④参照）など他のレーティング事例の設計・運用の参考とされていることから、ここでは敢えてレーティングの一事例として紹介している。

なおプログラム・レビュー・テストは恒久的ではない一時的な取組みであったが、その後もこの考え方はカナダ政府内にて継承されて、予算編成における重要な考え方となっている。

④カナダ（戦略的見直し）

2007年に導入された新たな「支出マネジメント・システム（Expenditure Management System: EMS）」の一環として、カナダ連邦政府は、上記のプログラム・レビュー・テストの発想を継承した新たな取組み「戦略的見直し（Strategic Review）」の実施を2007年に開始した。この取組みは、連邦政府の優先事項に合致し、かつ業績をあげる施策に重点的に資源配分を行うことで、支出の拡大を抑制することを目的としている。

各省は、「政府の優先事項」（優先事項との整合性、カナダ国民のニーズに照らした妥当性、等）と「業績」という2軸の視点からの評価によって、

表8 プログラム・レビュー・テストの6つのテスト

テスト	概要
1) 公共の利益 (Public Interest) テスト	このプログラム・活動には、公共の利益があるか?
2) 政府の役割 (Role of Government) テスト	プログラム分野や活動について、政府の正当かつ必要な役割があるか?
3) 連邦性 (Federalism) テスト	連邦政府が果たす現在の役割は適切か? 州政府へと再編成する候補ではないか?
4) パートナシップ (Partnership) テスト	全体もしくは一部を、民間ボランティア・セクターに移管すべき(しうる)か?
5) 効率性 (Efficiency) テスト	このプログラム・活動を継続する場合、その効率性は改善しうるか?
6) 実施可能性 (Affordability) テスト	残ったプログラム・活動は、財政制約の下でも実施可能か? そうでない場合どのプログラム・活動を廃止すべきか?

(出所) Bourgon (2009)より作成

表9 「戦略的見直し」における判断基準

	低い業績	高い業績
高い優先度	改善／再投資の候補	再投資の候補
低い優先度	再配分の第1候補	再配分の第2候補

(注) 再投資とは、再び予算を投じること。再配分とは、別の施策に予算を付け替えること。
(出所) Wilson (2009)

「優先度が低く業績の悪い施策」を特定することが求められる。その上で、特定された施策の内、予算額の5%に相当する施策への予算配分を取り止めることが求められる。その削減分を、より優先度が高くかつ業績の高い案件に再配分する仕組みである。但しその再配分は、同省の案件になされるとは限らない。廃止対象の5%相当の施策の特定には、表9の表が適用される。

毎年、直接支出の1/4の案件が評価対象となり、4年間ですべての直接支出が評価されることになる。現在は、この拡張版である「戦略的かつ運用上の見直し (Strategic and Operational Review)」が実施されており、間接支出にまで評価対象が拡大されている。

⑤兵庫県川西市

川西市では、総合計画、予算、政策評価の統合を推進しており、政策評価は毎年度の決算プロセスの一部に組み込まれ、Plan-Do-See (PDS) サイクルのSの一環として実施されている。具体的に、毎年8月に作成する「決算成果報告書」にて、行政サービスを、施策と事業の2段階で評価しており、そのうち事業を評価する「事業別行政サービス成果表」にてレーティングが用いられている。なおレーティングは、事業の担当部長による自己評価として位置づけられている。

レーティングでは、各事業が「妥当性」「効率性」「有効性」の視点から、それぞれ「適正である (3点)」「検討余地あり (2点)」「改善すべき (1点)」の3段階の何れかに評価され、その単純合計の得点によって、9-8点：A、7-6点：B、5-3点：C、とレーティングされる。ウェイトは設定されていない。

この評価結果を踏まえて、「今後の方向性、見直し等」の欄が記述される。なおこの欄では、<

参考>として、前年度の同欄の内容も併せて掲載され、過去1年間でどのような状況・環境の変化があるかの比較確認できるようになっている。

このレーティング結果を含む決算情報は、翌々年度の予算編成プロセスである冬のヒアリングにて用いられる。この点について同市の財政政策部財政室長 (出版当時) の松木茂弘は、毎年12月に実施される翌年度当初予算のヒアリングでの前年度決算情報の取扱いについて、「川西市では、8月の業務で説明しました施策別行政サービス成果表と事業別行政サービス成果表をベースに、めざすべき成果に対してどうなのか、また、その達成度合いと課題は何かについて確認するようにしています。… (中略) …また、予算要求書にもその要素を転記することで事業担当者、財政担当者の双方が認識できるようにしています。… (中略) … (事業別予算要求書 (予算計画統括表) の一番下段の「4.平成20年度決算を踏まえた予算計画の方針」) のところで決算成果報告書の一部をコピーして貼り付けるようにしています。ここでSEEからPLANへの連動を図っているところです。」と説明している (松木 2011)。

(3) レーティングにおける「ウェイト」の考え方

ウェイトは、レーティングの結果に影響を及ぼす重要な要素である。

①公共経営上の価値観の反映

ウェイトにて「どの視点を」「どの程度」重視するのにかについては、各行政機関の公共経営上の価値観が反映される。レーティングの結果と資源配分の判断をより直接的に結び付けようとする場合には、ウェイトの有りようが政府としての判断の有りように直結する。従って、レー

ティング手法の設計段階において、当該機関の価値観を的確に反映するようにウェイトが設定されているかを、厳密に確認する必要がある。そうでないと、自らの意図とは異なるレーティング結果を使用して意思決定を下してしまうことになる。

そのような厳密性の一方で、どの程度重視するかは具体的な数値は、客観的・合理的にというよりは、ざっくりと分かりやすいところで決められる、という側面もある。例えば「有効性」の配分が、49%や51%ではなく50%ちょうどであるのは、政府・政権の価値観を踏まえた市民や政府職員へのメッセージとして、区切り良く分かりやすいところにて具体的な比率が決められていることによる。

②重視される視点とその程度

本稿にて紹介する事例では、米国PARTにて、有効性（成果）に最も多い50%の配分が割り振られている。韓国のSABPの場合にも、同様に50%が配分されている。行政機関が成果重視の経営を遂行して、社会における政策課題の解決などの成果を実現させていくために、有効性に重きをおいたレーティングを採用することには意義がある。

次にカナダの「戦略的見直し」の場合には、どの程度かは議論しにくいだが、常に優先性が有効性よりも重視される仕組みとなっている。限られた資源を効果的・効率的に活用するために、優先性（優先事項との整合性、カナダ国民のニーズに照らした妥当性、等）により重きが置かれることにも意義がある。

そして日本の川西市の場合には、評価の視点が均等に重視される構造となっている。このように特定の要素にウェイト付けをしない場合にも、つまり全ての要素を同じ比重で勘案する場合でも、実は「全ての要素が同程度に重要だと判断している」ことになる（Hatry 2006）との指摘は重要である。統計的な分析ではなく印象論に過ぎないが、わが国の政策評価の実務におけるレーティング事例には、同市のように特定の要素にウェイト付けをしていないものが多いように思われる。特定の要素を重視することを表明しにくい背景が有るのかも知れないが、Hatryの指摘のように、それでもウェイト付けを避けたことにはならないことに

は留意が必要であろう。

(4) レーティング結果の予算編成への反映

このようなレーティング結果が、予算編成に如何に反映されているのかを次に確認する。

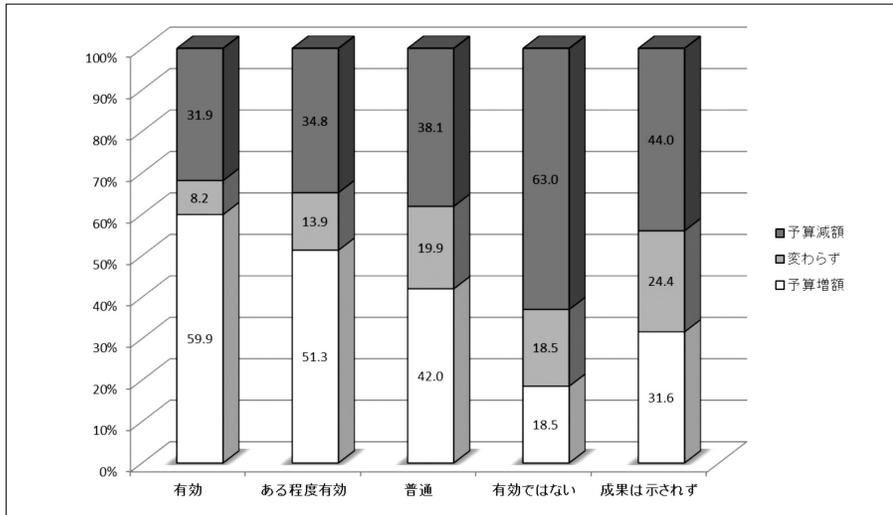
米国PARTにおけるレーティング結果の大統領予算への反映状況を詳細に分析している左近（2008）によれば、レーティング結果が良かったプログラムは相対的に予算額が増額される傾向にあり、逆に悪かったプログラムは減額される傾向にある（図2参照）。またBlanchard（2008）も、別の分析手法により、大統領予算及び議会承認された予算の双方において、PARTのレーティング結果が高いプログラムは予算が増加され、逆に、低いプログラムは減額される傾向にあることを実証的に分析している。

またPark（2010）によると、韓国のSABPでも同様の傾向がみられる。「有効」「ある程度有効」と判断されたプログラムの予算配分は増額され、「普通」「有効ではない」と判断されたプログラムの予算配分は減額されている（表10参照）。

定義が示すように、業績と予算額との直接の関連性がないのがPIBであるが、これら事例においては、有効性（成果）及びそれ以外の情報により総合的に判断されたレーティング結果が、予算編成のプロセスにて、より直接的に反映される傾向があることが分かる。レーティング結果の予算増減へのストレートな反映は、予算の要求側・査定側の双方にとって正当化しやすく、その結果として説明責任も果たしやすい側面があるのも確かである。

なおカナダ連邦政府の2つの取組み（プログラム・レビュー・テスト、戦略的見直し）は、前者は支出削減を目的としたものであり、また後者は再配分のための財務余力をつくることを目指したものであるから、予算編成への反映については、主にマクロ面での反映となる。プログラム・レビュー・テストの結果として、政府全体で支出が10%以上削減され、また省庁別では50%以上の支出削減がなされた省も複数見られた（Bourgon 2009）。また、戦略的見直しについては、2007～2010年の4年間の見直し結果として、計28億カナダドルの歳出削減がなされ、これらが再配分及び

図2 米国PARTの結果の予算(大統領予算)への反映状況(2009年度大統領予算)



(単位) % (件数の%)

(注) 左近 (2008) によると、同様の傾向が2004年度大統領予算から継続してみられる。

(出所) 左近 (2008) より作成

表10 韓国SABPの結果の予算への反映状況

<2005年SABP>

	予算要求時	計画予算省査定時	最終
有効	9,348	3,355	6,889
ある程度有効	11,342	14,062	12,768
普通	-3,593	-12,441	-6,852
有効ではない	-4,437	-4,074	-4,576

<2006年SABP>

	予算要求時	計画予算省査定時	最終
有効	575	446	-19
ある程度有効	2,545	2,208	2,498
普通	-411	-6,699	-7,211
有効ではない	-5,392	-6,031	-6,051

<2007年SABP>

	予算要求時	計画予算省査定時	最終
有効	1,099	391	—
ある程度有効	29,070	25,268	—
普通	6,992	-3,990	—
有効ではない	-414	-805	—

(注) 前年度予算額からの増減を示す (単位: 100,000ウォン)。

(出所) Park (2010) より作成

財政赤字の縮減に用いられた（カナダ政府の公表数字）。

おわりに（むすびにかえて）

以上、本稿ではPIBの実践概要とそこで使用される情報及び手法について、先行研究や実践事例を基に整理してきた。本稿での議論を総括すると、以下ようになる。

①業績予算の一類型であるPIBは、その名前を使用しているか否かに関わらず、日本を含めてすでに多くの国にて、また政府のレベル（国・地方）を問わずに実施されている。PIBは、成果重視の経営の下での、業績マネジメント及びPEMに貢献しうるものであり、予算編成時の予算査定のみならず、予算プロセスの各段階で実施されうるものである。

②このようなPIBの実務においては、その定義も示すように業績情報は重要な情報として位置づけられるが、またその活用により個別の施策・事業の業績の良し悪しは把握できるが、それだけでは次期以降の予算増減の判断ができず、他の情報も合わせて初めて判断ができるようになる。具体的には政府・政権の優先順位についての情報や、各省庁としての戦略計画・業績計画、個別案件についての業績以外の情報も必要となる。その意味では、業績情報は重要ではあるが、必要となる情報の一部である。

言い換えれば、PIBでは省や政府全体の成果がまず重視され、その予算配分の効率性を高めるための個別の施策・事業の在り方を、業績情報や他の多様な情報を活用しながら判断している。

③その必要となる多様な情報の一端を把握する手段として、レーティングが活用されている。レーティングには、合計点の算定や総合判定など本質的な面での大きな制約や懸念があるものの、実務的に多くの効用があり政府によって使いやすいツールであるため、多くの政府にて導入されている。

④いくつかの主要な事例をみると、レーティングの結果は、やや直接的に予算の増減に反映される傾向がある。PIBにおいては、定義と実務の双

方にて、業績目標を達成しても予算が減額される（もしくは案件が廃止される）場合もあれば、業績目標が未達成でも予算が維持されたり増額されたりする場合もある。しかしその一方で、レーティング結果の予算増減へのストレートな反映は、予算の要求側・査定側の双方にとって正当化しやすく、その結果として説明責任も果たしやすい側面があるのも確かである。

レーティング結果の良い案件への継続的な予算配分によって今後の更なる成果発現が期待できるのであれば、またその点にて他案件よりも相対的に優れているのであれば、そのような案件が、次期以降の予算配分にて肯定的な判断を得ることは、成果重視の経営の下で、省や政策分野・政府全体の成果目標を確実にかつより高い水準で実現させるとの観点からは、むしろ当然ともいえるかもしれない。

今後の残された課題として、PIBの理論面での整理を更に進めることに加えて、PIBの実務における予算編成の詳細を解明することがあげられる。特に個別施策・事業の予算額の決定方法（増減幅の決定方法を含む）については、更なる研究が必要である。つまり、政策評価の結果を用いて今後の方向性について判断を下す際に、レーティングに組込まれる情報（業績及びその他の情報）などがどのように用いられ、どのようなロジックや理由づけによって判断を下しているのか、またどのように予算額を増減しているのかを実証的に確認することである。換言すれば、PIBが、個々の施策・事業の予算増減ではなくPEMの目的である省や政府全体の予算配分の効率性をまず重視する中で、「では個々の事業・施策の予算額は政策評価の結果との関連でどのように決定されるのか」を明らかにすることである。

その際には、レーティング手法が持つ根本的な制約を実務上で如何に克服しうるのかも、十分に検証する必要があるだろう。

またPIBの実務にて米国PARTなどにみられる、政策評価結果を用いた利害関係者間（例えば省担当官と予算査定官など）の「対話」について、その重要性を論じる先行研究（Moynihan 2008など）がわが国の実務にどのような意味を持つのかも、

残された検証課題である。

謝辞

本稿の基となった日本評価学会第12回全国大会（2011年、於：岩手県立大学）の研究発表の機会に大変貴重なご指摘とご示唆を頂いた田中啓先生、上野宏先生、佐々木亮先生、三好皓一先生（ご発言順）に、まずお礼を申し上げたい。そして本稿の査読に際しても、査読者の先生方から複数回にわたり、未熟な拙稿に対する大変的確なご指摘と懇切丁寧なご助言を頂戴した。この場を借りて、深くお礼を申し上げたい。

注記

- 1 本稿にて述べられている見解は、筆者個人としてのものであり、筆者が所属している組織の見解とは全く関係ない。
- 2 ここでは、国による政策評価、自治体における行政評価の双方を包括的に示している。また、政策評価の3方式の内、「実績評価」を想定して用いている。以下、同様。
- 3 OECD加盟諸国及びその他諸国の業績予算への取組み動向は、OECDの「業績と成果に関する上級予算担当者ネットワーク」の年次会合で紹介され議論されている。
<http://www.oecd.org/gov/budgeting/seniorbudgetofficialsnetworkonperformanceandresults.htm>
- 4 例えば、インドネシア政府がPIBの導入を進めており、World Bank、日本（JICA）、オーストラリアがその取組みを支援している。JICAによる支援（技術協力）である「業績評価に基づく予算編成能力向上支援プロジェクト」については、以下などを参照のこと。
<http://gwwweb.jica.go.jp/km/ProjectView.nsf/0/9cb334a17c2e80f6492576f6003ad795?OpenDocument>
- 5 田中（2011）は、「財政当局は、政策評価や業績評価を省庁に要求し、予算の効率化を図っている。この政策評価や業績評価は、必ずしも、財政当局の伝統的な仕事ではない。」「財政当局の評価機能は、…アメリカ、オーストラリア、カナダ、フランス、スウェーデン、オランダなどでも強化されている。」

と述べて、近年の傾向として主要国の財政当局が政策評価を重視して予算の効率化を図ろうとしていることを指摘している。

- 6 政府全体に係る省庁横断的な戦略計画や戦略目標は、ブッシュ政権時には存在していなかった。オバマ政権にて、そのような取組みの萌芽がみられる。
- 7 チェックリストの項目は、①デザインと計画：施策の目的、政府支出に対する根拠（正当性）、他の施策との重複、施策デザインの効率性、業績目的（objectives）及び指標の妥当性、業績目標（targets）の妥当性、②マネジメント：モニタリングの努力、施策実施の障害、計画に基づく実施、効率性の改善もしくは予算額の節約、③成果と説明責任：独立したプログラム評価、成果、市民満足度、評価結果の活用。

参考文献

- 左近靖博（2008）「米国連邦政府ブッシュ政権における「予算と業績の統合（BPI）」の結果考察 ～予算編成プロセスにおける業績情報活用の成果総括～」、『季刊 政策・経営研究』（三菱UFJリサーチ&コンサルティング）、4：91-111
- 田中秀明（2011）『財政規律と予算制度改革－なぜ日本は財政再建に失敗しているか』、日本評論社
- 松木茂弘（2011）『自治体財政の12か月』、学陽書房
- Blanchard, Lloyd A. (2008). PART and Performance Budgeting Effectiveness. In F. Stevens Redburn, Robert J. Shea, and Terry F. Buss (Eds.), *Performance Management and Budgeting: How Governments Can Learn from Experience*. Armonk, New York: M.E. Sharpe, Inc., 67-91.
- Bourgon, Jocelyne (2009). *Program Review: The Government of Canada's experience eliminating the deficit, 1994-99: a Canadian case study*. London: Institute of Government.
- Breul, Jonathan D., and Moravitz, Carl (2007). The Budget Office Tomorrow. In Jonathan D. Breul and Carl Moravitz (Eds.), *Integrating Performance and Budgets: The Budget Office of Tomorrow*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc., 1-17.
- Hatry, Harry P. (2006). *Performance Measurement: Getting Results 2nd edition*. Washington, D.C.: The Urban Institute.

- Joyce, Philip G. (2007). Linking Performance and Budgeting: Opportunities in the Federal Budget Process. In Jonathan D. Breul and Carl Moravitz (Eds.), *Integrating Performance and Budgets: The Budget Office of Tomorrow*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc., 19-69.
- Joyce, Philip G. (2008). Linking Performance and Budgeting: Opportunities for Federal Executives. In F. Stevens Redburn, Robert J. Shea, and Terry F. Buss (Eds.), *Performance Management and Budgeting: How Governments Can Learn from Experience*. Armonk, New York: M.E. Sharpe, Inc., 49-66.
- Melkers, Julia, and Wiloughby, Katherine (2007). Staying the Course: The Use of Performance Measurement in State Governments. In Jonathan D. Breul and Carl Moravitz (Eds.), *Integrating Performance and Budgets: The Budget Office of Tomorrow*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc., 71-105.
- Moynihan, Donald P. (2008). *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, D.C.: Georgetown University Press.
- OECD (2007). *Performance Budgeting in OECD Countries*. Paris: OECD. (OECD編著・茂木康俊・平井文三訳)
- (2010) 『世界の業績予算：政策評価・行政評価に基づく新たな予算編成システム』、OECD・財団法人行政管理研究センター)
- Osborne, David & Hutchinson, Peter (2004). *The Price of Government: Getting the Results We Need in an Age of Permanent Fiscal Crisis*. New York, New York: Basic Books.
- Park, Nowook (2007). Developing and Using Evaluation: Role of Spending Ministries and Agencies in Korea. OECD Senior Budget Officials (SBO) Network on Performance and Results.
- Park, Nowook (2008). Performance Budgeting in Korea: Observations and Lessons. (Presentation Slide).
- Park, Nowook (2010). Case Study 5.2 - Using Performance Information in Korea. In The World Bank, *Results, Performance Budgeting and Trust in Government*. Washington, D.C.: The World Bank, 181-189.
- Redburn, F. Stevens, Shea, Robert J., and Buss, Terry F. (2008). Performance-Based Management: How Government Can Learn from Experience. In F. Stevens Redburn, Robert J. Shea, and Terry F. Buss (Eds.), *Performance Management and Budgeting: How Governments Can Learn from Experience*. Armonk, New York: M.E. Sharpe, Inc., 3-18.
- Robinson, Marc (2007). Performance Budgeting Models and Mechanisms. In Marc Robinson (Ed.), *Performance Budgeting: Linking Funding and Results*. Washington D.C.: International Monetary Fund, 1-18.
- Robinson, Marc, and Last, Duncan (2009). *A Basic Model of Performance-Based Budgeting*. Washington, D.C.: International Monetary Fund.
- Schick, Allen (2001). The Changing Role of the Central Budget Office. *OECD Journal on Budgeting*, 1 (1), 9-26.
- Schick, Allen (2007). Performance Budgeting and Accrual Budgeting: Decision Rules or Analytic Tools?. *OECD Journal on Budgeting*, 7 (2), 109-138.
- The World Bank (2010). *Results, Performance Budgeting and Trust in Government*. Washington, D.C.: The World Bank.
- Wilson, Tim (2009). *Expenditure Management: The Canadian Perspective*. OECD Senior Budget Officials (SBO) Network on Performance and Results 6th Meeting. (2014.11.10受理)

Performance-informed Budgeting: Practices, and Information and Methodologies Utilized

Jun Arakawa

Aichi Prefectural Government
arakawa525jun@gmail.com

Abstract

This paper discusses how results of policy evaluation are utilized in resource allocation process of government, taking examples of Performance Budgeting practices. More specifically, the paper analyzes practices of “Performance-informed Budgeting (PIB),” and information (performance information and other ones) and methodologies utilized there, by discussing domestic and foreign countries’ examples.

Practices of PIB well show that, although performance information of specific program/project on effectiveness (results) is fundamental under “Results-based Management,” and therefore plays important role in making decisions on financial resource allocation, the information itself tells nothing about neither increase nor decrease of budget amount. Much other information such as government and/or department priorities, strategies, and more detailed situation of programs/projects is needed to make decision on resource allocation. In other words, under PIB, budget allocation for specific programs/projects is discussed in the context of “allocative efficiency” for realizing government and/or department-level outcomes.

In implementing PIB, many governments have been introducing “Rating” system that could effectively and efficiently analyze multiple evaluation criteria as a tool for utilizing performance and other information simultaneously in decision-making process. Rating system structurally contains various constraints, however its practical usefulness and effectiveness make it popular among many governments. Practical examples imply rather direct reflection of the rating results into budget decision.

Keywords

Performance Budgeting, Performance-informed Budgeting, Performance Information,
Rating System, Results-based Management

【実践・調査報告】

参加型モニタリング・評価手法MSC（Most Significant Change） ーバングラデシュNGOでの実践から4つの特色を考察するー

田中 博

参加型評価ファシリテーター

nepalippine@gmail.com

要 約

MSC（Most Significant Change）は、欧米の国際NGOが活用する参加型モニタリング・評価手法である。筆者は、MSCが日本のNGOにおいて学習目的のプロジェクト評価に効果的と考え、その特色を文献及びバングラデシュNGOでの実践において4つの視点から考察した。4視点とは、①モニタリング及び評価手法である、②参加型評価手法である、③質的分析手法である、④組織学習の手段である、の4つである。実践の結果、参加型評価の長所である利害関係者の意識啓発、能力開発、相互理解につながる学習効果が確認された。また参加型評価の課題である「妥当性」や「代表性」の問題を緩和する機能が一定程度観察された。効果的なMSC実施にあたっては、プロジェクト開始時からの継続の実施や、スタッフのインタビュー技術等の能力開発が求められる。日本のNGOが今後MSCを活用していくためには、MSCによる評価の意義のさらなる考察や、日本のNGOの評価の現状を把握し、導入の意義と可能性を検討していく必要がある。

キーワード

参加型評価、質的分析、学習、下向き説明責任、NGO

1. はじめに

筆者は長くNGO活動に携わり、各国で様々な団体の事業評価に関わってきた。その体験を通じて、NGO事業に対する「学習」目的の評価の重要性を痛感している。学習とは、評価を通じて得られた教訓を、将来の活動に反映させることである（三好 2008）。NGOが利害関係者の学習をねらいとした評価を行うことで、プロジェクトを改善、目標を達成し、NGOの存在意義である組織のミッションの実現に貢献することができる。

また学習目的の評価には、参加型評価の実施が

適していると考えている。NGOのプロジェクトは、多くはブループリント型ではなく、プロセス型だといわれている。ブループリント（青写真）型プロジェクトとは、計画（目標設定）と実施が直線的な因果関係で繋がっている事業である。手段と目的の関係が明確で成果を数量化しやすい特徴があり、インフラ・プロジェクト等、経済開発事業に多い。これに対してプロセス型（プロセス・ラーニング型）プロジェクトは、目標設定や計画があいまいだが、試行錯誤の中で効果的なアプローチを採用する。住民に働きかけて意識や行動を変化させる等、社会開発事業に多い。また成

果を数値で測ることが難しい場合がある。このようなプロセス型プロジェクトでは、利害関係者が行動しながら学習する態度が必要になる（アユス 2003）。

プロセス型社会開発プロジェクトにおける社会や人間等の変化は、複数の要因が複雑に関連しながら時間をかけて発生し、直線的な因果関係では説明できない想定外の変化も起こりうる¹。まさに「主体的参加型の開発は、進路が予測できない航海のようなもの（チェンバース2000）」なのである。従って、NGOが行うプロセス型社会開発事業の評価においては、利害関係者が事業を振り返って教訓を学び、それを活かして内容を柔軟に修正・改善することが求められる。そのためには、学習を通じて改善を促す、参加型評価の実施に利点があると考えられる。筆者がフィリピンやカンボジアでNGOプロジェクトの参加型評価をファシリテートした経験でも、利害関係者が評価に参加することで「学習」効果が発生し、事業の改善に貢献することを観察した（田中 2010, 2011）。

さて日本のNGOでは、プロジェクト評価が以前に比べ活発に行われるようになってきた。しかしながら、行政や財団等の資金提供者への「説明責任」の達成を重視した評価が中心なようだ。これに対して、受益者等、利害関係者への「学習」目的の評価の実施は不十分な現状がある。

筆者はNGOのプロジェクト評価において、資金提供者へ説明責任は重要と考える。しかし、それだけではプロジェクト目標の達成やミッションの実現に不十分である。利害関係者への学習を重視した評価が同時に求められるのではないか。そして、学習効果の高い参加型評価手法を、評価体制に導入・活用することが必要と考える。

MSC（Most Significant Change）は1994年にリック・デイビス（Rick Davies）によって考案された参加型モニタリング・評価（M/E）手法である。現場から「重大な変化の物語」を集め、組織的に「最も重大な変化」の選択を定期的に行うことが基本手順である。評価に参加する利害関係者への学習効果が高いといわれ、欧米の国際NGOによる開発プログラムを中心に、世界中で活用されている²。

MSCは日本ではまだ実質的に使われていない

が、筆者は今後日本のNGOが、これを導入・活用することで、学習目的の評価を推進し、ミッションの達成に貢献できるのではないかと考えている。その可能性を探る第一歩として、本稿では、MSCはどのような手法なのか、その特色を文献及び実践を通じて考察する。

本稿の構成を述べる。第2章では、参加型評価について紹介し、定義や特徴、長所と課題について解説する。第3章では、MSCがどのような手法なのか、考案された背景、実施方法と、4つの特色について述べる。MSCは参加型評価手法の一つといわれ、その性質を備えているが、同時に質的分析等、他の特色も合わせ持っていることを解説する。続く第4章では、筆者がバングラデシュNGOの事業評価にMSCを適用した報告を述べ、その結果から手法の特色を改めて論考する。最後の第5章では、日本のNGOが、MSCを活用していくための、次なる研究課題について述べる。

本稿で用いる用語を説明する。NGO（Non-Government Organization）とは、開発途上国で国際協力を行う民間非営利組織の総称である。「国際NGO」と限定する場合は、オックスファム、ケア・インターナショナル等、欧米発祥で国際協力を行っている民間非営利組織を指す。これに対して、日本の国際協力を行う民間非営利組織を「日本のNGO」と称する。また社会課題を解決するための、開始時点と終結時点のある社会的介入について、プロジェクトとプログラム、事業を同義語で用いる。デイビスの著作では、これはプログラムとされているのでこれに従う。

2. 参加型評価と利害関係者の学習

(1) 参加型評価とは何か

参加型評価は、評価活動に評価専門家以外の利害関係者が「参加」し、評価プロセスを共有することで、付加価値を高める評価である。評価専門家の出した評価結果による影響を重視する従来型評価と比較して、参加型評価は評価過程自体が評価に参加した利害関係者へ与える影響を重視した評価アプローチである（源 2007, 2008）。

参加型評価が登場した背景には、2つの流れが

ある。一つは北米において、専門家による事業評価の結果が十分に活用されておらず、評価基準が評価対象の固有の状況や文脈の多様性を反映していないという疑念から、評価の実用性を重視して行われるようになった流れである。もう一つは開発途上国等において、社会開発の当事者の自立やエンパワーメントを目指した参加型調査や、アクションリサーチ等の流れがある。また参加型評価には、利害関係者の参加の範囲や程度によって利害関係者評価、協働型評価、実用重視型評価、エンパワーメント評価等、様々な理論や方法論があるが、共通する定義は「利害関係者が評価活動に関わる評価」であり、「評価プロセスを活用して改善・変化を促す評価（源2008：99）」であるという。

また、このように利害関係者間の「対話」を重視した参加型評価は、「事実 (truth)」は社会の文脈の中で構成されるという社会構成主義の立場を取る（源 2008：98）。

(2) 参加型評価の長所

参加型評価の長所を述べる。まず、スタッフや受益者等、評価対象の利害関係者の評価への参加が、「学習」過程として作用し、彼等のプロジェクトに対する主体性や当事者意識の促進、態度や行動の変容、評価能力の向上等、能力開発に貢献することである。これを通じて、利害関係者のエンパワーメントや相互理解が発現し、評価結果の活用度合いが高まり、プロジェクトの改善に貢献する（三好・田中2001 アーユス2003 源 2008）。

また従来型評価が資金提供者への説明責任の確保やプログラムの継続可否の判断等に対して効果が期待されることに対して、参加型評価では、評価結果は評価に参加した関係者と共有し、次の行動につなげていくことが目指され、受益者を含む被援助国へのアカウントビリティ（説明責任）が重視される（2001 JICA:4）。

説明責任は、重要な評価目的の1つであるが、NGOには二種類の説明責任が存在する。資金提供者や組織上層部への報告である「上向き説明責任 (Upward Accountability)」と、プロジェクトの受益者への「下向き説明責任 (Downward Accountability)」である（Mango 2014）。参加型評価を通じて、評

表1 参加型評価の主な長所と課題

長所	課題
利害関係者の意識啓発・能力開発・相互理解	利害関係者の評価知識・技術が不十分である恐れ
評価結果が活用されやすい（事業の改善）	自己評価による主観やバイアス（妥当性や中立性）
受益者等への下向き説明責任に貢献	誰が参加するか、という代表制の問題

（出所）源（2008）JICA（2001）三好・田中（2001）を参考に筆者作成

価結果を利害関係者と共有して活用する「下向き説明責任」を推進することができる。この意味で下向き説明責任は学習効果があるといえる。

(3) 参加型評価の課題

一方、参加型評価には短所も指摘されている。まず、専門家でない利害関係者による評価であり、評価の知識や能力が必ずしも充分でないことが、あげられる。また、自己評価中心のため、当事者の主観やバイアスによる弊害の恐れである（源 2008, 三好・田中 2001, JICA 2001）。筆者の参加型評価のファシリテート経験でも、調査経験の乏しい現場スタッフが収集したデータの「妥当性や信頼性」、また当事者による分析・判断の「客観性や中立性」が疑わしい場面があった（田中 2010）。さらに、多様に存在する利害関係者の中から、参加すべき人々がきちんと参加しているか、排除されていないか、という「参加者の代表性」の問題がある（源 2008）。参加型評価では、この危険性を回避するため、「多様なデータ・情報間の整合性に注意することによって評価の偏りを回避する」必要があるという（三好・田中 2001）。参加型評価の主な長所と課題は、表1にまとめた。

3. MSCとは何か

(1) MSCが考案された背景

①NGOの評価に関する議論

MSC考案の経緯と背景を述べる。バングラデシュのNGOであるCCDB(Christian Commission for Development in Bangladesh)が実施する参加型地域

開発プログラムが多岐にわたり、成果も多様であった。モニタリング・評価体制づくりが困難を極め、その解決を探る中で、1994年にデイブースが考案した。(Davies & Dart 2005 : 9)。

このように、MSC考案された背景には、欧米の国際NGOとそのパートナーである途上国のNGOによる、より良い事業評価の探求から生まれた側面がある。2章において参加型評価が登場した背景に、開発途上国での社会開発における参加型調査・アクションリサーチの流れがあると述べた。国際NGOはプロジェクト評価に長い経験を持ち、理論や手法の研究も進んでいる。例えば、「NGO評価に関する国際会議」³が世界中のNGOの参加で何度も開催され、NGOが行う社会開発事業と評価について、1980年代後半から議論が続いてきた。その結果、NGOの事業評価においては、資金提供者への「上向き説明責任」は必須ではあるが、最優先の目的は事業を改善する教訓を「学習」することであり、また現場の当事者が事業を改善していくために情報を還元する「下向き説明責任」の確保が重要である、という共通認識が形成されてきた (Garbutt 2011, Mebrahtu, Pratt & Lonnqvist 2007)。

②LFAを補完する手法の模索

また80年代は、イギリス国際開発局 (DFID) 等、国家や国際機関が国際NGOに多額の公的資金投入を開始した時期でもある。これら資金提供者は国際NGOに資金の提供とともに、先進国の納税者への「上向き説明責任」の確保を求めた。そして、ロジカル・フレームワーク・アプローチ (Logical Framework Approach: LFA) に基づくプロジェクト管理を国際NGOに義務付けた (Mebrahtu, Pratt & Lonnqvist 2007, Taylor & Soal 2004)。LFAは米国で業績測定のお考え方に基づき発達した、プロジェクト管理方法である。世界の援助機関で、計画立案や事前審査、モニタリング・評価のために使用されてきた (河村2003)。LFAでは事業の目標と達成に至る方法を、手段と目的という直線的な因果関係において演繹的に考察する。また出来る限り客観的な指標を事前に設定し、定期的に測定することによって進捗や達成度を確認することで、資金提供者や意志決定者への報告 (上向き説明責任) が可能となる長所がある。

そうすると、LFAを採用した国際NGOから、「資金提供者から要求されるLFA中心の評価体制は、上向き説明責任に有用だが、現場での学習や下向き説明責任の確保は難しい」という意見が多く寄せられた。国際NGOが行う社会開発事業のモニタリング・評価においては、直線的因果関係で把握できない波及的な変化や質的な変化を把握・分析し、教訓を学び、それを現場に還元すること大切だが、LFA中心の体制ではその実現が困難だということ (Mebrahtu, Pratt & Lonnqvist 2007 Wallace & Chapman 2004)⁴。

このような議論を経て、LFAだけでなく複数の手法を併用することで、「上向き説明責任」だけでなく、「学習」や「下向き説明責任」にも貢献できるモニタリング・評価体制を整えていくべき、という共通認識が生まれた。そして、学習や下向き説明責任に重点をおいた参加型評価アプローチや手法が研究され、その活用が奨励されるようになった。現在、MSC やOM (Outcome Mapping)⁵ 等が活発に利用されている。

(2) MSCの実施方法

①MSC実施の10段階：

MSCの基本的な手順は、MSCのガイドブックである「Most Significant Change (MSC) Technique: A Guide to Its Use」⁶ に示された10のステップを繰り返して実施することである (Davies & Dart 2005)。表2を参照頂きたい。

「ステップ1 導入と注意喚起」では、組織やプ

表2 MSC手法実施の10段階 (ステップ)

1. 導入と注意喚起
2. 変化の領域を決める
3. 時間の範囲を決める
4. 重大な変化の物語を集める
5. 最も重大な変化の物語を選ぶ
6. 選択過程をフィードバックする
7. 物語が事実であるか検証 (事実確認)
8. 定量化
9. 二次分析及びメタ・モニタリング
10. システムの改定

(出所) Davies and Dart 2005

プログラムの利害関係者に手法を紹介し、関心を高め参加を促すところから始まる。「ステップ2 変化の領域 (domain) を決める」では、変化が起こる一定の分野を決定する。例えば「人々の生活における (変化)」という要領である。業績指標のように、厳密に設定する必要はなく、柔軟に決めて良い。「ステップ3 時間の範囲 (reporting period) を決める」は、「過去1年間で」とか「この3ヶ月の中で」といった様に、領域で起こる変化を調べる時間の頻度を定める。「ステップ4 重大な変化の物語を集める」では、主に受益者の住民や現場スタッフ等、プログラムに直接関わっている人々から集める。

続く「ステップ5 最も重大な変化の物語を選ぶ」では、集められた様々な物語の中から、プログラムや組織の階層を利用して選択、領域ごとに「最も重大な変化」を1つ選ぶ。選ばれた物語は、選択理由とともに、組織の上位階層に送られる。順番に各階層での選択を終え、最後に最上レベルで選ばれた1つの物語が全体の「最も重大な変化」となる (図1参照)。

「最も重大な変化」が選ばれたら、選択した物語と選んだ理由を、他の利害関係者に説明する「ステップ6 選択過程をフィードバックする」を行う。このプロセスを3ヶ月~6ヶ月の期間で繰り返し、一定の期間 (1年等) 経た後、選ばれた全

ての物語とその理由をまとめて文書化する。この評価レポートが組織の意志決定等に利用されるという流れである。

その後必要に応じて、現場訪問等を行い「ステップ7 事実確認」が行われる。物語の内容は通常質的データが中心であるが、変化に関わった人数等の量的データを分析する「ステップ8 定量化」や、収集したデータ全体を俯瞰し、変化の過程や背景等を考察する「ステップ9 二次分析及びメタ・モニタリング」の実施も可能である。最後の「ステップ10 システムの改定」は、MSCのプロセスがここで終わるので、やり方や体制を振り返って、改良していくことである。

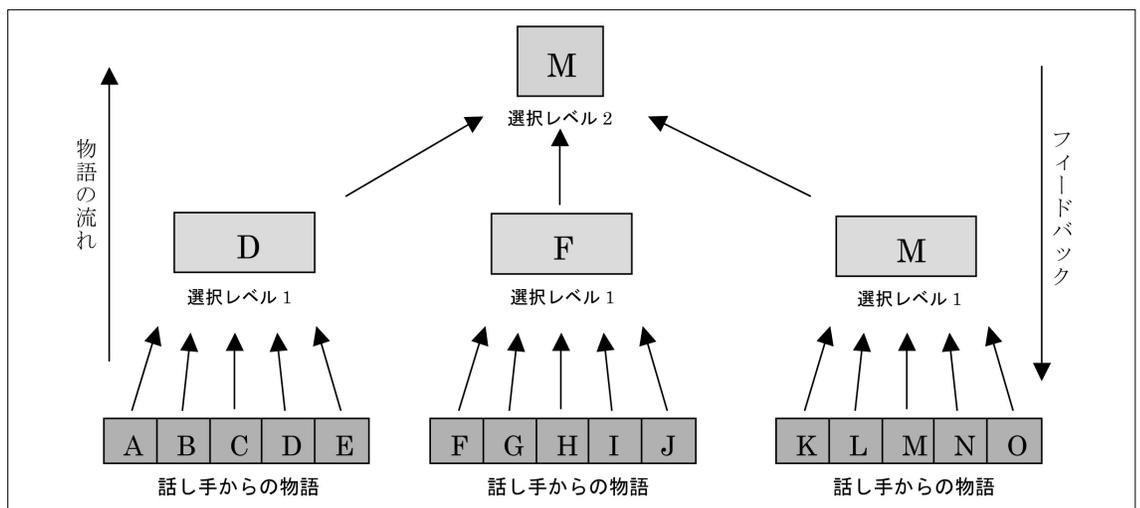
以上が基本の10ステップであるが、デイビースら (2005 : 15) は、ステップ4・5・6がMSCの本質的な特色を表すという。そこで、この3段階を詳しく説明したい。

②ステップ4：重大な変化の物語を集める

ステップ4は、プログラムに直接関わった現場の受益者等から物語7を集めるデータ収集の段階である。訊き方の基本は以下のオープン・クエスチョンを行うことである (Davies & Dart 2005 23-28)。

1. 特定の期間を限定する (例：先月を思い出して)
2. 相手自身の意見を聞く (例：あなたの考えでは)
3. 変化の起こる範囲である「領域」を絞る (例：

図1 MSCにおける物語とフィードバックの流れ



(出所) Davies & Dart 2005

人々の生活に質に関して)

4. 必要に応じてさらに領域を絞る (例: この村においての)
5. 物語を1つに絞る (例: 最も重大な)
6. 漠然とした状況ではなく、変化を聞く (例: 変化は何ですか)

物語の集め方は、現場のフィールドワーカーが日常活動で見聞きした経験をまとめる、人々にインタビューする、グループ討論等の選択肢から、事業や組織の実情に合わせて決める。収集すべき情報は「誰が物語を集めて、いつ出来事が起こったか」に関する情報、「物語自体の説明 (何が起こったか)」、そして、なぜそれを重要と感じたか、話し手にとっての「物語の重要性」が説明されなくてはならない。

③ステップ5: 最も重大な変化の物語を選ぶ

複数の変化の物語とその理由が揃ったら、利害関係者のグループ等で、領域毎に物語の選択作業 (データ分析) を行う。基本は以下の討論であるが、投票等を併用して決めることもできる。

1. グループ皆で物語を読み込む。
2. グループで、どの物語を選ぶのか、深く議論する。
3. どの物語が最も重大に感じたかを決める。
4. それを選んだ理由を書き留める。

通常、組織やプログラムには、活動現場から上位の意志決定者まで、いくつかの階層がある。下のレベルで選択された物語は、より上位の階層に提出、同様に精査される (図1)。従って、選択過程を通じて物語の数は減っていく。また各階層の選択を終える毎に、選んだ理由が記録され、他の利害関係者に共有される。即ち、上位の選択過程は下位の選択過程における判断理由を踏まえて実施されることになり、これにより、上に行くにつれて、正しい判断へと調整されていく仕組みとなっている。このプロセスを「選択による要約」と呼ぶ。選択を繰り返した後、最上位の階層の「最も重大な変化」と選択理由を、領域ごとの文書にまとめる (Davies & Dart 2005 28-34)。

④ステップ6: 選択過程をフィードバックする

全体の「最も重大な変化」が1つ決まったら、フィードバックに移る。選択の結果を、物語を集めたプロジェクトの現場や選択を行った組織内各

階層の関係者に、文書や口頭で伝える。このフィードバックは「下向き説明責任」の効果があるという。フィードバックで最低限に報告・共有すべき事項は、選ばれた「最も重大な変化」の物語と、それを選んだ理由である (Davies & Dart 2005 34-35)。物語の選択と情報還元が行われるイメージは、図1を参照していただきたい。

(3) MSCの特色

MSCの特色に関しては、考案者デイベースや、共にMSCを実践しているジェス・ダート (Jess Dart) が複数の文献で述べている。その中で最も包括的で詳細なものが、前述のガイドブック冒頭の説明である。多くの組織がMSCを有効であると考え理由、として以下の7項目を述べている。

1. 当初予測できなかった変化を把握することができる。
2. 組織内に存在する価値観を明らかにし、その中でどの価値観が最も重要なのかについて、実際的な議論を行うことができる。
3. 特別な技術を必要としない、参加型のモニタリング手法である。他のモニタリング手法と比べて、異なる文化圏でも理解しやすい。
4. 自分は何ぞある出来事の方が別の出来事よりも重要と思うか、説明しなければならぬことから、データ収集およびデータ分析を促す手法である。
5. データ分析力やインパクトを言葉で表現する能力を強化するための人材育成に有効である。
6. 組織的・社会的・経済的变化が数値化され、簡素化した表現となるのではなく、何が起きているのかについて、生き活きと表現することができる。
7. あらかじめアウトカムを設定していない、ボトムアップの取り組みであっても、モニタリング・評価することができる

(Davies & Dart 2005:12)。

7項目を概観すると、人材育成に有効等、参加型評価の特徴を中心に、他の性質も含まれ、MSCに多様な特色があることを理解できる。その一方で7項目は、近似している性質が混在していたり、ここに含まれていない特色がガイドブッ

クの別部分に書かれていたりする。その結果、手法の理論的特色に関しては、ガイドブックを読んでも全体像を把握することがやや困難になっている⁸。MSCは開発現場の試行錯誤の中で考案されたため、理論的解釈や体系化が後追いになっているのかもしれない。

筆者は、MSCの特色をさらに整理する必要があると考え、ガイドブックを中心にデイビスとダートの著作から見解を抽出し、類似内容ごとに分類を試みた。それらを参加型評価や質的分析等、評価に関連する既存の理論や概念に対応させ、以下の4つのカテゴリーに大別した。

- ①モニタリング手法及び評価手法であるMSC
- ②参加型評価手法としてのMSC
- ③質的分析手法としてのMSC
- ④組織学習の手段としてのMSC

この4つの視点に従って、MSCの特色を順番に説明していきたい。

①モニタリング手法及び評価手法であるMSC

考案者デイビスら(2005:58)は、MSCは「モニタリング手法及び評価手法」であるという。彼らはモニタリングを「主にプログラム管理を目的とした、進行中である情報収集のプロセス」と、評価を「アウトカムとインパクトに焦点を当てた、頻度の低い情報収集のプロセス」と定義し、いずれも達成度についての判断を含むが、評価はプログラム全体を広範に観察し、長期間に及ぶ傾向がある、と述べている。そして「プログラム・サイクルの全工程で実施され、プログラムを実施管理するうえで有効な情報を提供する」ことから、モニタリング手法であり、さらに、「プログラム全体の業績を評価する」うえで評価手法としても有用、としている。

MSCは基本ステップを定期的に繰り返し実施する中で、モニタリングと同様にプログラム管理のための進捗状況を提供する。さらに、「重大な変化の物語」を集め(ステップ4)、「最も重大な変化の物語」を選ぶ過程では(ステップ5)、アウトカムとインパクトに着目するため、評価の性質も持っている。MSCは、モニタリング手法と評価手法の双方の性質を備えているようだ⁹。

②参加型評価手法としてのMSC

MSCは参加型評価手法の1つとして紹介される

ことが多い。MSCと参加型評価について考える。

デイビスら(2005:8)によれば、MSCは「記録すべき変化の決定やデータ分析の過程に多くの利害関係者が関与することから、参加型」である。MSCはプログラムの組織階層を利用して、受益者から意志決定者まで、幅広い利害関係者が参加するしくみがある(図1)。

MSCのデータ収集は、評価専門家ではなく、受益者や現場スタッフ等利害関係者が、自分達の視点に基づいて、物語を集める形式で行う(ステップ4)。収集データの分析も、専門家が情報を集約して行うのではない。利害関係者が、物語を選択する過程を通じて行う(ステップ5)。その際、なぜある出来事の方が、別の出来事よりも重要と思うか、説明しなければならない。この過程を通じて、利害関係者によるデータ収集およびデータ分析を促進し、その結果、データ分析やインパクトを表現する能力を強化する人材育成に有効になる。また最も重大な変化と選択理由を現場にフィードバック(ステップ6)することで、次の報告で類似の物語等、変化を探すのに役立つことや、組織内コミュニケーションを通じて、継続的な対話を作ることが可能になる(Davies & Dart 2005 Dart 2000)。

このようにMSCは、専門家でない利害関係者が、データ収集・分析・フィードバック等、評価プロセスへの参加を通じて「学習」し、意識啓発・能力開発・相互理解につながる参加型評価の長所を備えていると考えられる。しかし、これはエンパワーメント評価等他の参加型評価手法の特徴とも共通している。それではMSCの独自性は何か。デイビスら(2005:72)によれば、MSCはデータが文章(テキスト)による説明である点が、他の参加型モニタリング・評価手法と異なるという。参加型評価の短所とMSCについては後述する。

③質的分析手法としてのMSC

NGOの行うプロセス型社会開発プロジェクトは、住民意識や行動の変容等、成果を数量化できないことが多い。MSCは、変化を数値・簡素化するのではなく、複雑な質的变化をいきいきと描写できる、質的分析手法の側面があるという(Davies & Dart 2005:12 Dart 2000:128)。

質的分析とは、文章を中心とした数量化されない質的データから、どのような意味が見いだせるか、帰納的な分析を行うことである。対象のひとつひとつを事例として、存在する文脈に関してできるだけ再現可能でかつ妥当な推論をする。例えば、「状況を説明する・カテゴリー化する・情報間の関連性をみる」解釈を行う。質的分析では、対象の変化を深く理解して、達成に影響を与えた要因を把握できる、特色がある（長尾 2009：14-15）。また、MSCは妥当性の高い質的データの提供ために、「分厚い記述」を採用する。これはローカルな文脈によく関連付けて配置された出来事の説明であり、ケース・スタディ等の質的分析で使われる（Davies & Dart 2005：67）¹⁰。

質的分析手法としてのMSCには、国際NGOの実践報告に多く取り上げられている特質がある。それは、事前設定の指標を用いないため、直線的因果関係で推測されない、「想定外」の変化を把握できる柔軟性である。国際NGOのMS（Mellefolkelig Samvirk）¹¹は、タンザニアにおいてMSCでプロジェクト評価を行った結果、想定外の成果を把握した事例を紹介している。MSは評価にLFAに準ずる手法を使用してきた。収集データの大部分が効果的に活用されなかったため、LFAを補完する手法としてMSCを導入した。農村での収入向上事業のグループ討論では、男性に混じって若い女性から「私が男性と一緒に座って発言している。こんなことはこれまでなかった。これこそ支援のお陰です」と、女性のエンパワーメントに関する「重大な変化」が語られた。評価スタッフが、収入向上の指標に固執していたら、見落としてしまった変化かもしれない。これ以外にも予測しなかった変化や質的な情報が沢山もたらされ、得た教訓が次年度の計画に反映された（Sigsgaard 2004）。

デイビスら（2005：59）は、以下のように説明している。

「指標によるモニタリングは、『発生すると思われる事象に関する既存の概念・理論から導かれる、知る必要があると思われる情報』のみを示し、それを越える情報を示すものではない。MSCは従来のモニタリング・評価の枠組みでは見落とされがちな活動の複雑な変化～期待していなかった

無形、あるいは間接的結果を含む～モニタリングできるようになる。定期的にこのような情報を得て、またその意味を熟考することによって、求めているアウトカムをより達成できるようになる。」

④組織学習の手段としてのMSC

NGOにおける組織学習は、「利害関係者の充足をめざす組織行為の継続的变化の為、個人・集団の学習過程を内部活用すること（Britton 2005：8）」である。事業を改善する教訓を得る学習ではなく、事業を実施している組織自体を改善するための「学習」である。

参加型モニタリング・評価は組織学習に有効であるという。利害関係者が事業だけでなく、事業を行う組織の価値観や力量を問い直すことにより学習し、組織強化の効果がある（Estrella & Gaventa 1998）。MSCも同様に、物語を選ぶプロセスを通じて組織の価値観が明らかになり、それを議論したり変化させることができ、プログラムの具体的な活動とパフォーマンスの領域を越えて組織全体が成長することが可能である、という（Davies & Dart 2005：63）。

4. バングラデシュにおけるMSCの実践

(1) 対象NGOとプロジェクト

MSCの4つの特色を文献から考察した。これは実際のプロジェクト評価で、どのような形で発現するのであろうか。また実践する上でMSCを活かす留意点は何なのだろうか。

筆者は、2014年2月にバングラデシュでNGOが実施中のプロジェクト評価で、MSCを適用した¹²。対象組織はバングラデシュの現地NGOである自然資源研究センター（CNRS：Center for Natural Resource Studies）である。CNRSは1993年に設立され、全国で住民参加型の自然資源管理プロジェクトを実施している。MSCの適用対象はCREL（Climate Resilient Ecosystems and Livelihoods）プロジェクトとした。これは、CNRSが米国国際開発庁（USAID）の支援を受け、バングラデシュ東北部のスリモンゴルにおいて2012年10月から5カ年計画で行政・研究機関等と協働で実施中である。プロジェクト目標は「生物多様性への脅威を

削減、地球規模の気候変動に適応して生計を向上させる」であり、主な活動は「参加型で森林や湿地を保全しながら住民の生計を向上させる、代替生計手段の導入」である (USAID 2013a)。

CNRSは参加型開発に重点をおき、CRELプロジェクトは住民参加型の社会開発事業として計画・実施されている。モニタリング・評価に関しては、LFAの指標 (大部分は量的指標) を設定 (USAID 2013b)、それに従って参加型というよりは、内部専門家によるインタビューで進捗確認が行われている。

(2) MSC実施のプロセス

①MSCの講義とデザイン

CNRSスリモンゴル事務所において、MSCを実施する主体としてMSCチームを結成した。構成は、筆者とローカルスタッフと6名である。次にチームで話し合い、MSC適用の目的を、「CRELプロジェクト評価にMSCを適用して、手法の特色を考察し有効な実施方法を探る」とした。

筆者がチームへMSC概論を講義し、基礎知識を頭に入れた上で、チームで進め方を議論した。MSCは10ステップ (表2) の実施が基本だが、試験の実施であるため、ステップ2から6までを実行することにした。「ステップ2 変化の領域を決める」に関しては、MSC経験が少ない段階で複数の領域を事前に設定することは困難と判断し、暫定的に「プロジェクトに関する変化」と幅の広いものとした¹³。「ステップ3 時間の範囲を決める」は、プロジェクトを開始して1年余りであること、また受益者の変化に対する記憶が鮮明なうちに情報を収集する理由で「過去6ヶ月間に起こった変化」とした。その結果、受益者に訊く設問は以下の2つとなった。

- ・あなたにとって、過去6ヶ月でプロジェクトに関する最も重大な変化は何ですか。
- ・あなたがそう思う理由を教えてください。

質問をする対象は、CRELプロジェクトの典型的な受益者から公平に選ぶため、現場に詳しいチームメンバーと相談し、男性によるボロガンギナ村の住民組織であるRMO (Resource Management Organization)および、女性中心のシラジナガル村のVCF (Village Conservation Forum)の2カ村とし

た。

②ステップ4：重大な変化の物語を集める

チームで2カ村を訪れ、チームメンバーが分担して住民への2つの質問を訊くインタビュー調査を行った。ボロガンギナ村では男性8名から物語を集め、シラジナガル村では女性を中心に11名から物語を集めた。CNRSはこの地域でCRELプロジェクト以前から活動しており、住民と信頼関係が構築されていた。そのため住民はリラックスした雰囲気の中、自由に意見を表明していた。

収集が終わりチームで事務所に戻り、集めた多数の物語全体を概観したところ、内容的に幾つかの範疇に分類できることがわかった。そこで改めて、変化の領域を「自然資源管理における変化」「社会経済的变化 (個人の変化含む)」「社会的責任に関する変化」「その他」と細かく再設定して、各物語の内容に従って4つのいずれかに割り当てた。

③ステップ5：最も重大な変化の物語を選ぶ

各メンバーが集めた物語を清書した後、チームで物語を選択する会合をもった。各変化の内容と話し手が重要と考える理由について、順番に読み上げ全員で共有した。続いて物語を順番に精査して、選択作業を行った。選択結果全ての紹介は字数的に困難なので、代表的な例を以下に述べる。

・ボロガンギナ村：「自然資源管理」領域では「プロジェクトによって渡り鳥が増えた」等5つの物語が集まった。この中から「湿地帯での商業的な魚の養殖が環境を脅かしている」ことを述べた物語を「最も重大な変化 (MSC)」として選んだ。プロジェクトによって起きた変化ではなく、周辺環境に起こった変化であり、LFAの指標にはない。選択理由は「否定的な変化だが、この問題に対処せずに放っておけば、プロジェクトの成果が水泡に帰す脅威となる。そういう訳で、この変化を把握する意味がある」であった。またチームメンバーから「プロジェクトにより住民の環境への意識が高まり、これまで気にしていなかった環境悪化への懸念を示すようになった。『住民の意識変化』として、良い変化と考える」とコメントがあった。

・シラジナガル村：女性による「社会・経済的变化」領域の7つの物語から比較検討の末、「私の孫はクラスメイト」と題された物語を選択した。

「非識字者であった年配女性が、プロジェクトの識字教室で自信をつけ、自分の孫と一緒に勉強することに喜びを感じている。それが孫にも良い影響を与えている」が選択理由である。他に「経理を覚え夫から馬鹿にされなくなった」「野菜栽培を始めて収入増加が期待される」等、女性のエンパワーメントに関する複数の変化があった。これらの中でどれをMSCに選ぶかで、真剣な意見交換が展開された。最終的に「私の孫はクラスメイト」が選ばれた理由は「この物語では女性の変化が若い世代に好影響を与えているが、他は変化が本人だけに留まっている」であった。

筆者はチームの選択過程を観察していたが、3つの点を把握した。第1に選択過程が、チームメンバー（スタッフ）参加の質的分析であったことである。物語はほとんどが個人や社会の質的变化の叙述である。共通点はあるものの、それぞれに独自の背景と経過がある。皆が納得できるように1つを選択するには、各変化の詳細を把握しながら相互に比較検討をしなくてはならない。これは各対象を事例として、文脈に関して再現可能で妥当な推論をし、状況を説明する・カテゴリー化する・情報間の関連性をみる、質的分析を行うものであった。

2番目は、選択過程を通じて、スタッフのプロジェクトや受益者に対する理解と共感が高まることである。チームの大半はCERLプロジェクトに従事しており、事業について理解している認識だったようだ。しかしMSCを通じて、気づかなかった受益者の意識変化等を知り、驚きとともに視野が広がり、住民への愛着が高まったようだ。これはスタッフのモチベーション向上につながる。

第3は選択を通じて、情報の検証・修正が行われることである。物語をチームで精査すると、一部事実と異なると思われる情報があった。記憶違いや計算間違い等である。選択する側もスタッフで現場の実情に精通しているので、「この発言は以前の記録と矛盾するので、確認する必要がある」という具合に、不確かな点を発見・修正できるのである。物語相互の比較検討においても、複数の目で解釈しながら、1つの結論を導かなくてはならない。この過程で視点が偏っていたり、根拠が不十分な変化は、徐々に排除されていった。

④ステップ6：選択過程をフィードバックする

数日後チームは2カ村を再訪してフィードバックを行った。住民に再度集まってもらい、「最も重大な変化」に決まった物語と、選んだ理由を順番に説明した。両村の住民は熱心に耳を傾け、結果に同意した。関連して変化に関する追加情報を提供してくれる場合もあった。チームは村を再訪する機会を使用して、不確かな情報の真偽を直接住民に確認することもできた。

筆者はチームが選んだ物語が住民に受け入れられるかどうか関心があった。実際のところ、質疑応答の後、彼らは自然に結果を受け入れた。CNRSが住民と信頼関係を築いていることもあるが、チームが選択過程を丁寧に説明したことが良かったと思われる。住民は自分の物語に愛着があり、また他の物語にも興味があるだろう。結果として自分の物語が選ばれれば嬉しいし、そうでなくても公平かつ適切に選択されたことを理解したため、結果に同意したと思われる。また、この過程が住民のプロジェクトに対する記憶を刺激し、関連状況を提供する姿勢につながったと思った。

ステップ6が終了した段階で、住民の女性リーダー1名に、MSCに受益者として参加した感想を聞いた。彼女は、「物語を話すのは、皆と気持ちを共有できて楽しかったです。最初、話す事が難しかったです。段々上手にできるようになりました。女性の教育は重要と思うので、チームの選択には賛成します。このような過程を、将来支援がなくても、自分達だけでできるようになりたいです」と、しっかりと意見を表明した。彼女はリーダーとして、プロジェクトへの参加は既に高かったと思われるが、MSCで学びを共有することで、更に当事者意識が高まったと考える。

⑤振り返り

MSC実践が終了後、チームで振り返りの会合を持ち、質問紙に意見を記入してもらった(表3)。参加型評価の特長である能力開発に対しては、「伝統的手法(LFA)はデータを上層部に送るのみである。MSCはスタッフがそれを活用できる」、「人々が進捗を知り、実績を分析できる」、等のコメントがあり、質的分析や指標に捕らわれない柔軟性に関しては、「伝統的手法はプロジェクト目標に沿った変化のみを把握するが、MSCは目標

表3 MSCチーム振り返りの結果

MSCの特徴	チームメンバーからの意見
利害関係者の人材育成に関して	<ul style="list-style-type: none"> ・ 伝統的手法 (LFA) はデータを上層部に送るのみである。MSCはスタッフがそれを活用できる ・ 人々が進捗を知り、業績を分析できる ・ スタッフや受益者が知識や意識を向上できる
質的分析や想定外の変化を把握する柔軟性に関して	<ul style="list-style-type: none"> ・ 伝統的手法はプロジェクト目標に沿った変化のみを把握するが、MSCは目標の外にある変化を把握できる ・ 受益者は (指標に捕らわれず) 自由に話すことが可能になる ・ 徐々に発生・展開する質的变化を文書で捉えることができる ・ MSCは質的变化を参加型で捉える
その他コメント	<ul style="list-style-type: none"> ・ インタビューを通じて住民から適切に物語を聞き取り、わかりやすく文書化することが難しかった ・ プロジェクト開始時から、定期的の実施することが望ましい ・ 継続的なMSC実施が、より良い運営につながる ・ 異なる変化を見ることで『何が起きているか』が明確にわかる

(出所) MSC実施プロセスに従って筆者作成

の外にある変化を把握できる」、「受益者は指標に捕らわれず、自由に話すことが可能になる」、「MSCは質的变化を参加型で捉える」といった感想があった。チームの結論は、「CNRSはMSCを正式に採用するべきである」であった。

5. 考察と結論

(1) 4種類の特色から見た考察

バングラデシュでの実践を踏まえて、MSCの4つの特色に準じて考察していきたい。

①モニタリング及び評価手法であるMSC

今回のMSC実施は単発であり、モニタリングというよりは評価の形態であった。学習効果はあったが、振り返りでは「MSCをプロジェクト開始時から、定期的の実施することが望ましい」との感想があった(表3)。本来MSCは、基本サイクルをモニタリング的に3ヶ月から半年毎に行うことが奨励されている。意識・行動変容のような、時間をかけて複雑な過程を経て発現する変化(アウトカム)は、単発でなく時間をかけてデータ収集・分析を繰り返すことが望ましいかもしれない。「環境への住民意識」や、「女性のエンパワーメント」事例は、その後どのように移り変わるのか。そうなる要因は何か。新しい変化はあるのか。複数変化の相互関係は何か。MSCを継続的・定

期的に実施することで、時間の経過に沿って深い質的分析が可能になる。またフィードバックに関係者が繰り返し参加することで、意識啓発や能力開発に一層寄与することが期待できる。

②参加型評価手法としてのMSC

参加型評価の長所に、利害関係者の意識啓発や相互理解・能力開発の促進がある。今回MSCで得た情報や教訓はスタッフのレベルに留まるだけでなく、すぐさまフィードバックで受益者に還元された。これまでモニタリング・評価で得た情報は、組織上部や資金提供者に送られたが、その後どう活用されたか、現場スタッフや受益者が知る機会は限られていた。また女性リーダーはMSCを通じて、変化や教訓を共有する喜びと識字教育の重要性を再認識した、と語り、今後MSCを主体的に活用したい意志を表明した。参加型であるCRELプロジェクトにおいて受益者が、活動へ理解を深め当事者意識を高めること、また評価結果を共有することは、事業改善と目標達成に貢献するだろう。フィードバック(ステップ6)が「下向き説明責任」の役割を担っている。

また筆者はMSCに、参加型評価の短所を補う仕組みが内在していると思った。まず「データの妥当性が劣る」点についてだが、当初物語には事実関係が不明な点が見つかったが、選択過程で修正され、不明な点はフィードバックの際に、直接話し手に確認することで、真偽の検証ができた¹⁴⁾。

また「判断の客観性」では、物語の選択過程でチーム全員が多様な視点から議論を行うことで、バランスの取れた結論に導かれた。次いで「参加者の代表性」の問題だが、今回は試験的实施でありプロジェクト関係者全てが参加したわけではなく、受益者の一部分の2カ村であった。しかし2カ村のグループは全員が平等に物語を表明し、スタッフが全ての物語を公平に審議した。筆者の経験では参加型評価といえども、意見を表明するのは一部の住民に限られたり、データを分析・解釈するのは当事者でなく専門家であったりする場合が多い。MSCは、NGOの組織構造を利用した参加で、「代表性」の問題を一定程度緩和できるのではない。

MSCは受益者から組織の上層部まで、広範な利害関係者が参加し、多様な視点によって集約・識別する機能があるという。プログラム受益者の70%以上がMSCに参加した例をあげて、ランダムなサンプル調査よりもずっと幅広い参加者からの報告を得られる場合があり、調査対象として小さな部分集合に集中しないという (Davies & Dart 2005:69)。これは偏った情報による判断の危険を避ける意味がある。「多様なデータ・情報間の整合性に注意することによって評価の偏りを回避する」機能が、MSCに内包されていると考える。この点が、MSCが他の参加型評価手法と異なる独自性の1つではないか。しかし一度の実践でMSCが参加型評価の弱点を、十分に補うことができるかと結論するのは早急であろう。

③質的分析手法としてのMSC

CRELプロジェクトにおける「女性のエンパワーメント」に対しては、研修参加人数や世帯収入が関連の量的指標である。スタッフはその進捗状況は知っていた。しかし女性1人1人がどのように技能を身につけ、自信をつけたかという各プロセスの詳細は、MSCを行うことで初めて理解したようだ。「私の孫はクラスメイト」をはじめ、多様な質的变化を生き活きと描写できるMSCは、スタッフにとって新鮮であった。得た情報を今後のプロジェクト運営に活かすことで、事業を改善できる可能性がある。

また質的分析は、現場での臨機応変な対応等、慣れと経験が必要な「職人芸」といわれ (藤掛

2008 古賀 2008)、概して専門家の評価になりがちである。今回、データの質的分析はスタッフが、物語の選択において意見交換を通して行った。専門家ではないスタッフによって、一定レベルの質的分析ができた。試験的实施であり組織の上層部まで巻き込んで参加は実現しなかったが、MSCは参加型で質的分析が可能になる手法と思われた。

デイビスら (2005:68) は、代表的な質的分析手法であるケース・スタディでは、専門家が情報の収集・分析をする。またその根拠や過程は一般に利害関係者には知らされない。これに対してMSCは、利害関係者の参加で収集・分析を行い、判断理由も選択過程やフィードバックを通じて他の関係者に共有される透明性がある、としている。

また、CRELプロジェクトのモニタリング・評価はLFAの指標の確認のみが行われていた。収集データは、指標に関連した内容に限られる傾向があった。今回MSCで把握した「受益者の意識変化」は、指標に書かれていない変化であり、否定的な変化でもあった。しかしこの情報を認識することで、将来プロジェクトに否定的影響を与える課題を認識し、適切な対策を講じることができる。住民が「指標に捕らわれず自由に語れる」ため、これまでのモニタリング・評価体制では、見落とされがちな波及的な変化を、柔軟に把握することができた。

④組織学習の手段としてのMSC

「組織の価値観を明らかにして議論・変化させる」特色については、単発のMSC実施では判断材料を得る事はできなかった。しかし個人的に印象に残ったことがある。CNRSは住民参加を重視するNGOだが、スタッフと接する中で、官僚的ではなく一人一人を大切にしている組織文化を感じた。筆者がMSC実施中に彼らを観察すると、物語の選択過程で社会やコミュニティの変化よりも、個人のエンパワーメントに関する変化を優先して選ぶ傾向がみられた。これはCNRSが持っている価値観が表れているのではないか。このように組織の価値観を明らかにしていく特色は観察できた。これが時間の経過や社会環境の変化ともどのように変化してゆくかは興味深い。それを知るにはMSCを継続的に実施して観察する必要がある。

(2) 実施上の課題と提言

バングラデシュの実践に基づき、MSCを実施する上で、課題になる点と提言を述べる。

①スタッフへの研修・能力開発が必要

村の女性リーダーは「最初、話す事が難しかった」と物語を聞かれた際の感想を述べた。また振り返りでチームメンバーからは、「住民から適切に物語を聞き取り、わかりやすく文書化することが難しかった」という声があった(表3)。集めた物語を精査すると、複数の変化を混同したり、選択理由を適切に記述できていなかったりした。実際のところ、途上国ばかりでなく教育等が普及した先進国においても、人々に突然、「変化は何か」、「その理由は」と訊いても、すぐに適切な回答は難しいかもしれない。またスタッフに文書作成の経験が少なければ、限られた時間で「分厚い記述」として記録をまとめることは困難であろう。訊く側であるスタッフ等に、住民から適切に情報を引き出す技術や、情報をわかりやすく文書化する能力開発が求められる。この課題は考案者にも認識されており、効果的なMSC実施のためは、スタッフ対象のインタビューやファシリテーションに関して、能力開発が必要と述べている(Davies & Dart 2005:54-57)。

②継続的・反復的なMSC実施が望ましい

振り返りでは、「MSCをプロジェクト開始時からモニタリング的に実施することが望ましい」との感想があった。考案者はMSCを3ヶ月から半年毎に繰り返すことが効果的であり、他のモニタリング制度に平行してMSCを導入・実施することを提唱している(Davies & Dart 2005)。NGOがMSCを定期的に繰り返し実施することで、教訓を学び活かすPDCAサイクルが、より効果的に実現する可能性がある。その一方で、「MSCの単発の評価での成功例が増えている」との意見もあり、今後の実践研究が待たれる¹⁵。

(3) 実践からの考察のまとめ

①4つの特色と学習

冒頭で述べたように、筆者は日本のNGOがMSCを導入・活用することで、学習目的の評価を推進し、ミッションの達成に貢献する可能性が高まる、と考えている。そして本稿では、その第

一步として、MSCの4つの特色を文献と実践から考察した。その結果をまとめる。

「モニタリング及び評価手法であるMSC」に関しては、今回は評価として単発のMSCを活用し、基本ステップを実施することで、利害関係者への学習効果があきらかになった。どのような学習があったかだが、「参加型評価手法としてのMSC」の視点からは、受益者やスタッフがデータ収集(ステップ4)、やデータ分析(ステップ5)に参加する中で、彼らの意識啓発や能力開発・相互理解の貢献することが確認された。特にフィードバック(ステップ6)によって下向き説明責任が確保されていることがわかった。また妥当性・代表性に伴う参加型評価の短所を、一部補う機能がみられた。また「質的分析手法としてのMSC」の観点では、MSCによって、専門家でなくスタッフの参加による質的分析が可能になり、特に指標から予測しにくい変化を把握・分析する特質が観察できた。最後の「組織学習の手段としてのMSC」に関しては、対象となる組織の価値観の反映をある程度観察することはできたが、それが今後どう変化するかは、継続的な考察が必要である。

効果的な実施に向けての課題としては、スタッフのインタビューやファシリテーションに対する能力開発の必要性がある。また単発での実施でなく、MSCをモニタリング的に継続することで、より深い学びを得る示唆を得た。

バングラデシュでの実践を通じて、程度の差はあるが、MSCの4つの特色の観点から、学習効果の発現を確認することができた。得た情報や教訓を今後の活動に反映させることで、プロジェクト目標やミッションの達成に貢献できると考える。一事例ではあるが、NGOの社会開発事業対象の評価にMSCを活用することで、利害関係者の学習を推進できることが明らかになった。

②NGOがMSCを活用する理由

現場での実践を通じて筆者は、NGOがMSCを積極的に活用する理由を理解した。国際NGOだけでなく、CNRSのような実績ある現地NGOは支援を受ける場合、通常は指標によるモニタリング・評価が義務付けられる。だが現場スタッフの立場にたつと、その作業だけを繰り返しても、事業を改善する教訓を学ぶことは難しい。これに対

してMSCを併用することで、現場で起こる様々な変化を参加型で分析して教訓を学び、受益者と共有できる。また自分や受益者の意識啓発や能力開発を実現できる。現場スタッフが、この学習効果を有益と捉えていることが、多くのNGOで採用され、報告に取り上げられる理由ではないか。

(4) 今後のMSCの研究課題

MSCが日本のNGOで利用されるためには、以下の2点に着目した研究が必要と考える。

①MSCを用いた評価の意義

MSCは、NGOの評価に本当に有益なのか、3つの方向性で研究を進めたい。

・複数事例の比較と中長期的な考察：バングラデシュの事例のみの実践で、MSCの学習効果について詳細が判明したとはいえないであろう。社会開発事業といっても、実施国や組織、目標によって内容は様々である。複数事例の比較・検討による分析が求められる。また単発ではなく、MSCのモニタリング的な継続活用を含めて、中長期的な時間経過を通じた観察を行いたい。これによって4つの特色に関しても、より深く検討できると考える。

・他の手法・アプローチとの比較

NGOの学習目的の評価にはMSCが有効である、というのが筆者の主張である。しかし利害関係者の学習効果等MSCの特色は、エンパワーメント評価等他の参加型評価とも共通している。NGOには、他の手法ではなくなぜMSCが相応しいのか。その独自性に関しての考察は不十分である。3章で、MSCと他の参加型手法の違いは、テキスト（文章）を用いること、というデイベースの指摘を述べた。また今回の実践で、MSCは利害関係者の組織的な参加で「代表性」の課題に対応する機能があること等を、一定程度観察したのみである。MSCの学習目的評価の意義を考察する上で、理論及び実践において、他手法との比較検討が有益であろう。

・MSCと上向き説明責任との関係

NGOにとっても資金提供者等への「上向き説明責任」は重要である。MSCは下向き説明責任に貢献するが、[上向き]説明責任目的の評価には不向きと言われる（Davies & Dart 2005：13）。

しかし、筆者は必ずしもそうではないと考えている。複雑な過程を経て時間をかけて発現する社会開発事業の成果を、文章を通じて質的に把握し参加型で分析するMSCは、事業の達成度だけでなくプロセスの詳細を説明・理解する意味で、上向き説明責任の達成にも利点があると考えられるからである。

②日本のNGOにおけるMSC導入の意義と可能性

MSCによる評価の意義が確認されても、そのまま日本のNGOで機能するとは限らない。MSCを受け入れる日本のNGO側の現状を知る必要がある。日本のNGOのプロジェクトは、国際NGOと同様に教育や人材育成、保健医療等の社会開発が多く、プロセス型だと思われる。筆者は、これらの事業の学習目的評価に、MSCが貢献すると考えている。しかし、そう結論する前に、日本のNGOの事業内容やマネジメントに関する詳細な分析が必要だろう。

加えて、NGO以外の広い範囲の社会・非営利活動、例えば福祉や教育等国内NPOの事業評価にも、MSCを活用する可能性があると考えられる。この点は、現状では筆者の能力の限界を越えるので、長期的な計画で検討したい。また、他の研究者・実践家によるMSC事例報告を期待している。

謝辞

本稿の執筆にあたり、査読者をはじめ有益なコメントを頂いた皆様に感謝を申しあげたい。

注記

- 1 佐藤（2007）は、因果律に基づいてプロジェクトの貢献度が測りにくいこと、目標が曖昧なので、数量化した評価になじまないことを理由に、社会開発的な介入の評価は困難であると指摘している。
- 2 2004までにMSCが実施されたこれまでの実績：フィリピン、バングラデシュ、インド、ラオス、タイ（以上アジア）、マラウィ、エチオピア、モザンビーク、ガーナ、タンザニア（アフリカ）、太平洋諸島、パプアニューギニア、中央アメリカ、セルビア、オーストラリア、複数の先進国（Davies & Dart 2005）
- 3 英国能力開発NGOであるINTRAC（International

- NGO Training and Research Centre) が主催する、“Evaluation Conference” が、1989年から2011年まで、先進国・途上国のNGOスタッフや研究者の出席で6回行われている。
- 4 このように先進国のドナー(資金提供者)による評価基準は、「援助する側の論理構成に基づいた評価」になりがちで、被援助国側(本稿ではNGOの開発現場)の関心とは必ずしも一致しない(源 2007:74)
 - 5 Outcome Mappingについては、以下に詳しい。Earl, Sarah., Carden, Fred., & Smutylo, Terry, Foreword by Michel Quinn Patton (2001). Outcome Mapping: Building Learning and Reflection into Development Programs, International Development Research Centre
 - 6 日本語訳である「モスト・シグニフィカント・チェンジ(MSC)手法・実施の手引き(日本語版2013)」は、<http://blog.livedoor.jp/sankagatahyouka/MSCGuide-Japanese-2013.pdf> でダウンロードできる。
 - 7 MSCでいうところの「物語」はanecdotes(逸話)やvignettes(短い描写)に類似するという(Davies & Dart 2005:72)。筆者は、質的分析でいうnarrativeや、日常語でいうエピソードに近い概念と考える。
 - 8 その一方で、このガイドブックは、MSC実施の手順やトラブルシューティングに関する記述など、全体を通して実践家向けのマニュアルの内容が充実している。
 - 9 なお、国際NGOでは参加型モニタリングと参加型評価はほぼ同義語で使われ、明確に区別することは難しいという(Estrella 2000 p5)。国際NGOは通常、両者をまとめて参加型モニタリング・評価(Participatory Monitoring and Evaluation: PME)と呼ぶ。
 - 10 分厚い記述とは、人類学者クリフォード・ギアツが提唱した、すぐれた質的研究の報告書に盛り込まれている、研究対象や調査現場の状況に関するリアルできめ細かい記述を指す(佐藤 2008:4)。
 - 11 MS (Mellempfolkeligt Samvirk/Action Aid Denmark)は1944年に第二次世界大戦からの復興を目的に設立され、現在はアフリカやアジア、中央アメリカで、現地パートナー組織を通じて開発プロジェクトを実施している。国際NGOネットワークである、アクションエイド・インターナショナルのメンバーでもある。<http://www.ms.dk/en/>
 - 12 バングラデシュでのMSC実施には、日本の国際協力NGOの人材育成を目的とした、外務省主催・国

際協力NGOセンター事務局の「平成25年度・海外スタディ」制度を活用した。

- 13 MSCの基本ステップによると、物語を集める(ステップ4)前段階のステップ2で、「変化の領域」を決めることが基本である。しかし事前に決めることが困難な場合は、最初は領域に分類せず、データ収集後にデータ内容に合わせて領域を決定することも可能であるという(Davies & Dart 2005:20)。
- 14 デイビースら(2005:67)はMSCの妥当性(validity)について、物語の「分厚い記述」による詳細な説明や、構造的な選択プロセス、透明性、事実確認(ステップ7)、参加、仲間による検証、を通じて、収集データが取舍選択、修正されて、妥当性を確保できるとしている。
- 15 デイビース(2014)が主催する、MSC意見交換メーリングリストの議論より。

参考文献

- アユス(2003)『国際協力プロジェクト評価』、国際開発ジャーナル社
- 河村能夫(2003)「住民参加型農村開発のための計画立案手法」、『参加型開発～貧しい人々が主役になる開発へ向けて』、日本評論社
- 国際協力事業団(JICA)(2001)「参加型評価とは何か」『国際協力と参加型評価』
- 佐藤郁哉(2008)『質的調査分析法～原理・方法・実践』、新曜社
- 佐藤寛(2007)『テキスト社会開発』、日本評論社
- 古賀正義(2008)『質的調査法を学ぶ人のために』、世界思想社
- 田中 博(2010)「フィリピンでのNGO教育事業参加型評価～ファシリテーターの役割とステークホルダーのエンパワーメント」、『日本評価研究』、10(1): 75-90
- 田中 博(2011)「市民社会におけるNGO/NPOと評価の役割～マネジメント能力を高め、NGO/NPOの進化を加速させる参加型評価」、『日本評価研究』、11(1): 75-90
- チェンバース、ロバート(2000)『参加型開発と国際協力』、明石書店
- 長尾真文(2009)「データ収集・分析」、『第三期評価士養成講座テキスト』、14-15

- 藤掛洋子 (2008) 『評価論を学ぶ人のために』、世界思想社
- 源由理子 (2008) 『評価論を学ぶ人のために』、世界思想社
- 源由理子 (2007) 「ノンフォーマル教育援助における参加型評価手法の活用－『利害関係者が評価過程に評価主体として関わること』の意義」、『日本評価研究』、7 (1) : 73-86
- 三好皓一 (2008) 『評価論を学ぶ人のために』、世界思想社
- 三好皓一・田中弥生 (2001) 「参加型評価の将来性～参加型評価の概念と実践についての一考察」、『日本評価研究』、1 (1) : 66-69
- Britton, B (2005). *Organisational Learning In NGOs*, Praxis Paper 3, INTRAC
- Dart, J. Jess, (2000) . *Stories for Change : ANew Model of Evaluation for Agricultural, Extension Projects in Australia*
- Davies, Rick (2014) . *The "Most Significant Changes" approach*, <https://groups.yahoo.com/neo/groups/MostSignificantChanges/info> (2014/9/24アクセス)
- Davies, Rick & Dart, Jess (2005) . *Most Significant Change (MSC) Technique: A Guide to Its Use*, <http://www.mande.co.uk/docs/MSCGuide.htm> (2014/7/8アクセス)
- Estrella, Marisol (2000) . *Learning from Change, Learning from Change: Issues And experiences in participatory monitoring and evaluation*, ITDG Publishing and International Development Resource Centre
- Estrella-Gaventa (1998) . *Who Counts Really? Participatory Monitoring and evaluation: literature review*, IDS
- Garbutt, Anne (2011) . *Background Paper: INTRAC 7th Evaluation Conference, Monitoring and Evaluation: new development and challenges*, INTRAC, Oxford
- MANGO (2014) . *Accountability*, <http://www.mango.org.uk/guide/accountability> (2014/9/26アクセス)
- Mebrahtu, Esther., Pratt, Brian.& Lonnqvist, Linda (2007) . *Rethinking Monitoring and Evaluation: Challenged And Prospects in the Changing Aid Environment*, INTRAC, Oxford
- Sigsgaard, Peter (2004) . *Doing away with Predetermined Indicators: Monitoring Using the Most Significant Changes Approach, Creativity and Constraint: Grassroots Monitoring and Evaluation And the International Aid arena*, INTRAC, Oxford, 125-136
- Taylor, James and Soal, Sue (2004) . *Measurement in Development Practice: From the Mundane to the Transformational, Creativity and Constraint: Grassroots Monitoring and Evaluation and the International Aid arena*, INTRAC, Oxford, 95-109
- USAID (2013a) . *USAID Launches new environment initiative to improve Bangladesh's Resilience to Climate Change*, USAID
- USAID (2013b) . *M&E-Baselines, Climate Resilient Ecosystems and Livelihoods Project*
- Wallace, Tina & Chapman, Jennifer (2004) . *An Investigation into the Reality: Behind NGO Rhetoric of Downward Accountability, Creativity and Constraint: Grassroots Monitoring and Evaluation and the International Aid arena*, INTRAC, Oxford, 33-44 (2014.11.18受理)

MSC (Most Significant Change), Participatory Monitoring and Evaluation Tool: Consideration to Four Characteristics of MSC Based on its Implementation in Bangladesh NGO

Hiroshi Tanaka
Participatory Evaluation Facilitator
nepalippine@gmail.com

Abstract

MSC (Most Significant Change) is a participatory monitoring and evaluation tool widely utilized by international NGOs. Author has analyzed four characteristics of MSC based on literature review and MSC implementation in Bangladesh NGO, considering MSC is effective for learning purpose evaluations of Japanese NGO projects. The four characteristics are as follows: 1) a monitoring and evaluation tool, 2) a participatory evaluation tool, 3) a qualitative analysis tool and 4) means for organizational learning.

After the MSC practice in Bangladesh, it has been observed that MSC has advantages of participatory evaluation, such as awareness raising, capacity building and mutual understanding of stakeholders and some functions to compensate for limitations of participatory evaluation, such as “validity” and “representativeness” to some extent.

Furthermore, regular implementation from the beginning of project and staff capacity building like interview skill training are important for better MSC implementations.

In order to make MSC more utilized by Japanese NGOs in future, further study is needed, such as to consider about importance of evaluation using MSC, present situation of NGO evaluation in Japan and merits and possibilities of MSC introduction to Japanese NGO.

Keywords

Participatory Evaluation, Qualitative Analysis, Learning, Downward Accountability, NGO

日本評価学会誌刊行規定

2005.2.15改訂

2002.9.18改訂

2001.9.9改訂

(目的および名称)

1. 日本評価学会（以下、「学会」という）は、評価に関する研究および実践的活動の成果を国内外の学界をはじめ評価に関心をもつ個人および機関に広く公表し、評価慣行の向上と普及に資することを目的として、「日本評価研究（仮名）」（英文仮名：“The Japanese Journal of Evaluation Studies”、以下、「評価研究」という）を刊行する。

(編集委員会)

2. 「評価研究」の編集は、後で定める「編集方針」にもとづいて編集委員会が行う。
3. 編集委員会は、学会会員20名以内をもって構成し、委員は学会理事会が選任する。編集委員の任期は2年とし、再任を妨げないものとする。
4. 編集委員会は、互選により委員長1名、副委員長2名および常任編集委員若干名を選出する。
5. 編集委員会は、最低年1回編集委員会を開き、編集方針、編集委員会企画、その他について協議するものとする。
6. 編集委員会は、その活動等について、随時理事会へ報告し、承認を受けるとともに、毎年1回学会年次大会の場で、過去1年の活動成果と翌年の活動計画に関する報告を行う。
7. 委員長、副委員長および常任編集委員は、常任編集委員会を構成し、常時、編集実務に当たる。

(編集方針)

8. 「評価研究」は、原則として、年2回刊行する。
9. 「評価研究」の体裁は、B5版とし、和文又は英文とする。
10. 「評価研究」に掲載する原稿（以下「論文等」という）の分類は、以下の5カテゴリーからなるものとする。
 - (1) 総説
 - (2) 研究論文
 - (3) 研究ノート
 - (4) 実践・調査報告
 - (5) その他
11. 「評価研究」への投稿有資格者は、学会会員および常任編集委員会が投稿を依頼した者とする。学会会員による連名での投稿および学会会員を主筆者とする非会員との連名での投稿は、これを認める。編集委員による投稿はこれを認める。
12. 投稿原稿を上記分類のどのカテゴリーとして扱うかは、投稿者の申請等をもとに常任編集委員会が、下記の「作業指針」に従って決定する。
 - (1) 「総説」は、評価の理論あるいは慣行について概観する論文とし、その掲載については編集委員会が企画・決定する。
 - (2) 「研究論文」は、評価の理論構築あるいは慣行の理解について重要な学問的貢献となると認められる論文とし、その採否については次項に定める査読プロセスを経て常任編集委員会が決定する。
 - (3) 「研究ノート」は、「研究論文」作成過程での理論的あるいは経験的な研究の中間的成果物に相当する論考で、その採否については次項に定める査読プロセスを経て常任編集委員会が決定する。
 - (4) 「実践・調査報告」は、評価事業の実践あるいは評価にかかわる調査の報告で、その採否については次項に定める査読プロセスを経て常任編集委員会が決定する。

(5)「その他」には、編集委員会が独自に企画する特集に掲載する依頼原稿や学会誌の刊行に関する編集委員会からの学会会員への連絡等が含まれる。

13. 論文等は2名の査読者により査読することとし、その人選は編集委員会が行う。「研究論文」については、査読結果と編集委員会が査読者とは別に指名する担当編集委員1名の参考意見をもとに、編集委員会が掲載に関する決定を行う。「総説」、「研究ノート」、「実践・調査報告」および「その他」の論文については、査読結果にもとづき編集委員会が掲載に関する決定を行う。
14. 編集委員が「評価研究」に投稿した場合には、当該委員はその投稿に係わる常任編集委員会あるいは編集委員会の議事に一切参加しないものとする。
15. 上記いずれのカテゴリの投稿についても、常任編集委員会による掲載の判断は可・不可の二者択一で行うこととする。但し、場合によっては編集委員会の判断で、小規模の修正による掲載も認める。「研究論文」としての掲載が適当でないと判断された場合でも、投稿者が希望すれば、常任編集委員会は「研究ノート」あるいは「実践・調査報告」としての掲載を決定できる。

(投稿要領の作成公表)

16. 編集委員会は、上記の編集方針にもとづき投稿要領を作成し、理事会の承認を得て、広く公表する。

(配布先)

17. 「評価研究」は、学会会員に無償で配布するほか、非会員に有償で提供する。

(抜刷の配付)

18. 「評価研究」掲載論文等の抜刷り30部を、投稿者（原著者）に無料で配布する。それ以上の部数を希望する場合は投稿者（原著者）の自己負担とする。

(インターネット上の公開)

19. 「評価研究」掲載論文等は、投稿者（原著者）の了承を得て全文をインターネット上で公開する。

(著作権)

20. 「評価研究」に掲載された論文等の著作権は各投稿者（原著者）に帰属するものとし、編集権は本学会に帰属するものとする。

(事務局)

21. 「評価研究」編集及び配布の事務は、それに関連する会計も含めて学会事務局が担当する。

(以上)

『日本評価研究』投稿規定

2008.9.29改訂

2003.4.18改訂

2002.3.25改訂

2001.9.9改訂

1. 『日本評価研究』(The Japanese Journal of Evaluation Studies)は、評価に関する論文、論考、調査報告等を掲載する。
2. 『日本評価研究』は、会員間の研究成果交流の場を提供し、内外における評価研究の一層の発展に資することを主目的として発行されており、原則として会員による寄稿を掲載する。なお、依頼原稿を除き、ファーストオーサーは学会員でなければならない。また、投稿は、一時に一原稿に限るとともに、他学会誌などへ二重に投稿などのない未発表のものとする。
3. 投稿された原稿は、編集委員会の責任において審査を行ない、採否を決定する。審査にあたっては、1原稿毎に2名の査読者を選定し、査読結果を参考にする。(査読者には、投稿者名を伏せて査読を依頼する。)
4. 原稿料は支払わない。
5. 『日本評価研究』に掲載された論文等は、その全文をインターネット上の本学会のホームページに掲載する。
6. 投稿にあたっては、投稿原稿が、①研究論文、②総説、③研究ノート、④実践・調査報告、⑤その他のうち、どのカテゴリーに入るかを明記する。ただし、カテゴリーについての最終判断は、編集委員会で行なう。「研究論文」は評価の理論構築あるいは慣行の理解について重要な学問的貢献となると認められる論文、「総説」は、評価の理論あるいは慣行について概観する論文、「研究ノート」は「研究論文」作成過程での理論的あるいは経験的な研究の中間的成果物に相当する論考、「実践・調査報告」は評価事業の実践あるいは評価にかかわる調査の報告、「その他」は編集委員会が独自に企画する特集に掲載する依頼原稿等である。
7. 投稿方法
 - (1) 使用言語は日本語又は英語とする。
 - (2) 著者校正は原則として第一校までとする。
 - (3) 英文原稿については、ネイティブスピーカーによる英文チェックを済ませ、完全な英文にして投稿すること。
 - (4) ハードコピー4部(A4版)を提出する。その際、連絡先(住所、Tel、Fax、Email)と原稿の種類を明記すること。掲載可と判断された原稿については、必要なリایتを経た後に、最終原稿のハードコピー2部とDOS/Vフォーマットのフロッピーを用いたTEXTファイルを提出する。その際、オリジナル図表を添付すること。
 - (5) 刷り上がりは最大14ページとする。これを超える場合は、その経費は著者負担とする。

- (6) 日本語原稿の最大文字数は以下のとおり。①研究論文20,000字、②総説15,000字、③研究ノート15,000字、④実践・調査報告20,000字、⑤その他適宜。それぞれ和文要旨を400字程度、英文要旨を150words程度、及び和文・英文でキーワード（5つ以内）を別に添付する。印刷は1ページ、20字×43行×2段（1,720字）とする。20,000字の原稿の場合、単純計算では英文要旨1ページを加えて合計13ページとなるが、図表の量によっては、それ以上のページ数となり得るので、注意すること。
- (7) 英文ではA4版用紙に左右マージン30mmをとり、10ポイントフォントを使用し、1ページ43行のレイアウトとする（1ページ約500words）。論文冒頭に150words程度のAbstractをつける。14ページでは、7,000words相当になるが、タイトルヘッド等を考慮して、最大語数を約6,000words（図表、注、文献込み）とする。図表の量によっては、ページ数が予想以上に増える場合もあり得るので、注意すること。

8. 送付先

〒108-0075 東京都港区港南1-6-41 品川クリスタルスクエア12階
一般財団法人 国際開発センター内 日本評価学会事務局
TEL: 03-6718-5931
FAX: 03-6718-1651
E-mail: jes.info@evaluationjp.org

『日本評価研究』執筆要領

2002.9.18改訂

2002.3.25改訂

1. 本文、図表、注記、参考文献等

(1) 論文等の記載は次の順序とする。

日本語原稿の場合

第1ページ：表題、著者名、所属先、E-mail、和文要約（400字程度）、和文キーワード（5つ以内）

第2ページ以下：本文、謝辞あるいは付記、注記、参考文献

最終ページ：英文表題、英文著者名、英文所属先、E-mail、英文要約（150words程度）、英文キーワード（5つ以内）

英文原稿の場合

第1ページ：Title; the author's name; Affiliation; E-mail address; Abstract (150 words); Keywords (5 words)

第2ページ以下：The main text; acknowledgement; notes; references

(2) 本文の区分は以下のようにする。

例1（日本語）

1.

(1)

①

(2)

(3)

例2（英文）

1.

1.1

1.1.1

1.1.2

(3) 図表については、出所を明確にする。図表は原則として、筆者提出のものをそのまま写真製版するので、原図を明確に作成すること。写真は図として扱う。

例1：日本語原稿の場合

図1 ○○州における生徒数の推移



(注)

(出所)

表1 ○○州における事故件数



(注)
(出所)

例2：英文原稿の場合

Figure 1 Number of Students in the State of ○○



Note:
Source:

Table 1 Number of Accidents in the State of ○○



Note:
Source:

(4) 本文における文献引用は、「……である (阿部 1995、 p.36)。」あるいは「……である (阿部 1995)。」のようにする。英文では、(Abe 1995, p.36) あるいは(Abe 1995)とする。

- (5) 本文における注記の付け方は、(…である¹。)とする。英文の場合は、(…¹)とする。
 (6) 注記、参考文献は論文末に一括掲載する。

注記

1 ……。

2 ……。

- (7) 参考文献は、日本語文献は著者の五十音順、外国語文献は著者のアルファベット順に記し、年代順に記載。参考文献の書き方については以下のようにする。

日本語単行本：著者（発行年）『書名』、発行所

（例）日本太郎（1999）『これからの評価手法』、日本出版社

日本語雑誌論文：著者（発行年）「題名」、『雑誌名』、巻（号）：頁－頁

（例）日本太郎（1999）「評価手法の改善に向けて」、『日本評価研究』、1（2）：3-4

日本語単行本中の論文：著者（発行年）「題名」、編者『書名』、発行所、頁－頁

（例）日本太郎（2002）「行政評価」、日本花子『評価入門』、日本出版社、16-28

複数の著者による日本語文献：著者・著者（発行年）『書名』、発行所

（例）日本太郎・日本花子（2002）『政策評価』、日本出版社

英文単行本：著者（発行年）. 書名. 発行地：発行所.

（例）Rossi, P. H. (1999). *Evaluation: A Systematic Approach 6th edition*. Beverly Hills, Calif: Sage Publications.

英語雑誌論文：著者（発行年）. 題名. 雑誌名, 巻（号）, 頁－頁.

（例）Rossi, P. H. (1999). Measuring social judgements. *American Journal of Evaluation*, 15(2), 35-57.

英語単行本中の論文：著者（発行年）. 題名. In 編者 (Eds.), 書名. 発行地：発行所, 頁－頁.

（例）DeMaio, T. J., and Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco, Calif: Jossey-Bass, 177-196.

2名の著者による英語文献：姓, 名, and 姓, 名（発行年）. 書名. 発行地：発行所.

（例）Peters, T., and Waterman, R. (1982). *In Search of Excellence: Lessons from America's Best Run Companies*. New York: Harper & Row.

3名以上の著者による英語文献：姓, 名, 姓, 名, and 姓, 名（発行年）. 書名. 発行地：発行所.

（例）Morley, E., Bryant, S. P., and Hatry, H. P. (2000). *Comparative Performance Measurement*. Washington: Urban Institute.

（注1）同一著者名、同一発行年が複数ある場合は、(1999a)、(1999b)のようにa,b,cを付加して区別する。

（注2）2行にわたる場合は2行目移以降を全角1文字（英数3文字）おとして記述する。

『日本評価研究』査読要領

日本評価学会 『日本評価研究』編集委員会
2005年9月10日決定

1. 本査読要領の趣旨

本査読要領は、『日本評価研究』における掲載論文等の審査の要である査読手続きについて、投稿する会員及び査読を依頼される会員に対して解説を行い、審査手続きを効率的かつ効果的に行うことを目的として、定めるものです。

2. 査読の目的と投稿者の責任

査読は、投稿原稿が『日本評価研究』に掲載される論文等としてふさわしいものであるか否かについての判定を当編集委員会が行う上で必要とされるものです。

査読に伴って見いだされた疑問や不明な事項について、必要な場合は修正意見をつけて、修正を求めることがあります。査読は、その意味で、投稿原稿の改善に資するものでもあります。ただし、修正が求められた場合においても、論文等の内容に関する責任は著者が負うべきものであり、査読者の責に帰するものではありません。

査読者は2名で、編集委員会において学会会員の中から当該分野の専門家を選び依頼されますが、学会会員以外に依頼することもあります。

3. 査読の視点

査読は、以下の5つの視点によりますが、投稿原稿の種類によって、重点が異なります。

- (1) テーマの重要性・有用度
- (2) 研究の独自性
- (3) 論理の構成
- (4) 実証法・方法論の妥当性
- (5) 評価理論・実践への貢献

- ・研究論文の査読については、上記の5項目全てに配慮する。
- ・研究ノートの査読については、上記5項目のうち、特に (1), (2), (3), (4) の諸項目に配慮する。
- ・実践・調査報告の査読については、上記5項目のうち、特に (1), (3), (5) の諸項目に配慮する。
- ・総説の査読については、上記5項目のうち、特に (3) と (5) の諸項目に配慮する。

4. 投稿に当たっての留意点

2. に掲げた査読の視点以外に、基本的な論文の完成度の問題があります。例えば、

- 論文等として体裁が整っているか、
- 執筆要領にしたがっているか、
- 簡潔明瞭に記述されているか、
- 実証的なデータは適切に位置づけられているか、
- 注や参考文献は本文と対応しているか、
- 専門用語の使用は適切か、
- 語句や文法的な誤りがないか、
- 誤字脱字はないか、
- 句読点に誤りはないか、
- 英文要約などの英文表現は適切か、(必ずしも和文要約の直訳である必要はなく、英文としてまとまっていること)
- 字数は規程に従っているか、

など、内容及び形式に関する留意点があります。

大学院生及び実務家の投稿において、論文としての体裁が整わないまま送付されている例があり、査読そのものに至らないものもあるので、しかるべき指導を受けた後に投稿されるよう強く勧めます。

5. 査読にあたっての判断事例

(1) 完成度において不十分であるが掲載を考慮できる場合

萌芽的な研究、発展が期待できる論文等は評価論の発展のためにできるだけ評価してください。

- 検証は十分とはいえないが、理論や定式化が学問の発展に有用である。
- 考察は十分とはいえないが、新たな理論の形成・促進に有用である。
- 文献調査は十分とはいえないが、研究の位置づけは明確である。
- 比較研究は十分とはいえないが、適用例としては意義がある。
- 考察は十分とはいえないが、社会的、または、歴史的に重要な事例の評価として意義がある。
- 考察は十分とはいえないが、特定の社会活動の評価として意義がある。
- 論文の構成や表現は適切とはいえないが、内容は評価できるものがある。
- 論理性は十分とはいえないが、実務上の有用性がある。
- 有意義な実践・調査報告である。

(2) 掲載を考慮するのが困難と判断される事例

- 問題意識や問題の設定が不明確。
- 基本的な用語の概念の理解や分析枠組が不明確または不適切。
- 論拠とするデータ等の信頼性が乏しい。
- 論旨の明確さや論証の適切さがない。
- 論文の構成、表現（用語、引用、図表等）が適切でない（または整合性がとれていない）。

6. 判定

掲載についての判定は以下の4つの類型に分かれ、最終的に常任編集委員会において決定します。ただし、これらの判定は、評価できる項目や問題のある項目の多少によるものではありません。(3)及び(4)にあるように、投稿論文の種類以外であれば、掲載を考慮できるとする場合があります。別の種類となる場合、字数の関係で、大幅に修正を要することがあります。

- (1) 掲載可とする。
- (2) 小規模の修正による掲載可とする。
- (3) 大幅な修正による掲載可とする。
但し、(総説／研究論文／研究ノート／実践・調査報告)として掲載を考慮できる。
- (4) 掲載不可とする。
但し、(総説／研究論文／研究ノート／実践・調査報告)として掲載を考慮できる。

～投稿案内～

日本評価学会では、「日本評価研究」掲載のための投稿原稿を募集しております。随時、投稿を受け付けておりますので、ご興味をお持ちの方は投稿規定・執筆要領をご参照のうえ、奮ってご投稿ください。

投稿先：特定非営利活動法人日本評価学会 事務局 投稿窓口
〒108-0075 東京都港区港南1-6-41品川クリスタルスクエア12階
一般財団法人国際開発センター内
e-mail：jes.info@evaluationsjp.org

『日本評価研究』第14巻第2号
2014年12月26日

編集・発行 特定非営利活動法人 日本評価学会
〒108-0075 東京都港区港南1-6-41
品川クリスタルスクエア12階
一般財団法人国際開発センター内
TEL: 03-6718-5931 FAX: 03-6718-1651

印 刷 株式会社 研恒社

©日本評価学会

本誌に掲載されたすべての内容は、日本評価学会の許可なく転載・複写はできません。

Japanese Journal of Evaluation Studies

Vol. 14, No. 2, December 2014

CONTENTS

Special Issue: Current state and future prospects of program evaluation

The Role of Program Evaluation as Complement to Performance Measurement:
The Case of GPRAMA in the United States

Satoko Tanabe

Present situation and its social conditions that inhibit evidenced-based practices and their
formative approaches from established in the Japanese society: Implications for improvement

Iwao Oshima

Article

Financial Analysts' and Evaluators' Efforts to Improve Ethics:
Comparative Study of Professional Ethics

Nobuyuki Kobayashi

Research note

Performance-informed Budgeting: Practices, and Information and Methodologies Utilized

Jun Arakawa

Report

MSC (Most Significant Change), Participatory Monitoring and Evaluation Tool:
Consideration to Four Characteristics of MSC Based on its Implementation in Bangladesh NGO

Hiroshi Tanaka