日本評価研究

The Japanese Journal of Evaluation Studies

Vol. 3, No. 2, September 2003

特集 評価手法の現状と課題

特集にあたって 佐々木 亮

Using Policy Evaluation to Support Decisionmaking and Improve Program Performance
Joseph S. Wholey (和文要約:田中 啓)

系統的レビューに基づく社会政策を目指して:キャンベル共同計画の取組み 津富 宏

わが国評価におけるより適切なプログラム・セオリーの構築を目指して

- 国際協力評価と政策評価に焦点を当てて - 三好 皓一 森田 智 藍澤 淑雄

実用重視評価の理論と課題 長尾 眞文

エンパワメント評価の特徴と適用の可能性

~ Fettermanによる「エンパワメント評価」の理論を中心に~源 由理子

ドラッカーによる成果重視の自己評価手法

~評価領域拡大への示唆~ 田中 弥生 伊永 大輔

戦略計画と一体となった評価

~戦略計画を政策評価にどう活かすか~ 龍 慶昭 佐々木 亮

研究論文

「市民参加型」二次評価の試み - JICAの「インドネシア国スラウェシ貧困対策支援村落 開発計画」プロジェクト評価を事例として - 森田 智 高松 香奈 三好 皓ー Incidences of Public Education Spending in Developing Countries Takako Yuki

研究ノート

The Tao of Evaluation Craig Russon

社会指標による政策効果測定に関する一考察

- 時系列指標を比較する方法 - 小野 達也

地方自治体における行政評価の実際と新しい評価者モデル

- エンパワメント型行政評価者 - 西出 順郎

住民意識の分析による水循環の施策と評価について 陸路 正昭

実践・調査報告

インドネシアにおけるジェンダー統計分野への協力の評価 浜野 敏子

日本評価学会

The Japan Evaluation Society

『日本評価研究』編集委員会

Editorial Board

編集委員長 三好 皓一(立命館アジア太平洋大学)

Editor-in-chief Koichi MIYOSHI

副委員長 西野 桂子 (グローバル・リンク・マネージメント)

Vice-Editors-in-chief Keiko NISHINO

古川 俊一(筑波大学) Shun'ichi FURUKAWA

常任編集委員 牟田 博光(東京工業大学) 山谷 清志(外務省)

Standing Editors Hiromitsu MUTA Kiyoshi YAMAYA

編集委員 石原 俊彦(関西学院大学) 大沢 真理(東京大学)

Editors Toshihiko ISHIHARA Mari OSAWA

賀来 公寛(東洋大学) 河野 善彦(国際協力銀行)

Kimihiro KAKU Yoshihiko KONO

佐々木 亮(国際開発センター) 田中 弥生(東京大学(2003.10~))

Ryo SASAKI Yayoi TANAKA

長尾 眞文(広島大学) 松岡 俊二(広島大学)
Masafumi NAGAO Shunji MATSUOKA

村松 安子(東京女子大学) 森 茂子(日本大学)

Yasuko MURAMATSU Sigeko MORI

事務局 〒135-0047 東京都江東区富岡2-9-11 京福ビル

Office 財団法人国際開発センター内

日本評価学会事務局

TEL: 03-3630-8031, FAX: 03-3630-8095

E-mail:yabuta.m@idcj.or.jp

日本評価研究

第3巻 第2号 2003年9月

目 次

[:	特集 評価手法の現状と課題] - 佐々木 亮
	特集にあたって
	Joseph S. Wholey(和文要約:田中 啓) Using Policy Evaluation to Support Decisionmaking and Improve Program Performance 6
	津富 宏 系統的レビューに基づく社会政策を目指して:キャンベル共同計画の取組み23
	三好 皓一 森田 智 藍澤 淑雄 わが国評価におけるより適切なプログラム・セオリーの構築を目指して - 国際協力評価と政策評価に焦点を当てて・40
	長尾 眞文 実用重視評価の理論と課題57
	源 由理子 エンパワメント評価の特徴と適用の可能性 ~ Fettermanによる「エンパワメント評価」の理論を中心に~70
	田中 弥生 伊永 大輔 ドラッカーによる成果重視の自己評価手法~評価領域拡大への示唆~87
	龍 慶昭 佐々木 亮 戦略計画と一体となった評価~戦略計画を政策評価にどう活かすか~108
矽	开究論文
	森田 智 高松 香奈 三好 皓一 「市民参加型」二次評価の試み - JICAの「インドネシア国スラウェシ貧困対策支援村落 開発計画」プロジェクト評価を事例として
	Takako Yuki Incidences of Public Education Spending in Developing Countries

研究.	ノート
-----	-----

	Craig Russon	
	The Tao of Evaluation	159
	小野 達也	
	社会指標による政策効果測定に関する一考察 - 時系列指標を比較する方法	169
	西出 順郎	
	地方自治体における行政評価の実際と新しい評価者モデル	
	- エンパワメント型行政評価者	179
	陸路 正昭	
	住民意識の分析による水循環の施策と評価について	188
	践・調査報告	
	浜野 敏子	
	インドネシアにおけるジェンダー統計分野への協力の評価	203
_	- ロ 人 バチャ	
	過去を表現しています。 はない はいまた はいまた はいまた はいまた はいまた はいまた はいまた はいま	040
	研修委員会	
	国際交流委員会	
	第4回全国大会のお知らせ	220
	★顿伍 兴 春也的	222
	本評価学会規約	
	本評価研究刊行規定	
	本評価研究投稿規定	
ㅁ	本評価研究執筆要領	231

【巻頭言】

特集:評価手法の現状と課題

特集にあたって

佐々木 亮 日本評価学会学会誌編集委員

評価手法について議論するとはいったいどういうことなのだろうか。手法はしょせん手法に過ぎないとして話を片付けることは簡単だろう。しかし、なぜアメリカでは繰り返し繰り返し評価手法が議論されて来たのであろうか。そして時代とともに新しい評価手法が提案されてまた新たな論争を生んで来たのであろうか。それは評価手法を論じることが、評価の目的や評価結果の利用の仕方を論じることと密接に関係しているからである。もっと率直に言えば、評価手法を論じるということは、すなわち評価の本質について議論するということでもあるということだ。

評価に関して標準的なテキストと見られている本のひとつに、アメリカの評価研究者のピーター・ロッシ(Peter H. Rossi)とハワード・フリーマン(Howard E Freeman)が書いた「評価:体系的アプローチ」(Evaluation: A Systematic Approach)があるが、その中で彼等は、「科学的評価」対「実践的評価」(Scientific Versus Pragmatic Evaluation Postures)に関する論争を紹介している。この論争は、評価手法を巡る論争であると同時に、そもそも評価とは何なのか、そして評価はどうあるべきかという根本的な問いに対する論争でもあった。もちろん決着がつくはずはない論争であるが、評価を考える際の出発点として、評価研究者の間で共有されるべき論争の一つであろう。若干長くなるが、以下に彼等の解説を、敢えてそのまま引用したい。

「たぶん、評価研究の世界で、もっとも影響の大きかった論文とは、キャンベル(Donald T. Campbell)が1969年に発表したものだろう。この論文は、キャンベルが何十年かにわたって主張してきた見方を示している。それは、政策や施策の決定は、社会状況を改善する方法をテストする継続的な社会的実験の結果に基づくべきである。それだけではなく、社会調査の技術は、「実験する社会」を実際に実現するため利用可能だと述べた。そしてキャンベルは、社会心理学において彼が学び、実際に適用した手法である実験モデルを、評価調査にも適用することを追求した。彼の後年の著作では、いくぶんその立場を和らげているが、キャンベルは、科学的調査のパラダイムに評価調査をあてはめようとした人物とみなすのがフェアであろう。

一方、キャンベルの立場は、評価のフィールドにおけるもう一人の巨人であるクロンバック (Lee J. Cronbach)によって挑戦されることとなった。調査の手続きと同じ考え方で科学的調査・評価は、使えることもあるかも知れないと断りながら、クロンバックは、評価の目的は、科学的調査の目的とははっきり違うと主張した。彼の見方では、評価は科学というよりもアート(=技法)であり、全ての評価は、意思決定者や利害関係者のニーズに合うように形作られるべきであるとする。それゆえ、科学的研究が基本的には調査のスタンダードを満たしているかどうかに奮闘する一方で、評価は、政治的環境や、施策の制約、そして利用可能な資源の枠のなかで、意思決定者に最大限に役立つ情報を提供することに貢献すべきであると主張した。』

つまりこの論争の本質は、評価は、有史以来人類が築き上げてきた科学の発展に貢献すべきなのか、それとも我々が生活する実社会の改善に奉仕すべきなのか、いったいどちらが評価のあるべき姿なのかという壮大な論争である。

そのほか評価手法に関しては様々な論争が行われてきた。一例をあげれば、定性的評価手法と定量的評価手法に関する論争、第三者評価と自己評価に関する論争、独立した評価と参加型評価に関する論争、そして最近活発になってきた業績測定とプログラム評価の対立あるいは相互補完関係に関する論争などがある。また、評価の視点を(広義の)結果(Results)や成果(Outcomes)に重きを置くべきという主張と、実施過程(Process)により重きを置くべきという主張の間の論争もある。さらに細分化された議論の一例としては、定量的評価手法のなかでも、実験モデルと計量経済学的手法のどちらが優れているかという論争がある。その一方、定性的評価手法のなかでも、認識したことをそのまま記載すべきという反応的評価(Responsive Evaluation)の立場と、事前に設定した問いに答えるべきという立場の間の論争などがある。その他、評価研究者は、改善効果の有無やその程度の評価にフォーカスすべきであり、かかった費用と実現した改善効果を貨幣価値に換算して比較する効率性評価までやるべきではないという主張がある一方、日々政策選択せねばならない行政官に資するためには効率性評価までやるべきだという主張もあり、やはりその間に論争が見られる。

これらは、一見すると手法に関する論争に見えるが、じつはそれぞれが、評価とは何なのか、そして評価は何を目的として行われるべきなのかという根本的な問いに対する論争であると言える。

この特集号では様々な評価手法が取り上げられて論じられている。この特集号で展開される議論は、定量的手法と定性的手法のどちらが優れているかというある意味で教条主義的な論争でもないし、第三者評価と自己評価のどちらが適切かという単純化された論争でもない。世界の評価研究の最近の議論を見れば、単一で最上の評価手法が存在するわけではなく、評価の目的やニーズによって、用いられるべき評価手法は選択されるべきだという論調にある。そして言うまでもなく、そのためには幅広い評価手法を知っている必要があるわけである。この特集号に掲載されたそれぞれの論文は、過去の論争を十分に咀嚼した上でそれぞれの評価手法について論じているとともに、その手法の効用や限界、そして日本における適用可能性について論じている。

この特集号は、評価理論の研究者にとっては、様々な手法を網羅した大きな「見取り図」としての意義があるだろう。一方、実務者にとっては、直面する課題に対処するための評価手法の「メニュー」として役立つであろう。この特集号が、日本における様々な評価手法の定着とそのあり方に関する議論に貢献することを期待している。

以下、簡単にそれぞれの論文の内容について紹介したい。

アメリカにおいて評価研究の議論をリードするジョー・ホーリィ (Joe S. Wholey) 氏が論文を寄せてくれた。それが「意思決定の支援とプログラムのパフォーマンスの改善のために利用される政策評価」であり、和文の解説を田中啓氏が執筆している。ホーリィ氏は、日本でも近年急速に注目を浴びるようになった「業績測定」(Performance Measurement) と行政における

「結果重視の管理」(Result-oriented Management)の発想と仕組みの洗練と定着を他の研究者とともに主導して来た第一人者である。ホーリィ氏は自身が主導した業績測定と結果重視の管理の実践に基づく経験から、結果重視の管理の導入に関する最近の課題を3点、そしてそれに伴う評価実施者の新たな役割を3点解説している。そして最後に、実績測定と評価の相互補完関係を指摘したあと、今後の展開に関する提言を述べている。また日本で制定された「政策評価法」に対してホーリィ氏がはじめてコメントを述べて議論している。

上記のホーリィ氏が提案する手法が実践的な評価の系統に属すると見ることができるとすれば、その次に掲載された津富宏氏寄稿の「系統的レビューに基づく社会政策を目指して:キャンベル共同計画の取組み」は、科学的な評価を志向する立場からの主張であると見ることができるであろう。実験モデルは実はいろいろな呼び方があり、実験デザインやランダム化比較実験などの呼称があるが、この手法を適用してなされた評価結果を収集して系統的レビューを加えてデータベース化する試みが、各分野で進められている。津富氏の論文では、先駆的な試みとして医療分野のコクラン共同計画がまず紹介されている。そしてその実績を踏まえて社会施策の各分野を網羅するキャンベル共同計画という名の試みが続けられていることを紹介して、その有用性や日本での展開の可能性について議論している。もちろんキャンベル共同計画の名前は、科学的評価の父と呼ばれる先述のドナルド・キャンベルから採られている。また、最近ではアメリカ教育省でも同様の系統的レビューの試みが開始されている。そして津富氏の論文では、こうした系統的レビューとデータベース化の試みが、最新かつ最善の「何が機能するのか」(What Works)の情報を提供する貴重な社会インフラとして、キャンベルが長年主張してきた、実験する社会、(Experimenting Society)の実現を支えることになるだろうと結んでいる。

三好皓一氏、森田智氏、藍澤淑雄氏からは、「わが国評価におけるより適切なプログラム・セオリーの構築を目指して - 国際協力評価と政策評価に焦点を当てて - 」と題する論文を寄稿していただいた。プログラム・セオリーは、別名ロジック・モデルとも呼称されるが、もともと施策の効果を評価する前に、施策が想定した構成要因の間の因果関係を明確化しておこうという意図から始まり、現在では独立した評価あるいは評価手法として認識され実際に適用されている。日本においてはプログラム・セオリーを明確化する作業が十分には行われて来なかったという認識から、日本のODA評価及び国内行政に関する政策評価の効果をあげていくために、より明確かつより適切なプログラム・セオリーの構築に注力することは重要かつ不可欠な課題であるという指摘が、本論文ではなされている。

長尾眞文氏には、「実用重視評価の理論と課題」という論文を寄稿していただいた。長い間続けられてきた定量的手法対定性的手法のパラダイム論争においては、科学性、客観性、一般化可能性に関してどちらが優れているかという点が議論の主題であったわけであるが、マイケル・パットン(Michael Q. Patton)は「評価の活用」(Utilization)という視点から両者の実践的な統合を提起したと本論文では解説されている。そして、パットンが提唱する実用重視評価の理論では、「評価の有用性は評価結果の活用で測られるべきである」という主張がなされていると解説している。最後に、実用重視評価の基本的な考え方の紹介や実践の例示および問題点の指摘をしたあと、日本における実用重視評価の活用に関する課題について論じている。また冒頭で、実用重視評価の誕生の背景となった定量的手法対定性的手法に関する論争が整理されている。評価理論を研究する者にとっては必読である。

4 佐々木 亮

上記の実用重視評価は、それまで評価実施の外部者とみなされてきたステークホルダー(利害関係者)を積極的に巻き込み、彼等が知りたいと欲する情報を評価によって提供することを主眼としていたが、それをさらに一歩進めたと見ることができるのがエンパワメント評価である。源由理子氏から寄稿していただいた「エンパワメント評価の特徴と適用の可能性~Fettermanによる『エンパワメント評価』の理論を中心に~」と題した論文では、その本質と具体的な適用例が解説されている。エンパワメントとは「力をつける」ということであり、デービッド・フェッターマン(David Fetterman)は、そのためには参加者のトレーニングによる評価の内在化、啓蒙、精神的解放などがキーワードとなると指摘していると解説されている。なお、エンパワメント評価の具体的な作業として(1)ミッションの確認、(2)現状の評価、(3)将来計画の策定、(4)進ちょく状況のモニターの4段階の作業がフェッターマンによって整理されていると同論文では解説されているが、これは経営学からのフィードバックであると見ることができる。エンパワメント(力をつける)とは、自らの活動(それは企業活動に限らない)を自らが経営できるようになるということに他ならないということであろう。これらを踏まえて源氏は、日本における適用の範囲と可能性について論じ、ODA分野、NPO活動、そして行政が行う協働事業で適用可能性が高いと指摘している。

経営学の大家ピーター・ドラッカー (Peter F. Drucker) が主張する成果を重視する自己評価手法について論じたのが、次に掲載された田中弥生氏と伊永大輔氏からの寄稿である「ドラッカーによる成果重視の自己評価手法~評価領域拡大への示唆~」である。評価に経営学的な視点を取り入れる傾向が近年強まっているが、同論文では、ドラッカーが主張する成果重視の自己評価手法と、ホーリィの主張とほぼ重なり合う成果重視のマネジメント/業績測定を比較して、その違いを指摘している。結論として、成果重視の自己評価手法は、仮説の検証に執着せず、環境の変化に反応して仮説そのものを変更していくことを特徴としていることを指摘し、評価及び評価研究が新たな領域へと突入していく可能性を示唆している。

評価に経営学的な視点を取り入れる立場からのもうひとつの評価のあり方を主張する論文が、
龍慶昭氏と佐々木亮の「戦略計画と一体になった評価~戦略計画を政策評価にどう活かすか~」である。アメリカの公共組織は、民間企業経営に用いられる戦略計画 (Strategic Planning)とその実践から多くを学んだ。ところが、アメリカでは業績測定と一対のものとして用いられることが多い戦略計画の導入が、日本ではあまり省みられて来なかった。そのことが逆に、日本において政策評価がスムーズに導入され定着することにつながったと言える一方で、一連の行政改革が目指した理念や目的を達成することに大きな貢献をして来なかったことにつながったと同論文は指摘している。そして、今後行われる政策評価法の見直し作業においては、事業評価、実績評価、総合評価の3つの手法の横並びを見直し、「戦略計画の策定と実績評価のセットでの実施」を全省庁で全省的に導入することを提案している。

以上の7論文において展開された議論や提言は本当にバラエティに富んでいるが、すべての論文に共通する特徴は、日本国内の議論よりもそれぞれがむしろ世界的な議論を踏まえて書かれたと言えることである。そしてこの前書きで解説したように、それぞれの論文及び主張は相互に関連している。この特集号がきっかけとなって、日本の評価研究者及び実務者が評価手法に関する世界の議論を吸収し、それが日本の土壌に則した議論の萌芽となり、そしてその芽が大きく育って日本における様々な評価手法の実践と定着に結びつくことを期待している。

注記

- 1 第6版から、新たにリプセイ(Mark Lipsay)が共著者として加わっている。
- 2 Campbell D. T. (1969). 'Reform as Experiments'. American Psychologist, 24-409. Cronbach L. J. (1982). Designing Evaluation of Educational and Social Programs. San Francisco, CA: Jossey-Bass, 1-2.

Rossi P.H., Freeman H.E. and Lipsay M.W. (1999). *Evaluation: A Systematic Approach* 6th edition. Newbury Park, CA: Sage publications, 29-30.

【研究論文:依頼原稿】

Using Policy Evaluation to Support Decisionmaking and Improve Program Performance

Joseph S. Wholey

University of Southern California wholey@usc.edu

Abstract

Evaluators can and should help policymakers and managers to overcome the challenges in use of performance information to improve policy decisionmaking and program performance. Results-oriented management is likely to increase the demand for and supply of evaluation studies.

Keywords

results-oriented management, managing for results, performance measurement, performance monitoring, use of evaluation information, comprehensive evaluation, evaluation studies, performance evaluation, policy evaluation, Government Policy Evaluations Act, Government Performance and Results Act

Using Policy Evaluation to Support Decisionmaking and Improve Program Performance¹

My interest is in the use of policy evaluation to improve the performance of public and non-governmental organizations. In this paper, policy evaluation (or evaluation) means ex post (retrospective) assessment of the performance of programs that have been implemented by public and non-governmental organizations. A program may be any policy, program, project, function, agency, bureau, or activity that has an identifiable purpose or set of objectives (see U. S. General Accounting Office, 1998). The primary focus of evaluation is on the program's outcomes (effects or results), though

evaluation may also assess program inputs, program activities, or program outputs (products and services delivered by the program). The purposes of evaluation are to strengthen accountability, to improve program performance, and to support resource allocation and other policy decisionmaking.

As Scriven has suggested, evaluation necessarily involves the selection of criteria and standards (see Scriven, 1980; Shadish, Cook, and Leviton, 1991). My questions have been: "Whose criteria...? Whose standards?" Different policymakers, managers, and other key stakeholders may have different views as to the meaning of program performance. From the perspectives of different stakeholders, program performance may focus on program inputs, on

program activities, on program outputs (products and services delivered), on clients and populations served, on service quality or efficiency, or on outcomes (results): intermediate outcomes, end outcomes, or unintended outcomes. To enhance the likelihood that evaluation will be used to improve policy decisionmaking and program performance, I am especially interested in planning and evaluation approaches that help develop key stakeholders' agreement on program goals and provide timely feedback on program performance in terms of criteria that key stakeholders consider relevant.

I have been impressed with the potential value of a simple and relatively inexpensive form of program evaluation: performance monitoring. Performance monitoring (or performance measurement or performance evaluation) is the regular measurement of process and outcomes in terms of performance indicators developed with the participation of managers, staff, and other key stakeholders (see Wholey and Hatry, 1992)2. Early work with stakeholders could clarify evaluation criteria. Performance monitoring systems could be tested and refined over time, and performance standards (goals or targets) could be established after baseline and trend data were available. Regular monitoring of program activities and outcomes could improve the relevance and usefulness of evaluation work. Timely monitoring and evaluation could strengthen accountability and encourage policymakers, managers, and other stakeholders to use performance information to improve policies and programs.

A new leadership and management approach – results-oriented management – now requires public and non-governmental organizations to develop outcome-oriented goals and to measure program outcomes on a regular basis. This paper examines results-oriented management and discusses roles that evaluators can play in overcoming challenges that arise at

each stage in results-oriented management efforts. It suggests that involvement of evaluators in results-oriented management is likely to be helpful in moving public and non-governmental organizations to a results orientation and thus improving the lives of those served by those organizations.

Managing for Results

To protect societal values and meet public needs, democratic societies develop complex arrays of laws, regulations, agencies, and programs. Throughout the world, however, agencies and programs often fall short of the performance needed to meet public needs and earn public support. Those within and outside public and non-governmental organizations too often see poor leadership and management, inefficiency, low-quality services, and ineffective performance. Perceptions of poor management and poor performance limit the resources made available and thus limit the contributions of public and non-governmental organizations.

As Drucker (1974) noted, government agencies tend to be misdirected because they are supported by budget allocations rather than being paid for results, and such agencies tend to fragment themselves by trying to please everyone. Many nonprofit organizations face similar problems because of their dependence on grants and contracts from multiple sources. Today global and domestic forces demand more effective management and higher levels of organizational performance, however. Leaders and managers in public and non-governmental organizations face demanding constituencies, higher public expectations, and aggressive media scrutiny. Public and non-governmental organizations must find ways to focus a reasonable fraction of their resources and energies in specific directions and get external support for movement in those directions.

Results-oriented management has emerged as a common element in current reform efforts. After three years of pilot projects, full-scale implementation of the Government Performance and Results Act has been under way in the United States government since 1997. With related efforts under way in other countries, at state and local levels, and in foundations and other funding agencies, pressures to demonstrate results are being felt throughout the public and not-for-profit sectors. Since 2002, the Government Policy Evaluations Act (GPEA) has required Japanese ministries to evaluate the effects of policies under their jurisdiction, publish evaluation reports, and report on the use of evaluation results. The meaning of accountability is changing as agencies are asked to take responsibility for outcomes that can only be achieved with the cooperation and collaboration of other organizations.

Results-oriented management (or performance-based management) is the purposeful use of resources and information in efforts to achieve and demonstrate measurable progress toward outcome-related agency and program goals. Vedung (1997) likens results-oriented management to management by objectives, which he sees as incorporating three features thought to constitute good management practice: setting clear goals that refer to results, involving managers and staff in decisionmaking, and frequently monitoring and evaluating the results.

Results-oriented management is accomplished through a three-step process, each of which typically requires a series of iterations: (1) developing a reasonable level of agreement among key stakeholders on missions, outcomeoriented goals, and strategies to achieve the goals; (2) monitoring and evaluating performance (in particular, outcomes achieved) on a regular basis; (3) using performance information in efforts to improve program effectiveness and

strengthen accountability to key stakeholders and the public. The first step is often accomplished through strategic planning that involves key stakeholders within and outside the organization and may result in agreements to provide greater managerial flexibility in return for accountability for results. All three steps provide opportunities to communicate the value of agency and program activities to stakeholders and the public. GAO's <u>Executive Guide</u> describes the three steps in results-oriented management, identifies practices that are important at each step, and illustrates effective implementation of those practices in public agencies (U.S General Accounting Office, 1996).

Results-oriented management systems are typically developed over a number of policy and management cycles as policymakers, managers, and their staffs develop and refine goals and strategies, identify factors likely to affect organizational performance and ways to leverage or mitigate the influence of such factors, implement performance monitoring systems, and use performance information. Goals, strategies, and monitoring systems may be revised to reflect changes in policies and resources, experience in implementing strategies, changes in priorities, new technologies, or changes in the availability of information on performance. When priorities change or results are unacceptable, new goals or strategies may be adopted and unproductive activities may be abandoned.

Results-oriented management aims to shift management's focus from inputs and process to results. Its purposes are to: improve service delivery and program effectiveness, strengthen accountability to key stakeholders and the public, support resource allocation and other policy decisionmaking, and improve public confidence and support. Regular monitoring of program outcomes and use of the resulting information are central features of results-oriented management systems. Shifting to a focus on outcomes,

however, confronts managers with problems whose resolution may require sophisticated evaluation approaches.

Challenges to Results-Oriented Management

Though results-oriented management approaches may be fine in theory, a host of institutional, organizational, and technical challenges must be overcome if results-oriented management is to achieve its promise. Resultsoriented management may be hampered by fragmentation of power and conflict over agency and program goals, legal and regulatory requirements, organization cultures and interorganizational capacities, overlapping information demands from key stakeholders, problems in measuring performance, lack of information on how to improve performance, and concerns over possible misuse of performance information.

Competing Values, Constitutional Protections, and Institutional Factors

Good management is important, but other things may be equally or more important. Efforts to move to a results orientation may come into conflict with national cultures or with values like freedom, decentralization, equity, and privacy. Efforts to manage for results may run into problems if they are seen as conflicting with values that oppose unfunded mandates or favor "equitable" or "needs-based" resource allocation. Value conflicts may be magnified by institutional and political conflicts.

In the United States, in particular, separation of powers makes results-oriented management difficult. Power is widely fragmented: between the Federal Government and state governments; among the legislature, the executive branch, and the courts; and between

public and private institutions (including interest groups and the media). Key stakeholders often differ over goals, over strategies for achieving goals, or over the importance of unintended consequences of agency and program activities.

In a book intended to refute the notion of government bureaucracy as a monolith, Wilson (1989) identified political factors that affect the activities of government executives, managers, and staff. In an admitted oversimplification, he identified four political environments in which an agency might be situated. Wilson suggested that an agency might be the product of either (1) client politics, where agencies like the Department of Veterans Affairs face a dominant interest group that favors agency goals; (2) entrepreneurial politics, where agencies like the Environmental Protection Agency and the Food and Drug Administration face a dominant interest group like the manufacturing or pharmaceutical industry that is hostile to agency goals; (3) interest group politics, where agencies like the Occupational Safety and Health Administration face two or more interest groups like corporations and unions that are in conflict over its goals; or (4) majoritarian politics, where the agency faces no important interest group; (pp. 76-83, 248-251). Wilson's categorization suggests that results-oriented management is likely to be more difficult when there is conflict over agency goals.

In addition, since most outcomes are influenced by the activities of more than a single agency, results-oriented management often requires interagency coordination, collaboration, and partnership efforts to achieve the intended results. Results-oriented management is especially challenging in decentralized systems.

Organizational Culture and Capacity

Most public and nonprofit organizations focus on delivery of products and services, and

focus less energy on results that lie beyond the control of agency staff. In both public and non-governmental organizations, executives and manages may be so bound by internal and external rules and constraints that they have little freedom to manage for results. Agency management typically is process-oriented, and agency cultures are often dominated by standard operating procedures. Furthermore, agency executives, managers, and staff may lack the skills or the time and other resources needed to get agreement on results-oriented goals and strategies, to measure and evaluate performance, or to use performance information effectively.

Technical Factors

It can be costly and difficult to get a reasonable level of agreement on outcomeoriented goals and strategies, and to get performance information that is sufficiently timely, complete, accurate, and consistent to document performance and support decision making. Resource constraints may make it difficult to obtain sufficiently accurate data on results. Further challenges will present themselves when results occur over extended periods of time.

Wilson (1989) calls particular attention to two factors affecting the cost and feasibility of managing for results: the extent to which agency goals can be precisely specified, and the extent to which progress toward agency goals can be reliably measured. Using these factors, Wilson categorizes agencies as either (1) production agencies like the Social Security Administration and the Postal Service, where staff activities and the results of those activities can be observed; (2) procedural agencies like the Occupational Safety and Health Administration and army units in peacetime, where staff activities can be observed but not the outcomes that result from those activities; (3) craft organization like detective

bureaus and army units in wartime, where staff activities cannot be observed but results are relatively easy to evaluate; or (4) coping organizations like universities and the diplomatic corps, where neither staff activities nor results are easily observed and both process-oriented management and results-oriented management are likely to be difficult. Wilson's categorization suggests that results-oriented management is more likely to be difficult outside production agencies that provide direct services. Radin (1998) has suggested that, "In many ways, the [results-oriented management] process is designed for agencies that actually deliver service...; have relatively stable histories that are amenable to a planning approach; have cultures of data production (with agreement on typologies and belief in the accuracy of the information); and have manageable levels of conflict between external actors (or stakeholders)" (p.309).

Meeting the Challenges in Results-Oriented Management: Roles for Evaluators

Public and non-governmental organizations use several approaches to overcome institutional and organizational challenges to results-oriented management: (1) using nonfinancial incentives including the time and attention of high-level officials to build senior management commitment to results-oriented goals and strategies, (2) using research and evaluation studies to assess the results of current and potential strategies, (3) negotiating with key stakeholders and if necessary revising goals and strategies to get sufficient levels of support, (4) delegating authority in return for accountability for results, (5) using nonfinancial and financial incentives to build staff commitment to results-oriented goals and strategies, (6) developing partnerships to achieve results beyond the control of any one agency, (7) redesigning procurement and grants

management systems to focus on performance, and (8) developing systems for using performance information to support resource allocation and decisionmaking. organizations use at least four other approaches to overcome technical challenges to resultsoriented management: (9) using intermediate goals and measures (outputs and outcomes) to show progress or contributions to intended results; (10) using qualitative goals and measures when appropriate; (11) disaggregating and reporting performance information in ways useful to management; (12) using regression models to take into account the influence of external factors on program outcomes.

Though some believe that evaluators' roles should be confined to assessment of program performance, evaluators can play important roles at each stage in the results-oriented management process.

Developing Agreement on Goals and Strategies

Performance data and evaluation studies can be important sources of information for policymakers and managers in their efforts to develop agreement on outcome-oriented goals and strategies to achieve the goals. Evaluators can assist policymakers and managers in identifying intended outcomes, identifying factors that could affect achievement of the goals, and developing strategies for achieving the goals; in particular, developing partnership efforts to leverage or mitigate the influence of external factors that are likely to affect results. Evaluators may use logic models - flow charts that identify program inputs, activities, outputs (products and services delivered), intermediate and end outcomes (results), and assumed causal linkages - to facilitate stakeholder agreement on goals, strategies, and performance measures.

Hatry's (1999) book on performance

measurement presents many examples of logic models used by public and non-governmental organizations. The Office of National Drug Control Policy brought together interagency working groups to help develop strategic- and program-level logic models, the National Drug Control Strategy, and action plans specifying the activities and programs to be used to achieve specific performance goals (see, Office of National Drug Control Policy, 1999, 2000). Many United Ways and United Way-funded agencies have used logic models and outcomefocused performance monitoring systems to clarify the purposes of funded programs, to help focus staff efforts on common goals, and to communicate the value of funded programs to potential funders and donors.

Logic models can help organizations to identify intermediate goals (outputs and intermediate outcomes) that are within the agency's span of influence and can reasonably be expected to lead to intended results. Agreement on such goals is a key ingredient in most resultsoriented management efforts. A disease prevention program's goals might be stated in terms of increases in amounts of vaccines distributed and increases in immunization rates, for example, since such increases are within the program's span of influence and can be expected to help reduce communicable diseases. Highway safety programs now reward states for increasing penalties for driving under the influence of alcohol and for reducing the percentage of alcohol-related fatalities, both of which can be expected to help reduce the total number of highway fatalities.

Since few outcomes can be achieved through the efforts of a single agency, it is likely that interest will intensify in development of results-oriented management systems for <u>crosscutting programs</u> (or <u>horizontal programs</u>) that reflect the contributions of multiple agencies to common goals. Evaluators can make significant

contributions in such complex management environments, when even the definition of performance constitutes a major hurdle.

Monitoring and Evaluating Performance

Performance monitoring systems play a central role in results-oriented management by providing regular feedback on the extent to which strategies have been implemented and the extent to which goals have been or are being achieved. Evaluators may assist in designing and redesigning performance monitoring systems; for example, by helping identify output and intermediate outcome measures to show progress or contributions to intended results. Evaluators may assist in developing sample surveys to get valid, reliable information on program outputs and outcomes.

Rarely will a single performance monitoring system meet the information needs of all key stakeholders. It will usually be necessary to develop hierarchies of performance measures to meet information needs at different levels. As the General Accounting Office stated after studying experience in leading agencies in the United States and abroad, "As the leading organizations we studied strive to align their activities and resources to achieve missionrelated goals, they also strive to establish clear hierarchies of performance goals and measures. Under these hierarchies, the organizations try to link the goals and performance measures for each organizational level to successive levels and ultimately to the organization's strategic goals....The experiences of leading state, foreign, and federal governments show that at least four characteristics are common to successful hierarchies of performance measures.... [Performance measures should] demonstrate results....The number of measures for each goal at each organizational level should be limited to

the vital few.... [Performance measures should] respond to multiple priorities....Performance measures should be linked directly to the offices that have responsibility for making programs work." Performance data should be sufficiently timely, complete, accurate, and consistent "to document performance and support decisionmaking at various organizational levels" (U. S. General Accounting Office, 1996, pp. 24-27).

To provide more valid measures of program performance, evaluators may suggest that agencies disaggregate performance data to reflect the degree of difficulty in achieving intended outcomes, as the Occupational Safety and Health Administration has done, or statistically adjust performance standards to reflect the influence of client characteristics, economic conditions, or other degree-of-difficulty factors on client outcomes as the Employment and Training Administration has done (see Barnow, 1992; U.S. General Accounting Office, 1999).

Agencies typically use numerical performance measures, but may also use peer review assessment of performance or some combination of quantitative and qualitative performance measures. The National Science Foundation (NSF), for example, uses assessments by external experts to monitor the agency's success in achieving annual performance goals relating to the effectiveness of NSF-supported research and education activities (National Science Foundation, 2000). The General Accounting Office uses both quantitative and qualitative performance goals and measures to assess its performance (U.S. General Accounting Office, 2003).

Evaluators can evaluate performance monitoring systems in order to assist agencies in validating performance data and improving performance monitoring systems. Evaluations of performance monitoring systems should focus both on the technical quality of the monitoring system and on the extent to which performance information is used in managing to achieve performance goals and in providing accountability to key stakeholders and the public (Wholey, 1999a, 1999b). External evaluation can be important in ensuring the accuracy and credibility of performance information intended for dissemination to policy levels and to the public.

Performance information may come from monitoring systems that regularly compare outputs and intermediate outcomes with performance targets, or from less frequent but more comprehensive evaluation studies. Evaluation studies may: measure program implementation, test assumptions connecting program activities to results, explain current performance levels, measure unintended outcomes, assess the cost-effectiveness of current strategies, estimate the causal impact or net benefits of program activities, or measure other hard-to-measure outcomes. Evaluation studies can provide a fuller and more accurate picture of program performance than the rough sketch obtainable through typical performance monitoring systems, and may include recommendations intended to improve performance.

As Weiss and Morrill (1998) noted, learning organizations use multiple feedback loops and many types of information to inform policy formulation, program improvement, and program redesign. Performance information may come from performance monitoring systems, audits, case studies, benchmarking comparisons, basic and applied research, evaluation studies, and experiments. Recent GAO publications explore interrelationships between performance monitoring systems and evaluation studies, and show how evaluation studies can supplement and improve the data produced by performance monitoring systems (U.S. General Accounting

Office, 1998, 1999, 2000a).

In the Upward Bound program, for example, the U.S. Department of Education used evaluation studies to learn about high school courses and grades, educational expectations, high school completion, and college enrollment of disadvantaged students during the summer and mentoring and tutoring over the school year in the 9th through 12th grades (U.S. General Accounting Office, 2000a). On the basis of the Upward Bound program's random assignment of students to experimental and control groups, the evaluation estimated the program's net impact in terms of college enrollment.

Using Performance Monitoring and Evaluation Information

Within an agency or program, policymakers and managers may use performance information in reallocating resources or redirecting activities to improve performance, may redesign personnel management and other central management systems to focus on organizational performance, and may use performance information to create financial and nonfinancial incentives for improved performance. The Veterans Health Administration and the U.S. Department of Transportation, for example, have used performance agreements to focus executives' attention on agency performance goals (U.S. General Accounting Office, 2000b; Hatry, Morley, Rossman, and Wholey, 2003). Agencies may also use performance information in developing partnerships with other public or private agencies to improve performance.

Evaluators can assist policymakers and managers in using performance monitoring and evaluation to improve service quality and program effectiveness. Evaluators can assist policymakers by conducting credible evaluation studies and by helping design institutions that stimulate and support the use of performance information in policy decisionmaking. Evaluators can assist managers in redesigning program management systems, budget systems, personnel management systems, procurement systems, and grants management systems to focus on results, and may assist in developing performance-oriented partnerships with other agencies that have common or complementary goals.

While performance monitoring systems limit themselves to assessing the extent of progress toward agency and program goals, evaluation studies typically identify opportunities to improve performance and may offer options or recommendations toward that end. In fiscal year 1999, for example, when its weekly performance reports revealed a sudden outbreak of Medflies in Mexico on its border with Guatemala border (and Medfly outbreaks occurred in California and Florida, the U.S. Department of Agriculture used an evaluation study to determine why program performance had declined and how their trapping and spraying programs could be improved to control and eradicate fruit flies and thus help protect agriculture in the United States (U.S. General Accounting Office, 2000a; Hatry, Morley, Rossman, and Wholey, 2003).

The U. S. Coast Guard has used performance monitoring systems and evaluation studies to improve its performance as well as its accountability to key stakeholders and the public. The Coast Guard set 5-year performance goals related to deaths, injuries, and environmental damage; for example, reduce accidental deaths and injuries from marine casualties by 20%, reduce the amount of oil and chemicals going into the water from marine sources by 20%. With the help of an ad hoc program evaluation group from throughout the agency, the Coast Guard then developed and refined systems for monitoring and reporting on performance; for

example, monitoring worker fatalities per 100,00 workers. The evaluation group disaggregated performance data to help program managers understand trends and risk. By disaggregating fatality data, for example, the Coast Guard found that commercial towing was even less safe than fishing. The Coast Guard then entered into a partnership with the towing industry to address the problem. A collaborative study showed that, "[M]ost of the casualties were deckhands, and most of them were falling overboard – a problem which did not lend itself to an off-the-shelf regulatory program for towboats" (U.S. Coast Guard, 1996. p. 8). As GAO noted, the Coast Guard and the towing industry developed training and guidelines to reduce the causes of fatalities. "This joint effort contributed to a significant decline in the towing industry fatality rate, from 91 per 100,000 employees in 1990 to 27 per 100,000 in 1995" (U.S. General Accounting Office, 1996, p. 37). The U.S. Department of Transportation has been doing similar studies to determine the causes of truck crashes that cost 5000 lives each year in the United States (Downey, 2001).

Evaluators can also assist agencies in using performance information externally, to strengthen accountability and support policy decisionmaking. Many public and nongovernmental organizations use logic models to communicate to external stakeholders, in graphic form, relevant elements of program design: key inputs, activities and processes, outputs and intermediate outcomes that reflect agency contributions, end outcomes, and causal linkages among inputs, activities, outputs, intermediate outcomes, and end outcomes. Use of logic models may strengthen agency budget requests or grant applications by helping clarify agency and program contributions to intended outcomes. In Canada, logic models have been used to demonstrate linkages among program inputs, activities, outputs, and outcomes (see Auditor General of Canada, 2000). United Way of America has produced a number of guides that can be used to introduce policymakers, managers, and staff to the use of logic models and outcome data (United Way of America, 1996a, 1996b, 1996c).

As is suggested by experiences in the Job Corps and the Special Supplement Food Program for Women, Infants, and Children (WIC) program, impact evaluations may strongly influence policy decisionmaking if the evaluations include cost-benefit analyses that compare program costs with cost savings and other economic benefits associated with the program's net effects. In the Reagan years, for example, an impact evaluation found that the Job Corps had a positive net effect on participants (employment service applicants who were high school dropouts) and estimated that the program's benefits to the society were greater than its costs (Mallar and others, 1980, 1982). The findings of the 1980 interim evaluation were used by Senator Hatch to help persuade his colleagues to maintain the Job Corps at a time when other employment and training programs were being eliminated or sharply reduced. Conservative Republicans, who are ordinarily skeptical of federal social programs, gave considerable weight to evaluation findings that demonstrated the effectiveness of the Job Corps and its value to society (Wholey, 1986a). The Job Corps, a program that had little in the way of a political constituency, now has an annual budget in the \$1 billion range.

Similarly, WIC program evaluations have been used in congressional decisionmaking for many years (Wholey, 1986b). The Special Supplemental Food Program for Women, Infants, and Children (WIC) provides vouchers for food supplements, and nutrition education and counseling, to low income pregnant women nursing mothers, infants, and children judged to be at nutritional risk. A number of evaluation

studies – some based on randomized experiments – have demonstrated that the WIC program significantly increases birthweight, significantly reduces the percentage of low-birth weight infants, and saves more money in other programs than the program costs (see, for example, U.S. General Accounting Office, 1992, which combined evaluation synthesis and benefit-cost comparisons). WIC's annual budget now exceeds \$4 billion.

Reinforcing Roles for Performance Monitoring and Evaluation

Given the pressures facing communities and nations across the globe and given the challenges that arise as public and nongovernmental organizations work to make effective use of performance monitoring and results-oriented management, there is likely to be increasing demand for those who understand qualitative and quantitative measurement and evaluation methods. Interest in performance monitoring and results-oriented management is likely to grow, building on the recent progress that has been made in many public and nongovernmental organizations. In this context, evaluators will be asked to help design and implement performance monitoring systems, to conduct evaluation studies to explain reasons for current performance levels and measure hard-tomeasure outcomes, and to assist in translating performance monitoring and evaluation information into actions to improve the performance and credibility of public and nongovernmental organizations.

Since few outcomes can be achieved through the efforts of a single agency, it is likely that interest will intensify in development of goals, strategies, and results-oriented management systems for cross-cutting programs (or horizontal programs) that reflect the contributions of multiple agencies to common

goals. Evaluators can make significant contributions in such complex management environments.

Effective implementation of performance monitoring and results-oriented management systems will typically require training: in strategic planning, in involvement of stakeholders in development of goals and performance measures, in outcome monitoring, in data analysis, in process and impact evaluation, and in use of performance monitoring and evaluation information. Much of the needed content for such training should come from case studies. cross-case analyses, evaluation studies, and applied research on efforts to use results-oriented management systems in public nongovernmental organizations. Evaluators should be well-equipped to provide such training and to perform such studies.

Performance monitoring systems and evaluation studies can and should be mutually reinforcing. Both logic and experiences in this country and abroad suggest that the demand for and the supply of evaluation studies are likely to increase after agencies begin to measure program outcomes on a regular basis. When policymakers and managers are regularly provided information on a program's outcomes, they are likely to want to know why the outcomes have occurred (what difference the program has made) and how performance can be improved. Such interest is likely to stimulate evaluation studies to answer the "why" and "how" questions.

In addition, evaluation studies should be more feasible, less costly, and more useful when performance criteria have been clarified and performance data have been collected. When agencies are held accountable for improved performance, executives and managers will be more likely to use evaluation information to improve program performance and to communicate the value of agency and program activities to policy levels, other stakeholders, and

the public, and policymakers will be more likely to use evaluation information in their own decisionmaking.

Progress, Problems, and Next Steps

A number of years will typically be needed to achieve effective implementation of results-oriented management systems. Improvements in program performance can take longer. Over the past several years, many public and nongovernmental organizations have made progress in results-oriented management but, even among leaders, progress has often been followed by setbacks and disappointments in the face of political, organizational, and technical challenges (see Aristigueta, 1999).

At the federal level, ten years after passage of the Government Performance and Results Act and six years into its government-wide implementation, federal agencies now have experience in strategic planning and performance monitoring. The General Accounting Office (GAO) concluded that, "New and valuable information on the plans, goals, and strategies of federal agencies has been provided....Issuance of agencies' performance reports each March now represents a new and potentially more substantive stage in the implementation of GPRA...." GAO identified several continuing implementation challenges, however; for example, articulating and reinforcing a results orientation (only half of federal managers perceived strong leadership commitment to achieving results); coordinating cross-cutting programs; and building capacity to gather and use performance information (U. S. General Accounting Office, 2001, p. 11).

A focus group of current and former senior government officials and public management experts concluded that tools such as the Government Performance and Results Act were focusing Congressional and Executive Branch attention on performance and results, but that

major management problems remained. The focus group noted that some agencies and programs had improved performance. Most programs had not developed and used outcomeoriented performance measures to track and report results, however, and collecting valid and reliable performance data remained a major challenge for most agencies (Reason Public Policy Institute, 2000). Similar groups convened by the National Academy of Public Administration suggested that performance-based government is important because it helps to make government more transparent, and because it provides a management framework through which leadership can implement its program agenda. They noted that the Government Performance and Results Act had increased accountability and documenting of goals and results, and had focused attention on crosscutting programs and the need to coordinate them (National Academy of Public Administration, 2001). Building on agency experience under the Government Performance and Results Act, the Bush Administration has initiated an ambitious effort to use performance monitoring and evaluation information to support budget development, policy development, and management reform (U. S. Office of Management and Budget, 2001, 2002, 2003a, 2003b, 2003c).

After several years of effort in the Canadian government, a report from the Auditor General concluded that, while the concept of managing for results was widely accepted and supported among senior managers and performance information was used in some programs, actual performance monitoring and use of performance information had made only limited progress. Concluding that managing for results was still not an integral part of managing federal departments and that the evaluation function had regressed, the Auditor General called upon the Treasury Board Secretariat to

assemble and communicate successful practices and suggested that it might be time for the government to consider introducing accountability legislation. The Auditor General found growing attention to horizontal issues that cross departmental boundaries, but found that government was still at the start of efforts to report and use performance information in managing such issues for results (Auditor General of Canada, 2000).

In the not-for-profit sector, a three-year evaluation of the use of program outcome information in seven leading United Way communities concluded that most funded agencies believed that implementation of outcome monitoring was helpful in focusing staff efforts on common goals, communicating results to stakeholders, and competing for resources. Program directors believed that implementation of outcome monitoring had a positive effect on service delivery and effectiveness. Most program directors indicated that they would recommend that directors of similar programs consider implementing outcome measurement. The limited scope and duration of the evaluation prevented the evaluators from providing information on whether clients were in fact benefiting from the programs' outcome monitoring efforts, however. Most program directors remained concerned about the cost of outcome monitoring, and indicated that staff turnover resulted in a continuing need for training (see United Way of America, 2000).

My assessment of evidence to date, from different levels of government and from non-governmental organizations, suggests that results-oriented management holds great promise but that its ultimate value is still uncertain. Progress has been made in implementing results-oriented management in many public and non-governmental organizations. Even the initial step of establishing outcome-oriented goals and strategies has improved accountability in many

agencies and programs. Performance information has been used to improve performance in specific agencies and programs. Many agency plans and reports document results and describe strategies to improve results, but many reveal difficulties in obtaining timely and valid data on program outcomes. As Greiner has suggested in discussing performance monitoring, research is needed on the feasibility, cost, and value of results-oriented management in various circumstances, including research on successful and unsuccessful efforts and "best practices" approaches (Greiner, 1996). We need studies that compare agencies and programs in which strategic planning and performance monitoring have been undertaken with similar agencies and programs that have not attempted to use these management. In various policy areas, we should undertake experiments in which policymakers invest resources in resultsoriented management and evaluators assess the implementation, outcomes, and impacts of those efforts.

Given the pressures facing public and nonprofit organizations and the potential value of results-oriented management, my view is that evaluators should be involved in efforts to move public and non-governmental organizations to a results orientation. My expectation is that evaluators' involvement in such efforts will pay off in better management, improved services, more effective programs, better lives for those served, better communities, and increased public confidence in institutions that all of us need in our increasingly urbanized societies.

Notes

1 An earlier version of this paper was presented at the Stauffer Symposium at Claremont Graduate University in February 2001 (see Wholey, 2003). It draws on earlier and more recent papers (Wholey, 1999a, 1999b, 2001, 2002, forthcoming).

2 In the Summary of Standard Guidelines for Policy Evaluation that guide implementation of the Government Policy Evaluations Act (GPEA), performance monitoring is referred to as performance evaluation.

References

- Aristigueta, M. P. (1999). *Managing for Results In State Government* (Westport, Conn.: Quorum).
- Auditor General of Canada (2000). "Managing Departments for Results and Managing Horizontal Issues for Results," *Report of the Auditor General of Canada*, Chapter 20 (Ottawa: Author).
- Barnow, B. S. (1992). "The Effects of Performance Standards in State and Local Programs: Lessons for the JOBS Program." In Manski, C. F., and Garfinkel (eds.), *Evaluation* of Welfare Training Programs (Cambridge: Harvard University Press), pp. 277-309.
- Downey, M. (2001). *Departmental Management: Experiences at DOT*, Presentation at the National Academy of Public Administration (Washington, D. C., February).
- Drucker, P. F. (1974). *Management: Tasks, Responsibilities, Practices* (New York: Harper and Row).
- Government Policy Evaluations Act
- [http://www.soumu.go.jp/English/kansatu/evaluation/evaluation_07.html]
- Greiner, J. M. (1996). "Positioning Performance Measurement for the Twenty-First Century." In Halachmi, A., and Bouckaert (Eds.), Organizational Performance and Measurement in the Public Sector (Westport, Conn. Quorum), 11-50.
- Hatry, H. P. (1999). *Performance Measurement: Getting Results* (Washington, D. C.: Urban Institute Press).
- Hatry, H. P., Morley, E., Rossman, S. B., and Wholey, J. S. (2003). *How Federal Programs*

- Use Outcome information: Opportunities for Federal Managers (Arlington, Va.: IBM Endowment for the Business of Government).
- Mallar, C., and others (1980). Evaluation of the Economic Impact of the Job Corps Program: Second Follow-up Report, Report prepared for the U. S. Department of Labor, Employment and Training Administration (Princeton, N. J.: Mathematica Policy Research).
- Mallar, C., and others (1982). Evaluation of the Economic Impact of the Job Corps Program: Third Follow-up Report, Report prepared for the U. S. Department of Labor, Employment and Training Administration (Princeton, N. J.: Mathematica Policy Research).
- National Academy of Public Administration, Center for Improving Government Performance (2001). *Transition Briefing Papers* (Washington, D. C.: Author).
- National Science Foundation (2000). *GPRA Performance Report: FY 1999* (Washington, D. C.: Author).
- Office of National Drug Control Policy (1999).

 Performance Measures of Effectiveness:

 Implementation and Findings (Washington, D.
 C.: Executive Office of the President).
- Office of National Drug Control Policy (2000). Performance Measures of Effectiveness: 2000 Report (Washington, D. C.: Executive Office of the President).
- Radin, B. A. (1998). "The Government Performance and Results Act: Hydra-headed Monster or Flexible Management Tool?" *Public Administration Review*, vol. 38, pp. 307-316.
- Reason Public Policy Institute (2000). *The Transition Dialogue Series* (Washington, D. C.: Author.).
- Scriven, M. (1980). *The Logic of Evaluation* (Inverness, Calif.: Edgepress).
- Shadish, W. R., Jr., Cook, T. C., and Leviton, L. C. (1991). *Foundations of Program Evaluation* (Newbury Park, Calif.: Sage).

- Summary of Standard Guidelines for Policy Evaluation
 - [http://www.soumu.go.jp/English/kansatu/eval uation/evaluation_04.html]
- United Way of America (1996a). Focusing on Program Outcomes: Summary Guide (Alexandria, Va.: Author).
- United Way of America (1996b). *Measuring Program Outcomes: A Guide for United Ways* (Alexandria, Va.: Author).
- United Way of America (1996c). *Measuring Program Outcomes: A Practical Approach* (Alexandria, Va.: Author).
- United Way of America (2000). Agency Experiences with Outcome Measurement: Survey Findings (Alexandria, Va.: Author).
- U. S. Coast Guard (1996). Using Outcome Information to Redirect Programs (Washington, D. C.: American Society for Public Administration).
- U. S. General Accounting Office (1992). Early Intervention: Federal Investments Like WIC Can Produce Savings (Washington, D. C.: Author).
- U. S. General Accounting Office (1996). Executive Guide: Effectively Implementing the Government Performance and Results Act (Washington, D. C.: Author).
- U. S. General Accounting Office (1998).
 Performance Measurement and Evaluation:
 Definitions and Relationships (Washington, D. C.: Author).
- U. S. General Accounting Office (1999).

 Managing for Results: Measuring Results That

 Are Under Limited Federal Control

 (Washington, D. C.: Author).
- U. S. General Accounting Office (2000a). Program Evaluation: Studies Helped Agencies Measure or Explain Performance (Washington, D. C.: Author).
- U. S. General Accounting Office (2000b).

 Managing for Results: Emerging Benefits from
 Selected Agencies' Use of Performance

- Agreements (Washington, D. C.: Author).
- U. S. General Accounting Office (2001). Major
 Management Challenges and Program Risks:
 A Governmentwide Perspective (Washington,
 D. C.: Author).
- U. S. General Accounting Office (2003). Performance and Accountability Report: Fiscal 2002 (Washington, D. C.: Author).
- U. S. Office of Management and Budget (2001).President's Management Agenda (Washington,D. C.: Executive Office of the President).
- U. S. Office of Management and Budget (2002). Preparation and Submission of Strategic Plans, Annual Performance Plans, and Annual Performance Reports, Circular A-11, Part 6 (Washington, D. C.: Executive Office of the President).
- U. S. Office of Management and Budget (2003a). Budget of the United States Government, Fiscal Year 2004 (Washington, D. C.: Executive Office of the President).
- U. S. Office of Management and Budget (2003b). Analytical Perspectives, Budget of the United States Government, Fiscal Year 2004 (Washington, D. C.: Executive Office of the President).
- U. S. Office of Management and Budget (2003c). Performance and Management Assessments, Budget of the United States Government, Fiscal Year 2004 (Washington, D. C.: Executive Office of the President).
- Vedung, E. (1997). Public Policy and Program Evaluation (New Brunswick, N. J.: Transaction).
- Weiss, H. B., and Morrill, W. A. (1998). "Useful Learning for Public Action." Paper presented at the American Society for Public Administration national conference (Seattle, April).
- Wholey, J. S. (1986a). "The Job Corps: Congressional Use of Evaluation Findings." In Wholey, J. S., Abramson, M. A., and Bellavita, C., *Performance and Credibility* (Lexington,

- Mass.: D. C. Heath), pp. 245-255.
- Wholey, J. S. (1986b). "WIC: Positive Outcomes for a Demonstrably Effective Program." In Wholey, J. S., Abramson, M. A., and Bellavita, C., *Performance and Credibility* (Lexington, Mass.: D. C. Heath), pp. 271-283.
- Wholey, J. S. (1998). "Assessing Performance Measurement Systems," *The Public Manager*, vol. 27, no. 3), p. 23.
- Wholey, J. S. (1999a). "Performance-Based Management; Responding to the Challenges," *Public Productivity and Management Review*, vol. 22, no. 3, pp. 288-307.
- Wholey, J. S. (1999b). "Quality Control: Assessing the Accuracy and Usefulness of Performance Measurement Systems." In Hatry, H. P., *Performance Measurement: Getting Results* (Washington, D. C.: Urban Institute Press, 1999), pp. 217-237.
- Wholey, J. S. (2001). "Managing for Results: Roles for Evaluators in a New Management Era," *American Journal of Evaluation*, vol. 22 (2001), pp. 343-347.
- Wholey, J. S. (2002). "Making Results Count in Public and Nonprofit Organizations." In Newcomer, K., Jennings, E. T., Jr., Broom, C., and Lomax, A. (Eds.), Meeting the Challenges in Performance-Oriented Government (Washington, D. C.: American Society for Public Administration), pp. 13-36.
- Wholey, J. S. (2003). "Improving Performance and Accountability: Responding to Emerging Management Challenges." In Donaldson, S. I., and Scriven, M., *Evaluating Social Programs* and Problems (Mahwah, N. J.: Erlbaum), pp. 43-61.
- Wholey, J. S. (forthcoming). "Using Evaluation to Improve Performance and Support Policy Decisionmaking. In Alkin, M. C., *Evaluation Roots* (Thousand Oaks, Calif.: Sage).
- Wholey, J. S., and Hatry, H. P. (1992). "The Case for Performance Monitoring," *Public Administration Review*, vol. 52, no. 6, pp. 604-

610.

Wilson, J. Q. (1989), Bureaucracy: What Government Agencies Do and Why They Do It (New York: Basic Books).

(2003.8.19受理)

(1 te w 1 of R. Busic Books).

意志決定の支援とプログラムのパフォーマンスの改善のために利用される政策評価 (要約)

ジョー・ホーリィ(Joseph S. Wholey)は米国におけるプログラム評価と業績測定 (Performance Measurement)の権威の一人であり、永年にわたりその発展と普及に指導的役割を果たしてきた。数々の連邦政府機関や公的機関において評価の実務に携わった経験から、政策の立案者や実施者にとって有用な評価のあり方を追求してきたのがホーリィの一貫した姿勢である。特に、近年日本でも注目を集めている業績測定の発展において、アーバン・インスティテュート(The Urban Institute)のハリー・ハトリー(Harry P. Hatry)とホーリィの大きな貢献があったことはよく知られている。今回、そのホーリィから寄稿を得たことは、日本の評価研究者のみならず評価に関心のある実務家にとっても朗報であろう。

ホーリィの論文は平易かつ明快であり、敢えて説明の必要もないように思われるが、日本の読者の便宜のために、その要点を整理した上で日本にとって参考となる点を指摘したい。

ホーリィ論文の主題は公的機関(行政機関や非政府機関など)における成果志向の管理 (results-oriented management)と、それにおける評価者(evaluators)の役割である。

成果志向の管理とは、ホーリィによれば、現在多くの国が取り組んでいる公共部門の改革に 共通するアプローチである。成果志向の管理には、(1)目標や戦略に関する合意形成、(2)定 期的な業績の監視と評価、(3)業績情報の利用、という3つのステップが含まれる。

成果志向の管理は、その名が示すとおり、公的機関の経営を従来の資源投入や手続きを重視するやり方から、活動の結果を重視する手法に変えることを目指すものである。その実践に当たっては、価値観、法制度、組織文化、技術要因などさまざまな面で困難や課題に直面することになる。

このような困難を克服する上で、評価者の役割は単に「評価する」ことのみに留まらない、というのがホーリィの主張である。例えば、前述の成果志向の管理における3つののステップのうち、(1)目標や戦略に関する合意形成においては、評価者はロジック・モデル(logic model)を用いて、施策の目標や目標達成の手段について当事者間の合意形成を支援する役割があるという。2番目の(2)業績の監視と評価では、評価者は業績監視システム(performance monitoring system)の設計・改訂や業績情報を収集するための調査の企画における貢献が期待される。最後の(3)業績情報の利用については、評価者は政策立案者や実施者が業績情報を利用して施策の有効性を改善することを支援したり、組織内のさまざまなシステム(施策管理システム、予算システム、人事管理システム、等)を変更したり、他の機関と業績志向のパートナーシップ(performance-oriented partnership)を締結するのをサポートしたりすることが可能だとしている。

ホーリィは、成果志向の管理への関心は今後も高まっていくと予想しており、それに伴って 評価者に対する需要とその役割もさらに拡大していくものとみている。このような環境の下で、 評価者には(政策立案者や実施者に対して)成果志向の管理に必要なスキルの訓練を提供する 能力を備えることと、実際に成果志向の管理を用いた取り組みについての評価・研究を進める ことが必要だと主張している。

成果志向の管理に対する期待は大きいものの、米国においてさえまだ十分に機能しているとは言い難く、その有用性について将来的にどのような評価が下されるかは明らかではない。重要なことは、成果志向の管理について引き続きその実行可能性、実施コスト、有用性等についての研究を進めることであるとホーリィは述べている。

最後に結論として、ホーリィは成果志向の管理に評価者が参画することの必要性を強調した上で、そのような取り組みが最終的には公的機関の業績向上や市民の生活向上など望ましい変化につながるだろうという期待を示している。

以上のホーリィの主張は、そのまま日本の現状に当てはめてもほとんど違和感はない。呼び方はさまざまであるが、国、地方を問わず国内の多くの行政機関が取り組んでいる改革は、ホーリィが指摘する成果志向の管理の特徴を備えているし、それらの取り組みの中で、評価の専門家は現実的には評価に留まらない幅広い役割を既に担っている。

成果志向の管理において評価者が重要な役割を果たすというホーリィの主張は、評価の専門家にとっては心強いものであるとともに、日本の公共部門に対して大きな課題を突きつけているとも言えよう。評価者はプログラム評価(program evaluation)業績測定、戦略計画(strategic planning)などさまざまな手法に通じている必要があることをホーリィは示唆しているが、日本ではこのような人材がまだ少ないことから、公共部門としてこのような専門家を養成していく方法を検討する必要がある。またホーリィが指摘するように、成果志向の管理の有用性はまだ明らかにはされていないことから、評価に関わる研究者は成果志向の管理の有用性やそれが有効に機能する条件などについて研究を進めていくことが必要であろう。

(富士通総研 経済研究所 主任研究員 田中 啓)

【研究ノート:依頼原稿】

系統的レビューに基づく社会政策を目指して: キャンベル共同計画の取組み

津富宏

静岡県立大学 tsutomi@u-shizuoka-ken.ac.jp

要約

系統的レビューは、「リサーチクエッションの確定 文献の探索 文献のスクリーニング 文献のコーディング 統計分析 定型的な報告」という、一定の科学的手続きに従うレビューである。医学分野においては、この系統的レビューに基づく「エビデンス」を生み出し、伝え、絶えず更新しようという国際ネットワークである、コクラン共同計画が発展した。コクラン共同計画の成功に学び、社会政策決定への情報提供を行うことを目指して、キャンベル共同計画が設立された。キャンベル共同計画の生み出したエビデンスはデータベース化されて、電子媒体によって配布され、政策決定、サービス供給の両面にわたって、利用されることが期待される。キャンベル共同計画は、その時点における最善の「What works」を提供する社会インフラとして、Experimenting Societyを支えることになろう。

キーワード

系統的レビュー、エビデンス、キャンベル共同計画、社会政策、

1. 犯罪者処遇の評価研究

筆者は、長らく法務省で非行少年の処遇に 携わってきた。そこから、非行少年を更生さ せるための方法としてもっとも有効な方法は なんだろうか、すなわち、犯罪者・非行少年 の更生に関するエビデンスはなんだろうかと いう関心が芽生えた'。

犯罪者の処遇については、犯罪者処遇効果研究という評価研究の一分野が存在する。インプット(介入)が「処遇」、アウトカムが「再犯率」である一連の研究である。評価研

究の代表的なテキストブックである、Rossi and Freeman²のEvaluationの早期の版³が、刑事政策の効果研究を事例として多く用いていることに表れているように、犯罪者処遇効果研究を含む刑事政策の効果研究は、評価研究をリードしてきた分野の一つである。現在、刑事政策の効果研究は、英語文献を中心として、RCT(無作為化割付実験)が2000件超あると言われている⁴。

筆者は1980年代後半から上記の関心に基づ き、RCTを中心として犯罪者処遇効果研究を 追っていたが、文献が増えるにつれ、困難な 24 津富 宏

問題が生じてきた。それは、その捕捉が困難になるばかりでなく、特定の処遇方法(例えば、カウンセリング)に関する、複数の処遇効果研究が、必ずしも同じ「結果」(例えば、「処遇が有効であること」)を見出すいることと」)を見出すであることと」)を見出す研究もあれば、有意な増加を見出す研究もあれば、有意な増加を見出すいるといいの事態に頻繁に出するという事態に頻繁に出するという事態に頻繁に出するという事態に頻繁を見出すが多々存在するといるのであるという事態に頻繁に出するといるのであるという事態に頻繁を見出すがあるという事態に頻繁に出するといるのであるという。

このような際、どのような場合に、効果が 「ある」(あるいは「ない」、さらには「逆効 果である」)と言い切ってよいかははっきり しない。例えば、ある処遇方法に関する10本 の研究中、5本が有意な減少を見出し、3本は 有意な変化を見出さず、2本が有意な増加を 見出すという場合について、どのように判断 をしたらよいのだろうか。そもそも、私が文 献を網羅的に把握しているという確証もな い。ところが、レビューを見ると、その道 の「専門家」であるはずのレビューワが、自 己の主張に合った研究成果だけを引用してい るということすら少なくない。これでは、良 質の処遇効果研究が発表されたとしても、レ ビューワの主観によって、総体的な結論がゆ がめられてしまう。

このような悩みに答えを与えたのが、メタ・アナリシスである。個別研究の効果値を統合する統計手法は長い歴史を持つが、メタ・アナリシスの語はGlass (1976)がアメリカ教育研究学会の会長講演で始めて用い、以降定着したものである。1980年代からメタ・アナリシスは犯罪者処遇効果研究の統合にタ・アナリシスが画期的な影響力を持つに至った、その理由は、従前のメタ・アナリシスに比べ規模が大きかっただけではなく、犯罪者処遇の分野で長らく唱えられてきた犯罪者処遇不信論(Martinson 1974)を明確に否定す

るエビデンスとして提出されたからである。

筆者は、このメタ・アナリシスに出会い、個別の処遇効果研究レベルのエビデンスをいくら読んでも(私自身の)主観によるバイアスを乗り越えることはできないのだから、特定の処遇方法が有効かどうかの判断はメタ・レベルのエビデンスを提供するメタ・アナリシスによらなければならないと考えるようになった。

しかしながら、Andrews et. al (1990)の ように、学術論文として発表されたメタ・ア ナリシスには次の3つの決定的欠点がある。 一つは、個別の学術論文として発表されたメ タ・アナリシスについては、十分な品質保証 がなされていないことである。学術論文とし て発表されたメタ・アナリシスの手法の洗練 の度合は、(論文の査読制度を考慮しても) 本質的には当該論文の著者に委ねられてお り、学界としての統一的な見解を踏まえたも のとは言えない。 また、メタ・アナリシスの 対象とする一次研究(個々の評価研究)の渉 猟に当たっても、個別の研究者が行える努力 には限度があり、そのトピックに関心を持つ 研究者が協力し合って文献を把握した方が漏 れが少ない。この欠点は、研究者のネットワ ークにより、メタ・アナリシスの品質管理を 行うシステムを要請する。

さらに一つは、一次研究の発表に応じて更

新されないことである。つまり、ある時点において、メタ・アナリシスが、もっとも統合的な知見を提供していても、新しい一次研究を踏まえなければ、その内容は陳腐化した。 まう。しかし、学術論文は、いったん発ない。な訂は、審査を経て文章を確定し公表という形態をとる学術雑誌にはなじまず、個人の研究者には負担が重すぎるといえよう。つまり、この欠点もまた、ITの利用と、研究者のおり、この欠点もまた、ITの利用と、研究者のらい、

これら3つの欠点を乗り越えるものとして、 私は、まず、系統的レビューに出会うことに なった。 系統的 レビューとの出会いは、 Evidence-based medicineに始まった。

2. Evidence-based medicine (EBM)

Evidence-based medicine (エビデンスに基づいた医療)は、1991年にカナダのマックマスター大学のGuyattが始めて唱えた(林2000)、医療行為を、科学的エビデンスに基づいて行おうというムーブメントである。EBMの中核的な主張の一つは、エビデンスには質の差があるというものである(表1)。

表1に見るように、EBMでは最上位のエビ デンスは、系統的レビュー(systematic review)である。系統的レビューは、「リサ ーチクエッションの確定 それにかかる文献 の探索 文献のスクリーニング 文献のコー ディング (通常メタ・アナリシスによる) 統計分析 定型的な報告」という一連の、個 別研究の手続きと同様の、一定の科学手続き に従うレビューである。EBMで、「エビデン スに根ざした (Evidence-based)」と言うとき の「エビデンス」とは、通常、この系統的レ ビューによって生み出されたエビデンスを含 意している。なお、系統的レビューの語は、 EBMによって医学を中心に広く使われるよ うになったが、社会科学においては、研究統 合 (research synthesis) という呼び方も有力

である (Cooper and Hedges 1994)。以下、系統的レビューについて簡単に述べる。

系統的レビュー

個別研究においては、サンプリング、概念を反映した尺度構成、測定、そして適切な統計手法の適用といった厳密な手法が適用が求められる一方、これらの個別研究の概観であるレビューにおいては、どのような手法がふさわしいかという議論すら行われてこなかったことに対する反省にレビューの手法が、系統的レビューである。実際、従来のレビュー(非系統的レビュー)は、こうした一定の手続きを踏んでおらず、その結果、良質の個別研究をレビューした場合であっても、無価値な結論を下してきたのである。系統的レビューは、以下のステップから成る。

1) リサーチクエッションの明確化

個々の研究と同様、系統的レビューはリサーチクエッション(仮説)の明確化から始まる。つまり、系統的レビューは、個々の研究(一次研究)をサンプルとし、その属性をデータとして用いる、仮説検証である。であるからこそ、系統的レビューの知見は、「エビデンス」と呼ぶことができる。非系統的レビューは、仮説すら明確にせず、研究の出発点から誤っていることが少なくない。

2) 文献の探索

系統的レビューにおいては、個々の研究における分析単位(unit of analysis)に相当するのは、一次研究である。確定したリサーチクエッションを吟味している、すべての一次研究を、文献が公表されているか否かに関係なく、データベースのチェック、専門家からの情報収集、ハンドサーチなどあらゆる方法を用いて探し出す。

3) 文献のスクリーニング

あるリサーチクエッションを吟味している一次研究を徹底的に探し出したら、その一次研究の用いている研究方法の質を踏まえて、文献のスクリーニングを行う。通常は、RCTのみを残すことになるが、特定のリサーチクエッションについてRCTが十分な数がない場合には、より緩和した基準を用いることもある。このステップは、一次研究におけるサンプルの確定に当たる。

4) 文献のコーディング

個別研究における、データの収集・測定に当たる。コード票にしたがって、一次研究の属性を記

26 津富 宏

録し、計量的な操作の対象となるデータセットを 作成する。個人を対象にした社会調査で言えば質 問項目にあたる、コーディングの項目を十分に吟 味される必要がある。コーディングは、通常2人 以上のコーダーが行い、その一致度が検討される。 5) 統計分析

作成されたデータセットについて、できうる限り量的な分析を行うこととなるが、その中核的な手法がメタ・アナリシスである。系統的レビューとは、上記の過程を通じて、ある特定の仮説を吟味した一次研究について収集された、良質の偏らないデータを用いて行われる、メタ・アナリシスであるといってもよい。

6) 定型的な報告

個々の実証研究が、通常、「背景 仮説 サンプル データの測定 統計分析 結論 考察」といった一連のフォーマットで報告されるように、系統的レビューも「問題の所在 仮説 文献収集の方法 文献選択の基準 コーディングの手法統計分析 結論 考察」といった定型的な様式で報告される。

以上、見てきたように、系統的レビューは、 実証研究の手法によるレビューであり、従来 の非系統的レビューとはまったく異なってい る。系統的レビューの登場により、個々の一 次研究を一覧して、「専門家」としての見識 をもって、エビデンスを「解釈」するという 無意味な行為が姿を消すこととなったのであ る。

EBMは、このような系統的レビューのもたらしたエビデンスに基づいて医療行為を行うべきであると主張するムーブメントであるが、一般臨床医が行っている実際の医療行為は必ずしもエビデンスに基づいているわけではない。そこで、次に、EBMが取り組んだのは、良質のエビデンスを生み出して維持し、そのエビデンスを一般臨床医が使えるようにするためのシステム作りである。

しかしながら、すでに指摘したように、学術雑誌はこのようなシステムのインフラとしては適していない。品質管理に限界があり、利用者が限定され、更新が難しいからである。こうして、EBMを推進する新たな仕組みと

して、ITを媒体として活用する、研究者の国際ネットワークが提案された。これがコクラン共同計画である。

3. コクラン共同計画

疫学者故Archie Cochraneの名を冠する、 「コクラン共同計画」は、客観性を担保した 系統的レビューを作成し、これを電子情報と して提供する国際プロジェクトである。1993 年に誕生したコクラン共同計画は、時代的要 請にこたえることにより、まさにEBMを支 える中核的インフラとなった。コクラン共 同計画の主体は、この系統的レビューを生み 出すために、領域(例えば、皮膚疾患)ごと に結成されている、レビュー・グループであ る。レビューグループは、ある領域に関心の ある研究者、治療者、医療の受け手などが集 まって、その領域における系統的レビューを 推進する役目を担っている。現在、レビュー グループはおよそ50あり、すでに総計2000近 くの系統的レビュー(コクラン・レビュー) が生み出されている。

コクラン共同計画は、学術論文として発表されるメタ・アナリシス(ないし系統的レビュー)のもつ、3つの欠点を以下のように乗り越えている。

まず、系統的レビューの品質管理がなされている。コクラン共同計画には、方法論グループの方法論があり、このグループが系統的レビューの方法論について研究を行っており、このグループが開発したガイドラインに従って、国際的ネットワークであるコクラン共同計画は、クラン・レビューが作成される。また、国際的ネットワークであるコクラン共同計画は、分別のデータベースを維持し、良質の系統的レビューの産出を援助している。

ついで、利用可能性が担保されている。生み出されたエビデンスは一般臨床医が利用可能でなければならないが、コクラン共同計画は、コクラン・レビューをインターネットあ

るいはCD-ROMを通じて提供することでこれを可能としている¹⁰。原則は有償だが、レビューの要約はインターネットを通じて無償で閲覧することができるし、国によってはオーストラリアのように政府機関が団体契約することで全文を無償で利用できるところもある¹¹。また、コクラン・レビューは、学術論文と異なり、定型的な読みやすいフォーマットで書かれている。

さらに、絶えず更新されている。コクラン 共同計画は、レビューワがコクラン・レビューを行いたいと申し出た時点で、2年ごとに アップデートする旨の意思表明を行うことに なっており、最新の一次研究の成果や方 なっており、最新の一次研究の成果や方 なっており、最新の一次研究の成果や方 なっており、最新の一次研究の成果や方 なっており、最新の一次研究の成果や方 は は いる。しかも、コクラン・レビューは公開 されているため、利用者から更新の必要性に ついて指摘を受けることができる。また、電 子媒体上の情報の更新は、紙媒体の学術論文 と異なり、技術的に容易である。

コクラン共同計画は、研究者がIT環境を通じて国際的に協働することで、エビデンスに基づく医療を実現するために、現在も拡大を続けているが、コクラン共同計画に刺激を受け、同様の成果を社会政策の分野にもたらすために設立されたのが、キャンベル共同計画である。

4. キャンベル共同計画

キャンベル共同計画¹³は、社会政策、教育政策などの分野において、「何が有効か」(What works)に関する「系統的レビュー」をつくり、更新し、伝えることによって、エビデンスに基づいた意思決定を援助することを目的とする国際的組織である。キャンベルの名は、社会は、絶えざる社会実験に基づき、社会政策やプログラムが行われることにより進歩していくべきであるという、実験する社会(Experimenting Society)のコンセプトを主張した、評価研究の祖でもある、心理学者Donald T. Campbellに由来する。

キャンベル共同計画は、1999年7月にロンドンで準備会合が開かれ、2000年2月のペンシルベニア大学における会合で正式に発足した。現在は、運営委員会のもとに、教育部会、刑事司法部会14、社会福祉部会の3つの部会があり、それぞれの分野において重複のないようレビュー・トピックを立てて、レビューワやレビューのレフリーなどを確保して、系統的レビューの産出を担っている。なお、筆者は、刑事司法部会の運営委員の一人として2001年5月の運営委員会会合から参加をしている。

これらの3部会に対して、系統的レビューのガイドラインを提供しているのが方法論部会であり、すべてのレビュー・プロトコル(レビューの計画書に相当するもの)は、この部会の審査を経て始めて認められることになっている。方法論部会は、その中に、統計グループ、擬似実験グループ、実施過程グループ、情報収集グループ、研修グループの5グループを持ち、積極的に活動を展開している。

さらに、これらのグループが生み出した、系統的レビューの成果の伝達を担当しているのがコミュニケーション・広報部会であり、ホームページの運用など、キャンベル共同計画の鍵を握るITの活用を担っている¹⁵。キャンベル共同計画は、それが提供する情報の質を保障しつつ、容易にアクセスできるようにしなければならないが、その過程で生じる、口ゴの使用や著作権・翻訳権などに関する技術的な問題もこの部会が扱っている。

これらの部会のほか、図1にあるように、ペンシルベニア大学に、共同計画全体の事務局が置かれており、全体の調整を行っている。キャンベル共同計画は自発的な個人のネットワークであるが、法律上は、ペンシルベニア州に登録した非営利団体であり、事務局は非営利団体の本部として収支報告を行っている。

また、これらの活動を地域ベースで支援するものとして、「地域センター」がある。2002年の11月には、デンマーク政府の助成を

28 津富 宏

得て、北欧を対象とする地域センターである、 ノルディック・キャンベル・センターが誕生 した¹⁶。同センターは、レビューワ・グループを立ち上げるための金銭的な支援、レビューワ・グルーー・プロトコルやメタ・アナリシスに関する研修など、活発な活動を始めており、今後は、政策決定者や実務家などレビューのユーザを巻き込んで地域のニーズに合ったレビューをといったと関定し、実際に活用されるエビデンスの産出を行うという計画を持っている(Konnerup 2003)。コクラン共同計画には世界13カ国に15の地域センターがあり、キャンベル共同計画についても同様の展開が予測される。

(1) キャンベル共同計画の活動原則

キャンベル共同計画の9つの活動原則は以下のとおりである。この活動原則は、コクラン共同計画の活動原則をそのまま利用しており¹⁷、キャンベル共同計画はその運営ノウハウの多くをコクラン共同計画に負っていると言ってよい。

- 1. 共同計画の内部及び外部において、良好なコミュニケーション、オープンな意思決定とチームワークをはぐくむことにより協働する。
- 2. 異なるスキルと背景を持つ人々を巻き込み 支えることで、個人の熱意に拠って立つ。
- 3.無駄な努力をしないよう、よい運営と連携 によって、不必要な重複を避ける
- 4. 科学的エビデンスに関する高い基準に従う、 広範な参加を求める、利害の矛盾を避ける といった、さまざまな方法によって、バイ アスをできる限り小さくする。
- 5. 新たなエビデンスを見つけ含めることによってキャンベル・レビューを維持するというコミットメントを行い、常に最新であり続ける。
- 6.人々にとって意味のあるアウトカムを用いた政策と実務の評価を促進することで、有意義であろうとする。
- 7. 戦略的連携の長所を生かして共同計画のアウトプットを広く頒布し、世界中のユーザ

- のニーズにあった適切な価格、内容、媒体 を推進することで、アクセスを促す。
- 8.批判に対してオープンに応じ、方法論の進歩を適用し、質的向上のシステムを開発することで、質を保証する。
- 9. レビューに対する責任、編集のプロセス及び共同計画の中心機能を、維持し更新することによって継続する。

(2) キャンベル・ライブラリー

こうした活動から生み出されるのが、キャ ンベル・ライブラリーである。キャンベル・ ライブラリーは2つのデータベースから成る。 一つは、キャンベル共同計画の成果を発表 するC2-RIPE (The C2 Reviews or Interventions for Policy Effectiveness) である。現時点では C2-RIPEは誕生しておらず、これに代わり、 ウェブサイト上で直接、部会ごとに承認され た、タイトル、プロトコル、そして、キャン ベル・レビューが公開されている。レビュー の完了以前に、登録された時点でタイトルを、 そして、承認された時点でプロトコルを公開 するのは、レビュー・トピックの重複を防ぐ とともに、レビューの手法についての批判や 助言を得るのが目的である。今のところ、現 在登録されているタイトルは70件(教育15; 刑事司法37;社会福祉17;方法論1) 提出さ

れているプロトコルは7件(刑事司法6;方法 論1) 完了しているレビューは1件(刑事司 法1)である(表2参照)⁹。

もう一つは、ランダム化実験のデータベースである、C2-SPECTR (The C2 Social、Psychological、Educational and Criminological Trials Register)である(Petrosino et. al. 2000)。これには、教育、社会福祉、刑事政策の分野における、1万件を超える無作為化割付(と思われる)実験が含まれており、現在も更新が続けられている。C2-SPECTRは、系統的レビューを行う際の基礎的データベースであるが、その価値を高めるには、研究終了後にC2-SPECTRに登録するのではなく、研究開始前に事前登録する仕組みに移行する必要がある²⁰。

キャンベル・ライブラリーは、キャンベル 共同計画のサイトから提供され、政策決定、 サービス供給の両面にわたって、利用される ことが期待される。このようにして、キャン ベル共同計画は、その時点における最善の 「What works」を提供する社会インフラとし て、Experimenting Societyを支えることにな る。

その一例として、現時点では唯一のキャン ベル・レビューである、「少年非行の防止に ための『スケアード・ストレート』等少年自 覚プログラム」を紹介しよう。「スケアード・ ストレート」は、字義通り、「怖がらせて、 真っ直ぐに立ち直らせる」という意味で、非 行少年を刑務所に招き、受刑者が刑務所生活 の恐ろしさなどを語ることで立ち直らせよう というプログラムである21。1976年に、アメ リカ合衆国ニュージャー州ラシュウェイ刑務 所で誕生したこのプログラムを終了した17人 の少年の更生振りを描いたテレビドキュメン タリーがアカデミー賞とエミー賞を受賞する など、新聞、雑誌、テレビなど各メディアの 絶賛を受けて、政治家にも支持され、多くの 州で取り入れられることとなった。その後、 アメリカ合衆国に留まらず、カナダ、イギリ ス、デンマーク、ノルウェイなど各国にも広 まり、少年非行防止の花形プログラムの一つ となった22。

しかしながら、無作為割付を用いた評価研究は、いずれも、このプログラムが期待通りの効果をあげていないことを示していた。そこで、キャンベル・レビューは、これらの評価研究についての系統的レビューとが期待であると、が期待であると、があると、処遇群の再犯率は80~85%に達強力によって、非行を行うオッズ比を「1.6から1.7」増加させるという劇的なものであった。これは、統制群の再犯率を50%に達強力に、なり、次アード・ストレートが再犯を強力に「増加」させる処遇であることを示している(別添の要旨参照 デ。

読者は、実験デザインに基づかない、単な る「数値目標」を掲げるだけを評価と呼ぶの なら、行政が「スケアード・ストレートの実 施率90%にする」といったばかげた数値目標 すら掲げかねないことに気づいておられるで あろう²⁴。キャンベル共同計画の目指すもの は、行政がエビデンスに基づくことであって、 行政の指標化や数値化ではない。Donald T. Campbellの伝統における「評価研究」におい ては、エビデンスを生み出すものだけが評価 研究だからである。キャンベルの伝統におけ る「評価研究」(一次研究)の成果を統合す るのが系統的レビューであり、系統的レビュ 一の与えるエビデンスこそ、私たちが「何が 有効か」について知っていることの集大成で ある。

しかしながら、これまで実証研究の生み出すエビデンスが政策決定を必ずしもリードしてこなかった経緯を踏まえると、キャンベ直ちに政策立案や実務に反映されると考えるのは、と見である。キャンベル共同計画は、その点、個々のレビューが社会的影響力を持つととそのではなく、レビューの蓄積とそのインと考えている。それは、コクラン共同計画がレビューの頒布を通じて、その影響力を幾何級

30 津富 宏

数的に高めていった経緯があるからである²³。 つまり、キャンベル共同計画のメンバーは、 より多くのキャンベル・レビューさえ産出さ れれば、同計画が必然的に大きな影響力を持 つことになると確信している。

もちろん、キャンベル共同計画は、単にインターネットを通じた成果提供を行うだけではなく、その進度を加速するため、系統値について、積極的なロビーイング活動をも行っている。例えば、刑事司法部会は毎年ワシ、2日間にわたって、議会スタッフを対象とし、まつにわたって、高部会の成果を伝達きした、その利用について議会スタッフと意見としている。というの結果、教育部会は、米う育なロビーイングの結果、教育部会は、米う育なロビーイングの結果、教育部会は、米う育なロビーイングの結果、教育部会は、米う育なロビーイングの結果、教育部会は、米う育なロビーイングの結果、教育の助成を獲得している。

いずれにせよ、今後、エビデンスに基づく 社会を建設するために重要な課題は、すでに 医療ではベッドサイドにおける意思決定に直 ちにエビデンスを活用することが始まってい るように、社会政策の決定や実施においてエ ビデンス利用をビルトインすることである。 そのために、Sherman (2003)は、次に掲げ るDRIVERモデルを推奨している(例として、 Hot Spot Policing (犯罪集中地域に対する警 察活動の強化)を付す。

Diagnosis: 診断

例 犯罪の多い箇所に関するその地域の データ

Revision: 研究のレビューに基づく、政策の修正

例 犯罪多発地域に警察活動を集中させ る活動に関する決定的なレビューの 参照

Implementation: 指標設定した上での政策実施

例 警察官のパトロール要員分配のモニ タリング

Vindication: リスク要因を踏まえたうえで

の結果の確認

例 リスク要因から導かれる期待値と実 測値の比較

Evaluation: 非実験デザインによるエビデンスに対する修正の評価

例 アウトカムやコストの評価

Revision: 実験デザインによるエビデンス の産出ないし非実験デザインによるエビ デンスを応用するための、政策の再修正 例 政策のさまざまな側面の修正

DRIVERモデルは、無作為割付実験ないし系統的レビューの生み出すエビデンスによって、介入を繰り返し、それを指標によって確認するというモデルであり、キャンベル共同計画の目指す、Experimenting Societyにほかならない²⁸。Shermanはその例として、ニュータ市警の犯罪管理システムであるCompStat (McDonald 2002)やメリーランド州ボルチモア市の行政管理システムであるCitiStat²⁹をあげている。これらはいずれも、トレンド・モニタリングを基盤とし、成功したプラクティスを直ちに追試する仕組みである。

医学領域では、ベッドサイドの臨床判断において、エビデンスを参照することが始まりつつあるが、今後、重要なのは、エビデンスが、政策担当者や実務家に対して利用可能な形で提供されることである。この例でわかるように、エビデンスは直ちに入手可能でなければならない。

5. わが国におけるキャンベル共同 計画

2002年から急速に、わが国でもキャンベル 共同計画を導入しようという動きが展開して いる。例えば、筆者は2002年の9月に、刑事 司法部会の日本語ウェブサイトを立ち上げ、 さらに、2003年夏には、キャンベル共同計画 事務局の日本語ウェブサイトを立ち上げる予 定である。ウェブサイトの開設は、わが国に おけるキャンベル共同計画への関心を高めるとともに、今後加速度的に産出されてくるであろう、キャンベル・レビューの日本語への翻訳体制のプラットフォームを整えることに目的がある。

もちろん、このような活動は、教育、社会福祉、刑事司法といった、異なる分野の研究者・実務家の協働を要請する。つまり、キャンベル共同計画の成長に伴い、わが国においても、キャンベル共同計画に貢献する個人から成るネットワークが成長していくことが必要である。わが国においては、人文・社会分野におけるRCT、ひいては、エビデンスが乏しい(例えば、小林他 2003)が、そのようなわが国におけるキャンベル共同計画の課題は大きく分けて2つあると思われる。

- (1) キャンベル共同計画によって生み出さ れたエビデンスの頒布
 - ・キャンベル共同計画についての広報: 研究者、政策決定者、実務家等を対象
 - ・キャンベル・レビューの翻訳:テクニカルで正確な翻訳
 - ・キャンベル・レビューの日本語データ ベースの作成、頒布
- (2)わが国発のエビデンスの産出
 - ・エビデンスの産出に関する研修:無作 為化割付実験、系統的レビューの手法 について
 - ・系統的レビュー産出の支援体制の充実: 研究報告の方法の構造化、文献 探索体制の強化
 - ・一次研究の実施・英語への翻訳
 - ・系統的レビューの実施・英語への翻訳

このような活動を進めていくためには、安定した体制の構築が必要であることは言うまでもない。そのためには、先述した、ノルディック・キャンベル・センターの活動が参考になろう。

キャンベル共同計画は、系統的レビューを 行う能力のある研究者を中心として、運営が 進められているが、そのような研究者を持つ 国は世界的に見れば多くはない。エビデンス を生み出すための系統的レビューを行うためには、RCTである一次研究の存在が不可欠だからである。医学分野においては、系統的レビューの発達が、RCTの増加に結びついてきた。それは、系統的レビューが、NRCTに対するRCTの優位性を明らかにしてきたからである。よって、社会政策の分野においても、キャンベル共同計画の進展が、RCTの増加をもたらすことは想像に難くない。

エビデンスを尊重する文化自体を根付かせること自体も、キャンベル共同計画の重要な使命である。その意味で、わが国には、他の「エビデンス途上国」とともに、キャンベル共同計画に参加し、キャンベル共同計画が「エビデンス途上国」にとっても真に有用なものとなるよう、積極的に役割を果たしていくことが求められている。

注記

- 1 ここでいうエビデンスは、RCT(randomized controlled trial:無作為化割付実験)を頂点とする評価研究をいう。RCTとは介入を受けるか否かの割付を無作為に行い、介入の効果を見る実験である。
- 2 現在の版の著者としては、この二人に加え、Mark Lipseyが加わっている。Lipsey も、犯罪者処遇効果研究のメタアナリシスで知られている。
- 3 例えば、Rossi and Freeman (1989)。
- 4 NRCT(非無作為化割付実験)は再犯率 の予測変数を十分に統制したものに限っ ても数万件あろう。
- 5 犯罪者処遇効果研究は、犯罪学、社会学、 心理学、精神医学、社会福祉、刑事政策 などの多岐の分野の雑誌で発表されてい る上、学術雑誌以外の発表も多く、見つ け出すのが非常に困難である。
- 6 このような際には、基準を緩めることが どのような効果をもたらすかを吟味する ため、分析段階において感受性分析を行 うことが必要となる。
- 7 メタ・アナリシスとは、簡単には、それ

32 津富 宏

ぞれの一次研究から得られる効果の推定値(ES_i)を統合した推定値 \overline{ES} を求める統計手法である。推定値は、その精度において異なっているので、統合を行うに当たっては、より精度の高い推定値により大きな重みを付けて統合する必要がある。すなわち、重み(W_i)は、推定値の分

散の逆数として定義される($w_i = \frac{1}{SE_i^2}$)。 すなわち、 $\overline{ES} = \frac{\sum (w_i ES_i)}{\sum w_i}$ である。メタ

・アナリシスに関する理解を深めたい読者は、メタ・アナリシスに関するテキストブック (Cooper and Hedges 1994; Cooper 1998; Lipsey and Wilson 2001)を参照されたい。

- 8 http://www.cochrane.org/
- 9 方法論グループのガイドライン自体、方法論の進歩に応じて絶えずアップデートされる。すなわち、品質の絶えざる向上が保証されている。
- 10 これは、コクランライブラリーといわれる電子媒体化されたデータベースの一部として提供されており、コクランレビューのほか、コクラン・レビューのデータベースだけでなく、コクラン・レビュー以外の系統的レビュー、治験、コクランの方法論グループの成果物、方法論研究の一覧などを含んでいる。
- 11 http://www.abc.net.au/science/news/health/ HealthRepublish 731690.htm
- 12 http://www.cochrane-net.org/openlearning/ HTML/mod19-2.htm
- 13 http://www.campbellcollaboration.org/
- 14 http://www.aic.gov.au/campbellcj/
- 15 2003年2月に開催された、キャンベル共同 計画の年次総会である、第3回キャンベル コロキアムでは、ユーザー/消費者グル ープが追加されている。
- 16 http://www.sfi.dk/sw1271.asp
- 17 コクラン共同計画の活動原則は改訂され、 今は、10点ある。
- 18 これは、質の保障を目的とした仕組みで

- あるが、キャンベル・レビューの産出速 度を抑える原因ともなっている。
- 19 現在では、タイトル(70件)はキャンベル共同計画のページに列挙されており、そこからキャンベルレビュー(1件) プロトコル(7件)にリンクが貼られている。
- 20 事前登録制にすることにより、特定の偏りをもった(例えば、有意な結果を見出した)研究成果の方が公表されやすいために生じる「公表バイアス」を防ぐことができる(cf. Chalmers 2000)。
- 21 http://www.wild-side.com/scaredstraight.html
- 22 筆者との面談の際、非行防止のためのボランティア団体として知られる、ガーディアン・エンジェルスのわが国における代表者もスケアード・ストレートを高く評価し、日本に導入すべきだと述べていた。この事例は、わが国に、「専門家」の意見ではなく、エビデンスを尊重する文化を浸透させる必要があることを示している。
- 23 評者から、本レビューの手法、特にメ タ・アナリシスについて詳述するように 求められたが、本レビューは、キャンベ ル・レビューの一例に過ぎず、詳述する ことは、本稿の目的を逸脱することにな る。なぜなら、キャンベル共同計画の本 質はレビューの品質管理(さらには品質 認証)にあり、本稿は、個々のレビュー の品質ではなく、品質管理の体制につい て述べなければならない。そのために用 意されたのが図2である。また、キャンベ ル共同計画のレビューのガイドラインの 骨子は、本稿で概説した系統的レビュー の手順であるが、詳細は、ウェブサイト (http://www.campbellcollaboration.org/Fragu idelines.html) に掲載されている。興味の ある読者はぜひ参照されたい。

また、メタ・アナリシスについては、注7 において簡単な説明を付したが、系統的 レビューの一部分を担う統計手法に過ぎ ない。つまり、質の高いエビデンスの産 出を保証するのは、系統的レビューの全 体であって、メタ・アナリシスではない。 別添1に示すように、本レビューではオッズ 比が統合されているが、オッズ比の統合 に当たっては、a,b,c,dを2×2表の各セル の頻度とすると、最初に対数変換を行い、

$$ES_{\log} = \log_e(ES)$$
 (ただし、 $ES = \frac{ad}{bc}$) 及び $SE_{\log} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$ として統合し、最終的に、指数変換してオッズ比に戻すという手法がとられる。

一応、本レビューについてコメントす れば、1) 文献探索が1945年以降の文献に ついてハンドサーチ、電子データベース サーチ及び専門家からの聞取りによって 行われており(このトピックが刑事司法 分野において比較的関心を持たれてきた こともあって)網羅的である点、2)無作 為(及びほぼ無作為の)割付実験である9 件の研究が選別されたが、これは、メ タ・アナリシスを行うには十分な件数で ある点、) 一次研究が共通の効果値が持つ ことを仮定する固定効果モデルと一次研 究ごとにランダムな効果値を持つことを 仮定するランダム効果モデルの両方が推 定され、いずれにおいても同じ結果が示 されたことから、良質なものであると考 えられる。

また、メタ・アナリシスは、系統的レビューの一部分を担う、統計手法に過ぎまい。つまり、質の高いエビデンスの全を保証するのは、系統的レビューの全体であって、メタ・アナリシスではない。メタ・アナリシスについては、注7に成はで簡単な説明を付したが、すでに成におれた統計手法の紹介を行わないのと同様の手法の説明を行わないのと同様にその手法の説明を行わないのと同様である。

24 その理由は、評価研究の基本モデルである、「インプット アウトプット アウトカム」モデルは本質的に因果モデルであり、実験デザインによってしか担保できないことにある。数値目標モデルは、こ

のモデルを、表面上模しただけで、実質 的には何も担保していない。本文中の例では、スケアード・ストレートは単にアウトプットの指標でしかない。数値目標で デルが、目標を設定して「手段 目標」 構造を取り入れ、一見、因果関係を取り 入れているように見せかけているだけで、 その実、行政が伝統的に負ってきた法り なマンデートを、言語から数字に変換 ただけという、論理的にはすり替えのモ デルであるからである。

- 25 コクラン共同計画では成果の提供は有償であるが、キャンベル共同計画では無償である。このように、キャンベル共同計画は、コクラン共同計画以上に、エビデンスの入手可能性に配意している。
- 26 スケアード・ストレートに関しては、キ ャンベル・レビュー以前からの研究の蓄 積もあって、刑事司法を専門とする実務 家には無効であるという認識はすでに定 着しており、その活用は縮小傾向に入っ ているが、政策決定者やマスコミには依 然として有効論者が少なくない。このよ うに、社会政策の評価においては、実務 家の認識と、政策決定者の認識のずれが しばしば問題となるが、その理由は、前 者は政策目標の達成をアウトカムに置く のに対し、後者は国民(=有権者)感情 の満足にアウトカムを置くからである。 よって、キャンベル共同計画の将来は、 国民に対し、エビデンスの意義をどれだ け理解させることができるかにかかって いる。この点については、実務家や政策 決定者と二人三脚で進めていくことにな ろう。例えば、本文中のWhat Works Clearninghouseは、そのための取組みの好 例である。
- 27 http://www.w-w-c.org/
- 28 Shermanは、このような社会を建設するための行政を行う政府を、エビデンスに基づく政府(Evidence-Based Government)と呼ぶ。
- 29 http://www.baltimorecity.gov/news/citistat/

34 津富 宏

- 30 http://fuji.u-shizuoka-ken.ac.jp/ campbellcj/
- 31 担当アドバイザは、部会が承認する。
- 32 この段階において、担当アドバイザは、 プロトコルの下書きに対して、コメント と批判を行う、編集チームを設けること を強く要請される。編集チームは、部会 運営委員会、方法論の研究者、各分野の 専門家などから成る。担当アドバイザは、 方法論部会のコーディネータないし議長 と連絡を取り、方法論部会のメンバーを 編集チームに含める必要がある。
- 33 プロトコルのために編集チームが設けられた場合、同じメンバーがレビューの下書きも担当する。
- 34 C2-RIPEの維持及びチェックは、キャンベル共同計画事務局が責任を持つ。

参考文献

- 小林秀資・加藤則子・三砂ちづる・津富宏・ 正木朋也(2003)「青少年暴力に関わる国 内研究動向の文献的検討」、『厚生労働科学 補助金(特別研究事業)「思春期における 暴力行為の原因究明と対策に関する研究」 分担研究報告書』
- 津富宏(2000)「EBP(エビデンス・ベイスト・プラクティス)への道 根拠に基づいた実務を行うために 」、『犯罪と非行』124、67-99.
- 林謙治 (2000)「根拠に基づく健康政策への アプローチ」、『公衆衛生研究』、49(4)、 346-53.
- Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., and Cullen, F. T. (1990). Does Correctional Treatment Work? A Clinically Informed Meta-Analysis. *Criminology*, 28, 369-404.
- Chalmers, I. (2000). Current Controlled Trials: an opportunity to help improve the quality of clinical research. *Current Controlled Trials in Cardiovascular Medicine* 1:3-8.
- Cooper, H. (1998). Synthesizing Research: A Guide for Literature Reviews. Thousand Oaks,

- CA: Sage.
- Cooper, H. and Hedges, L. V. (1994). *The Handbook of Research Synthesis*. New York: Russell Sage.
- Fincknauer, J. O. and Gavin P. W. (with Hovland A. and Storvoll E.) (1999). Scared Straight and the Panacea Phenomenon. Revisited. Prospet Heithts, IL: Waveland Press.
- Fincknauer, J. O. (1982). Scared Straight and the Panacea Phenomenon. NJ: Prentice-Hall.
- Glass, G. (1976). Primary, Secondary, and Meta-Analysis of Research. *The Educational Researcher*, 10, 3-8.
- Konnerup, M. (2003). Personal Communication at the Campbell Crime and Justice Group Steering Committee meeting held in Pairs, May 2003.
- Lipsey, M, and Wilson D. B. (2001). *Practical Meta-Analysis*. Thousand Oaks, CA: Sage.
- Martinson, R. (1974). What works?.....Questions and answers about prison reform. *The Public Interest*, 35, 22-54.
- McDonald, P. P. (2002). Managing Police Operations.: Implementing the New York Crime Control Model---Compstat. Belmont, CA: Wadsworth/Thomson Learning.
- Petrosino, A., Boruch, R. F., Rounding, C., McDonald, S. and Chalmers I. (2000). The Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register (C2-SPECTR) to Facilitate the Preparation and Maintenance of Systematic Reviews of Social and Educational Interventions. *Evaluation Research in Education*, 14.
- Rossi, P. H. and Freeman, H. E. (1989). Evaluation: A Systematic Approach (Fourth edition). Newbury Park, Calif.: Sage.
- Sherman, Lawrence W.(2003) New Approaches to Evidence Assessment: The "Booksite "and the "Driver" model. Presentation at the Smith Richardson Foundation's Conference on Testbed Reviews: PowerPoint File.

表1	エビデンスの質	
衣门	エピアノ人の買	

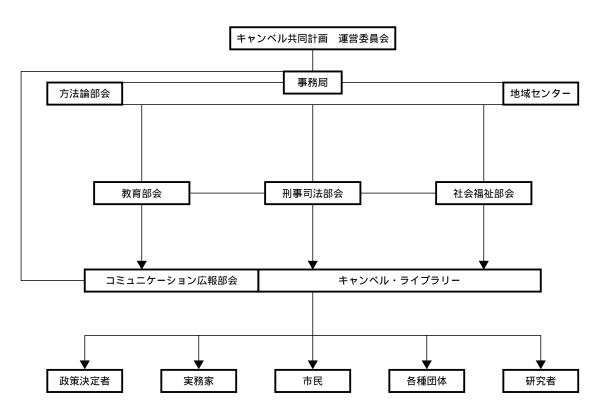
上	上	RCTØSR		
	中	(信頼区間の狭い)個々のRCT		
	下	その治療方法が用いられたときに全員が治癒し、用いられなかったときに全員が治 癒しなかったことを示す、一連の症例研究		
中	上の上	コホート研究のSR		
	上の中	個々のコホート研究・質の低 i NRCT		
	上の下	個々の「結果」研究		
	下の上	ケースコントロール研究のSR		
	下の下	個々のケースコントロール研究		
下		一連の症例研究(及び質の低いコホート研究あるいはケースコントロール研究)		
下の下		批判的吟味を欠いた、あるいは、意見等に基づく専門家の意見		

RCT: randomized controlled trial (無作為統制実験)

SR: Systematic Review (系統的レビュー)

(出所)津富(2000)

図1 キャンベル共同計画の組織図



(出所)キャンベル・コロキアム配布資料(2003)

36 津富 宏

表2 キャンベル共同計画に登録されているタイトルの一覧

- ✓ レビュープロトコルが提出されているもの
- ▶ レビューが完成しているもの

教育グループ

英語を第二言語とする生徒の学習指導 小学校における,級友の支援による学習 児童の構音指導の有効性

怠学を減少させるための介入

ボランティア個人教授プログラムの効果に関する エビデンス

教師用導入研修プログラムの有効性

小学校児童の学業成績に対する親の関与の有効性 問題基盤学習の有効性

学校を基盤とする社会問題解決的介入の攻撃行動 への効果

高校中退の防止

両親と親の連携による介入 (FAST)

成績不良者に対する特別グループ・個人教授ない し統合指導

反社会的行動に対する集団を基盤とした指導ない しその代替策の教育及び犯罪に対する効果 課外プログラムの効果

社会福祉グループ

社会福祉改革の家族構造に対する効果の系統的レ ビュー

住居改良の健康に対する効果の評価 性的虐待を受けた子どもに対する認知行動療法 道路建設の健康や社会的影響

両親と親の連携による介入(FAST)の効果 10代の親とその子に対する,心理社会的アウトカムを向上させるための個人・集団による育児訓練幼稚園児及び小学生を対象とする,(個別栄養状態を含む)栄養状態,成長,認知,成績に対する,学校を基盤とする栄養・給食介入の効果

里親の問題行動への対処を援助するための認知行 動療法訓練による介入

社会的にハンディキャップを負った母親に対する 家庭を基盤とする支援

思春期拒食症に対する心理社会的介入

重度精神障害者に対する住居支援

学習障害を持った性的犯罪者への介入

10歳から17歳の行為障害及び非行のある児童・少年に対する家族・親を通じた介入

0歳から3歳の子どもの情緒的・行動的適応を向上させるための集団を基盤とした親の訓練プログラム 児童の逸脱行動に対するメディアを用いた行動プログラム

母親の心理社会的健康を向上させるための親の訓練プログラム

社会的扶助受給者に対する職業・活動プログラム 刑事司法グループ

拘禁期間の長さ

> 少年非行の防止のための「スケアード・ストレート」 等少年自覚プログラム

再被害化防止プログラム

夫婦間暴力への介入

性犯罪者の処遇

拘禁された暴力非行少年への介入

刑務所における薬物処遇プログラム

銃犯罪に対する警察の介入

✓ 犯罪者の労働市場への復帰

ドラッグコート

修復的司法

監視カメラ

街路灯

少年に対する門限条例

✓ 電子監視

早期幼児期における親に対する支援と訓練 児童の技能訓練

犯罪者に対する認知行動プログラム 犯罪者に対する信仰に基づく宗教プログラム

- / 非行少年及び犯罪者に対するブートキャンプ
- ✓ 犯罪多発地域における集中的警察活動

人格障害に対する非薬物治療 少年に対するコミュニティを基盤とするプログラ ム

企業犯罪の抑止

家族を基盤としたプログラム

家族以外に対する暴力への介入

司法精神サービスに委ねられた精神障害と診断された人々を対象とする暴力行動への介入

近隣による監視

薬物犯罪者に対する通所治療

商業航空機におけるあるいは対する犯罪の防止 司法精神障害者に対するリスクアセスメントの手 法

✓ 収容中の非行少年の自殺リスクのスクリーニング 検査

判決の費用効果分析

再犯に対する,収容(対非収容)判決の効果 非行少年の出所後の処遇

<u>方法論グループ</u>

✓ 教育,訓練,雇用サービスにおける非実験による 追試

図2 キャンベル・レビューが産出される過程

系統的レビューを行うことに関心のある者(レビューワ)ないし各部会が、レビュー・トピックを設定す る

レビューワが、行いたいレビュー・タイトルとレビュー完了予定期限を、部会コーディネータに提出する

部会ないし部会コーディネータが、担当アドバイザ (Principal Advisor) を割当てる担当アドバイザの援助を得て、レビューワがレビュープロトコルの下書きを完了する

担当アドバイザ(ないし、編集チーム)が、レビュープロトコルの下書きを審査する³²

担当アドバイザが、プロトコルを承認し、C2-RIPEに提出する

レビューワが、レビューの下書きを完了、担当アドバイザに提出する

担当アドバイザが、編集チームから下書きについてコメント・批判を得る33

レビューワが、編集チームからのフィードバックを得て書直す(このステップは繰り返されうる)

レビューワが、最終レビューを提出する

レビューが、C2-RIPEに発表される³⁴

多様な読者に合わせ、さまざまなバージョンのレビューが発表される。

(出所)キャンベル・コロキアム配布資料(2003)

38 津富 宏

別添 「少年非行の防止のための『スケアード・ストレート』等少年自覚プログラム」に 関するキャンベルレビューの要旨

著 者: Petrosino, A; Turpin-Petrosino, C; Buehler, J.

タイトル: 少年非行の防止のための「スケアード・ストレート」等少年自覚プログラム

出 典: コクラン・系統的レビューデータベース、2002年、第4版

要 旨:

背 景: 「スケアード・ストレート」等のプログラムは、非行少年ないし犯罪行動を行うリスクのある少年による刑務所訪問プログラムである。このプログラムは、参加者に刑務所の生活を直接観察させ、成人受刑者と接させることにより、将来、犯罪を行わせないことを目的としている。このプログラムはその効果を疑わせる研究やレビューが発表されているにもかかわらず、世界的に用いられている。

目 的: (少年裁判所により正式に判決が下された、あるいは、有罪宣告を受けた)非行少年ない し(問題は起こしているが正式に非行少年としての判決を受けていない)非行を行うおそれ のある少年に犯罪行動を行わせないことを目的として、彼らに刑務所を訪問させるプログラムの効果を評価する。

文献探索手法: 第一著者が1945年から1993年の間の犯罪学における無作為化割付実験をハンドサーチし、このハンドサーチをキャンベルのSPECTRデータベース及びコクランのCCTRを含む、16の電子データベースの構造化された探索によって補った。本分野の専門家に対して照会を行い、引用文献のフォローアップも行った。

選択基準: 非行少年ないし非行を行うおそれのある少年による刑事施設の訪問プログラムの効果を検証した研究を含んだ。少年と若い成人の両方にまたがるサンプル(例、14歳から20歳)を持つ研究を含んだ。参加者を条件に無作為ないしほぼ無作為に(例えば、交互に)割り付けた研究のみを考慮した。研究は、処遇を受けない対照条件群及び刑務所訪問後の犯罪行動のアウトカムを最低一つ持っていなければならない。

データ収集と分析: 基準を満たした9つの研究について概略を述べた。公的データを用いた介入後の 犯罪率に関するメタ・アナリシスを行った。その他のデータ(例、自己申告)はいくつかの研 究になく、あるいは、重要な情報(例、標準誤差)が落ちていた。介入直後効果(当初の効 果)をそれぞれの群における再犯者の比率からオッズ比を計算し、固定効果モデルとランダ ム効果モデルを仮定してそれぞれ分析を行った。

主たる結果: 分析は、無効であるというより、有害であることを示した。固定効果モデルと仮定しようともランダム効果モデルを仮定しようとも、メタアナリシスの手法に関わりなく、プログラム効果はほぼ等しくマイナスだった。

結 論: 「スケアード・ストレート」のようなプログラムは、同じ青少年に対して、何もしないことに比べ、有害な効果をもたらし非行を増加させると結論する。この結果を踏まえ、この種のプログラムを行っている機関は、プログラムが(非行を減少させるという)本来の目的を果たしていることだけでなく、少なくとも害を及ぼしていないことを確認するために、プログラムの厳密な評価を行うべきである。

Towards Social Policy-making Based on Systematic Reviews : Challenges of the Campbell Collaboration

Hiroshi Tsutomi

University of Shizuoka tsutomi@u-shizuoka-ken.ac.jp

Abstract

A systematic review is a literature review following a scientific protocol beginning with the identification of a research question, literature search, screening and coding of literature and statistical analysis, and ending with structured reporting.. In medicine, the Cochrane Collaboration, an international network of individuals, which aims to produce, disseminate and update systematic reviews of evidence, has evolved. Leaning from its success, the Campbell Collaboration was established to inform social policy-making. Evidence generated by the Campbell Collaboration will be distributed electronically and utilized by policy makers as well as service providers. The Campbell Collaboration which will provide the best available evidence on "what works" will form a social infrastructure of the experimenting society.

Keywords

systematic review, evidence, Campbell Collaboration, social policy

【研究論文】

わが国評価におけるより適切なプログラム・セオリーの構築を目指して 国際協力評価と政策評価に焦点を当てて

三好 皓一

森田 智

藍澤 淑雄

立命館アジア太平洋大学 miyoshik@apu.ac.jp 東京大学大学院

(財)国際開発センター

要約

本稿は、「プログラム・セオリー評価」の視点に基づき、わが国の国際協力評価と政策評価についてプログラム・セオリーに焦点を当て考察を試みる。わが国の国際協力評価及び政策評価は実務経験を踏まえ実践的な手法として発展してきた経緯があり、プログラム・セオリーを必ずしも明確に提示しえていない。その結果として評価の有効性はそがれることになる。こうした状況を考慮すれば、国際協力評価及び政策評価の効果をあげていくために、より明確な、より適切なプログラム・セオリーの構築方法に注力することは重要かつ不可欠な課題である。

キーワード

プログラム・セオリー評価、国際協力評価、政策評価、政策体系、成果 / アウトカム

1. はじめに

全ての行政活動は目的を有し、政策決定者や行政担当者はその目的を達成するために、種々の行政手段を駆使して対応する。それゆえに、行政活動は、明示的であれ、暗黙的であれ、政策決定者や行政活動と手段の連鎖関係、言い換えれるの目的と手段の連鎖関係、言い換えれば、行政活動には、行政活動には、行政活動には、行政活動には、行政活動には、行政活動には、行政活動には、行政活動には、行政活動には、行政活動には、行政活動には、行政活動を支えることになる。このような関

連は一般にプログラム・セオリーと呼ぶことができる。

評価では、過去30年間このような目的と手段からなるプログラム・セオリーを確認または想定し、このプログラム・セオリーを評価の案内役として活用し、行政活動を評価しようとする試みが行われてきた(Rogers他 2000)。このような評価は、「プログラム・セオリー評価(Program Theory Evaluation)」と呼ばれ、評価の中で主要な地位を占めるようになっている。実際、60年代には、既に「プログラムの構成要因における因果関係を明確する」という考え方は、欧米の専門家の間で定着し、以来、専門家の間では、プログラム・セオリーを、成果

階層モデル、アクション・セオリー、プログラム・ロジックなど、異なった表現を使いながらかかる議論が重ねてられてきた(Rogers他 2000)。これらは表現に違いがあるものの、プログラムの構成要因における因果関係を明確化するという点で共通している。

プログラム・セオリー評価は、プログラ ム・セオリーを評価すること、すなわち、 プログラムの構成要因である投入や中間的 な成果が、期待される最終成果をもたらす ために適切であるか、あるいはこれら構成 要因の論理構成が適正な因果関係に基づい ているかを検証することである。そのため に、プログラム・セオリーを明確にすると ともに、プログラムを構成する一連の目的 と手段の連鎖関係のそれぞれについて因果 関係を調べることになる。プログラムの論 理構成の適正さを検証することはプログラ ム・デザインの改善に大きく貢献すること から、プログラム・セオリー評価はこの点 においてプログラムの有効性を高めるため の評価と位置づけることも可能となる。

近年、プログラム・セオリー評価の定義が明確化されるに従い、評価手法の一つされてますますその役割が重要視されてきなり、関連する議論も活発化している。特にプログラム・セオリー評価を、行政活動が意図された、または、観察された成果(アウトカム)をいかに生じしめるかについてのセオリー、または、モデルと、少を基にして評価するという考え方を満たすものとなる(Rogers他 2000)。

国際協力分野では、このようなプログラム・セオリーに基づいた評価が従来からロジカル・フレームワーク(LF)の活用の名のもとに実施されてきている。国際協力事業団等の援助機関における国際協力評価では、従来から上位目標・プロジェクト目標・結果・活動・投入からなる一連の目的と手段の連鎖関係を想定したLFやプロジェ

クト・デザイン・マトリックス(PDM)が 使用されており、プログラム・セオリー評 価を基本とした評価活動が行われている。

他方、わが国では、1990年代後半以来、中央政府や地方公共団体への評価の導入が進められる一方、最近においては特殊法人や大学などの独立行政法人化に伴って評価についての議論が活発に行われている。これらの評価は、政策・施策・事務事業からなる目的と手段の一連の連鎖関係を想定した政策体系を基に評価することを基本としており、プログラム・セオリー評価として扱いえるものである。

本稿の目的は、このような認識の下に、 わが国の国際協力評価と政策評価について、 「プログラム・セオリー評価」の基になるプ ログラム・セオリーに焦点を当て考察を試 みるものである。国際協力評価ではLFや PDMとしての目的と手段の因果関係を基に 評価を行うこととしているが、これらにつ いてプログラム・セオリーとしての特性と 発展を考察する。また、政策評価では、政 策・施策・事務事業からなる一連の政策体 系を業績評価、事業評価、総合評価として 評価することになるが、これらの評価にお けるプログラム・セオリーとしての政策体 系の一層の明確化の必要性について考察す る。従来、このような評価活動をプログラ ム・セオリー評価の観点から議論したもの は少ないことに鑑みれば、議論することに よって理論と実践の双方においてより一層 の発展が期待しえると筆者は考える。

2. ロジカル・フレームワーク(LF) としてのプログラム・セオリー の発展

(1) LF発展の歴史

LFの起源

現在、LFあるいはPDMとして知られるこの手法は、1960年代後半に米国国際開発庁

(USAID)が議会予算要求資料作成のために 開発したツールである(国際協力事業団 1999)。LFは通常、「上位目標 (Goal)」、「プ ロジェクト目標(Purpose)」、「結果(Outputs)」、 「活動 (Activities)」、「投入 (Inputs)」の5レ ベルからなる階層型モデルであり、隣接す るレベル間においてはそれぞれ、原因・結 果(手段・目的)の因果関係を仮定として 成立している。ここでは、一般に、これら の5つの項目内容が、外部条件となる仮定と ともに特定された後、各レベルにおいて、 それぞれ達成度を測るための検証可能な指 標が設定され、さらに対象となるターゲッ ト・グループの明示化と、指標データの検 証手段が特定される(Binnendijk 2001)。表1 は、LFの具体例を示す。

LFの発展

1970年代後半以降、UNDPやUNICEFをはじめとする多くの国際機関ではLFを導入し、また欧米では民間企業においても採用され、事業の運営管理に用いられてきた(国際開発高等教育機構 2001)。また1980年代になると、ドイツ技術協力公社(GTZ)がLFの作成過程に「参加型」の概念を取り入れることで「目的指向型プロジェクト立案(ZOPP)手法」を開発し、ヨーロッパ諸国の二国間援助機関やNGOにおいてZOPP手法の考え方が導入され、結果としてLFがより広範に普及していった(国際開発高等教育機構2001;国際協力事業団1999)。

尚、LFの構造は組織によって必ずしも同一ではなく、例えば、いくつかの機関では活動を除いた4つのレベルしかなかったり、またGoalとPurposeの代わりにそれぞれImpactと(Intermediate)Outcomeを用いたりしているなど、用語の使用状況にも違いがみられる(Binnendijk 2001)。

(2) 日本の開発援助におけるLFの導入

PCM手法の開発と運用状況 日本では1990年以降、国際開発高等教育

機構(FASID)がZOPP手法に基づき、ODA 事業の質の向上という観点から「プロジェ クト・サイクル・マネジメント (PCM)手 法」の開発に努めてきた。このPCM手法は、 「プロジェクトの発掘、形成を含む計画、プ ロジェクトの審査、実施、モニタリング、 評価とそのフィードバックまでの一連の事 業サイクルを運営管理する手法」(国際協力 事業団 1998) とされており、参加型計画手 法とモニタリング・評価手法からなる。前 者に関しては、参加者分析・問題分析・目 的分析・プロジェクト選択の4つの分析段階 と、PDM及び活動計画表の立案段階から構 成され、後者に関しては、作成されたPDM に基づき、妥当性、有効性、効率性、イン パクト、自立発展性からなるDAC評価5項目 を用いたモニタリング・評価が実施される ことで計画修正が検討され、その都度PDM も修正される(国際開発高等教育機構2001)。

例えばJICAでは1994年以降、全ての技術協力プロジェクト(旧プロジェクト方式技術協力)においてこのPCM手法を取り入れる動きがみられたが、実際の現場においては一般に、PCM手法のうち4つの分析段階を省略した案件実施開始時におけるPDMの作成と、中間評価等の調査団派遣時のPDMの修正、そして終了時の評価用PDM(PDME)の作成とDAC評価5項目を用いた評価がこれまで多くの場合において実施されてきたのツールであるという認識のもと、ついてあるという認識のもと、のツールであるという認識のもといったまできた。

PDMの特徴

PDMによる評価は、評価用のPDM_Eの作成、評価グリッドの作成、実績・実施プロセスの確認、評価5項目による価値判断と効果発現・阻害要因の分析からなる(三好 2002a)。PDMは前述したように、「上位目標」、「プロジェクト目標」、「結果」、「活動」、「投入」及びプロジェクトを取り巻く外部条件の論理的な相関関係としてのプログラム・セオリーをマトリックスとして示したものであ

夷1	Ι	F	П	概	粟
1.3.1	_	.ı \	"	TIM.	~

プロジェクト要約 (Narrative Summary)	客観的に検証可能な指標 (Objectively Verifiable Indicators)	検証手段 (Means of Verification)	重要な仮定 (Important Assumptions)
<u>上位目標(Goal)</u> : 当該プロジェクトのモデルがタイ 全国に普及し、交通事故外傷によ る死亡率が低下する	・タイ全国の交通事故外傷 による死亡率 ・モデルを適用した県数	保険省の評価シート	・政府からの継続的な政策的支援
プロジェクト目標 (Purpose): コンケン県において交通事故外傷 による死亡率が低下する	・コンケン県の交通事故外 傷による死亡率	コンケン病院記録	・政府からの継続 的な政策的支援
結果 (Outputs):1 病院における外傷患者ケアを改善する2 外傷のプレホスピタルケアが効果的になる3 交通事故予防・軽傷化の活動が促進される	・回避可能な死亡率 ・通報後に患者を病院へ搬 送するまでの時間 ・ヘルメット・シートベル ト着用率 	コンケン病院記録 	・通報システムの整備 ・交通規則の強制 力強化 ・道路環境の向上
活動 (Activities): 1 病院ケアのための諸活動 ・ガイドライン整備 ・スタッフ訓練 ・ 2 プレホスピタルケア ・住民に対する応急措置訓練 ・広報 ・			・ベースラインデータ収集 ・利害関係者の プロジェクトへ の協力 ・外傷センターの サービス開始
投入(Inputs): 			

(注)筆者により一部改変

(出所) Binnendijk (2001) 原口 (2001) を基に筆者作成 (一部抜粋、邦訳は筆者)

る(三好 2002a)。すなわち、あるプロジェクトを想定した場合、投入を効果的に用いて活動を実施することで結果が得られ、またここで外部条件が満たされた時にPDMの論理構成に基づいてプロジェクト目標と上位目標が達成される。

PDMの制約と使用上の問題点

PDMは個別プロジェクトを対象としているため、当該案件に関してプログラム・セオリーを明確化するには理論上有効である。

しかし、その一方で、プロジェクト目標や 上位目標は個別単独事業では目標を達成で きない場合が多いことに加え、PDMでは他 事業との関係が明示されないことにより、 プロジェクト目標や上位目標への関心、意 識は希薄になる。結果として実際のプロジェクト評価では、投入・活動・結果に焦点 が当てられ、プロジェクト目標と上位目標 が軽視されるため、評価5項目による分析が 適切に行われず因果関係が不明瞭となるこ とが多い(三好 2002a、三好 2002b)

また、JICAにおいては実施上困難を伴う ことからこれまでPCM手法のうち4つの分析 段階が多くのプロジェクトで省略されてき た。結果として、援助事業が援助対象国の 政策体系への介入であるのもかかわらず、 相手国にとっての事業実施の妥当性と、相 手国の政策体系との事業の整合性を十分に 検討し得なかったというPDMの使用上の問 題点があった。このような問題に対しては、 近年のJICAにおける事前評価の導入によっ て介入の明確化が図られると考えられる。 しかし、安易な事前調査が実施されるなら ば、計画者の経験や直感に基づいて因果関 係あるいはプログラム・セオリーが適用さ れるという問題点 (Weiss 2000) を排除する ことがより困難になってしまう可能性を否 定できない。

また、プログラム・セオリー評価におい ては、投入と成果の因果関係が成立する過 程において影響を及ぼした要因を全て特定 することは難しく、また、適用されるプロ グラム・セオリーの各レベル間における因 果関係を精査する際、時間軸の観念が無視 される場合が多い(Cook 2000)。 これらの点 はPDMを用いる際にも当てはまる。こうし た問題点に対処するためには、事業実施中 の各段階において作成されたPDMを時系列 に並べ、各時点における論理を比較検証す ること、また事業実施のプロセスにおいて 影響を及ぼした要因を時間軸に沿って整理 し直し、因果関係の変化を追うということ も理論上可能である。しかし、JICAがこれ までPDMを用いる際には、本来PDMに組み 込まれている数値を用いた業績指標とモニ タリングの機能が実際には重視されてこな かったため、評価5項目を用いても、結果と して事業の効果を把握しきれなかったとい う側面が挙げられる。

(3) 結果フレームワーク (RF) の開発

RFの起源

米国では1990年代に、公的機関が実施す

る事業の効率化向上に対する機運が高まり、1993年には「政府の実績と結果に関する法律」(GPRA)が制定され、また同時にアメリカ合衆国議会がアメリカ合衆国国際開発庁(USAID)の活動をアメリカ政府の外交政策の遂行手段として位置づけた(藤村2001)こと等を背景に、USAIDでは1995年からRFを用い始めた。現在ではUNDPでも国別協力フレームワーク(CCF)という名において同様のツールが採用されている。

RFの特徴

Binnendijk (2001) によると、RFは、ある援 助対象国における政策体系を明確にし、一 般にセクターあるいはサブセクターレベル における特定の開発目標を達成するための 戦略を階層図で示すためのツールである。 通常、あるセクター / サブセクターにおけ る一つの 戦略開発目標 Strategic Development Objective)」の下に、複数の「中間結果/成 果 (Intermediate Results/Outcome)」と、さら にその下位レベルに「(活動の)結果 (Activity Outputs)」があり、計3レベルから なる。隣接するレベル間においてはそれぞ れ、原因 - 結果(手段 - 目的)の因果関係 が仮定として成立しており、基本的な論理 構造はLFと同じであるが、個別の活動の投 入や実施プロセスではなく、より高次での インパクトの達成に焦点を当てているため、 多くの関連プロジェクトや活動の成果 (Outcome)が含まれる点において異なる。 成果は一レベルのみとは限らず、数レベル になる場合もあり、またそれらの成果どう しの間においても因果関係が見出されるこ ともあり得る。その後で、結果の達成に影 響するかもしれない外部条件となる仮定が 同定される。結果フレームワークが作成さ れた後には、全ての階層レベルにおいて業 績指標及び数値目標が設定される (Binnendijk 2001)。図1は、ある対象国にお いて「国内市場における食料流通量の増加」 を仮定として開発目標に据えたRFの一例で ある。

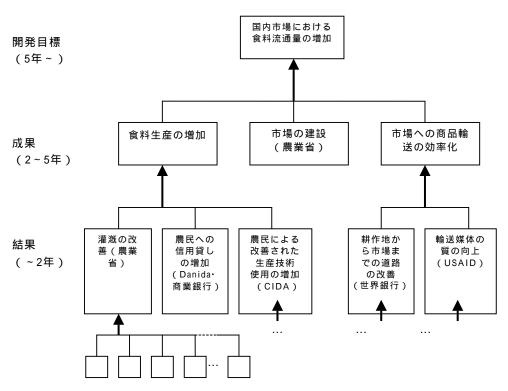
RFの制約と使用上の問題点

RFにおける各階層では、達成までに要す るとされる時間枠や対象範囲が異なるため、 上位レベルになるほどより長期間となり、 またその対象範囲も広がり、個別のプロジ ェクトないし活動とかけ離れていく。さら に、RFでは個別のプロジェクトが基本的に 列挙されない(藤村 2001)ため、上位の開 発目標においては、必ずしも結果の達成度 と個別のプロジェクトあるいは活動の効果 を関連づけられるとは限らず、それゆえ、 個別の関係機関の結果達成への寄与度を測 定するのは難しいという制約が存在する (Binnendijk 2001)。すなわち、事業実施の際 に想定した援助対象国の政策体系において、 政策及び施策と事業との間の因果関係が不 明瞭となり、乖離が生じてしまう。結果と

して、事業実施におけるマネジメント上の意思決定が恣意的にならざるを得ないことも否定できない。加えて、全ての階層レベルで数値を用いた業績測定を想定するとかなり複雑となり、時間や費用の面においてデータ収集が容易ではない点も指摘される(Binnendijk 2001)。

そのことと関連して、RFを実際に用いて体系図を作成する際に、援助対象国の政府が先導をとり、援助供与機関のみならず援助受益者、その国の関係政府機関、民間セクター、市民社会の機関、NGO、他の利害関係者等が協調を図りつつ、その作成過程に関り、また役割分担することが想定されている(Binnendijk 2001)。ここで、実際のプログラムの論理構成要素に関しては、一つの成果に対し一つの要因に帰結するとは





(出所)Binnendijk (2001) (邦訳は筆者)

限らずその因果関係が複雑であり(Roger 2000) 本来はそれらを包括的に検証する必 要があるもののコストが高くつくため、セ オリーの選択を行うことが不可避となる (Weiss 2000)。しかしRFでは、上述したよう に役割分担を行うことで、各利害関係者が 好都合なセオリーを一つ選択し、また各々 が負担するコストが少なくて済むという利 点も存在する。さらに、LFに関して述べた のと同様に、より多くの利害関係者が体系 図の作成に加わることで、少数の場合に比 べて、因果関係あるいはプログラム・セオ リーが個人の経験や直感、または既に立証 された因果関係等に誘導される可能性が理 論上は低くなる。しかし、例えばドナー機 関の間では、現時点において援助協調が円 滑に実施できるような体制にはなっていな い。RFは比較的新しいツールであり、その 使用に関してはまだ多くの機関に採用され ていないこととも関連して、経験の蓄積が 少なくまた解決すべき点も多い。

(4) プログラム・セオリー・マトリックス (PTM)の有用性

PTMの特徴

近年、JICA事業を事例として、PTMのツ ールを用いたプログラム評価の事例が報告 されている。PTMは、「最終成果(End Outcome)」、「中間成果 (Intermediate Outcome)」、「結果 (Outputs)」、「活動 (Activities)」、「投入(Inputs)」を並列に並べ、 同一のマトリックス内で事業を分析するた めのツールである(三好 2003)。 隣接するレ ベル間においてはそれぞれ、原因 - 結果 (手段 - 目標)の因果関係、すなわちプログ ラム・セオリーが仮定として成立している。 以下に示すのが、JICAが実施してきたフィ リピン国感染症対策分野での協力に関する プログラム評価の事例研究において、個々 のプロジェクトを実際に統合し作成された PTMの概要である。

実際の行政活動では政策・施策・事業の3

層の区別が明確でない場合も少なくないが、PTMを用いることで、この3層の混在したかたまりを整理して視覚化し、政策・施策・事業が相互に目的と手段の因果関係を保を理することが可能となる。基本的な論理することが可能となるが、基本的な論理はやはりLFやRFと同じであるが、上述した3層に加え、活動及び投入を全て包括した3層に加え、活動及び投入を全て包括って3層に加え、活動及び投入を全て包括って3層に加え、活動及び投入を全て包括って3層に加え、活動及び投入を全て包括って3層に加え、活動及び投入を全て包括って3層に加え、活動及び投入を全て包括した3層に加え、活動及び投入を全て包括の対策はある。他方、このようなの場合である。

PTMにおいては、最終成果が政策レベル、 中間成果が施策レベル、また結果が事業レ ベルに対応していることが分かる。ここで の結果(Outputs)とは、施策や事業におい て提供された生産物、資本財、サービスを 全て包括したものが想定され、行政の直接 的なコントロールの範囲内である一方、達 成すべき成果(Outcome)とは、ターゲッ ト・グループや社会の変化を表し、行政の 直接的なコントロールの範囲外とされる。 またPTMでは、これらの3レベルにおいて、 業績指標と数値を用いた目標値がそれぞれ 設定されるため、PTMに基づいて評価を行 う際には、測定された業績値が用いられる ことで、定性的評価ともに定量的評価を行 うことが理論上可能となる。

また、PTMにおいては、以下に示すように、DAC評価5項目と各階層の関連性も明確に視覚化されている。評価5項目の特徴は、政策体系を構成する最終成果と活動/投入の関係を特定した上で、それぞれインパターを表したことで、対したことで、対したのでであり出したが作りの変化のプロセスにある。とが、ターゲループの変化のプロセスになった、ターゲループの変化のプロセスに察がある。とを可能にする。このよう

表2	DT	MA	一例
オマム	-	כטועו	ויער

政策 / 最終成果 (End Outcome)	施策 / 中間成果 (Intermediate Outcome)	事業/結果 (Outputs)	活動 (Activities)	投入 (Inputs)	
対象地域におけ	対象地域において公共 セクターがカバーする「個 別施策層」(vulnerable group)のHIV/STIs感染率 を抑える/性感染症罹患率 を減少させる (<u>指標</u> ・ラオ	1) HIV/STIsプログラムのた	医療機材供与		
るHIV/STIs感染 率を抑える/性 感染症罹患率を 減少させる		めの組織基盤・機能を整備する (指標例) ・ラボラトリーにおける検査実施可能数/日	中央共同ラボ ラトリーへの 技術移転		
HIV感染による 社会・保健問題			HIV/AIDSの検 査診断技術の 移転		
の減少	・対象地域の公共セクター				
	での「個別施策層」 (vulnerable group)のHIV/ STIs感染率	(vulnerable group)のHIV/ 2) 体に区域 り こへには自 のHIV/STIsに関する知	スタッフへの 研修開催		
			啓蒙普及活動 への支援実施		
		3)「個別施策層」(vulnerable group)のHIV/STIsに対する啓蒙活動			
	対象地域において民間 セクターがカバーする「個 別施策層(vulnerable group) のHIV/STIs感染率を抑え る/性感染症罹患率を減少 させる	セクターがカバーする「個 別施策層(vulnerable group) 2)人材の能力向上	1)組織の機能整備		
			2)人材の能力向上		
		3)「個別施策層」(vulnerable group) への啓蒙活動			

(注)筆者により一部改変

(出所)三好他(2003)を基に筆者作成(一部抜粋)

な文脈の中では、妥当性はどのような社会を作るかという基本的な課題を考察することになり、また自立発展性に関しては、事業を成り立たせている因果関係の連鎖関係が維持され発展していくかを議論することになる。

実際にPTMを用いることで、政策立案の適否や施策の整合性といった、個々の事業評価の場合には不可能だった多くの分析が可能となるという利点が指摘される(三好2003)。個々の事業におけるセオリーに基づいた個別のPTMを一つに統合することで、重要な開発目標の達成を目的とした政策・施策の枠組みの中において、各事業の関連性が見出しやすくなる。すなわち、PTMを

用いることで、一つの政策体系内において、政策・施策・事務事業のあり方、すなわちずログラム・セオリーについての分析が可能となる(三好 2003)。また、ターゲット・グループを村、市まなど検討した、ターゲット・範囲について特定するとはよってとは高することに対応する社会の変化を議論するまともらに、なりないで、全体的は大き、1998)を実施するのにPTMは向いて、といわゆるメタ・アナリシス(Weiss 1998)を実施するのにPTMは向いて、以も当り、を実施するのにPTMは向いて、以も対象・アナリシスを行うことで、シャルも大きくなる(Pertrosino 2000;三好 2002a)。

評価項目	最終成果	中間成果	結果	活動 / 投入
妥当性	* * * * * * * *			
インパクト	* * * * * * * * *	* * * * * *		
有効性		* * * * * * * * *	* * * * * *	
効率性		•	* * * * * * * * *	* * * * * * *
自立発展性	* * * * * * * * *	* * * * * * * * * * *	* * * * * * * * *	* * * * * * * *

表3 政策体系とDAC評価5項目

(出所)三好他(2003)を基に筆者作成

PTMの使用状況と展望

PTMを用いた事例研究の数はまだ数える ほどしか存在せず、これまで主にJICA事業 の国別援助実施計画や国別プログラムを対 象として行われた事後的な評価のみに用い られてきた。LFやRFとは異なり、事業実施 中における使用が念頭に置かれていた訳で は必ずしもないため、事後的な評価におい て用いた場合、特に業績に関する必要な数 値データが存在しない場合が少なくないこ とや、対象が広範囲にわたるためデータ収 集のコストが高くつく点も指摘される。さ らに、これまでツール使用の長年にわたる 経験の蓄積があるLFや、経験が徐々に蓄積 されつつあるRFとは異なり、PTMに関して はその適切な使用法が未だ確立されていな いため、特にこれまで述べてきたセオリー の選択や因果関係の誘導性といった観点か ら、事業実施における使用法について議論 を深める必要がある。

(5) まとめ

本節でみてきたように、国際協力評価でのプログラム・セオリーの発展は、それを具体化・視覚化するためのツールの発展の中に見てとることができる。これまでの論点をまとめると、LFあるいはPDMでは、個別事業においてより焦点が当てられ、施策・政策レベルでの因果関係が明確でない。また、これらのツールを用いた評価結果は、施策や政策へのフィードバックがされにく

い。RFでは、施策・政策レベルでの因果関係は明確であるが、個別の事業が必ずしも含まれないため、施策・政策と個別事業との因果関係が不明確となり、乖離が生じてしまう。

PTMでは、政策・施策・事業の3層におい て因果関係がより明確となり、また個々の 事業におけるセオリーを統合することで、 一つの政策体系内において、政策・施策・ 事業のあり方、すなわちプログラム・セオ リーについての分析が理論上可能となる。 具体的には、政策・施策の枠内において事 業の位置づけや各事業の関連性がより明確 となるのみならず、施策の整合性や政策立 案の適否を分析するといったことも可能と なる。さらに、事業実施主体において評価 結果のフィードバックが有効に機能してい るという前提のもとでは、政策決定者によ る評価結果の活用ポテンシャルが大きくな るため、このようなツールを用いてプログ ラム・セオリー評価を実施することの意義 は大きい。

他方、実際にプログラム・セオリー評価を行う際には、それぞれのツールにおける使用状況の中にも見て取れるように、さまざまな制約や問題点が存在する。これらの影響力を最小限にとどめ、あるいは可能な限り排除するためにも、ツールは適切な方法において使用される必要があり、各ツールの使用法に関して今後議論がより一層深まることが期待される。

3. 政策評価におけるプログラム・セオリーの明確化の必要性

(1) 政策評価制度導入の背景と目的

近年の日本の行政改革として、2001年1月の中央省庁等の再編に伴い、2002年4月には「行政機関が行う政策の評価に関する法律」(以下、評価法)が施行された。政府レベルでは通称「政策評価」と呼ばれている本制度導入の目的は、「国の行政機関が主体となり、政策の効果等に関し、測定又は分析し、一定の尺度に照らして客観的な判断を行うことにより、政策の企画立案やそれに基づく実施を的確に行うことに資する情報を提供すること」(総務省 2001)とされる。

尚、評価法の施行に先立ち、1999年8月に 総務庁行政監察局(当時)の研究会として 発足した「政策評価の手法等に関する研究 会」により、2001年1月には「政策評価に関 する各府省共通のルールを定めるもの」(政 策評価の手法等に関する研究会2000)とし て「政策評価に関する標準的ガイドライン」 (以下、ガイドライン)が公表されている。 また、2001年12月には、「政策評価に関する 基本方針」(以下、基本方針)が閣議決定された。さらに、総務省行政評価局に関策と 2003年7月には「各府省が実施した政策評価についての審査・平成14年度総括・」(以下、審査報告書)が公表されている。本節での 考察は主にこれらの文献に依拠する。

(2) 政策評価における「政策」の概念とプログラム・セオリーの重要性

行政側からみた行政機能を大別すると、政策(Policy)・施策(Program)・事務事業(Project)の3層構造となるが(大住1999;上山1999)、総務省が定めるところの「政策評価」とは、ここでの政策(Policy)を対象とした評価のことではなく、この3層全てを包括する(広義の)政策を対象とする評価のことを指す。ガイドラインでは、この3層に

ついて以下のように定義されている(総務省2001)。

「政策(狭義)」:特定の行政課題に 対応するための基本的な方針の実現を 目的とする行政活動の大きなまとまり。 「施策」:上記の「基本的な方針」に基 づく具体的な方針の実現を目的とする 行政活動のまとまりであり、「政策(狭 義)」を実現するための具体的な方策や 対策ととらえられるもの。「事務事 業」:上記の「具体的な方策や対策」 を具現化するための個々の行政手段と しての事務及び事業であり、行政活動 の基礎的な単位となるもの。

尚、本稿では、広義の政策を指す場合には「政策」、狭義の政策を指す場合には「(狭義の)政策」のように区別する。

行政活動においては、「行政の大局的な目 的や方向性を示すもの」(行政クリアリング ハウス)として策定された(狭義の)政策を 実現する手段として施策が策定され、また 施策における目標を実現するため、それを 具現化する手段として、施策の枠内におい て事務事業が計画・実施されるという論理 が成り立つ。ここでは、(狭義の)政策・施 策・事務事業が相互に目的と手段の連鎖関 係を保ちつつ、全体として一つの政策体系 を形成していることが想定されている。す なわち、これら3層からなる政策体系はプロ グラム・セオリーの考え方が見てとれる。ま た、第2節での国際協力評価についての議論 に基づくと、(狭義の)政策が最終成果(End Outcome) 施策が中間成果 (Intermediate Outcome) さらに事務事業が結果 (Outputs) と同じレベルにそれぞれ対応する。ここで は、行政サービスの受益者である国民のた めの成果重視の視点から、事業実施による 結果の達成よりはむしろ、社会やターゲッ ト・グループの変化としての最終成果や中 間成果を達成すること、すなわち(狭義の) 政策あるいは施策目標の実現が重要といえ

る。そうした見地からも、政策評価では、これら3層の論理構成が適切な因果関係に基づき、結果の達成が中間成果及び最終成果の達成へと寄与しているかを確認すること、すなわち講義の政策である政策体系を構成するプログラム・セオリーを検証することが重要となる。以下、政策評価の実施状況を踏まえ政策評価におけるプログラム・セオリーのより一層の明確化の必要性を考察する。

(3) 政策評価の標準的方式と現状

ガイドラインによると、政策評価における標準的な評価の方式としては、実績評価、事業評価、総合評価の3つが挙げられ、各府省は所掌する政策の特性等に応じて、これら3つの標準的な評価方式や、これらの主要な要素を組み合わせるなどして適切な評価の方式を採用するとされている(総務省 2001)。

実績評価

実績評価は、「政策を決定した後に、政策 の不断の見直しや改善に資する見地から、 あらかじめ政策効果に着目した達成すべき 目標を設定し、これに対する実績を定期 的・継続的に測定するとともに、目標期間 が終了した時点で目標期間全体における取 組や最終的な実績等を総括し、目標の達成 度合いについて評価する方式」(総務省行政 評価局2003a)とされている。ガイドライン では当初、施策を中心とした評価とされて いたが、審査報告書では、政策効果として の成果 / アウトカムに着目した評価と記さ れている。実際に各府省が行った実績評価 の状況をみると、(狭義の)政策及び施策が 実績評価の対象に含まれている。また、目標 の達成度を測定するため、目標設定はでき る限り定量的に行うとされており(総務省 行政評価局 2003a) 実績評価は主に(狭義の) 政策・施策レベルの業績測定(Performance Measurement)を想定した評価となる。

他方、実績評価は、(狭義の)政策・施策 レベルでのアウトカムに着目しており、そ のために、まず「目的 - 手段」としての (狭義の)政策と施策、あるいは施策と事務 事業の因果関係、すなわちプログラム・セ オリーを明確にすることで、既存の政策体 系の整合性を判断することが重視される。 さらにその上で、政策成果としてのアウト カムに関する指標と数値目標、さらに目標 達成までの期間(目標期間)を設定するこ とで、アウトカムの達成度を測定し判断す る。逆に言うと、政策効果としてのアウト カムを適切に把握できなければ、プログラ ム・セオリーの検証及び修正がより難しく なり、特に(狭義の)政策・施策レベルで の政策立案への評価結果のフィードバック が適切に行われ得なくなる。

審査報告書によると、平成14年度分。に関しては8府省。から計471件の評価書が総務大臣に送付されたが、これらの8府省のうち、政策における目的と手段の関係、すなわち(狭義の)政策と施策、あるいは施策と事務事業の因果関係を明示的に整理しているのは2省でのみとなっている(総務省行政評価局2003a)。また、上述した8府省のうち農林水産省を除く7府省において、目標期間(基準年次及び達成年次)が明確に設定されていない政策が多く、評価の対象とされる政策を全体のうち目標期間が設定されていないものの割合が6割強であることも指摘される(総務省行政評価局2003a)。

実際の行政活動では、分野によっては (狭義の)政策・施策・事務事業の3層の因 果関係を明確にしにくい場合や、政策にお ける数値目標及び指標の設定が難しい場合 があることも否定できない。しかしながら、 審査報告書にも述べられているように、実 績評価の主旨に沿って評価実施のあり方を 改善し、目標設定の際には目標期間を定め、 できる限りアウトプットではなくアウトカ ムベースで指標や目標を設定することが必 要となる。

事業評価

事業評価は、「政策を決定する前に、その 採否、選択等に資する見地から、あらかじ め期待される政策効果やそれらに要する費 用等を推計・計測し、政策の妥当性、必要 性、効率性などの観点から評価するととも に、必要に応じ事後の時点で事前の時点に 行った評価内容を踏まえ検証する方式」(総 務省行政評価局 2003b) であり、個々の事業 や施策の実施を目的とする政策を対象とし た評価である。事業評価は政策決定の時期 を境にして事前評価と事後評価に分かれる。 事前評価は、主に予算概算要求の際に実施 される。一方、事後評価は、政策決定後に 政策効果を把握し、政策の見直しや政策立 案にフィードバックするための情報を得る ことを目的とする。公共事業の事後評価で はさらに、事業採択後から事業完了前に行 われる「再評価」(事業未着手のものも含む) と、事業完了後の評価とに分かれる。(総務 省行政評価局 2003a)。

事前評価は、2003年3月の時点で、一定以上の事業規模を有する個々の研究開発、公共事業及びODA事業®の実施を目的とする政策を決定する際に、実施が義務付けられている。その一方で、それら以外の政策では実施が義務付けられていないものの、「政策評価に関する基本方針」によると各府省は順次実施に向けて取り組むこととされる。

事業評価の実施状況を公共事業評価に例をとると、審査報告書によると、平成14年度分としては3省(農林水産省、経済産業省、国土交通省)から計9,560件の評価書が総務大臣に送付された。そのうち事前評価は7割強の6,934件であり、評価手法は、上述した3省のうち前2者は費用便益分析を活用するともにその他の指標を用いた総合的なにでいる。一方、後者は費用便益分析を実施したものと、費用便益分析を実施したものと、費用便益分析を実施したものと、費用便益分析を対応される。また、再評価は1,607件が報告されている。その手法は、上述した3府省のの他の後2者は費用便益分析を活用しつつその他の

指標を用いた総合的な評価を実施している。一方、前者は基本計画上、費用便益分析を実施せず¹⁰、費用対効果分析の算定基礎となった要因の変化等を総合的に評価する方式をとっている。完了後の評価に関しては1,008件全てが農林水産省から報告されているが、再評価と同様の手法が用いられている(総務省行政評価局 2003a)。

事業評価では政策の妥当性、必要性、効率性などの観点から評価するとされており、公共事業評価は、費用便益分析等の定量的な手法が基本的に用いられているものの、一部では定性的な分析も合わせて、これらの3つの観点について包括的に評価しているものと理解される。その一方で、各府省において評価の仕組みの一層の改善が必要であることも指摘される(総務省行政評価局2003b)。

では、どのような仕組みの改善が考えら れるであろうか。前節の国際協力評価の考 察を踏まえれば、筆者は、主に行政サービ スの受益者である国民の視点から、社会の 変化としてどのような最終成果が期待され ているのかをあらかじめ明確にしておいた 上で、その目的を達成するための手段とし て、(狭義の)政策・施策の枠内において 個々の事業を位置づけることが必要となる と考える。また、そうした政策体系を想定 することで、政策の妥当性や有効性、イン パクトを、例えばDAC評価5項目の観点から より体系的に分析し、費用便益分析等の手 法と合わせて検証することも可能となる。 その際には、事前評価・再評価・完了後の 評価を合わせて実施していくことで、政策 体系に基づき選択したセオリーを一貫して 検証することが可能となり、またそれらの 評価結果が後の政策立案へと適切にフィー ドバックされることが期待される。

総合評価

総合評価は、「政策の決定から一定期間を 経過した後を中心に、特定のテーマについ て、当該テーマに係る政策効果の発現状況

を様々な角度から掘り下げて分析し、政策 に係る問題点を把握するとともにその原因 を分析するなど総合的に評価する方式」(総 務省行政評価局 2003b) とされる。審査報告 書における各府省の基本計画の概要をみる と、(狭義の)政策・施策を対象とした Policy/Program Evaluationにほぼ対応している ことがわかる1。総合評価は、主に事後評価 が中心であり、特定テーマに関する政策の 見直しや改善が主眼とされている。それゆ えに、特に(狭義の)政策・施策レベルで の因果関係と期待されるアウトカムをあら かじめ明確にし、政策体系を想定しておい た上で、プログラム・セオリーに基づきい ろいろな手法を組み合わせた形で評価を実 施するのが有効となると考える。

審査報告書によると、平成14年度分に関しては4省から計188件の評価書が総務大臣に送付されたが、これらのうち外務省をは3省からの29件¹²についてみてみるととず防衛庁の16件に関しては、テーマが防衛庁の16件に関してはり説策の見を述べることにより説策の見を立とを主眼として実施されたのでは発生を主まが指する。その一方で、文部科学省および国土交通省の計13件に関しいに許している。

総合評価の実施にあたっては、審査報告書によると、各府省が問題意識に基づいてとの必要性が指摘されている。その一方方で、評価の方法に関しては、また実際にはテーロの方法に関しては、また実際にはでいるとされるなど、現時点でに統領をとされるなど、現時点でに統領をとされるなど、平成14年度省の大大政評価局2003a)。実際、平成14年度省の大大政評価を実施する予定としていた12府付はの方ち、平成15年5月末までに評価書を送付のしたのは4府省にとどまっているなど、現のがでは経験がそれほど蓄積されていないのが

現状である。

(4) まとめ

本節におけるこれまでの要点をまとめる と、まず事業評価は、一定以上の規模を有 する事業に対して実施が義務付けられてお り、政策評価の主な目的の一つとされる国 民本位の視点からは、事業実施におけるタ ーゲット・グループや社会の変化としての アウトカムをあらかじめ明確にすることが 重要となる。例えば、公共事業においては 主に費用便益分析等の定量的な評価手法の 開発が進んでいるとされるが、公共事業評 価において政策の妥当性や有効性、インパ クト等を定性的な手法と合わせてより体系 的に分析するためには、(狭義の)政策・施 策の枠内において個々の事業を位置づけ、 プログラム・セオリーの視点から検証する ことが有効である。実績評価は、政策・施 策レベルの業績測定を想定した評価と考え られるが、まずは(狭義の)政策・施策・ 事務事業の3層の因果関係を明確にした上 で、政策成果としてのアウトカムに関する 指標と数値目標、さらに目標達成までの期 間(目標期間)を設定することで、アウト カムの達成度を可能な限り定量的に測定す る必要がある。またそうすることで、アウ トカムの達成度を判断することが可能とな る。総合評価は、特定テーマに関する政策 の見直しや改善を主眼として、総合的な方 式で行うとされているため、特に(狭義の) 政策・施策レベルでの因果関係と期待され るアウトカムをあらかじめ明確にし、政策 体系を想定しておいた上で、プログラム・ セオリーに基づきいろいろな手法を組み合 わせた形で評価を実施するのが有効となる。

以上を勘案すると、政策評価において重要なのは、その方式の区分に関らず、(狭義の)政策・施策・事務事業の構成要素間の 因果関係をまずは明確にし、その論理構成 の適正さを適宜検証することといえる。そ のためには、適切な手法及びツールを用い、 あるいはそれらを組み合わせて評価設計を 適切に行った上で評価を実施することが重 要となってくる。いずれにせよ、政策評価 は、今後、更なる検証が必要であり、また、 事例研究を重ねることが必要である。その 際には、それに伴い日本の行政におけるプ ログラム・セオリーの明確化が期待される。

4. おわりに

国際協力評価及びわが国の政策評価は実務的な要請を背景として開始されており、 実務経験を踏まえ実践的な手法として発展 してきた経緯がある。それゆえ、いまだ理 論的な位置づけが弱い。

他方、政策評価では行政機能における (狭義の)政策・施策・事務事業からなる3 層のプログラム・セオリーの考え方に基づ き、事業評価、実績評価、総合評価の3方式 による評価が行われている。しかし、実態 面では、政策、施策、事務事業の因果関係 を明確にすべき評価対象としてのプログラ ム・セオリーの考え方が的確に反映されて いるとは言い得ない。

こうした状況を考慮すれば、より有効な プログラム・セオリーの構築が、国際協力 評価及び政策評価をより有効にするために は重要かつ不可欠であり、その構築方法に 注力することは緊急な課題であると考える。 そのためには、PTMの活用などを踏まえた 実践などの経験を積み上げていくとともに、 プログラム・セオリーを基にした評価手法 を検討していくことが有効であり必要であ る。

注記

- 1 Chen, Weiss, Rogers, Petrosino, Huebner, Hacsiをはじめプログラム・セオリー評価を定義した多くの評価専門家は、プログラム・セオリー評価をフォーマティヴな評価であると性格付けている。
- 2 「Systematic Approach」でRossi、Freeman、 Lipseyも、1999年に出版された第6版から は、プログラム・セオリー評価に係る単 独の章を追加した。
- 3 各用語の定義に関しては、例えば行政クリアリングハウス「用語集」(2003年3月31日アクセス)や開発援助委員会援助評価作業部会(2003)を参照されたい。
- 4 「行政機関が行う政策の評価に関する法律」(第2条第2項)を参照されたい。
- 5 実際には、平成14年4月1日から平成15年5 月31日までの期間に送付されたものを指 す。以下同様。
- 6 ここでの8府省とは、総務省、財務省、文 部科学省、農林水産省、環境省、金融庁、 公正取引委員会、厚生労働省を指す。
- 7 ここでの2府省とは、農林水産省と環境省を指す。
- 8「一定以上の事業規模」とは事業費10億円以上を指すが、ODA事業に関しては、無償資金協力は供与限度額が10億円以上、有償資金協力は150億円以上のものが評価対象となる(総務省行政評価局 2003a)。尚、平成14年度分のODA事業評価に関しては報告がされていないため、本稿では扱わない。
- 9 同一の事業について再度評価が実施され た38件も含む(総務省行政評価局 2003a)。 以下同様。

- 10 但し、林野公共事業及び水産関係公共事業については試行的に費用便益分析を実施している。
- 11 各用語の定義や手法に関しては、例えば 行政クリアリングハウス「用語集」(2003 年3月31日アクセス) 上山(2002) 三好 他(2003) 龍・佐々木(2000)を参照さ れたい。
- 12 外務省からの159件に関しては、送付時期 が平成15年5月末ということもあり、審査 報告書では扱われていないため、本稿に おいても扱わないこととする。

参考文献

- 上山信一(1999)『「行政経営」の時代 評 価から実践へ』NTT出版
 - (2002) 『日本の行政評価 総括と 展望』第一法規
- 大住荘四郎 (1999) 『ニュー・パブリック・マネジメント 理念・ビジョン・戦略』 日本評論社
- 開発援助委員会援助評価作業部会(外務 省・国際協力事業団・国際協力銀行・日 本評価学会訳)(2003)『評価および結果 重視マネジメントにおける基本用語集』 外務省
- 国際開発高等教育機構(2001)「第1章 PCM手法の方法論と特色」PCM読本編集 委員会編『PCM手法の理論と活用』国際 開発高等教育機構
- 国際協力事業団編(1998)『国際協力用語集 [第2版]』国際開発ジャーナル社
- 政策評価制度の法制化に関する研究会 (2000)『政策評価制度の法制化に関する 研究会報告』総務庁
- 政策評価の手法等に関する研究会(2000) 『政策評価制度の在り方に関する最終報 告』総務庁
- 原口孝子(2001)「第3章 JICAプロジェクトにおけるPCM参加型計画手法の適用事例」PCM読本編集委員会編『PCM手法の理論と活用』国際開発高等教育機構

- 藤村建夫(2001)「第8章 海外援助機関に おける開発協力事業の質的管理」PCM読 本編集委員会編『PCM手法の理論と活用』 国際開発高等教育機構
- 三好皓一(2002a)「プログラム・セオリー・マトリックスの活用について 評価における分析の有用性を高めるために 」『日本評価研究』2(1)
 - (2002b)「政策体系におけるプログラム評価の位置づけ:プログラム評価とプロジェクト評価」『日本評価学会第3回全国大会報告論文集』
- 三好皓一・平田慈花・和田知代・中澤哉・ 喜多悦子(2003)「プログラム評価の試 み・フィリピン感染症対策分野の評価を 事例として・」『日本評価研究』3(1)
- 龍慶昭・佐々木亮(2000)『「政策評価」の 理論と技法』多賀出版
- Binnendijk, Annette (2001). Results Based Management in the Development Co-operation Agencies: A Review of Experience-Background Report. Organization for Economic Cooperation and Development: Paris, France.
- Chen, Huey-Tsyh. (1990). *Theory-Driven* Evaluations. Newbury Park, California: Sage.
- Cook, T. D. (2000). The False Choice Between Theory-Based Evaluation and Experimentation. Program Theory in Evaluation: Challenges and Opportunities. *New Directions For Evaluation*. No 87, Fall.
- Davidson, E. J. (2000). Ascertaining Causality in Theory-Based Evaluation. Program Theory in Evaluation: Challenges and Opportunities. *New Directions For Evaluation*. No 87, Fall.
- Hacsi, T. A. (2000). Using Program Theory to Replicate Successful Programs. Program Theory in Evaluation: Challenges and Opportunities. New Directions For Evaluation. No 87, Fall.
- Petrosino, A.(2000). Whether and Why? Potential Benefits of Including Program Theory Evaluation in Meta-Analysis. Program

Theory in Evaluation: Challenges and Opportunities. *New Directions For Evaluation*. No 87, Fall.

Rogers, J. P. (2000). Causal Models in Program Theory Evaluation. Program Theory in Evaluation: Challenges and Opportunities. *New Directions For Evaluation*. No 87, Fall.

Rogers, J. P., Petrosino, A., Huebner, T. A., and Hacsi, T. A. (2000). Program Theory Evaluation: Practice, Promise, and Problem. Program Theory in Evaluation: Challenges and Opportunities. *New Directions For Evaluation*. No 87, Fall.

Rogers, Patrica, J., A. Pertosino, T. Hueber, (2000), Program Theory Evaluation: Practice, Promise, and Problems. In Rogers, Patrica, J., T. Hacsi, A. Pertrosino, T. Huebner (Eds.), Program Theory in Evaluation: Challenges and Opportunities, San Francisco, Jossey-Bass

Stufflebeam, Daniel L., (2001), Evaluation Models, *New Direction for Evaluation*, AEA, San Francisco, Jossey-Bass

Rossi, P. H., Freeman, H. E. and Lipsey, M. W. (1999). *Program Evaluation: A Systematic Approach*. Thousand Oaks, California: Sage.

Weiss, C. H. (2000). Which Links in Which Theories Shall We Evaluate? Program Theory in Evaluation: Challenges and Opportunities. *New Directions For Evaluation*. No 87, Fall.

Weiss, C. H. (1998). *Evaluation*. New Jersey: Prentice Hall.

行政評価クリアリングハウス「用語集」 http://www.gyousei.gr.jp/reports/words.html (2003年3月31日アクセス)

総務省(2001)「政策評価に関する標準的ガ

イドライン」

http://www.soumu.go.jp/kansatu/gaido-gaidorain1.htm

(2002年10月16日アクセス)

総務省行政評価局(2002)「各府省が実施した政策評価についての審査の状況 - 平成14年度 第1次分 - 」

http://www.soumu.go.jp/s-news/2002/ 021205_3_01.html

(2003年7月9日アクセス)

(2003a)「各府省が実施した政策評価についての審査 - 平成14年度総括 - 」

h t t p : // w w w . s o u m u . g o . j p / s - news/2003/030711 2.html

(2003年8月1日アクセス)

(2003b)「各府省が実施した政策評価についての審査 - 平成14年度総括 - (報道資料)」

http://www.soumu.go.jp/s-news/2003/ 030711_2_yosi.pdf

(2003年8月1日アクセス)

(2003c)「各府省が実施した政策評価についての審査の状況 - 平成14年度第2次分 - 」

http://www.soumu.go.jp/s-news/2003/ 030407_2_01-00.html

(2003年7月9日アクセス)

内閣府 (2001) 「国の研究開発評価に関する 大綱的指針 (平成13年11月28日 内閣総 理大臣決定)」

http://www8.cao.go.jp/cstp/hyoukasisi.pdf (2003年7月17日アクセス)

(2003.8.14受理)

Toward Constructing More Suitable Program Theory for Japan's Evaluation -- Focusing on International Cooperation Evaluation and Policy Evaluation --

Koichi Miyoshi

Ritsumeikan Asia Pacific University miyoshik@apu.ac.jp

Satoshi Morita

Yoshio Aizawa

Graduate School at the University of Tokyo

International Development Center of Japan

Abstract

The purpose of this paper is to discuss Japan's international cooperation evaluation and policy evaluation focusing on program theory from the viewpoint of program theory evaluation. Methodology of Japan's international cooperation evaluation and policy evaluation have been developed heavily depending on actual administrative works so that they are weak at providing clear and accurate program theory. Consequently the utilities of evaluations are reduced. Considering this situation, to advance studies on how to construct more accurate and suitable program theory is indispensable and an acute issue for improving the effects in Japan's international cooperation evaluation and policy evaluation.

Keywords

program theory evaluation, international cooperation evaluation, policy evaluation, policy structure, outcome

【研究論文:依頼原稿】

実用重視評価の理論と課題

長尾 眞文

広島大学 nagaom@hiroshima-u.ac.jp

要約

事業評価の理論的関心は手法の選択に係わるもので、「定量的」対「定性的」手法のパラダイム論争がその中核を占めてきた。そこでの中心的課題は評価結果の科学性、客観性、一般化可能性であったが、評価の有用性の視点から両者の実践的な統合を提起したのがMichael Q. Pattonの実用重視評価理論である。この理論は、評価の有用性を結果の活用で測るものとし、評価者の基本的な役割として結果の活用に強い関心を持つ利害関係者の特定化と、評価プロセスへの参加を通じる当事者意識の向上を掲げる。本論文では、はじめに実用重視評価誕生の背景となる論争について概説した後で、実用重視評価の基本的な考え方の紹介、その実践の例示および問題点の指摘をする。最後に実用重視評価の日本における活用に関する課題に触れる。

キーワード

実用重視評価、定量的評価、定性的評価、評価結果の活用、評価可能性アセスメント

はじめに

事業評価は、社会科学の研究手法を用いて事業や組織の価値やメリットを判断することから始まった。もちろん当初から独自の学問分野を形成していたわけではなく、むしろ経済学、教育学、心理学、行政学、会計学をの各分野の応用的関心として成立していた。応用的関心は、大別すると3つで、経済開発における投資の意思決定、公共部門における投資の意思決定、公共部門におけるを立るとして教育、医療、福祉等の社会的・人的サービス分野における介入の効果測定であった。これらはいずれも個別具体的な事業を対象とし、投資決定は費用便益分析による

事前評価、監査は会計基準に基づく事後評価、 社会的・人的サービスの効果測定は介入の ザインに応じた事後評価が中心であった。事 業評価の学問的展開は、前二者については を形成したが、3番目の社会的・人のサービスの 事業の出発からがよって、分野に通したが、3番目の社会的・分野に 事業の出発から始まって、方法的な共通に り徐々に学問的領域を共有するようにとも た。また応用的関心が多岐にわたる事業 た。また応用の分析対象は画一のは事業で た。また応アの分析対象は両一のな事業 にとざまらず、事業開始が 中の評価も包含する多様な展開を遂げて 中の神では、この社会的・人的サービス事 58 長尾 眞文

業の評価に的を絞って、その理論的展開の中で重要視されるようになった「実用重視」の視点²を紹介し、主たる研究課題について述べることにする。

そもそもの出発が応用的関心にあったこと から、事業評価理論の展開の底流には常に実 践的な問題意識がある。この分野の代表的な 著作のタイトルにも「実践 (practice)」、「実 践的 (practical)」、「現実的 (realist)」といっ た形容のつくものや、「政策やプログラムの 検討」や「理解、指導、改善」と応用目的を 明示したものが目立つ。その中で、わざわ ざ「実用重視」と称されるような理論が誕生し たのには、この分野の理論的展開の場となっ た北米、特に米国における社会的状況とそれ に関連する評価手法をめぐる論争が重要な影 響をもった。基本的な問題は、多額の費用を かけた大規模な調査を通して実施した評価の 結果が必ずしも確定的な判断に導かず、しか も評価結果が活用されないばかりか悪用・誤 用されるという事実で (Weiss 1997; Patton 1997; House 2001) 1970年代から80年代にか けて評価の最も普及していた米国で浸透した この認識は評価結果の有用性や活用に対する 関心を拡大した。その中で評価者が評価結果 の利用者の関心や意向に配慮することにより 評価が実際に役立つことを主眼として展開さ れたのが実用重視評価の考え方である。評価 の活用関心に沿って展開された理論や技法は いろいろあるが、本稿では主としてそれを標 題とする著書を出している米国の評価専門家 Michael Q. Patton (1997) の理論に着目する ことにする。

1. 実用重視評価誕生の背景

(1)「定量的手法」対「定性的手法」

実用重視評価誕生の主たる背景は、評価を 行うのに「定量的手法」を用いるべきか、「定 性的手法」を用いるべきかの論争である⁵。定 量的手法とは、事業の想定している因果関係 について数字あるいは数量化された量的デー タと標準化された統計モデルを用いて演繹 的、仮説検証的に分析する手法である。数値 指標を使って介入の成果の事前・事後の比較 や、類似案件の成果との比較を行うことによ り客観的に検証する評価のことである。定性 的手法とは、調査対象に関する記述や直接の 会話から得られる文字情報を事業の全体的な 文脈の中で帰納的、内容主題検出的に分析す る手法である。定量的手法は、方法的により 客観的、科学的とされ、特に事業対象グルー プに対して比較(制御)グループを設ける実 験デザインに基づく評価では、結果の信頼性、 一般化可能性も保証される。反面、そもそも のモデルの設定に評価者の方法的バイアスが 影響する、適当な量的データが得られない場 合の対応、事業の文脈の特異性を反映できな い、方法に忠実になると費用が多額になる等 の問題を抱えている。一方、定性的手法は、 評価者が事業関係者と直接に接触することか ら、事業の固有な状況、多様性、事業期間中 の条件変化等を評価結果に反映できる。しか し、評価者の主観的立場からの制約、結果の 一般化の困難等の問題から抜けられない。

この論争に一定の終止符を打つ形で登場し たのがPattonに代表される実用重視評価論者 で、定性的評価に近い立場から、評価の「科 学性」にとってかわる「有用性」を根拠とす る実践理論を提示した。 事業評価で最も重 要なのは結果が役に立つことで、何が役に立 つかは、評価の目的による。事業の背景にあ る因果関係の想定の厳密な検証が必要な場合 にはCampbell (1969) 流の実験デザインを用 いた定量的手法が不可欠だが、事業をめぐる 環境条件の変化等で複雑な評価を迫られる場 合には、より柔軟性の高い定性的手法の方が 適切との立場である。しかも一般に、ほとん どの評価は事業や組織の改善が目的で、方法 的な科学性や客観性よりも、事業関係者の行 動の指針となるか否かの実践的価値が優先さ れる。従って、ひとつの評価の中で定量的手 法と定性的手法を組み合わせる実用的アプロ ーチとして主流になりつつある (Smith 2001: Mark et. al. 2001 》。 その典型は、事業評価のために行うアンケート調査で、数量的な回答を求める質問(例えば、ある事象に関する「良い」、「悪い」、「どちらとも言えない」の3段階評価)と自由記述の質問を並べることである(Patton 2002, p. 5)。

(2) 評価における価値の選択

評価は価値判断を伴う行為である。そこで評価の根拠となる価値をどこに求めるかが重要な課題となる。社会科学の応用として評価の慣行が始まった当初は、手法の科学性、中立性の陰に隠れていたが、有用性基準が提起されると、評価者と事業関係者の間であるいは社会一般の価値との関係で、どのような価値の選択が為されるべきかが問題として浮した。特に重要なのは、評価結果の影響を受けることが予想される利害関係者(Stakeholders)との間に評価者が置くべき距離である(Mark 2001)。

利害関係者が評価を受け入れ、評価プロセ スを理解し、結果を活用する可能性を高める ために、実用重視評価者は、利害関係者の関 心や意見を重視する。Pattonは、利害関係者 の中でも特に評価結果の想定利用者(例えば、 行政組織が行う補助金事業の評価の場合、評 価結果を事業の改善に活用する可能性のある 事業担当者)に注目するように説く(Patton 1997)。このような実用的関心の延長線上に、 参加型評価が登場する。評価者が利害関係 者に近づくに連れて、評価判断の独立性が問 題になる。実用評価論者の立場は、評価者が 評価学会等の定める評価および評価者の基準 を尊重することと、それらの基準に則したメ 夕評価を受け入れることで、倫理性を確保し ようとする。

(3) 評価の文脈の考慮

評価に当たって、事業の文脈を構成する諸要素(例えば、事業参加者の特徴、事業の地理的・歴史的条件、不測の事件といった介入

行為以外の条件)を考慮することが必要なこ とは言うまでもない。しかし、事業の背景に ある因果関係想定が実現したか検証しようと する評価では、当然文脈の特異性や特殊条件 の影響を可能な限り排除しようとする。そう することにより、評価結果が仮説の科学的な 検証に基づく一般化に支えられるからであ る。これに対して、定性的手法論者は、文脈 の有意性を主張し、評価は一般化よりも個々 の評価案件の特殊性に注目すべきであるとす る10。 実用重視評価者の立場は、両者の中間 にあり、無数に行われてきた評価の蓄積を生 かして、質の高い経験学習を重ね、「ベスト・ プラクティス」の知識を普及すべきと考える (Patton 2001)。この考え方は、経営分野で普 及している知識管理の流れを汲むもので、方 法的には特定分野で行われた複数の評価を比 較し、「教訓」を抽出するメタ分析が採用され る (Mark 2001)。

(4) 評価結果の活用

以上の背景の議論に共通しているのは、実 用重視理論の評価結果の活用へのこだわりで ある。評価結果の使い道としては、行動の指 針、既存の考えの追認・補強、啓蒙等がある (Weiss 1997)。 行動の指針は事業に関する政 策的判断に役立てるといった評価の積極的、 手段的な活用を意味している。それ以外は概 念的な活用である。前述したように、米国で は1970年代から80年代にかけて、評価結果が 役に立たないとの認識が浸透したが、それは 主として連邦政府が行った社会福祉、教育、 保健関連の事業について実施された大規模な 評価が政策的判断に活用されなかったからで ある。活用されなかった理由は多々あったが、 決定的要因は外部的、中立的、科学的配慮に 基づいて伝統的な定量的手法で実施された総 括評価が政策担当者や事業担当者の意見や意 向を反映しなかったからであった (Patton 1997)。その反省から生まれた実用重視評価 アプローチでは、評価者が意図的に事業関係 者に接近し、評価を開始する前に評価結果の 60 長尾 眞文

活用について協議することを旨とするととも に、評価作業の重点を既存事業の改善に向け た形成的評価にシフトした。

2. 実用重視評価:基本的な考え方

(1) 評価者の役割

実用重視評価で最も重要な原則は、個々の 評価の目的に従って結果の有用性を極大化す ることである。有用性は評価結果の活用によ って決まるから、評価者はまず評価の目的を 見極めるとともに、評価結果により影響を受 ける利害関係者、特に結果の活用に関心を持 つと想定される層を特定化することから開始 する。評価の利害関係者は、評価の委託者、 事業の責任者、管理者、実施担当者、参加者、 受益者等々多様であり、それぞれの評価に対 する関心はもちろん一様ではない。評価者は、 それぞれの関心を分析し、評価の目的別(例 えば、総括評価 (summative evaluation) 形 成評価 (formative evaluation)、知識創出評価 (knowledge-creating evaluation) 実施評価 (implementation evaluation) 等々)に応じて、 利害関係者の評価への関与のパターンを決定 する必要がある。Pattonによると、そこでの 評価者の基本的役割は「交渉者」的性格である (Patton 1997, p.122),

(2) 利害関係者の評価作業への参加

実用重視評価では、評価結果の想定利用者を評価作業に参加させることにより、関係者相互の事業に関する理解の共有、事業に対する当事者意識の向上、事業への関与の拡大を導き、結果として評価結果の活用を担保できると考える。参加は、評価の目的、方法、全体的枠組みの決定を行う当初の段階が最も戦略的に重要であるが、評価目的によっては(例えば、事業の改善を目的とする形成評価の場合など)、データ収集から分析、さらに結論・提言の導出の段階でも、効果的な結果

を生む可能性が高い。評価者の側からすると、初めは想定利用者に対して参加を働きかけ(能動)次に参加が生み出す状況に応じ(反応)そして適切な結果に導く(適応)プロセスとなる(Patton 1997, p.134)。参加する結果利用者からすると、対話、検討、設問、自答の4つの学習プロセスを経験することになる(Preskill and Torres 1999, pp.51-52)。

(3) 成果重視の評価デザイン

評価デザインは通常評価の目的に従って決 定される。実用重視評価においてもその通り であるが、その決定の際に事業の成果を特に 重視する。成果を評価結果の利用者が望まし いと考える形で規定し、それに基づいて評価 を組み立てることにより、結果の活用の確率 は高まると考えられるからである。成果重視 の枠組みは、事業対象者(受益者)グループ そのグループにとって を特定化した上で、 望ましい事業成果の定義、 それを表す成 果指標(および必要であれば達成目標)の設 データ収集方法の選択を決定すること 定、 で構成される。もちろんこの枠組みが完成す るには、介入と成果とを結びつける因果関係 に関する想定が不可欠である。評価者は、事 業関係者および利害関係者とこの枠組みにつ いて繰り返し協議することにより、彼らが頭 に描いている変化の理論(プログラム・セオ リー)を明示化させるだけでなく、内部的だ けでなく対外的にも説明可能な想定を導くよ うにする。また、この評価デザインから導か れる結論が活用される方法も同時に規定す る。協議は、評価のデザイン段階だけでなく 評価作業が進行中にも繰り返され、必要に応 じて改定される。

(4) データの分析、解釈、判定

実用重視評価においてもデータの分析、解 釈、判定のプロセスで実施する作業は基本的 に他の評価と内容的に差異はない。必要に応 じて定量的データ・手法も定性的データ・手法 も使うし、判定に何らかの比較を持ち出する 程に何らかの比較を持ち出する。 ただ実に何らかの比較を持ち出すの 程に結果の想定利用者を参加することである。 主を導出する。 主を持つないである。 を持つなが、が、 を事前になずるは、 を事がしている。 を事がしている。 を事がしている。 を事がしている。 を事がしている。 には、 を事がしている。 でいる。 には、 を事がしている。 には、 を事がしている。 には、 を事がしている。 には、 をない、 には、 にない、 にな

(5) 提言と報告

評価者の中にはScriven (1993) のように、 評価者は事業の過去について「判断を提供す る」のが仕事であって、将来に関する提言は すべきでない、と考える論者も居るが、実用 重視評価では結果の利用を最重要視している ことから、それとは正反対の立場をとる。し かし、実用重視評価でも、将来の不確実性や 予測不能性を消すことはできない。但し、提 言の仕向け先となる評価結果の想定利用者が 評価に参加することにより、提言が実行に移 される可能性は、より高くなる可能性がある。 その場合でも、複数のシナリオを想定するこ と等により、提言の活用に弾力性を与える工 夫が必要である。実用重視評価では、提言も 含めて報告に盛られる内容は、評価実施の過 程で想定利用者を中心に利害関係者に知らさ れているので、その意味では、通常の評価で 提言・報告が一種のクライマックスを意味す るようなことはない。むしろ結果の活用を促 進するための機能のひとつと考えられている (Patton 1997, p.329),

(6) 政治的圧力への対応

評価は政治的影響や圧力から無縁ではな

い。特に実用重視評価では、評価者が事業と 事業関係者に接触し、そこから生まれる情報 を活用するだけでなく、そこで活動する主体 と積極的に係わろうとする。既に述べたよう に、評価者は交渉者でなければならず、当然 折衝力も備えなければならない。その上で、 想定利用者の考える望ましい成果の実現に役 立つような評価の実施を心がけ、必要とあれ ば政治的ゲームにも主体として係わらなけれ ばならない。その際の指針として、Patton (1997, pp.356-357) は、次の7点を挙げる。

関係者の中に利害の対立がある場合には「両者勝ち」シナリオを探す、 主要な利害関係者が特定の評価結果の出方に固執しないようにする、 事業成果の短期的意味合いだけでなく長期的な意味合いもあわせて考える、 多様な視点・解釈を尊重する場を形成する、 想定受益者の成果に焦点を絞る、できるだけ多くの選択肢を用意し、多様な視点が結果に反映されるようにする、 政治

できるだけ多くの選択肢を用意し、多様な 視点が結果に反映されるようにする、 政治 的誤用は極力避ける、の7点である。これら の点の実践に当たっては、前述した評価と評 価者の基準が判断の尺度を提供してくれる。

3. 実用重視評価の実践

実用重視評価の理論の柔軟性と動態的特色を理解するには、その実践応用例を参考にするとよい。ここでは、(1)研修評価への適用、(2)評価可能性アセスメントの実施、(3)事業推進手段としての評価の活用の3つの例を簡単に紹介することにする。

(1) 研修評価への適用

研修ワークショップの評価に際して、標準的な計量手法による評価デザインを用いる場合と実用重視のデザインを用いる場合とを対比してみよう¹¹。計量的手法で研修ワークショップの成果を測るためには、研修受講生の知識、技能、経験が研修の前と後で変化したかどうか検証する必要がある。受講生には研

62 長尾 眞文

修の冒頭で「事前テスト」、終了時に「事後 テスト」を受けて貰う。厳密なデザインにす るには、「事前テスト」を受けずに「事後テ スト」だけを受けるグループ、さらに比較の ために研修を受けずに「事前・事後両方のテ スト」を受けるグループと「事後テスト」だ け受けるグループを設ける必要がある。各グ ループの構成は無作為の選択によるものと し、共通のテストを同時に同条件で行う等の 配慮も要求される。そうして得られたデータ を統計的に処理して、受講生グループと比較 グループとの間に有意な差異があるかどうか を見極める。得られる結果は、「科学的検証」 の結果としての信頼性を獲得する。ただし、 その実施には、比較グループの手配等も含め て、かなりの費用と手間が必要となる。

これに対して、ワークショップの評価の目 的が、成果の有無の厳密な検証よりもワーク ショップの内容の改善のための研修成果の大 まかな把握にある場合には、次のような実用 重視のデザインの活用が考えられる。まずワ ークショップ冒頭で、全ての研修受講生に、 研修テーマに関する知識、経験、態度につい て「自己診断」して貰う。このステップは、 受講生にとってこれから始まる研修に対して 意識を高める効果があると同時に、研修主宰 者にとっては受講生に関するデータ収集のス テップとなる。そして研修終了時には、研修 受講生が「何を研修で学んだか」を明らかに するために再び「自己診断」してもらう。こ の終了時診断は、受講生自身に研修成果につ いて考えさせるという学習効果を持つ。この 実用重視のデザインでも、事前・事後のデー タを収集するが、そのステップは研修事業の 一部となっており、また比較グループを設け ることもしないので、余分の費用も手間も掛 からない。もし評価の目的が、厳密な検証を 要求するのでなく、むしろ「合理的な範囲」 内での成果の判定にあるなら、このような実 用重視評価にも充分なメリットがある12。

(2) 評価可能性アセスメントの実施

事業が終了し外部評価を依頼された評価者 が、事業関係者に事業開始当初の目的、方法、 投入資源、成果目標等に関する想定を聞く時、 しばしば一貫した説明が得られないことがあ る。その理由は、ほとんどの場合、そもそも 事業立ち上げの際に、事業の有用性や実行可 能性を明確にすることを怠り、関係者間で事 業計画を十分に練るという基本的ステップを 踏まないからである。公共事業の多くが、実 際には政治的配慮や人間関係によって採択の 決定が為されことを考えれば、それは驚くに 当たらない13。それでも評価を実施しなけれ ばならないことがある。しかし、公的部門の 説明責任が当然のこととして問われるように なった今日、行政や民間公益団体の事業管理 者は、自らの関係する事業の経過や成果を気 にしないわけにはいかない。事業終了時に好 結果を確保したいのはもちろんであるが、願 わくば事業開始時からある程度成功する保証 が欲しい、というのが正直なところであろう。 そこで登場するのが、評価可能性アセスメン トである。

事業が成功するかどうかは、結果として 「良い評価を得ること」であり、事業開始時に 評価の見通しを立てることが評価可能性アセ スメントである。この概念と慣行を最初に提 唱したのは米国の公共政策シンクタンク Urban Instituteを拠点として事業評価の普及 に重要な役割を果たしたJoseph S. Wholey (1983, 1994)で、評価の安価で実用的な活用 方法として、事業開始時に評価者が管理者と の共同作業により事業の企図と実態的条件お よび成果の評価基準を明確にすることを目的 とする評価可能性アセスメントを説いた。こ の作業を行うことにより、事業関係者間で 事業目的、目標および成果指標が明確に定義 されかつ理解され、 事業目的・目標が達成 可能であることが確認され、 事業成果の達 成度(業績)データ収集の目処が立てられ、 得られる情報の活用(事業改善努力等)に

ついて合意が成立し、結果として事業が成功

する確率を高める。

Wholeyがこの実用的評価手法を考案した のは、1970年代前半の米国で多額の費用をか けて行われた大規模の評価作業が対象とした 事業に関して余り確定的な評価に結びつか ず、結果として役に立たなかったことに対す る反省からであった。評価可能性アセスメン トの作業では、事業関係者との作業で目的・ 目標の確定からスタートして、投入資源、活 動、成果目標の間の関係を明確にするのであ るが、それがさらに理論的に展開されたのが プログラム・セオリーに基づく評価の考え方 である (Rossi et. Al. 1999; House, 2001)。プ ログラム・セオリーの活用方法は多様である が、根底には可能な限り事業関係者の考え方 を事業の概念形成に反映する志向がある。そ の源泉は評価可能性アセスメントに始まる実 用重視評価の伝統と言ってよいであろう。

(3) 事業推進手段としての評価の活用

この事業は、理数科教育訓練要員の日本研修を起点とする「さみだれ式」研修の実施により、無資格・低資格の現職理数科教員の教科指導力の強化を図ろうというもの。事業の目標は、事業対象地域の各中等学校で、毎月1回理数科教員が授業改善のための自主研修を行うシステムを構築することである(長尾・又地 2002)。各校の理数科教科主任は、

教員訓練要員が組織する定期的な地域研修で 新しい教授法を学んで自校に持ち帰り、学校 ベースの教員自主研修を主宰して、それを同 僚理数科教員に伝達するのである。教科主任 は、毎回学校ベース研修を開くたびに、規定 の用紙に活動内容の簡単な記述と自己評価を 書込み、四半期ごとに集計して報告書を作成 して訓練要員に提出する。訓練要員はその集 計により、自主研修を行う学校ベースのシス テムができつつあるのかモニターすることが できる。また、各校の理数科教科主任が集う 地域研修の機会を利用して、その初日に各校 の校長から実施状況を報告してもらい、実施 上の問題・課題を議論するモニター評価を通 して、校長の事業に関する理解と支持を確保 するようにしている。さらに、訓練要員は、 各校の四半期報告書に添付して提出される各 回の研修内容を検討して、その中で他校にと って学習効果のあると思われるセッション情 報を普及させることにより、学校ベース研修 の質的改善につなげる。このように、事業実 施者自身による事業活動のモニター情報の共 有の仕方を工夫することで、事業内部に学校 ベースの研修をより頻繁に行う圧力と良い活 動事例の模倣・学習による事業の改善を可能 にするのである(長尾 2001)。

当初3年の実際の事業の運営では、モニター評価は学校ベース研修活動の拡大にはつながったが、良い事例の模倣・学習を通した情報共有による事業改善は、思ったほど実現ーズを実施するに当たり、学校同士のグルーピングによりクラスターを形成して研修の上方に設置することとした。それに対応するようにモニタリング評価のシステムも変更された(MSSI 2003; JICA 2002; JICA 2003)。このように実用重視評価の典型のひとつは、評価が事業推進手段として事業内部に位置づけられる場合である。

64 長尾 眞文

4. 実用重視評価の問題点

(1) 倫理的問題

実用重視評価で評価結果の想定利用者を重 視し、評価作業を一緒に行うことには当然倫 理的問題が生じる (Scriven 1991)。それは評 価作業の生み出す情報に関して、特定の個人 に対してより容易なアクセスを提供すること であり、腐敗の危険を招きかねないからであ る。前述した評価・評価者の基準の遵守とい うことではその意味する問題に対する十分な 解決とならない可能性もある。この点につい て、Patton (1997, p.364) は、 評価者は自 らの道徳的判断を意識的に行う必要があるこ と、評価する事業や一緒に作業する利害関 係者の選択に気をつけること、 評価の中で 異なる利益集団が評価目的に対してバランス よく代表されていること、の3点を説く。問 題の本質は、特定のグループの利益や関心に 対する傾斜とそれ以外の一般的・社会的利益 への配慮との間のバランスをどうとるかであ る。

(2) 評価活動の境界の問題

実用重視評価では、評価する事業によって 評価者と事業およびその関係者との距離が異 なる。それでは一貫した評価アプローチの確 立は難しいし、評価者の役割を明確にするこ とも困難であるとの批判が生まれる (Stufflebeam 1994)。後者については、評価 の目的に事業関係者の能力強化を明示的に含 めると、評価自体の定義もおかしくなるとの 批判が加わる。この問題は、実用重視評価の 意味する参加型評価がエンパワメント評価の ようにアドボカシーの様相を加えるにいたっ て、境界の説明がさらに難しくなっている。 例えば、参加型評価で、事業の評価と事業受 益者の評価プロセスへの参加による能力強化 の双方を目的として設定する時、実際に行う 評価活動のどこまでが評価でどこからが評価 能力育成なのか不明な状況が発生し得る。も し、仮に『事業の成果は満足できるものではなかった』という評価結果が出されたので、 事業関係者の納得のゆくものとならなかった (つまり評価によって事業は失敗したと結論された)としても、受益者が評価について理解を深めることができたら、それで参加型評価は「成功した」と言えるのか。それは評価の理論ではなく、評価人材育成の理論ではないか、との反論である。評価理論としての適用領域の境界線を示さないと、妥当性の確保は難しいのである(Mark 2001)。

(3) 評価の活用の意味

実用重視評価では、評価結果の活用を主眼 として評価アプローチを定めるが、それは評 価者に活用の促進のための評価結果の解釈や 提言を迫り、結果として自縄自縛に陥るので はないかとの批判もある(Mark 2000)。評価 の成功の基準は、評価結果の活用に求めずと も、良いのではないか、事業の意思決定に有 用な情報を提供するあるいは啓蒙的な知識の 普及に資するだけで十分ではないか、との議 論である(Weiss 1997)。それでも評価をめ ぐる議論の動向に関するレビューの多くは、 評価者が一層事業受益者の能力強化や地位向 上を念頭に入れた評価をするようになること を予想している (Wholey et. al. 1994; Worthen 1997; Smith 2001)。評価の活用の意味を問う 議論は、評価の倫理に対する関心の拡大とと もに、評価理論の展開において多様な視点の 間で価値のバランスを確保しようとする試み であろう。

(4) 評価者の質の問題

実用重視の評価は、評価者の知識、経験、能力に依存するところが大きい。伝統的な評価が手法の質を主要な課題としたのに対して、実用重視では評価者の質が問題となる。Pattonの言う「能動 反応 適応」を実践するには、単に評価手法の訓練を受けるだけでは到底不十分である。「交渉者」としての実際

の経験を積まなければならない。大学に籍を 置き、折々選択的に評価作業に従事できる専 門家は別として、民間のコンサルタント会社 や公的機関で働く評価者は、多数の評価案件 を抱え、評価理論にしっかりと根付き、利害 関係者と十分に連携するような評価作業を行 う余裕はないようである(Christie 2003)

5. 結びに代えて

以上、米国における実用重視評価理論の展 開と課題についてPatton (1997) を中心に紹 介したが、その日本における適用可能性はど うであろうか14。実用重視評価が日本で定着 することが考えられるだろうか。確かに、現 在日本で実施されている評価は、ODA評価、 政策・行政評価、教育・学校評価のいずれをと っても内部評価が中心であり、評価に従事す る者と評価結果の潜在的な利用者の距離は近 い。外部専門家による手法重視の本格的評価 の実施は極めて限られており、日本人に一般 的な実践重視の姿勢も考慮すると実用重視評 価の普及する余地はあるように思われる。し かも評価手法的には、ほとんどが実用重視評 価の流れを汲むパフォーマンス・メジャーメ ントかプログラム・セオリーに基づく評価で ある。

しかし、評価慣行一般の定着の課題を別として、実用重視の評価が日本国内で浸透していくには、いくつか超えなければならないハードルがある。その第1は、評価人材の育成である。約40の大学院で専門的な評価教育を行っている米国と異なり、日本では評価構動を設けている大学院はあってもプログラムを組むまでにはいたっていない。実用重視の基礎的な知識と技能を付与する訓練機会の提供が必要である。これは国内の大学が対応を考えるべき課題である。

第2に、評価の倫理と評価の質の課題に対する取り組みの必要がある。実用重視評価は評価者の行動から作業結果の質まで本人の判

断に依拠する部分がほとんどである。国内の評価の実施状況に合致した評価の基準と評価者の行動規範を設けることで、評価者の職業的責任感を下支えすることが可能であろう。また、評価の質については、メタ評価の慣行の普及が望まれる。これらはいずれも学会組織、特に日本評価学会が担うべき役割であろう。

最後に、現在内部評価を実施している中央 省庁、地方自治体、関係機関では、組織内で 評価の実施に関する十分な意思統一が図られ ていない状況のようである。同一組織内でれ 価担当部局と事業を担当する部局とが対立ていまでも、不協和音を響かせてい までいかないまでも、不協和音を響かせてい ようにうかがえる。実用重視評価の考えから は、評価者と事業管理者が対話すること指す。 は、評価者と事業管理者が対話すること指言 評価作業を始め、共に役に立つ評価を目指す。 省庁や自治体の評価担当部局で、各局の会な 開いてみてはどうだろうか。実用重視はあるう。 の得意分野である。試してみる価値はあろう。

注記

- 1 事業評価は、系統的に行われる評価の総称で、評価対象は個別プロジェクト、その集合のプログラム、プログラムを包含する政策、さらには組織と多様である。英文では通常program evaluationと称される。
- 2 本稿で取り上げる実用重視評価は、「実用 重視」の観点から体系的、理論的に組み立 てられた評価のことで、有識者に見ても らうとか、手元のデータだけで簡便に判 断するといった簡便でクイックな方法で 評価を実用的に行うことを意味している のではない。
- 3 Shadish et. Al. (1991)、Worthern et. Al. (1997)、 Wholey et. Al. (1994)、House(1993)等参照。これらの文献も含めて、米国を中心とする評価の理論と実践の展開に関する要約的紹介の書として龍・佐々木(2000)がある。

66 長尾 眞文

- 4 Patton以外の代表的実用重視論者としては、評価における科学性の重要性に疑問を提起したCronbach (1980) 評価の有用性に重点を置くAlkin (1985) 評価を組織改善の手段として組織内に取り込む活用を説くPreskill and Torres (1999) らが居る。また技法的にはこれら論者と異なるが、全世界的に行政改革手段として採用されているパーフォーマンス・メジャーメント手法の推進論者 (Wholey 1997等)も実用重視の考え方の持ち主である。
- 5 この論争については、Patton(1997, Chapt. 12)、Wholey et. Al. (1994, Part 1)、Weiss (1998, Chapt. 4)を参照。
- 6 Pattonは、元々定性的手法を得意とする社会学者で、前掲の実用重視評価に関する著書のほかに、やはり3版を重ねている定性的手法に関する著書Patton(2002)を出している。書名の変化(第1版(Patton 1980)がQualitative Methods、第2版(Patton 1990)がQualitative Evaluation and Research Methods、第3版がQualitative Research and Evaluation Methods)に明らかなように、Patton自身の関心も定性的評価・手法から実用重視評価の立場を支える定性的手法へとシフトしてきている。
- 7 組み合わせ方としては、定性的手法によるプロセスの評価と定量的手法による成果評価の実施、定性的手法によるパイロット評価に基づく定量的評価の実施、両手法の併用実施等が考えられるWeiss(1998, pp. 268-269)。
- 8 参加型評価には、このような実用的関心を超えて、事業の受益者の能力強化や社会的・政治的地位の改善により大きな比重を置くエンパワメント評価も含まれる。実用的関心からの参加型評価とこのようなアドボカシー・アプローチによる参加型評価の区別については、Cousins and Whitmore (1998)参照。
- 9 米国では、評価関係学会が共同で定めた 「有用性」、「実現可能性」、「正当性」、「正 確さ」の4基準から成る評価の基準(Joint

- Committee 1994)と、全米評価学会が定めた「系統的調査」、「作業能力」、「誠実な態度」、「人に対する敬意」、「一般・公共福祉に対する責任」に関する評価者の事件原則(AEA 1995)が職業倫理体系 20数種の評価についてメタ評価については、20数種の評価についてメタン・ミシガストが作成され、ウェスタン・ミシガストが作成され、ウェスタン・ミシガストが作成され、ウェスタン・ミシガストが作成され、ウェスタン・ミシガストが作成され、ウェスタン・ミシガストが作成され、ウェスタン・ミシガストが作成され、ウェスタン・ミシガストが作成され、ウェスタン・ミシガストが作成され、ウェスタン・ミシガストが作成され、ウェスタン・ミシガストが作成され、ウェスタン・ミシガストが作成され、ウェスタン・ミシガストが作成されている。
- 10 その典型は、事例研究による評価の優位 性を説くStake (2000)の立場である。
- 11 ここでの事例紹介は、Patton (1988) に依 拠している。
- 12 以上の例は、国内研修機関でルーティン的に行われている研修効果測定のための事前・事後テストに近い。この方法はあまり信頼性が高くないので、研修機関それぞれに多様な手法で補填している。ワークショップの内容の改善を目的として評価することも考えられている。例えば、平松(2001)参照。
- 13 筆者もそのような評価に関係したことが ある(森藤・古屋・長尾、2002)。
- 14 以下の記述は、山谷(1997、2002) 古川・北大路(2002)および筆者も関わった国際開発センター・広島大学教育開発 国際協力研究センター(2002)の調査に依拠している。

参考文献

- 長尾眞文・又地淳(2002)「教育分野における新たな技術協力モデル構築の試み 南アフリカ・ムプマランガ州中等理数科教員再訓練プロジェクトから」、『国際教育協力論集』5(1):83-100
- 長尾眞文(2001)「援助における評価の目的 と活用方法:南アフリカ理数科教育支援事 業による例示」、『国際教育協力論集』4 (1):89-100
- 古川俊一・北大路信郷(2002)『公共部門評

- 価の理論と実際。日本加除出版
- 森藤桂子、古屋亮太、長尾眞文(2002)「山 形県遊佐町の国際交流事業の評価」、『都市 問題』93(1):29-42
- 山谷清志(2002)「わが国の政策評価 1996 年から2002年までのレビュー 」『日本評 価研究』2(2):3-14
- 山谷清志 (1997) 『政策評価の理論とその展開』 晃洋書房
- 龍慶昭・佐々木亮 (2000)『「政策評価」の理 論と技法』、多賀出版
- Alkin, M. (1985). A Guide for Evaluation Decision Makers. Newbury Park, CA: Sage Publications.
- AEA Task Force on Guiding Principles for Evaluators (1995). Guiding Principles for Evaluators. New Directions for Program Evaluation. Summer. 19-34. (日本語訳につ いては、Patton (1997)の訳本の巻末資料 を参照)
- Campbell, D. (1969). Reforms as Experiments. *American Psychologist*. 24(4). 409-429.
- Christie, C. A. (2003). What Guides Evaluation? A Study of How Evaluation Practice Maps onto Evaluation Theory. In C. A. Christie, (Eds.). The Practice-Theory Relationship in Evaluation. *New Directions for Evaluation*. No. 97. Spring. San Frnaciso, CA: Jossey-Bass. 5-35.
- Cousins, J. B. and Whitmore, E. (1998). Framing Participatory Evaluation. In W. Whitmore (Ed), Understanding and Practicing Participatory Evaluation. New Directions for Evaluation. 80. San Francisco, CA: Jossey-Bass. 5-23.
- Cronbach, L. J. and Associates (1980). *Toward Reform of Program Evaluation*. San Francisco, CA: Jossey-Bass.
- House, E. R. (2001). Unfinished Business: Causes and Values. American Journal of Evaluation, 22(3), 309-315.
- House, E. (1993). *Professional Evaluation: social Impact and Political Consequences*. Newbury

- Park, CA: Sage Publications.
- Joint Committee on Standards for Educational Evaluation(1994). The Program Evaluation Standards. Thousand Oaks, CA: Sage Publications. (日本語訳については、 Patton(1997)の訳本の巻末資料を参照)
- Mark, M. Evaluation's Future: Furor, Futile, Or Fertile. *American Journal of Evaluation*. 22(3). 457-479.
- Mark, M. M., Henry, G. T. and Julnes, G. (2000). Evaluation: A Integrated Framework for Understanding, Guiding, and Improving Public and Nonprofit Policies and Programs. San Francisco, CA: Jossey-Bass.
- Patton, M. (2001). Evaluation, Knowledge Management, Best Practices, and High Quality Lessons Learned. *American Journal of Evaluation*. 22(3). 329-336.
- Patton, M.(1988)Integrating Evaluation into a Program for Increased Utility and Cost-Effectiveness. In J. A. McLaughlin, L. J. Webber, R. W. Covert and R. B. Ingle (Eds.). Evaluation Utilization. *New Directions for Program Evaluation*, No.39. Fall. 85-94.
- Patton, M. (1997). *Utilization-Focused Evaluation: The New Century Text*. 3rd Edition. Thousand Oaks, CA: Sage Publications. 日本語抄訳版:マイケル・クイン・パットン著(大森彌監修、山本泰・長尾眞文編)(2001)『実用重視の事業評価入門』清水弘文堂書房)
- Preskill, H. and Torres, R. T. (1999). *Evaluative Inquiry for Learning in Organizations*. Thousand Oaks, CA: Sage Publications.
- Rossi, P.H., Freeman, H.E. and Lipsey, M. W. (1999). *Evaluation: A Systemic Approach. 6th Edition*. Thousand Oaks, CA: Sage Publications.
- Scriven, M. (1993). Hard-Won Lessons in Program Evaluation. New Directions for Program Evaluation. No. 58. San Francisco, CA: Jossey-Bass.
- Scriven, M. (1991). Beyond Formative and

68 長尾 眞文

Summative Evaluation. In M. W. McLaughlin and D. C. Phillips (Eds). *Evaluation and Education: At Quarter Century*. 90th Yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press, 18-64.

- Shadish, W. R., Cook, T. D. and Leviton, L. C. (1991). Foundations of Program Evaluation: Theories of Practice. Newbury Park, CA:Sage Publications.
- Smith, M.F. (2001). Evaluation: Preview of The Future #2. *American Journal of Evaluation*. 22(3). 281-300.
- Stake, R. (2000).Case Studies. In Denzin, N. K. and Lincoln, Y. S. (Eds.). Handbook of Qualitative Research, 2nd Edition, Thousand Oaks, CA.: Sage Publishers, 435-454.
- Stufflebeam, D. (1994). Empowerment Evaluation, Objectivist Evaluation, and Evaluation Standards: Where the Future of Evaluation Should Not Go and Where I Needs to Go. *Evaluation Practice*, 15(2), 321-338.

- Weiss, C. (1997). *Evaluation*. 2nd Edition. Upper Saddle River, NJ: Prentice-Hall.
- Wholey, J.S. (1994). Assessing the Feasibility and Likely Usefulness of Evaluation. In J. Wholey, H., P. Hatry & K. N. Newcomer (Eds), *Handbook of Practical Program Evaluation*. San Francisco, Calif: Jossey-Bass, 15-39.
- Wholey, J., Hatry, H. P. and Newcomer, K. H. (Eds),(1994) Handbook of Practical Program Evaluation. San Francisco, Calif: Jossey-Bass
- Wholey, J. S. (1983). Evaluation and Effective Public Management. Boston, Mass: Little, Brown.
- Worthen, B.R., Sanders, J.R. and Fitzpatrick, J. L. (1997). *Program Evaluation: Alternative Approaches and Practical Gruidelines*. 2nd Edition. New York: Longman Publishers.

(2003.7.30受理)

Utilization-Focused Evaluation: Theory and Issues

Masafumi Nagao

Hiroshima University nagaom@hiroshima-u.ac.jp

Abstract

Theoretical concern in program evaluation has for long focused on the choice of evaluation method, with the quantitative-qualitative debate occupying its central place. The principal question had to do with scientific nature, objectivity and generalizability of evaluation findings. Michael Q. Patton joined the debate with a theory of utilization-focused evaluation, proposing a pragmatic synthesis between the two propositions. This theory suggests that the utility of evaluation should be measured in terms of utilization of evaluation findings, and that the key role of an evaluator should be to identify those stakeholders who possess strong interest in the use of the results and to strengthen their ownership of the evaluation process through their participation. This paper first outlines the debate which formed the background to the birth of the utilization-focused evaluation, and then presents its main ideas, illustrates its uses and discusses some questions and issues surrounding it. The paper concludes by commenting on issues that need to be tackled if the high potential it promises for the Japanese evaluation context is to be realized.

Keywords

Utilization-focused evaluation, quantitative method, qualitative method, utilization of evaluation findings, evaluability assessment

【研究論文:依頼原稿】

エンパワメント評価の特徴と適用の可能性 ~ Fettermanによる「エンパワメント評価」の理論を中心に~

源 由理子

国際開発コンサルタント yminamoto@mvd.biglobe.ne.jp

要約

Fettermanによるエンパワメント評価は、評価の概念と技術を使い「自己決定能力」を身につけるプロセスを提供し、変革を支援するものである。それは受益者を含む当事者グループが、自らが継続的に評価を行うことを通して、エンパワメントしていくプロセスである。そこでは、評価専門家は査定者ではなく、ファシリテーター、評価手法のトレーナー、人々の代弁者などの役割を担う。エンパワメント評価は変革とエンパワメントのアジェンダがある事業に適していると考えられ、開発援助の社会開発プロジェクトや国内のNPOによる公益事業に適用可能である。従来型の評価と併用することにより、社会の多元的な視点を取り入れた評価が可能になり、地縁、血縁を越えた自発的なコミュニティの発展を考えるひとつの契機となり得る。今後は我が国においても事例を積み重ねることにより、理論、方法論の更なる検証が必要になるであろう。

キーワード

エンパワメント評価、参加型評価、変革、NPO、参加型開発

1. はじめに

近年、我が国においては、「サービス(便益)を受ける側の人々」の眼から政策や公益事業を評価する参加型評価の必要性が広く認識されてきている。古川(2000)によればその背景には、政府と民間の関係性の変化、及びNPOなどの市民による公益活動を視野にいれた公共部門の統治(ガバナンス)の台頭があり、公共部門の正統性が合法性、適法性からさらに進んで、政策や施

策の実質的な効果によって問われるなかで、その測定(=評価)に市民が加わることで政府や政策の正統性がさらに強化されるとしている。行政が内部で行う判断は自己にの市民の側からの評価が不可欠になるの参加の方式には、「政府機関とNPOなどとの協関とNPOなどとの協関で評価が行われる方式」、「市民が政府機関から独立して評価を行う方式」(西尾 2000、

pp.30-31)などがあり、既にいくつかの試みが実践されてきている。市民による独立の評価の場合はもちろん、協働で評価を行う場合も、市民の側に評価をする能力が必要になるのであり、山谷(2000)は判断能力を市民自らがつけることを支援するとされている「エンパワメント評価」の意義をその先に見出している。

開発援助分野に眼を向けると、参加型評 価は現地の当事者を「評価者」の一部に加 えて評価を実施することで、ドナーの論理 とは違った当事者の視点からの「現場」に 近い声を聴くことができ、それが評価の質 を高めるとともに、現場における教訓の活 用度が高まるとされ、国際協力事業団を中 心に事例が積み重ねられつつある(国際協 力事業団 2001)。また、評価のプロセスに当 事者が参加することを通して当事者意識の 高揚や主体性が確保され、結果として事業 の改善につながるという効果と、「評価を実 施することによって社会に変革をもたらそ うとする、評価に対する新しいアプローチェ への期待もある(三好・田中 2001)。この考 えには、開発援助分野で広く議論され、実 践も行われている「参加型開発」の文脈と つながるものがある。参加型開発とは、「開 発の過程に当事者が参加して、主体的に関 わること」であり、そこでは住民のエンパ ワメントが重要な要素となっている。計画、 実施、評価といった事業のサイクルすべて に参加することによりその学習効果と社会 的変化が期待されている。開発援助の分野 では「エンパワメント評価」という言葉は 使われていないが、参加型開発や社会開発 の中で、人々のエンパワメントを支援する ための様々な調査手法が使われている。

このように市民や当事者が評価に参加する参加型評価の必要性への認識が高まる中、そのひとつの方法論としてエンパワメント評価が注目されている。本稿ではその適用の可能性を、その代表的方法論として知られる、David M. Fettermanが提唱する「エンパワメント評価 (Empowerment Evaluation)」

に焦点を当て考察するものである。まず、 Fettermanによるエンパワメント評価の概念 を整理し、次にその方法論と事例を紹介し、 最後にその適用の可能性と留意点について まとめる。

2. エンパワメント評価の理念と特徴

(1) エンパワメントの意味

Fetterman (1996, 2001) による「エンパワ メント評価」は、自らが関わる事業を改善 し、自発的に自らの状況を改革しようとす る人々(=グループ)に対し「自己評価 (self-evaluation)」と「反省 (reflection)」を通 して自己決定能力(self-determination)を身に つけるプロセスを提供することであると定 義されている。エンパワメント評価はエン パワメントするプロセスとその結果(アウ トカム)双方を対象としており、例えば、 エンパワメントのプロセスは自己決定能力 を身につけるプロセスであり、主体的に自 らの状況の改善を図ることができるように なることを指す。一方、エンパワメントし た結果(アウトカム)は、人々が対象地域 において何らかの影響力がある活動を始め ることや、その活動の結果としての効果を 指す。もうひとつの特徴は、エンパワメン ト評価の参加者が、評価対象となる事業の 実施者、協力者、サービス利用者などであ り、エンパワメントは彼らがひとつの「当 事者グループ」として力をつけていくこと を意味する点にある。その当事者達のうち 特に事業に直接的に関わっている人々は、 自発的で、何らかの変革の意思があること が前提とされているように思う。つまり事 業そのものに何らかのエンパワメントのア ジェンダが盛り込まれていることが、エン パワメント評価の条件となる(Patton 1997. pp.156-157).

エンパワメントという言葉は日本語になりにくい。時としているいろな意味に解釈

されるが、エンパワメントが単なる能力強 化 (Capacity building) ではないことに留意 すべきである。エンパワメントの概念は主 に、福祉・教育分野や開発援助の社会開発 分野で使われてきたが、開発援助において は、1970年代から80年代にかけて経済開発 に代わるパラダイムとして出てきた人間開 発の代表的な理念として登場した。すなわ ち、貧しい人々の真の生活向上のためには たんなる経済的な向上だけでは不十分で、 制度的・組織的に力を剥奪されてきたこと が問題であると捉え、エンパワメントとは 貧しい人々自身が資源へのアクセス機会を 得ることにより意思決定における自律性を 確保し貧困からの脱却を図ること、と定義 されている(フリードマン 1995、齋藤 2002)。 そのエンパワメントは、おそらく特定のし かも一定期間の事業の介入だけで達成しう るものではなく、途上国の住民をとりまく 社会、政治、経済、コミュニティの人間関 係、伝統的価値など多くの要素と直面しな ければならないであろう。つまり社会の変 革なしにエンパワメントは成し遂げられな いという意味では、政治的な過程である。

では、エンパワメント評価そのものはどうであろうか。Fettermanは、エンパワメント評価のアイディアは「Speaking the language of power: Communication, collaboration, and advocacy (1993)」の執筆中に生まれたものであるとし、その時の心境を次のように述べている(Fetterman 1996, pp.6-7)。

「私は評価専門家や社会科学者達が仕事で関わる人々の声を代弁し、人々が抱える課題を政策へつなげていくための方策を切けたかった。そして、社会問題に関心の関心の学者達が、彼ら自身の洞での登者達が、彼らは、の学者を政策決定者に盛んにつ多りを知った。彼らは、多分を知っていることを知った。彼らは、多分を知ることを知った。では、多の解している。では、アメリカンインディアン等に対象のに取り組んでいた。これら学者や実務

家は、効果的な戦略を探り、教訓を人々と 共有し、エリート層や影響力のある政策決 定機関とコミュニケートできるように人々 の能力を強化することを目指していた。そ れらのケースのレビューを通して、協働 (collaboration)、参加(participation)、及びエ ンパワメント(empowerment)のキーワード が浮かびあがり、一本の線でつながり、エ ンパワメント評価の基本的なアイディアが 生まれたのである。」

これらの記述からわかるように、エンパワメント評価はその出自からして社会運動性が強く、評価の延長上に社会変革を明確に意識している。Fettermanはエンパワメントを可能にするのは当事者のみである、という信念に立ち、外部の評価専門家の役割として、抑圧された人々がエンパワーするための支援者であり、また更に一歩踏み込んで当事者の「代弁者 (advocate)」であることも想定しているところが特徴的である。

(2) 対話のプロセス

他の参加型評価と同様、評価専門家がフ ァシリテーターとしての機能を持つエンパ ワメント評価は、対話によるプロセスを重 視するアプローチをとる。評価の手法とし ては定量、定性評価の手法を駆使するも、 客観性や厳密な科学的精査を重視するもの ではなく、「事実は構成される」という前提 のもと価値の多様性を尊重することから出 発している。すなわち、評価は「事実とし て存在するもの」を見つけることではなく、 評価者や利害関係者間の交渉のプロセスを 通して関係者が「納得する」状況として創 り出されるものであるとした構成主義 (Constructivism) の立場をとり、第4世代評 価 (fourth generation evaluation) の流れを汲 むものと考えられる (Guba & Lincoln 1989. p. 8)。従って、評価専門家や参加者が対話 を通して協働で評価を行うプロセスが非常 に重要になる。

また、対話重視のプロセスでは人々の間 の関係性を対等なものとして扱う。対等と は、多様性のもとにおいて平等であり、あ る事業を評価し社会の状況を改善するとい う目的のもとに協働することを意味し、参 加者は調査や実験の「対象」となることは ない(Guba & Lincoln 1989, p.11)。Hardina (2002)が指摘するように、従来型評価にお ける専門家の知識は持てる者と持たざる者 の間にある伝統的な権力構造をより強固な ものにする恐れが高いのであり、エンパワ メント評価は有用である。

(3) 5つのキーワード

Fetterman自身は、エンパワメント評価の 概念を説明する際に、 参加者のトレーニ ファシリテーション ング (Training)、 (Facilitation) アドボカシー(Advocacy) 啓蒙 (Illumination) 及び 精神の解放 (Liberation)の5つのキーワードを使ってい る (Fetterman 1996, pp.9-18)

まず第1に、「参加者のトレーニング」は、 参加者が自ら評価手法を使って評価を行う ため、その前段で評価専門家によるトレー ニングが必要であることを説明するもので ある。これは、事業の中に「評価を内在化 (internalize)」するプロセスである。従来型 評価の場合は評価終了とともに評価専門家 は離れるが、エンパワメント評価では事業 関係者が自らの手で評価を継続するように なることが重要とされている。その「評価 体制づくり」が「自己決定能力を向上して いくプロセス(エンパワメントのプロセス)」 の重要な部分を占めていると考えられる。 評価専門家はトレーナーとして参加者の評 価技術を高める任を担っているということ である。

第2のキーワードであるファシリテーショ ンについてFettermanは、人々は自らの力で しかエンパワーすることはできない、とい う信念に基づき、評価を行うプロセスにお いて外部専門家の役割はファシリテーター

やコーチであると明言している。具体的に 何を実施するのかについてはケース・バ イ・ケースであり、基本的な働きは、参加 者が自ら評価を実施する上で障害となって いると考えられる事柄への助言(例:コン フリクト、関係者間の誤解、議論の焦点が 絞れていない場合など)やプロセスを適切 に動かすためのヒント(例:グループ分け、 方法論の助言、リフレッシュ・セッション など)を適宜与えることである。しかし、 あくまでも評価を行う参加者がファシリテ ーター(=評価専門家)に依存することに ならないように細心の注意を払うべきであ るとしている。

第3に、ファシリテーターとしての役割に 加え、評価専門家は「直接的な代弁者 (direct advocate)」として、参加者(=エン パワメント評価の主体)が評価を通してエ ンパワーすることを促す役割がある。例え ば、ホームレスのように自らの状況を必ず しも適切に言語化できない人達の代弁者と なる。つまり、彼らの問題解決の方法や方 向性について提案し、あるべき社会変革が 起こるように導く役割である。更に、自己 評価結果は資金提供者に対する報告として も有効であるとし、評価結果を提示する代 弁者としての働きも強調されている。評価 結果を代弁する働きは、内部評価や参加型 評価の役割としては定着しつつあるが、評 価専門家が直接的に事業内容のアドボカシ ーに関わることは、議論を呼ぶ2。このよう な新しい役割は、エンパワメント評価が社 会運動的な性質をもつことの表われとして とらえることができる。

エンパワメント評価の特徴としてさらに、 啓蒙を促し()、精神の解放をもたらす () という流れがある。他の評価と同様、 エンパワメント評価は評価のプロセスで関 係者の気づきや啓蒙を促す。ただし、その 内容は一般的な気づきや能力強化だけでは なく、より具体的に問題の把握や改善方法、 新たな戦略、事業の運営方法を考え出す力 をつけることも含む。また、啓蒙を通して、

エンパワメント評価は精神の解放をもたらすものでもあるとされている。精神の解放とは、抑圧された状態もしくはそれまでの古い考えや制度から解き放された状態を指す。これは、エンパワメントの定義とも関連し、自己決定能力をつけるためのひとつの条件と考えられる。

(4) 他の参加型評価との比較

上記に述べた特徴を持つエンパワメント 評価は他の参加型評価と比べてどう異なる のであろうか。北米の評価分野では、1970 年代以降、事業の主要な利害関係者 (stakeholders)が評価へ参加することは評価 結果の質を高めることになるとして、利害 関係者を巻き込んだ評価の理論と方法が注 目されるようになった。利害関係者との協 働作業 (collaborative work) による評価は、 「参加型評価 (Participatory Evaluation)」とい うネーミングで広く知られるようになった が、重視する評価の目的やプロセスに応じ て、利害関係者評価(Stakeholder-Based Evaluation)、民主的評価(Democratic Evaluation) 実用性重視評価 (Utilizationfocused Evaluation) エンパワメント評価 (Empowerment Evaluation) など多岐にわたっ ており、一言で参加型評価といってもその 拠って立つ理論的背景と方法論は多様であ る (Cousins & Whitmore 1998, 山谷 2000, 三 好・田中2001 他)。

Cousins & Whitmore(1998)は、評価の目的によって参加型評価の種類を、事業改善や組織強化を目指す実用型の参加型評価(practical participatory evaluation)と、社会正義や弱者の自立を目指す変革型の参加型評価(transformative participatory evaluation)の二つに分類できるとしている。前者の起源は主にアメリカやカナダなどの北米にあり、評価の目的は政策、プログラム、組織運営に貢献することである。利害関係者が関わることによって評価結果の妥当性や評価結果に対するオーナーシップが高まり、

その結果として評価の活用度が高まること が期待されている。一方、変革型の参加型 評価は、主に開発途上国におけるコミュニ ティー開発や成人教育の分野から発展して きたもので、参加型研究やアクションリサ ーチの流れを組み、パウロ・フレイレの 「意識化」の過程(抑圧された人々が、自ら の状況を識字学習を通して理解し抑圧され た状況を主体的に変革していく過程)の理 論にも深く根ざすものでもある(Brisolara 1998, pp.31-32)。これらふたつの類型は、評 価の目的のほかに、「意思決定権を誰がもつ のか(=評価者は誰か)」、「誰が参加するの か」に関しても違いがある。実用型の参加 型評価では参加者が主に評価事業のスポン サーや実施者であり、評価専門家とのパー トナーシップのもと評価を行うのに対し、 変革型は主な利害関係者も対象とするもの の特に事業の受益者 (beneficiaries) が参加 することを重視し、評価専門家とのパート ナーシップのもとに評価を行いながら最終 的な決定権は参加者が担う。

このふたつの類型に照らし合わせると、 「エンパワメント評価」はその目的において は変革型の参加型評価に近い。参加者の範 囲は評価結果を直接利用する立場にいる人 で、その面で実用型に近いが、サービスの 利用者(受益者)も含んだより巾がある事 業関係者である点に特徴がある。決定権に 関しては参加者が完全にコントロールする ため、いずれの類型にもあてはまらない (表1参照)。評価専門家がサービス利用者 (受益者)の視点に立った情報を評価の質と 客観性を高めるために使うのではなく、参 加者が評価結果を完全にコントロールする のであり、そのためには参加者自身が評価 の眼を持つことができている状態が必要で ある。その前提として、「評価の内在化」が ある。Pattonは、エンパワメント評価の5つ のキーワードのうち、「ファシリテーション」 と「啓蒙」に関しては、他の参加型とも共 通するものであるが、「トレーニング(=評 価の内在化)」、「アドボカシー」及び「精神

の解放」はエンパワメント評価独特のもの ではないか、と指摘している(Patton 1997. p. 149)。評価専門家の役割から見ると、ト レーニングを行ったり、事業内容に関する 人々の代弁者としての働きを課するなど、 新しい側面が加わっている。

3. エンパワメント評価の方法論と 事例

以上に述べたような特徴を有するエンパ ワメント評価はどのように実施されるのか、 本節ではFettermanによる方法論と事例を紹 介する。

(1) 四つのステップ

エンパワメント評価の実施方法は四つの ステップから構成されている。まず、最初 に 評価対象事業のミッションやヴィジョ

ンを確認し(Mission)次に、 強みや弱点 を含め事業の現状を吟味もしくは評価し (Taking stock) さらに 将来に向けて目標 設定やそれを実現するための戦略を決め (Planning for the future) 最後に、 事業の 進捗状況をきちんとモニターするために必 要なデータ、その収集方法やドキュメンテ ーションの方法を検討する(Documentation) という流れである。これらのステップは事 業の継続中に繰り返し実施されることにな る。各ステップの内容と特徴を以下に示す (Fetterman 2001, pp.23-34),

ミッションの確認 (Mission)

まず、参加者間で事業のミッションを定 義することから始まる。ここではエンパワ メント評価の専門家がファシリテーターと して積極的に関わることが想定されている。 既にミッションが明示されている場合でも、 参加者各人によって理解の度合いが異なっ たり、立場によってミッションの解釈の仕

表1	参加型評	価の分類	とエンバワ	リメン	ト評価の特徴

参加型評価の分類	主な目的 / 機能	評価者(意思決定権)	参加者の範囲	参加の度合い
実用型の参加型評価	問題解決をサポー	バランス型: 評価専門家及び参加者によるパート ナーシップ	評価結果の直接利用者: スポンサー、事業マネージャー、開発・実施者	評価のすべての段
変革型の参加型評価	· ·	バランス型: 評価専門家及び参加者によるパートナーシップであるが、最終判断は参加者	主な利害関係者: 特に受益者	高い: 評価のすべての段 階に参加
エンパワメント評価	政治的: エンパワメント、 啓蒙、自己決定能 カ	参加者: 参加者がすべての 決定権を持つ	評価結果の直接利用者: 事業の主なメンバーとより広い範囲の関係者(サービス利用者、受益者等)	評価のすべての段

(出所) Cousins & Whitmore (1998), p. 12、Fetterman (1996)を基に筆者作成

方が異なったりする。それぞれがどのようなイメージを持っているのかを知ること生まり、新たなミッションのフレーズがまれるここで句に、これることを表してもなったが、ついてもなったが、で完全な合意を形成してもようでによったが、自行をひとつにするというといいである。これが、の意識化と捉えることができる。これが、エンパワメント評価の出発点となる。

現状の評価 (Taking Stock)

2つ目のステップには2つの作業が含まれる。まず、事業を実施する上で重要な活動と考えられるものをリストアップし、優先度が高いものを選択する作業である。とはいえ優先度の順位はここではあまり大きは問題ではない。このプロセスでは、なぜある活動は別の活動よりも価値があると判断されるのかといった深い議論を避けるのが特徴的である。あくまでも参加者が総体として合意した活動群を選ぶ作業である。

次に、選択された優先度が高い活動に評 点をつける作業がある。ただし、ここでも 評点を「正確に」つけることが最終的な目 的ではない。各活動ごとに参加者による10 段階評価が行われるが、ここでは活動に関 する事前情報は最小限のまま話合いが進行 される。重要なのは、評点をつけるプロセ スで、対話(なぜある参加者は3点をつけて、 ある参加者は8点をつけたのか等)が始まり、 根拠となる事柄が指摘され、内容の具体化 が進み、参加者間のコミュニケーションが 促進されることである。そしてその結果、 評点を変更することもできる。重要なのは 現状を把握し、将来計画の戦略や目標の測 定がどのように行われるのかについて検討 するベースを提供することであるとしてい る。

将来計画の策定 (Planning for the Future) このステップは、現状を踏まえどのように活動を継続したいのかを参加者に問うプロセスである。参加者は前段階で選択した活動ごとに目標や具体的な戦略を考えていく。目標の設定内容に関しては資金を提供する側からの了解も必要である。最終成果や長期目標を掲げる前に、まず、できるだけ参加者(受益者)が直接恩恵を受ける「中間成果」レベルのものを設定することから始めることが重要とされている。

進捗状況のモニター (Documentation)

将来計画の策定内容を受け、活動や目標をモニタリングするためのデータや根拠を検討し、継続的な評価が行えるように体制を整えていく。これは評価のプロセスを組み込む上で重要なステップに位置づけられている。

各ステップでは、従来の評価方法が随所 で使われる。例えば、事業評価、自己評価、 アウトカム評価、ロジック・モデル、コミ ュニティの質的インデックスなど様々な方 法を駆使する。また、各ステップごとに展 開される意見交換は、参加者のアイディア を視覚化して、ボードやフリップチャート に書き表しながら進められる。これらの方 法は、他の評価においても広く活用されて いるものであり、その意味では、エンパワ メント評価は新しい評価のツールを提供す るものではない。特徴的なのは、この四つ のステップを対話のプロセスを通し参加者 自身が行っていくことであり、その評価の プロセスが全体として大きな意味を持つの ではないかと考えられる。

(2) 事例紹介

本事例は、Fettermanの著作の中で紹介されている数ある事例の中のひとつである。おそらくそれらの中でも、エンパワメント評価がめざしているものに最も近い事例の

ひとつに位置付けられると思う。

Accelerated Schools Project:ASP (強みを引き 出す教育のためのプロジェクトナ

プロジェクトの概要

ASPは、様々な問題を抱える子供たち(atrisk students)に対する教育のあり方を見直 すものである。それまでの補修プログラム や特殊学級による教育は効果をあげるどこ ろかマイナスにはたらいているとの認識に 基づき、「補修ではなく良いところを引き出 す教育」を目指すべきだとした。遅れた生 徒を下位に位置付けるのではなくすべての 生徒が教育によって便益を得られるものと すべきであるとし、そのためには小手先の 教育内容の改訂では役に立たず、学校のカ リキュラム、教育戦略、教育内容、組織体 制などすべての見直しが迫られた。これは 「学校文化の変革」事業と位置付けられ、学 校の活動を再構成することによりASPの価値 と理念を内在化するための戦略が練られた。 1986年に2つのパイロット小学校でスタート したプロジェクトで、8年後の1994年には全 米37州において700の小学校と中学校で実施 されるようになった。

基本方針(Accelerated Principle)

a.「統一された目標」

これまでは生徒の学力に応じ教育の目標 を差別化してきたが、今後はすべての生徒 に対しひとつの目標 すなわち、生徒の強 みを引き出す教育 とする。また教える側 と学ぶ側に分け隔てることなく、「夢の学校」 に向って関係者(学校側、教師、生徒、親) が幅広くエンパワーすることを目標とする。

b.「責任を伴うエンパワメント」

これまでは教材や教育要領などを外の権 威 学校から遠い存在 に依存し、直接の 関係者である教師、生徒、親からのインプ ットは何もなかったが、今後は、学校スタ ッフ、教師、生徒、親が主な決定に責任を 持ち、彼らがコントロールできない力によ

ってすべてが決定されることがないような 体制を作る。教育のアウトカムにも責任を 持つとし、そのためには継続的なアセスメ ントと評価を行うことが必要である。

c.「強みの活用」

これまでのように、'低い'レベルの子供た ちを'高い'レベルに追いつかせようとするの ではなく、子供たちの'強み'を発展させ'弱 み'をカバーするというアプローチを取り、 学校スタッフ、親、教師に関しても同様に それぞれの'強み'を生かし、教育過程に貢献 することをめざす。つまり、すべての構成 メンバーがプロジェクトの成功に向ってそ れぞれの強みを生かすべきである。

エンパワメント評価の適用

このように、ASPでは学校文化の変革と当 事者グループ(学校スタッフ、教師、生徒、 親)のエンパワメントが大きな特徴である。 この事例では、以下のように「エンパワメ ントのプロセス」と「エンパワメントの評 価」のふたつに分けてエンパワメント評価 の方法論を適用している。

a. エンパワメントのプロセス

上記ミッションはすでに確認されていた 中で、まず、学校側の主導により参加者全 員による現状評価(taking a stock)が行われ、 ベースラインデータが得られたところで、 将来計画が練られた。将来計画の段階では 特に生徒の役割が大きかった。彼らが望ん でいる学校の姿 (dream school) は何なのか を、エッセイ、インタビュー、表現力など を通して浮き彫りにし、その結果はプロジ ェクトがめざすビジョンとしてまとめられ た。この作業には数ヵ月かかっている。こ のようなプロセスを通して、第三者に決め られた目標ではなく、スクール・コミュニ ティ全体として自分たちが目指すビジョン を明かにするという自己決定能力を高めて いった。またその時点で、ベースラインデ ータとめざすべきビジョンの比較が行われ、 その格差を明らかにするとともに、優先す

べき項目は何なのかの議論も行い、いくつ かの優先分野が選択された。そして各分野 ごとに担当者3~5人が割り当てられ(もち ろんそれぞれの意志により) より詳細な分 析が行われた。このプロセスで担当グルー プは調査方法を学び能力を高めていくこと になる。実施体制としてはその他に情報配 布、活動調整、活動のモニタリングなどを 行う運営管理委員会(担当グループの代表、 校長、教師、スタッフ、親などの代表)と、 意思決定機関としてのSAW(school-as-awhole)が設置された。SAWには誰でも参加 できる。彼らは事前にグループ・ダイナミ ックスや会議運営、問題解決方法のトレー ニングを受けている。このエンパワメント のプロセスは、事業のサイクルの中で定期 的に繰り返される。

b. エンパワメントの評価

「現状の評価」、「将来計画の評価」、「プ ロセスの評価」、及び「アウトカムの評価」 の四つに分けて評価が行われた。それぞれ 具体的な評価指標が考えられている。例え ば、「プロセスの評価」では「現状評価の際 にすべての参加者が関与していたか」、「学 校の重要な側面がすべて現状評価の対象と なっていたか、「情報収集の方法は適切で あったか」、「学校の将来像を描く際に適切 な構成メンバーが参加していたか、「学校 側は担当グループ、運営委員会、SAWの人 選を適切に行ったか、またそれらは適切に 機能しているか」、「活動計画は試行され決 定されたか」などである。このように、ASP の評価では、実施プロセスにおいて事業の 価値や哲学が包括されているかどうかを見 ることが重要であった。そうでなければ、 その結果としてASPが目指している学校文化 の変革や教育内容の変革(エンパワメント を通して強みを生かした学校教育をめざす こと) はあり得ないからである。もうひと つの例として「アウトカムの評価」では、 アウトカムに対して当事者グループが責任 を負う能力もエンパワメントとして捉える

べきであるという考えのもと、担当グルー プ、運営委員会、SAWの三者による評価が 行われた。担当グループは、各事業に関連 するベースラインデータをもとにどのよう な状態になれば成功といえるのかを考え 「評価システム」を構築し、常に進捗状況の 見直しを行うとともに、必要に応じ「評価 システム」そのものの見直しも行っている。 SAWは、定期的に学校がヴィジョンに向っ て成果をあげているかどうかを大局から評 価するにあたり、過去のプロセスに立ち帰 る(revisiting the process)という方法を取った。 例えば、第1回目の「現状評価」において集 められたベースラインと新たな年のものと を比較し、その差を見ることにより教育の アウトカムの進捗状況を捉えたりしている。

(3) 事例から示唆されるもの

エンパワメント評価の方法論は、どんな 長期的変化を期待している事業であるかを、 参加者間でミッションを議論することを通 して確認し、現状を評価し、その結果を受 けて将来の絵を描き、さらにモニタリング の指標を含めた具体的な運営を考える、と いう一連の流れから成る。エンパワメント 評価ではこの連続性が重要で、途中で中断 されることは自己決定能力の向上につなが らない、とされている(Fetterman 2001, p. 13)。 事例においても見られるように、エンパワ メント評価は、事業終了後の効果測定だけ ではなく、事業のサイクル全体を視野に入 れたものである。決められた調査期間に評 価の結論を出すといった評価調査ではなく、 長い時間をかけて実施される評価の継続的 なプロセスであることに大きな特徴がある。 このような方法論の特徴に加え、変革とエン パワメントの目的を内包したエンパワメン ト評価はその理念への理解が不可欠で、ツ ールだけを使いこなすことでは不十分であ る。これは参加型開発で、PRA (Participatory Rural Appraisal:参加型農村調査)やPLA (Participatory Learning and Action:参加型開発 実践学習)などの参加型開発アプローチのツールを使いさえすれば参加型開発が可能になるわけではない、という議論と同様、「このやり方に沿って評価を行えばエンパワメント評価ができる」という手法ではないと考えられる。それはエンパワメントはあらかじめ想定できる活動の実施によってであるがして実現するものであるからである。

また、この事例でわかるように、エンパ ワメント評価では、コミュニティ(ここで は学校区と考えられる)における当事者グ ループの形成がひとつのキーとなる。学校 側(教師、管理部門)、生徒、親がひとつの 「グループ」として当事者性を帯びることが 重要である。つまり、それまで「第三者」 であったかもしれない人々が(例えば学校 の管理部門や地域の人々など) 事業に何ら かの関わりを持つ「行為者」へと徐々に変 化していく現象である。エンパワメント評 価ではそのような「当事者の広がり」が大 きな特徴であると考えられる。これは、例 えば変革型の参加型評価がサービス利用者 (受益者)の参加に焦点を当て、個々人の事 業へのコミットメントや行動の変容をもた らすものとは異なる(表1参照)。エンパワ メント評価による当事者性の広がりがコミ ュニティの中での自発的な行動を継続させ ることになると考えられ、エンパワメント 評価の重要な側面であると思う。

もうひとつのポイントは、自己責任としてのアカウンタビリティの側面である。エンパワメント評価は自己評価であるので、アカウンタビリティの確保が難しいのではないとの議論もあるが、エンパワメント評価では自己説明責任(self-accountability)が最もレベルの高いアカウンタビリティであるとして、アカウンタビリティの関係が説明されている(Fetterman 2001, p. 118)。事例においても、自己の責任を明確にするアウトカム評価の方法が組み込まれている。エンパワメントのプロセスで、力をつけ決定

権を増していくことは、参加者(=当事者) の責任がより明確になることであるという 考え方が根底にある。当事者が自らで評価 を行うことの根源には、「自分たちの事業を どのように組みたてるかは自分たちの責任 ではないか」という問いがある。ただし、 この責任をどのように取るのか、その仕組 みを内在化させることは、評価の仕組みと ともに重要になる。もとより、評価の仕組 みだけで事業の成果が上がるのではなく、 組織の意志決定の仕組み、人事、予算など 様々な仕組みとの連携が必要になる。エン パワメントと社会変革が起きるためには、 エンパワメント評価はひとつの手段になり 得るが、それが全てはない点に留意する必 要があるであろう。

4. エンパワメント評価の適用への 示唆

これまでの概念整理と方法論、事例を踏まえ、エンパワメント評価がどのような事業に適用可能であるか、その留意点も含め 考察する。

(1) 適用の範囲と可能性

セスであるので、そのプロセスをとおして、 人々のエンパワメントが促進され、評価活動が内在化されることにより、評価対象と なった事業をエントリーポイントとして新たな事業の広がりを促す手段としての活用 も期待できよう。

例えば、開発援助の分野では、住民が主 体、もしくは住民が直接的な受益者となっ ている社会開発型プロジェクトの評価にお いて活用することができるだろう。彼ら自 身の学習と適切な事業運営を側面から支援 することが可能になるからである。これは 援助という外からの介入がフェーズアウト する際の手段としても有効かもしれない。 開発援助事業の評価基準である評価5項目6に 基づいて評価する方法は、援助する側にと ってはマネジメント・ツールとして有効かも しれないが、当の住民や現地NGOにとって はあまり意味を見出せないかもしれない。 なぜならば、第三者が設定した実証可能な 指標の達成度合いやドナーのロジックは、 当事者の現実からかけ離れているかもしれ ないからであるっ。エンパワメント評価を手 段として自らが学習し、選択し、決定し、 開発の効果を定着させていくことが現地当 事者の関心に合う。その場合のエンパワメ ント評価は、ドナーの側から見ると参加型 モニタリングの一環として位置付けること ができよう。これと5項目評価が併行して行 われた場合は、モニタリングで得られた情 報(プロセス評価、彼ら自身が設定した指 標の変化など)は、5項目評価の貴重な情報 源となるはずである。一方、エンパワメン ト評価を現場で導入することは、ドナーの コントロールを越えたところでプロジェク トが展開される可能性もあるということを 意味する。これは、開発援助という外から の介入が、真の意味での主体性を伴ったエ ンパワメントとどう向きあうのかという問 題を提起する。

次に、国内の事業で言えば、NPOによる 社会サービス事業の評価方法として活用で きるのではないかと考えられる(行政から

の委託も含む)。NPOは、経済的価値以外の 多様な価値観を前提とし、市民の自主的な 参加と支援によって運営されている。また、 自己管理、自己決定などをキーワードに社 会変革を促進する役割を持つNPOも多い (雨森 2002, pp.24-28)。NPOの中でも特に、 福祉分野(障害者、高齢者) まちづくり、 村おこし(過疎地の住民) 教育分野(不登 校児童)など、社会的に弱い立場の人々と ともに活動を展開しているNPOの評価方法 として、エンパワメント評価を適用する意 味は大きいと考えられる。なぜならば、お そらくそれらの活動の究極の目的は、サー ビスの利用者でもある人々の自立支援であ るからだ。例えば、高齢者介護サービスを 提供するNPOの例を考えてみよう。エンパ ワメント評価で事業のミッションを問いか けることは、「高齢者の自立とは何か」を考 えることでもあり、利用者の声に耳を傾け ることが不可欠になる。利用者のみならず、 その家族、サービスを提供している介護員、 地域のボランティアなどの参加も必要だ。 そして、どんなサービスを提供すると(利 用者から見ればどんなサービスを選ぶと入 描いていた自立が近づくのか、という視点 から参加者全員で現状を評価し、新たな将 来戦略を考えていく。そのプロセスで、高 齢者自身が新たな自立への一歩を踏み出す ことができると理想的である。また、地域 の人々も第三者としてではなく、地域の当 事者として高齢者の問題やNPOの運営にか かわるようになることが期待される。エン パワメント評価が重きを置く対話のプロセ スは、福祉介護の分野においてサービス提 供者と利用者(=弱者)の間に生まれ易い 権力関係(笹川平和財団 2001, p. 19)を見直 す契機ともなる。

行政との協働関係による事業では、市民の多様な価値を反映させ、公共的な意思を形成する手段としても活用できる。これは協働事業の基本である目標の共有のためには大きな意味があるし、NPOにとっても、行政の都合の良い委託先に甘んじるのでは

なく(NPO側から見ると運営を安定させる 資金源として歓迎するのではなく)、本来の 目的である市民社会の形成に向けて、ひと つの試みに値する評価手法であると思う。

(2) 課題及び留意点

さて、これらの評価を行う上での課題の ひとつとして、エンパワメント評価を実施 する評価専門家の認識や能力がある。査定 者としての評価者ではなく、ファシリテー ター、コーチとしての機能が求められるこ とは前述した通りであるが、従来の評価に 従事する場合とは異なった役割や能力が伴 うことに留意する必要がある。まず、自分 が持っている評価の専門知識を参加者に伝 え、それを参加者が使って評価できるよう にトレーナーとしてサポートする役割があ る。評価人材を養成する立場である。また、 どの参加者にとっても中立的な立場でワー クショップを進行し(つまり、いろいろな 意見を相対化できること) すべての参加者 が対等な立場で意見を述べ合うことができ るような様々な工夫やグループダイナミズ ムを生み出すスキルが必要である。ファシ リテーションのプロセスでは、参加者の合 意を形成することが最終成果では必ずしも なく、人々のコミュニケーションによる相 互作用を生み出すことが重要であると考え られる。評価のプロセスで展開されるコミ ュニケーションが、バイアスを取り除き、 自己説明責任を果たすことにつながるとし ているエンパワメント評価において、この ような評価専門家の役割は大きい。そして 何よりも、評価対象事業がめざす社会変革 やエンパワメントの方向性への理解が重要 であると考えられる。これは、概念的には 評価者自身を事業の真っ只中に位置付ける ことであり、評価専門家として事業の中で 評価並びにそれを内在化した事業を支援す ることになる。エンパワメント評価の大き な特徴のひとつは、おそらく、評価専門家 にこのような新たな役割を課したという点 にあるのではないだろうか。エンパワメント評価がめざす精神の解放、エンパワメント、さらには社会変革といった展開は、評価専門家に社会運動家としての側面を持つことを求めている。

この評価専門家の役割とエンパワメント 評価への関わりに関連していくつかの留意 点が浮かびあがる。いずれも実践の中で注 意しなければならない点として四つ指摘し たい。まず第1に、客観性の問題である。エ ンパワメント評価は第4世代評価の構成主義 に代表されるように、評価結果が「事実」 として存在していたものではなく、評価者 を含む関係者間の相互作用のプロセスで 「創られた」ものであるというスタンスは、 最初から客観的な存在を前提していないと 考えられる。その評価結果は多様な価値が ぶつかりあった結果生み出されたものであ り、科学的に証明されうる正解としての評 価結果ではなく、評価以降も関係者が交渉 を継続する内容として扱われるべきだとい う立場をとる。つまり、当事者でなければ 出来ない評価であるからこそ、普遍化は前 提としないことになる。「評価以降も関係者 が交渉を継続すること」は、評価の内在化 による繰り返しのプロセスを意味し、その システムを当事者グループの中に組み入れ ることが重要なポイントになるであろう。

第2に、バイアスの問題がある。エンパワメント評価では対話のプロセスの中で多れては対話のプロセスがイアスが相互監視することにより、バイアスができるだけ取り除かれるというスタンスすり、評価のプロセススでするという。つまり、コミュニケーシッとが表して信頼性のあるという(Fetterman 1996, p.23、とは、その議論の中できちんとした論の中できるとの仮定があるという(Fetterman 1996, p.23、またことの仮定があると共通した連続にできるとの仮定があるというには疑問を違いたが常にでは疑問もある。異なった利害では疑問もある。異なった利害では対話を通りについては疑問もある。異なった利害では対話を通りでは対しては疑問もある。異なった利害では対話を通りでは対しては疑問もある。異なった利害では対話を通りでは対しては対しては対しては対しては対しては対しては対しては対している。

が鋭く対立することも容易に想像できる。また、合意をつくろうという動機を持ちうる人だけが評価に参加できる、など参加者の選定にバイアスがかかる可能性もある。エンパワメント評価の最初に行う「ミッションの確認」は、合意を形成する動機付けがどこにあるのかをもう一度問い直すという意味において、バイアスの影響を軽減させるひとつの方法として位置付けることができるのではないだろうか。

第3には事業の代弁者として関わることに よる、評価者としての中立性の問題がある。 Patton (1997) は評価者が、自らが関わった 評価結果のアドボカシーを行うことは、参 加型評価においてはめずらしいことではな いが、エンパワメント評価ではさらに一歩 進んで、事業そのもののアドボカシーまで 踏み込んでいることが議論を呼ぶ点である と指摘している(pp.157-159)。対象事業の 代弁者となることにより評価者の中立性が 歪められる危険性があり、評価結果に対す る信頼性が損なわれるという指摘である。 評価専門家が評価結果のアドボカシーに積 極的に関わることは、内部評価においては 事業改善や組織改善を促す評価専門家の役 割として認識されており、それは公平な評 価の後の行為であり、評価の信頼性をなん ら損なうものではないとされている (Sonnichsen 2000)、Fetterman自身は、エンパ ワメント評価においても評価専門家はあく までも評価結果を受けて事業を代弁してい くのであり、評価如何に関わらず事業に肩 入れすることではないと強調しているが、 その一方で、自らをエンパワメントするこ とが難しい社会的弱者のグループに対して は直接エンパワメントのサポートを行う役 割も想定している (Fetterman 1996, pp.12-13, 同2001, pp.115-117)。エンパワメント評価が 従来の内部評価と異なることは、当事者が エンパワメントを実現するプロセスを提供 し、しかも当事者自らが評価することによ ってエンパワメントと変革が可能になる仕 組みである点だ。そのプロセスを有効に機 能させるためには、それに関わる評価専門家の「エンパワメント」への理解と事業のものへの長期的なコミットメントがの程度必要になるのではないかというのがまである。それが外部評価専門そる全国動家としての側面を求める背景になってはないかと思う。実際のあるに際しては、事業へ過度に肩入れするるに、操作的なファシリテーションをするだいないように十分留意する必要があるだろう。

第4の点として、エンパワメント評価は、 自己決定能力を高め何らかの変革をもたら すためのひとつの手段であることに留意す る必要がある。自己決定能力の強化がどの ようなメカニズムで社会変革につながって いくのかは未知である。また、事業全体の 目標達成のためには、当事者グループによ る意思決定の余地を残した事業運営体制に 加えて、人事、予算など組織の運営に関わ る様々な要素が適切に機能しなければなら ない。そのためには、経営者の視点からの 事業評価や第三者評価も必要であろう。組 織全体から捉えるとエンパワメント評価は、 関係者の間で共有できる客観的意味を形成 していくプロセスに光を当てている。今後、 我が国において事例を積み重ねることによ り、エンパワメント評価と組織・事業運営 に関するメカニズムの研究が進められる必 要があろう。

5. おわりに

エンパワメント評価は、Fetterman自身の言葉を借りると、「評価の概念と技術を使い改革と自己決定能力を強化すること」(Fetterman 1996, p. 6)である。また社会変革、対話プロセスの重視、アドボカシーにおける評価者の役割、評価活動の内在化、参加者の精神の解放などの新しい側面を持ち合わせた評価である。そこでは、当事者グループ(事業実施者、関係者を広く含む)が

自らエンパワメントしていくプロセスが重要であると同時に、エンパワメントのアウトカムも評価の対象となる。エンパワメントと社会の理念とを内包した、当事者による評価の手法である。そこでは評価専門家は評価者としての査定は行わない。つまり、見でである、評価専門家が当事者の主体的事業にファシリテーターとして「参加」する評価手法なのかもしれない。

評価はその目的によって適用する理論や 手法は異なるものであるから、Fetterman自 身も述べているようにエンパワメント評価 も他の従来型評価を排除するものではない。 例えば、NPOがサービスの利用者や関係者 とともにエンパワメント評価を行う一方で、 経営者としての事業評価は当然必要になる であろう。開発援助プロジェクトの場合も、 ドナーの視点に立った5項目評価とエンパワ メント評価は両立し得る。同一の事業を対 象にしている限り、それらの評価は相互に 関係し合っている。特に、他の参加型評価 とは異なり、エンパワメント評価は対象事 業の将来計画を関係者で共に考え、ロジッ ク・モデルなども活用する。また、継続的 なモニタリングや評価を行う指標やモニタ リング方法の検討も行われる。つまり、従 来型の評価を行う側にとっては、評価対象 となる事業の整理が参加型のアプローチに より既に行われるということになる。エン パワメント評価と従来型のマネジメントツ ールとしての評価の組み合わせは、「お手盛 り評価」が懸念される事業実施主体のみに よる内部評価の仕組みの中に、社会の多元 的な視点を総合的に取り入れる評価の形と して注目に値する。さらには、エンパワメ ント評価はコミュニティの市民活動や公益 事業を評価する上で、広く当事者意識を醸 成し、市民が行為者として変る機会ともな る。また、それは市民の側から見ると、そ のプロセスに関わることで自分自身が変っ ていくというプラスの経験をすることでも あり、市民の評価に対するインセンティブ が高まることも期待される。

Fettermanのエンパワメント評価は、「評価」 の技術がエンパワメントや社会変革と結び ついた方法である。もっと言えば、評価が 抑圧された人々の自立にどのように役に立 つかを問うている。もちろんエンパワメン ト評価だけで自立が達成されるわけではな い。今後事例を積み重ねていくことにより、 その理論、方法論を検証し、エンパワメン ト評価と社会変革、自立との関連や組織・ 事業運営のメカニズムを調べる作業が必要 になると考えられる。その作業は、また、 社会における公益事業・社会サービスのあ り方や、広い当事者意識に支えられた、地 縁、血縁を越えた自発的なコミュニティが どのように発展していくのかを考えるひと つの契機になるかもしれない。

謝辞

本論文の草稿に対して、立命館アジア太 平洋大学三好皓一教授及び査読者の方々か ら数々の有益なご指摘をいただきました。 この場を借りて、深く感謝申し上げます。

注記

- 1 例えば、PRA(Participatory Rural Appraisal:参加型農村調査)、PLA(Participatory Learning and Action:参加型開発実践学習)など。しかし、それらの活用に当たっては、それらのアプローチのツールだけを取り出して使い、参加型と銘打って単なる住民からの情報提供や意思決定権が伴わない表面的な参加に終わっている例が多いと指摘されている(佐藤 2003)。
- 2 例えば、Patton (1997) は、事業の代弁者 ともなる評価専門家が、信頼性、有用性 等の観点から適切な評価ができるかどう か疑問を投げかけている。
- 3 Patton (1997) は、エンパワメント評価の

事例として発表されているもののなかに は、Fettermanの提唱する5つのキーワードに適した事例と、エンパワメント価のを使いながらも他の参事例が見れている。このひと指摘していると指摘している。このひと指摘していると指摘している。このひとにかなった要はないのではないが、でする検証が可能になるのではないが、と述べている。

- 4 H.M.Levin (1996) による論文を要約した。
- 5 注記1参照。Fetterman自身もエンパワメント評価の方法と組織の自己評価をマッチングさせ、その評価結果を外部評価の情報源として活用した事例を紹介しているが、エンパワメント評価の特徴である「精神の解放」、「変革」、「エンパメント」などの要素は見られない(Fetterman 1996)。これはエンパワメント評価の「ツール」を応用した事例として捉えるべきではないかと思う。
- 6 政府開発援助事業の評価基準として、 1991年に経済協力開発機構(OECD)の 開発援助委員会(DAC)で提唱された5 項目が広く使われている。具体的には、 妥当性、有効性、効率性、インパクト並 びに自立発展性である。
- 7 例えば筆者が関わった、ケニア国NGOによる貧困層の生活改善プロジェクト(JICAのパートナーシップ事業)の参加型評価では、住民のエンパワメントが重的のひとつであったために、プロジックを当てはめることが難しかった。現場で起きていることは、ある活動をごり、ではないうりニアな関係性ではなく、活動したらある成果と指標が測定できるのロセスで様々な要素が相互に関連していることは、カラリニアな関係性ではなく、活動したのカラリニアな関係性ではなく、活動し合いカリニアな関係性ではなく、活動したのカラリニアな関係性ではなく、活動したのカラリニアな関係性ではないカラリニアな関係性ではないカラリニアな関係性ではないカラリニアな関係性ではないカラリニアな関係性ではないカラリニアな関係性ではないカラリニアな関係性であったが別連している。

- 評価方法としては、PDMを参考にしながらもNGOスタッフを中心に、プロジェクト全体の関係性をもう一度描き出し、それに基づいて彼ら自身の言葉によって特定されたいくつかのイシューを更に掘り下げるという作業を行った。
- 8 佐藤(2003)は参加型開発ではドナーが 方向性をコントロールすることが困難に なりがちであることを指摘し、経験的な 原則として プロジェクトは参加型にな ればなるほど、セクターを越境する、 プロジェクトは参加型になればなるほど、 村人の戦略が優越するという2点をあげ、 参加型開発のリスクを分析している (pp.25-27)。
- 9 エンパワメント評価におけるミッション や目標の確認、指標の検討などは、評価 可能性調査(Evaluability Assessment)の ー環としても捉えることができると思う。

参考文献

- 雨森孝悦(2002)『NPO論』、日本福祉大学 通信教育部
- 猪俣正雄(1992)『組織のコミュニケーション論』、中央経済社
- 国際協力事業団 (2001) 『国際協力と参加型 評価 』 国際協力総合研修所
- 齋藤文彦編著(2002)『参加型開発』、日本 評論社
- 笹川平和財団 (2001) 『評価のすすめ 福祉 分野で活躍する非営利組織のために』
- 佐藤寛編(2003)『参加型開発の再検討』、 アジア経済研究所
- 西尾勝(2000)「行政の評価方式の拡張をめ ざして」、『行政評価の潮流 参加型評価 システムの可能性 』、行政管理研究セン ター、pp.1-35
- フリードマン、ジョン (1995) (齋藤千宏・雨森孝悦監訳)『市民・政府・NGO 「力の剥奪」からエンパワメントへ』、新評社(原題: Empowerment-The Politics of Alternative Development, 1992)

- 古川俊一(2000)「市民参加の評価方式 制度化の可能性と課題」、『行政評価の潮流-参加型評価システムの可能性』、行政管理研究センター、pp.39-76
- 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):65-79
- 山谷清志(2000)「評価の多様性と市民 参加型評価の可能性」、『行政評価の潮流参加型評価システムの可能性』、行政管理研究センター、pp.77-108
- Brisolara, S.(1998). The History of Participatory Evaluation and Current Debates in the Field, Understanding and Practicing Participatory Evaluation, *New Direction for Evaluation*, American Evaluation Association, Jossey Bass, San Francisco pp.25-41.
- Cousin, J. B. and Whitmore, E. (1998). Framing Participatory Evaluation, *Understanding and Practicing Participatory Evaluation*, New Direction for Evaluation, American Evaluation Association, Jossey Bass, San Francisco pp.5-23.

- Fetterman, D.M. (2001). Foundation of Empowerment Evaluation, Sage. (1996). Empowerment Evaluation-An Introduction to Theory and Practice, Empowerment Evaluation-Knowledge and Tools for Self-Assessment & Accountability, Sage, pp.3-45.
- Guba, E.G. and Lincoln, Y. S.(1989). Fourth Generation Evaluation, Sage.
- Hardina, D.(2002). Analytical Skills for Community Organization Practice, Columbia University Press.
- Levin, H. M. (1996). Empowerment Evaluation and Accelerated Schools, Empowerment Evaluation-Knowledge and Tools for Self-Assessment & Accountability, Sage, pp.49-64
- Sonnichsen, R.C.(2000). High Impact Internal Evaluation -A Practitioner's Guide to Evaluating and Consulting Inside Organizations, Sage.
- Patton, M. Q. (1997). Toward Distinguishing Empowerment Evaluation and Placing It in A
- Larger Context, *Evaluation Practice*, Vol.18, No.2, pp.147-163, JAI Press.

(2003.8.4受理)

Characteristics of Empowerment Evaluation and Possibility of Its Application -- Focusing on Theory of Empowerment Evaluation by Dr. Fetterman --

Yuriko Minamoto
yuminamoto@mvd.biglobe.ne.jp

Abstract

Empowerment Evaluation by Dr. Fetterman is the use of evaluation concepts and techniques to foster transformation and self-determination. It provides the process for a group of participants concerned to evaluate by themselves as well as to provide internalization of evaluation system, so that they can continue self-evaluation. Major roles of evaluators include facilitator, trainer of evaluation know-how and even direct advocate to help the group be empowered. Empowerment Evaluation appears most appropriate when the program itself has agenda of empowerment and social change. The mixed use of Empowerment Evaluation and traditional types of evaluations would make it possible to implement value-pluralistic evaluations, and that would certainly provide us with opportunity to reconsider development of community by voluntary people. The accumulation of experiences of Empowerment Evaluation in Japan is expected for further testing and adjustment of its theory and methodology.

Keywords

Empowerment Evaluation, Participatory Evaluation, Transformation, NPO, Participatory Development

【研究論文:依頼原稿】

ドラッカーによる成果重視の自己評価手法 ~ 評価領域拡大への示唆 ~

田中 弥生

伊永 大輔

東京大学 yayoi_tanaka@ohriki.t.u_tokyo.ac.jp 慶應義塾大学大学院 korenaga@1999.jukuin.keio.ac.jp

要約

本稿は、P.F.ドラッカーの提唱する成果重視の自己評価手法のフレームワークと特徴を紹介するものである。ドラッカーの発想には、時々刻々と変化する環境へ対応することによって組織のイノベーションが可能になるという思想が根底にある。そのため、本評価手法は、マネジメントの改善のみならず、組織の使命やガバナンスの見直しを促すことを主目的としている。すなわち、自己評価手法はイノベーションとマネジメント改善のための手段として明確に位置付けられているのである。これらの特徴を浮き彫りにするために、成果重視を特徴とした評価手法として代表的なResult Based Management (RBM)と評価手法の比較検討を行う。RBMはPerformance Measurementによる業績情報の利用を取り入れた評価枠組みであり、事業効果の仮説と検証に基づく評価と捉えることができる。しかし、ドラッカーの評価手法は仮説の検証に執着せず、環境変化を察し仮説そのものを改変することを促している。この点はRBMのみならず従来型の評価と異なる特徴であり、そのため既存の評価領域を拡大する可能性を示唆しているといえる。

キーワード

P.F.ドラッカーの自己評価手法、成果重視、マネジメント、民間非営利組織、イノベーション

1. はじめに

民間非営利組織'(Non-profit Organization)の研究の多くが古典的な経済学を基礎としている中で、独自の視点で民間非営利セクターの重要性を説いたのが、マネジメントの権威として知られるP.F.ドラッカーである。ドラッカーは1960年代にその著書『断絶の時代』の中において、専門知識が個人および経済諸

活動の中心的資源となる社会である「知識社会」の到来を指摘しており、民間非営利組織に1980年代より着目し世界的な動向になることを早くから予測していた。その一方で、民間非営利組織の本質的問題である共通の業績判断基準の不在をも指摘していた。

このような背景をもとに、ドラッカーによって民間非営利組織のための自己評価手法² (Self-Assessment Tool)が開発されることに

日本評価学会『日本評価研究』第3巻第2号、2003年、pp.87-107

なる。この自己評価手法は、簡潔で分かりやすい設問によって構成されるが、従来の評価 手法との共通点をもっている。その意味で、 民間非営利組織のみに適用されるものではなく、行政機関や他の公的セクターに対しても 評価手法としての示唆をもつ。以下、ドラッカーによる非営利組織論から成果重視の自己 評価手法の背景を探り、評価手順を概観する。 そのうえで、従来の評価手法の評価目的や視点との相違点を浮き彫りにし、自己評価手法 からの示唆を考察する。

2. ドラッカーの思想と自己評価手法の背景

(1) ドラッカーの非営利組織論

「知識社会」の到来と問題意識

現代社会においては、古典的な経済学の示す土地、労働、資本という生産要素に比しって、典型的には技術革新(innovation)によって表現される専門知識の重要性は増大するっちである。労働の生産要素の個別的な内容を情である。労働の生産であるではなく、高度トワーク化されることによって、初めて社会的できる。その意味で、現代は優・大いの意味で、現代は優・大いの意味で、現代は優・大いの意味で、現代は優・大いの意味であり、社会経済の発展に対して、東門知識の要素を無視することはできない。

このように、各個人が所有する専門知識が 組織的に統合されること、つまり組織化され た専門知識への転換が「知識社会」を特徴づける。現代社会におけるニーズの多様化、そ して価値観の多様化がそれに応じた専門知識 の多様化を産むため、必然的に多数・多元の 組織が存在することになる。しかし、この多 元的な組織の存在は、同時に一つの大きな問題を生むことになる。知識が専門的であるが ゆえに、組織化し得ないあるいは断絶 (disintegration)への途を辿らざるを得ないと いう内在的な宿命ともいえる問題である。

「知識社会」の到来によって生じる、もう 1つの問題は「流動する知識ワーカー」と組 織の問題である。「知識社会」の主要な労働 力は知識ワーカーである。彼等の忠誠心は自 分の知識に対してあるのであって、組織に対 してあるのではない⁴。つまり、知識ワーカ ーは、組織のために働くのではなく、自分の 知識を生かすために組織で働くのである。 知識ワーカーの増加による組織離れは社会生 活を流動的にしているが、他方では、このよ うな知識ワーカーほど帰属意識と精神的な拠 リ所を求めている(Drucker, 田中訳 1995)。 ドラッカーは、このような拠り所となる場の 一つの答えがコミュニティであると述べてい る (Drucker 1995)。 ここで言うコミュニテ ィとは、行政区域のような地理的なものを示 すだけでなく、共通の関心や属性によって形 成される人々の集合体を指す。知識ワーカー の参加と帰属への欲求を満足させるために は、市民としてコミュニティ構築に参加する 機会を提供する組織が必要なのである。

民間非営利組織の重要性

これらの問題を解決するためには、専門知識という一種「タテ割り」の枠組みを超えた「ヨコ」の組織概念が必要とされる。つまり、市民社会(Civil Society)という、個としての専門知識を連携させる新しいパラダイムはの専門知識を連携させる新しいパラダイムは、の専門知識を連携させる新しいパラダイムは、の事とされる。このようなパラダイムは、する、は、かるやかは、明治では、例えば、高度情報では、の統合の代表の代表のではいが、民間でしかも非営利であるといってよいが、民間でしかも非営利であるといってよいが、民間でを遂げたのは記憶に新しい。

また、民間非営利組織は、人々の社会的・ 心理的欲求を満たしうるコミュニティの構築 という公益目的に向かった人々を統合し機能 させる組織として存在意義を持つ。つまり、 民間非営利組織の最も重要な役割の1つに「市民性の創造」があるといえる⁶。現代の社会と政治は、あまりにも巨大かつ複雑なものとなり、責任ある市民としての社会参加機会の多くは投票や納税によるもので、では、一人でしかない。してないでである。民間非営利組織のボランティアとしてなっている。してできる(Drucker,上田他訳 1993a)」のである。民間非営利組織は、市民に積極的・直接的な社会参加の機会を与えることのできる「市民性創造機関」であるといえる⁷。

(2) 評価手法開発の経緯

ドラッカーは、民間非営利組織の重要性を個々の組織レベル、セクターレベルの双方から説いているが、その本質ともいうべき課題も見逃していない。民間非営利組織はその使命と活動目的が善意に基づいているために、それだけで十分だと誤解しがちである。結果として、活動の焦点を絞り切ることができず、資金や人材を無駄に使ってしまうため、十分な成果が出ないことになる。そのため、民間非営利組織は活動と結果について責任を問う視点が不足していると、ドラッカーは指摘している。

このことは「あらゆる民間非営利組織が、最もよくマネジメントされている民間非営利 組織と同じように、自らをマネジメント言とのはればならない(Drucker 1992)」との言とのでまとめられている。非営利組織は企業ののに決算がなく、そのため業績判断の共が存在しないため、マネジメント。ま達ないがうのである(Drucker 1992)。まき、民間非営利組織特有の問題である令にな出来がであった。といずマーケティング、成果、そしてマネジストが不可欠であった。しかし当時はははに関非営利組織向けのマネジメント研究は緒に ついたばかりの時期であった。そのため、民間非営利組織のマネジメント上の課題と解決の方向性を*自己評価手法*というツールに表現しようとしたのである。

非営利組織のための自己評価手法の開発が始められたのは、非営利組織マネジメントのためのドラッカー財団(P.F. Drucker Foundation for Nonprofit Management)設立直前後の1990年代初頭からであった。この評価手法の背景にある問題意識は、民間非営利組織へのマネジメント導入とその改善の必要性である。そのため、専門家の手を借りずに、民間非営利組織がマネジメント改善のための自己診断をするツールを目指し、ビジネス・スクールや非営利組織のリーダーの助言を受けながら完成したのが「The Five Most Important Questions, the Drucker Foundation Self-Assessment Tool for Nonprofit Organizations (1993b)」であった。

3. ドラッカーの自己評価手法

(1) 5つの設問

自己評価手法は5つの簡潔な設問によって 構成されているが、これらにはドラッカーの 経営思想が集約されている。各設問に細目質 問と解答欄からなるワークシートが用意され ているが、設問を実際の活動レベルに落とし 込んで考えるように工夫されている。また、 組織内でワークシートを使いこなすための準 備と手順、役割分担まで具体的に解説されて いる。ここではドラッカーの経営思想をもと に、設問と細目質問を簡単な事例を用いなが ら解説する。

設問1「われわれの使命は何か?」

使命とは、組織の存在理由そのものであり、 組織の出発点であり終着点である。そのため、 活動の手段を述べるものではない。 使命によって、 はじめて組織の方針と行動が規定され るのである。 組織の構成員は、 使命に共感す るからこそ士気を持ちる。共通の目標にて、 で行動するとがでわかりでまれ、 もので理解でいたが表現では、 もので理解でのはい場合では、 もので理解でではいる。 の間でのでででではなるががのででででででででででででででででででででででででででいる。 はいるのででででではないでででででででででででででいる。 はいるのでででではいるでででででででいる。 はいるでででででいるでは、 はいるででででいる。 はいるででではいるでは、 はいるでででは、 はいるでででは、 はいるででは、 はいるでででは、 はいるででは、 はいるでは、 はいるででは、 はいるのででは、 はいるのででは、 はいるのででは、 はいるのででは、 はいるのででは、 はいるのででは、 はいるのででは、 はいずでは、 はいずでは、 はいずでは、 はいずでは、 はいずでは、 はいずでは、 はいるのででは、 はいずでは、 はいずでは、 はいずでは、 はいずでは、 はいずでは、 はいずでは、 はいるのででは、 はいずでは、 はいずでは、 はいずでは、 はいずでは、 はいがでは、 はいるのでは、 はいずでは、 はいるのでは、 はいずでは、 はいるのでは、 はいずでは、 はいるのでは、 はいるのでは、 はいるのでは、 はいずでは、 はいるのでは、 はいるので

< 設問1の事例 >

ある大学の設立趣意書には「大学は古来、『分明の伝承』『知的生産』『人格の陶冶』を目的としてきた」と記されている。これは大学の使命を示しているが、簡潔にまとめられた内容は普遍的であり、かつ研究、知的財産の蓄積と継承、人材育成というように使命達成のための手段や方向性も記されている。

第一の設問に附随するワークシートを表1 に示した。最初の質問は「現在の使命は何か」 である。これによって、現在における、記さ れている使命の意味と達成目標を確認するよ うに導いている。第2の質問は「われわれの 課題は何か」であり、環境変化を課題と捉え、 短期的、長期的視野から確認させようとして いる。これは「われわれの機会は何か」とい う次の質問に通じ、課題と機会は表裏一体で あるとして、課題の中に新たな機会を見出そ うとしているのである。そして機会の選択基 準として、支援者の意向、使命達成にとって の有効性を挙げ、最も有望な機会を決めるよ うに導いている。最後に「使命を見直す必要 があるか」という質問が用意され、その細目 を通して、使命の表現方法の適切さを確認さ せ、必要であれば見直しをさせようとしてい る。

設問2 「われわれの顧客は誰か?」

民間非営利組織には二種類の顧客が存在する。「第一の顧客」と「支援してくれる顧客」である。「第一の顧客」は非営利組織のサービスの対象となる人々のことであり、変革されうる人々、すなわち人間変革機関の対象者を表す(Drucker 1989)。「支援してくれる顧客」とは、寄付やボランティアによって組織を支える人々である。これは、非営利組織のもうひとつの重要な役割である「市民性創造」の対象となる人々のことである(Drucker 1993a)。

従来、非営利組織は対象となる人々を「受益者」と呼び、有り難くそのサービスを受けるべき人という固定観念を抱いていた。しかし、彼等もサービスの選択権を持ってなりでは、ではいるではないとさいるでは、あるでは、対象となるを関けしている。また、見せ掛け、対象となるが、顧客のでは、対象に取るが、がで見据えながら逆に対象を終る必要のほともで見ばながらい。そのためには、応じならないを持った優先順位をつけ、必要に心が幅を持った優先順位を可け、必要に心がになるがあるべき対象を見定めねばなられるがあるがあるがあるがにあるがあるがあるがあるがある。

< 設問2の事例 >

全国の社会福祉協議会はボランティア・センターを運営している。ボランティア希望者に仕事を斡旋することが主たる仕事である。以前、登録者の多くは中高年の主婦層であったが、近年では、中高年の男性が増えているという。理由を聞いてみると、退職後の第2の人生にボランティア活動を選択する人々が増えているかららしい。ボランティア・センターの顧客に変化が生じているのである。顧客の変化を察知したら、新たな顧客層の拡大状況を把握し、より成長の可能性

があると判断したら彼等の志向に合った ボランティア活動メニューを準備する必 要がある。

ワークシートでは、まず「第一の顧客」と

「支援してくれる顧客」の現状把握を確認させる。次に、「われわれの顧客はどのように変化しているのか」という質問に続き、それぞれの顧客について質、量の変化を確認するように細目が用意されている。その際、変化

表1 設問1 「われわれの使命は何か」に基づくワークシート

【ワークシート1】 現在の使命は何か?	●現行趣意書の添付●その使命はどのような意味をもっているか●どのような分野で、組織は使命を達成しようとしているのか
【ワークシート2】 われわれの課題は何か?	選択肢: 短期的、 長期的、 両方、 現在はない ・人口構成の変化 ・コミュニティの状態の変化 ・文化的・社会的トレンド ・経済的・財政的環境の変化 ・組織と組織が対象としている人々に影響を与える政治、法律、規制の動向 ・競争状況(同業者) ・新しい技術、モデル、方法 ・その他 ●以上のうち組織にとってもっとも重要なものはどれか。その理由はなにか
【ワークシート3】 われわれの機会は何か?	選択肢: 短期的、 長期的、 両方 ・差し迫った問題や状況に取り組むためにどのような機会を利用できるか ・組織はサービス分野のギャップを埋めるためにどのような機会を利用できるか ・組織はパートナー、または資金提供者となる可能性のある人の興味を満足させるために、どのような機会を利用できるのか ・新たに到達目標を設定するために、どのような機会を活用できるのか ● 組織にとって最も有望な機会とは何か。その理由はなにか
【ワークシート4】 使命を見直す必要があるか?	選択肢: 全くその通り、 ある程度そうだ、 全然そうではない ・使命は短く、焦点を絞ったものである ・明確でわかりやすい ・なぜわれわれはこの事業に従事しているのか、つまり組織が存在しているのかを定義している ・手段を規定していない ・広い視野をもっている ・方向づけがなされている ・ 機会について述べている ・ われわれの能力に合致している ・ われわれのやる気をおこさせるものである ・ 最終的にこの使命を理解し、人々の記憶にとどめておいてもらいたい ・ 使命は見直すべきだ ●もしそうなら、どのように変更すべきか

(出所) Drucker, 田中監訳 2000, p.14-18

の原因あるいはそう判断した理由も確認させ、直感のみに頼らぬように注意が払われている。そして、使命達成のために新規開拓すべき顧客の存在を確認させ、同時にサービスを止めるべき顧客とその理由を確認させる。中止すべき顧客対象を明らかにし、その分の資源を新たな顧客に充当すべきという、選択と集中の思想はここにも表れている。

設問3 「顧客は何を価値あるものと考えるか?」 第一の顧客、支援してくれる顧客のニーズ、 欲求、希望を確認する問いかけで、非営利組 織側が一方的に憶測してはならないと忠告す る。Kotlerが「多くの非営利組織は自分たち の価値については明確に理解しているが、顧 客の視点や価値についてはあまり良く理解し ていない (Drucker, 田中監訳 2000, p.31)」と 述べているように、使命が善意に基づくもの であるが故に自らの目標に埋没して肝心の対 象者の希望がなおざりにされるという指摘 は、民間非営利組織が時として独善に陥る危 険性と原因を端的に説明している(田中 1999)。これを回避する一つの方法は、顧客 に尋ねることである。非営利組織が予測して いた顧客の価値と顧客の回答の間にギャップ があれば、現在のサービスの成果が上がって いないことを示している。

< 設問3の事例 >

公民館では住民を中心としたサークル活動が行われているが、演劇、ダンス、合唱等熱心に稽古が行われている。そとで、稽古の成果を披露したくなったとが多いという。その都要もしたなることが多いという。その都要するのは毎週のように対しているが、施設の住民もうんざりしているが、施設の住民もうんざり価値に対るという。このサークルは顧客は全く活動の感情を抱いていた。博愛に基づく活動

は確かに尊いが、見誤ると有り難迷惑に なりがちである。

(出所)田中2000, p.4

ワークシートでは、まず非営利組織自身が考える顧客価値を列挙させることから始める。次に、「顧客から学ぶ」という表現を用いて、どの顧客から何を聞き出したいか考えさせる。そして、誰がどのような方法で顧客に尋ねるのかを検討するように、質問票や面接などの細目リストが記されている。

設問4 「われわれの成果は何か?」

< 設問4の事例 >

骨髄バンク(骨髄移植推進財団)は、 骨髄ドナーを開拓しひとつでも多くの移 植手術を実現し白血病患者の命を救うことを使命としている。ドナー開拓のため にメディアを使った大型キャンペースを 毎年実施しているが、その反響をよって制標として問合せ数と登録数によ果で 指標として問合せ数と登録数によって削 に判断材料とすることによって、グを 効率良く得ることができる。

(出所) Drucker, 田中監訳 2000, pp.154-162

ワークシートでは、「われわれは成果をど のように定義しているのか?」という質問が 提示されている。そして、現在の成果の定義、 進捗と業績のモニタリング方法、そして成果 目標の達成状況という順番で細目質問が行わ れる。次に「われわれは成功しているのか?」 という質問に基づき、第一の顧客、支援して くれる顧客、組織自身の立場から成功の度合 いを三段階の評定によって判断させる。さら に「成果をどのように定義すべきか?」の質 問によって、将来の成果とそのモニタリング 方法を模索するように導く。最後に「われわ れは何を強化し、何を放棄すべきか?」とい う質問のもとで、成長性のあるプログラムと 放棄すべきプログラムを明らかにさせる。そ して、財務や人事などの見直しが必要となる 内部システムについても評価対象として列挙 させる。

設問5 「われわれの計画は何か?」

計画には、使命、ビジョン、ゴール、目標、行動計画、予算、査定から構成される一つのサイクルから成り立っている。図1の循環図では査定という言葉が用いられているが、これは計画と目的との整合性と成果達成度を評価することを意味しており、その結果は使命改変と新たな計画策定に反映される。その意味で、自己評価のプロセスは終わりを迎えることはなく、常にどこの段階に属しているの

か意識しながら継続されている必要がある。

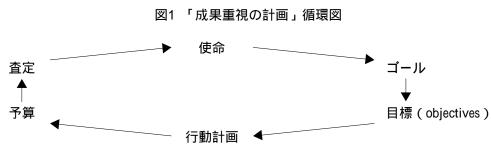
また、計画のポイントとして、放棄、集中、イノベーション、リスク、分析が挙げられている。特に、イノベーションは、変化する環境と課題の中に見出されるものであり、顧客の要望と価値観の中にヒントがあると述べている。そして、計画は技術ではなく責任であるという点は、民間非営利組織だけでなく組織一般に説得力がある指摘といえる。

< 設問5の事例 >

骨髄バンクはその使命に基づき10年間で10万人のドナー登録を達成することがゴールとしていた。そのために複数のログラムを作り行動計画に沿って活動した。その後10万人登録を達成した。その後10万人登録を達した。それに設定値を変更した。キー気軽に進むため、次の普及段階に進むためず、気軽にかっといる。さらにはでいる。さらにはできた。といるの登録に対応するためには従来の運営を新たに企画している。さらには従来の登録に対応するためには従来の運営体制では無理があるため、組織構成や人員配置も見直しを行った。

(出所) Drucker, 田中監訳 2000, pp.154-162

ワークシートでは、「使命を変えるべきか?」という質問が提示される。設問5に至るまでに顧客の変化や成果の達成状況を見て



(出所) Drucker, 田中監訳 2000, p.59

きたが、その結果を踏まえ、使命を変えるべきか再度問いかけている。そして必要であるならば、設問1で用意されたワークシートに戻って検討し、新たな使命とは何かを考えることを指示している。さらには「われわれのゴールは何か?」と問いかけるが、新たな使命に基づいて、ゴールを定義するよう示している。しかしながら、目標や行動計画に関する手引きが記されていないため、戦略的計画(Strategic Plan)などの技術を別途学ぶ必要があると思われる。

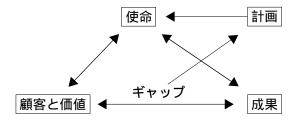
(2) 評価手法の構造

使命から使命への循環

自己評価手法は、使命、使命達成方法、業績について自己開示するためのプロセスを提示している。そして、これらのプロセスは常に組織外部の支援者やサービスを受ける人をの価値や満足、さらには、外部環境との照合を応じて行われる。さらには、外部環境に適適にはあために組織の使命と計画変更を意図的に組織の使命と計画で使命、顧客価値、成果のいずれをも照合し、どこかに対すれば新たな計画に反映している。使の2参照)

外部環境変化と組織変化 外部環境は、組織が直接サービスを提供す

図2 5つの設問:使命から使命の循環



(出所) Druckerの理論をもとに筆者が作成

る人々の環境、組織の活動を支援するタスク 環境など多様であり、直接あるいは間接的に 組織に影響を及ぼす組織と接する境界に存在 する。外部環境は組織の基本構造のみならず、 組織の構成員である個人の行動に影響や変化 を与えるが、それらを統制するために組織プ ロセスが機能することが肝要である。野中 (1974)は、市場多様性に対し、組織は「最 小有効多様性 を持って対応する必要があり、 それは組織がもつ諸制約のもとで組織構造を 変化させ、権力分布のパラメータを最適に調 整することによって実現されると述べてい る¹⁰。さらには「環境は組織の境界に存在し、 組織全体に直接的あるいは間接的に影響をお よぼす要素である(野中 1987, p.13)」として、 環境が組織全体に影響を及ぼす経路を「組織 現象の統合的コンティンジェンシー・モデ ル」"として示している。このモデルは、環 境、コンテクスト、組織構造、個人属性、組 織過程、組織有効性の諸要因12が如何に繋が り影響しているのかを示したものである。野 中が対象にしたのは、市場行動を行う営利企 業であるが、市場や他の外部環境は、組織の 戦略のみならず意思決定過程、組織風土や価 値観までにも影響を与え、それに応じて組織 が変化することを主張している。

ドラッカーが自己評価手法によって5つの 設問とワークシートの中で常に問いかけてい たのは、環境や顧客の変化を察知し課題と機 会を見出すことであった。そのために使命と 社会環境、過去に想定した顧客と現在の顧客、 成果と顧客の欲求の比較をして差異を見出す ように設問が導く構造となっている。この変 化と差異の発見こそがイノベーションの原点 であり、選択と集中をして組織を再編成する ことになって実現可能になる。これは野中が 説明する外部環境変化に対応する組織に通じ るものであると考えられる。そこで、野中の モデルを参考に*自己評価手法*における5つの 設問の図示を試みた(図3参照) 野中のモデ ルによれば、優れた組織は外部環境の変化に 諸要因が作用し組織が変化してゆく。ドラッ カーの自己評価は、このモデルに類似した経 路とその見直しの過程を意図的に作り、イノベーションの機会を意識的に掴むように導こうとしているのだと見るべきであろう。

4. 評価の視点からみたドラッカー の自己評価手法

ドラッカーの自己評価手法を解説した邦訳のタイトルには「成果重視のマネジメント」と銘打たれている。このタイトルに象徴されるように、自己評価手法は成果に基づくマネジメント判断のツールとして位置付けられている。この点、成果重視を特徴とした既存の評価手法として代表的なものにResult Based Management (RBM)が挙げられるが、そこでは成果の確認を客観的かつ適宜に行うことでマネジメント上の意思決定に関連させようとしている。しかし、同じ成果重視のマネジメントを目指すものであっても、自己評価手法とRBMは異なる評価手法であるといえる13。

そこで、RBMの沿革を説明するのに欠か せないPerformance Measurement (PM)を概 観し、RBMの評価枠組みを述べる。そのうえで、イノベーションとマネジメント改善のための手段として位置付けられている自己評価手法がRBMと比べ、どのような特徴をもっているかを比較分析しながら浮き彫りにしたい。

(1) 評価手法としてのPMとRBM

PMの意義と経緯

業績測定 (PM: Performance Measurement) とは、プログラムまたは社会サービス活動のアウトカム¹⁴や効率性¹⁶を定期的に測定する評価手法である。これまでの評価手法がサービスやプログラムのアウトプットの測定を中心としていたのに対し、顧客¹⁶の立場に立った成果主義のマネジメントに焦点を当て、特定のアウトカムの進歩状況を定期的に測定するところにこの手法の特徴がある。

PMの起源は1960年代に行われた米国防省の費用効果分析(Cost-Effective Analysis)にまでさかのぼる。当時は連邦も州もサービスの質やアウトカムに関する情報をほとんど持っていなかったため、顧客の限定された国防

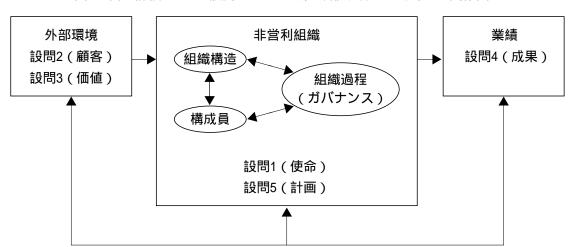


図3 自己評価「5つの設問」に基づく、外部環境と組織対応の関係図

(出所)田中2002, p.230

セクターの評価手法を多様な顧客層を持つセ クターにまで一般的に適用するには無理があ った。そのため、1990年代まではごく限られ た地域でのみPMが活用されてきたといえる。 しかしアウトプット指標による評価では、量 的に指標を把握しやすい半面、サービスの質 を指標として拾いきれないため、質の低下を 招きやすいという本質的問題が存在する。こ のような問題を克服し、より良いサービスを 顧客に提供するためには、現実の顧客の利害 と実際に提供しているサービスとの乖離を埋 めるための評価手法が必要であった。その後、 顧客に対するサービスの向上を目指し、定期 的なサービスの質と成果の測定が広く普及し ていったといえる17。このように、行政機関 を中心に導入されてきたPMではあるが、最 近では米国の民間非営利組織であるUnited Way (2003) も導入している。

PMの評価プロセス

PMの中心的機能は成果指標に基づいた明確なデータを定期的に与えてくれるところにある。このような一貫したPMの実施を行うためには、8名から12名の規模を目安とする作業部会によって、PMの作業工程を進めていく手法が一般的である¹⁸。

作業部会によって行われるPMの作業工程のうち、評価内容に大きな影響を与えるプロセスとして、() プログラムの使命(Mission)・目標(Objectives)・顧客(Clients)から、めざす成果を設定する、()成果と効率性を測る特定の指標を選定する、()がエグラム成果を比較するためのふさわしいベンチマークを確認する、()評価結果を業績情報の予算編成や事業改善などへの利用を確認するなどが挙げられる。これらの評価プロセスはPMの固有の機能を示す一方、PMの一般的な手法の枠組みだけでは有効な評価の射程範囲が広がらないことを意味する。すなわち、PMの本質的な限界がプロセスから導き出せる。

PMの限界

行政評価手法として発展してきたPMにとって、定期的な成果測定(regular tracking)は継続的なサービスの改善や事業の意義を確認するための不可欠な要素である。そのため、PMを行う負担が大きい小規模の組織においても、行いうる限りの範囲で定期的にサービスの質と成果を把握する必要がある。小規模の組織はコストのかからない方法を工夫する必要はあるものの、組織規模の大小はPMの限界を示すものではない¹⁹。

しかしながら、PMの本質的な問題は以下の3つの限界(Hatry 1999, p.256)から理解できる。このような構造上の限界は、実施者のみならず、PM情報の利用者にとっても、その用途を限定する上で重要な意義を持つ。

- a) 業績データ (Performance Data) だけで は成果との因果関係 (attribution) を説 明できない。
 - PMは成果を目に見えるデータ (descriptive data)として提示することができるが、どの程度プログラムが成果に 貢献したのかを明らかにすることはできない。この限界を補うため、業績の詳細を分析し、要因を説明する何らかの機会を組み入れる必要がある。また、この限界はアカウンタビリティの議論とも関係する。すなわち、プログラムの責任者はアからできない直接的な関連性を捉え、業績情報をプログラムの改善に役立てるべき責任を負うといえる。
- b) 直接的に測定できない成果もある。 例えば犯罪の防止など、暗数の把握が不 可能であるため予防成果を直接的に測定 するのが困難な分野がある。このような 場合は、関連現象を反映した代理指標で その成果を十分に捉えることが可能であ る²⁰。
- c) 意思決定に必要な情報の一部を提供する に過ぎない。

PMは過去のアウトカムの測定を通じて 課題や問題を提起することを目的とした 事後評価であり、将来に対する予測を示 唆するものではない。したがってPMそれ自体は、政策判断やリーダーシップ、 創造性に取って代わるものではなく、意 思決定の基準となり、将来予測の手がか りを与えてくれるものとして捉えるべき である。

これらの限界はPMの経緯や作業工程から必然的に導かれるものであり、評価手法としての役割を画する基準となっていると思われる。例えば、詳細なプログラム評価(indepth Program Evaluation)がPMを補完する評価手法として位置付けられるのは、PMのもつ成果分析の限界を補うからである(Hatry 1999, p.8, 259)。すなわち、インパクトの測定・確認などのプログラム評価を通じて成果への道筋を詳細に把握することは、特定の成果に対する実質的な因果関係を明らかにすることにつながる。そして、成果の要因分析を目的としたプログラム評価の場合は、その点でPMと役割分担しているといえる。

経緯・作業プロセス・限界を鑑みると、PMはアウトカムを測定することを最大の目的とした評価手法であると言ってよい。そのため、評価結果の利用に関しては十分な結論を有していないという側面がある。その意味において、PMにおける課題は他の評価手法との組み合わせにあるということができる。

RBMの意義

成果重視マネジメント(RBM: Result Based Management)とは、業績の改善によって行政機関の運用手法に重大な変化をもたらすことを目的とした広範なマネジメント戦略のことである(DAC 2001, p.3)。PMが業績情報を生み出すこと自体に関心があるのに対して、RBMはPMによって生み出された業績情報をもとに継続的な改善と成果を得るマネジメントに重点を置いている。

したがって、RBMの意義を探るうえでPMやEvaluationがどのように組み込まれているかが重要な意味を持つ。この点、DACはRBMの主な作業工程を以下のように整理し(DAC 2001, p.4)、RBMのプロセスのうち下

記(i)~(v)の実施段階がPMの手続に当たることになると説明している²¹。

- (i)ロジカル・フレームワークに裏打ちされた明確かつ測定可能な目標 (Objectives)と成果(Results)を確認する。
- (ii) 各目標に対する進行度合いを測定 するのための指標を選定する。
- (iii) 各指標に対して明確な目標値 (Targets)を設定する。
- (iv)実際の成果に基づいたデータを定期的に収集するための業績モニタ リング体制を構築する。
- (v)目標値と比較し実際の成果を検討、 分析、報告する。
- (vi)業績モニタリング体制による業績 情報を補完するためにEvaluationを 統合させる。
- (vii)内部マネジメントのアカウンタビリティ、学習や意思決定過程、利害関係者やパートナーに対する外部の業績報告のために業績情報を利用する。

RBMはもともと資源の効率的配分を目指す事業効率化の一環として注目されてきた経緯がある。そのため、適正なモニタリングによる事業評価を行い、得られた評価情報によってプロジェクト管理体制の強化に重点を置いている。

RBMにおけるEvaluationの役割

PMが導入された時期にはEvaluationは既に 定着しており、当初は両者の役割は区別され ていなかった。その後PMが評価手法として 洗練されていくにつれて、詳細でかつ時間の かかるEvaluationとの役割分担が強く意識さ れるようになる。

このようなPMとEvaluationの役割分担に応じて両者は使い分けられている。特にRBMの文脈においては、要因分析の必要に応じて補完的にEvaluationが用いられていることが

重要な点である。実際にRBMを導入した行政機関において、並存していたEvaluationはコスト上の問題から主にPMに代替され、Evaluationの実施数は大きく減少している(DAC 2001, pp.112-113)²。

(2) ドラッカーの*自己評価手法*とRBM

自己評価手法とRBMの比較

特に、両者の相違を明確にするものに、評価目的がある。ドラッカーの自己評価は、過去の教訓から改善を得るのみならず、現状を見据え、対応策を見出すことで、イノするとなる課題と機会を発見しており、形成評価としており、形成評価としており、形成評価としており、また、自己に対してある。また、評価の対象者に対し受益者を担いう言葉を用いており、選択権をもち満足させる対象という考え方を強調している点がRBMと大きく異なる。

さらに、自己評価手法は組織の使命とそれに基づく事業計画の見直しを目的とする中間評価を中心にしており、組織内の職員、理事や管理者、あるいはボランティア等の組織支援者の参加などフラットな組織内での協働による内部評価が基本となっている点が特徴的である。ただし、事業の受益者である「顧客」が評価に参加することについては言及してお

らず、評価主体にもRBMとの違いはない。

自己評価手法とRBMの相違点

評価設計の基本要素に従った自己評価手法とRBMの比較を行うことによって、成果を定量・定性データによって明らかにし、それを事業や組織のマネジメントに反映することに主眼が置かれている点で多くの共通点が見られた。ここでは、以下に示す項目に即し、表2をもとに両手法による相違点を検討していく。

「評価目的」

RBMはアカウンタビリティや業績改善など複数の目的をひとつの評価情報に求めようとしている。また、業績情報の利用を強化・促進するために様々な使途目的が列挙されている(DAC 2001, p.4-5)。その一方、自己評価手法の目的は改善とイノベーションに終られ、5つの設問はこの1点に集約するように設計されている。評価はあくまでも新たな機会計されている。評価はあくまでも新たな機会しつづける顧客や外部環境と現在の成果達成状況とのギャップを見出すことに注目する。

すなわち、RBMはPMを前提として、成果のチェック機能を果たすために開発された評価手法であると考えられ、自己評価手法はマネジメントを前提として革新し続けるために開発された評価手法であるといえる。

「顧客概念」

自己評価手法が開発された際の問題意識には、民間非営利組織が自らの善意と使命感に囚われ、真に重要な受益者の立場を見失いがちになるという点があった。このような問題意識は、対象者を顧客と捉えなおし、顧客を「第一の顧客」と「支援してくれる顧客」に分類する評価手法に反映されている。

これに対して、RBMは、顧客調査 (Customer Surveys)など顧客に対する配慮を 見せる一方、顧客の分類は曖昧で調査結果の 活用に重点が置かれているとは思われない。 そのため、顧客の変化を指標の変化で捉え、

表2 評価設計の基本要素に従った自己評価手法とRBMの比較

基本要素	RBM	自己評価手法
評価目的	形成評価、総括評価のいずれをもカバーしようとしている。組織内外のアカウンタビリティ、組織内での学習と知識の蓄積、事業改善、運営体制の見直し、意志決定支援などの複数目的をもつ。Pattonの区分によれば、アカウンタビリティと改善を目的とする。	形成評価。使命を終えたと判断し活動を終わらせるならば結果的に総括評価になるが、主眼は見直しである。Patton (1997)によれば評価目的は、アカウンタビリティの証明、改善、知識創造である。但し、イノベーションの契機となる課題と機会の発見が真の目的である。
評価対象	プロジェクト、プログラムが対象。開発援助の場合には、国別援助計画もプログラムに含まれる。	組織の使命、各プログラム、内部運営に至るまで組織全体をスコープに入れようとしている。顧客、成果に関する設問ではプロジェクト、プログラムが主たる評価対象となっているが、他の設問では外部環境や組織内部をも含めている。
成果	アウトプット、アウトカムが中心。インプットのみならずアウトカムレベルに注目することを奨励。顧客という表現を用いているが、成果対象に顧客の変化を含まないこともある。	成果という言葉で表現されているが、例示ではアウトカム、インパクトレベルの状態が示されている。また、成果は顧客の変化に在るという考え方をする。
タイミング	事業策定段階から指標に基づいたベースラインをとることが望ましく、その意味で事前評価、中間評価。事後のインパクトの評価はevaluationの領域であるとして、補足情報と位置付けている。	事業実施過程で実施するので中間評価。成果の達成状況の把握という意味で事後評価の要素も含まれる。設問5においては、レビューに基づいた新たな計画を策定するという意味で事前評価でもある。
評価主体	自己評価(内部評価)。組織内部の幹部、 職員が着手することが前提。	自己評価(内部評価)。組織内部の職員、 幹部が評価結果を自らの組織の使命や運営 に反映させることを目的としている。また、 同じ業務に従事するか、同じ身分の数人の グループでの作業が適当とする。
アプローチ	参加型評価が望ましい。内部評価ではあるが、事業の利害関係者が作業プロセスに参加することで、データ収集がより効率的で信頼できるものになる ²⁴ 。	組織内部において参加を促す一方で、作業の効率上、グループ区分をする。また、サービスの受益者は評価プロセスに入らず、評価のための調査の対象としている。
手法	インプットから目標に至るまでの道筋をロジカル・フレームワークで描く。そのログフレームをもとに、測定可能で現実的な指標選択と業績モニタリング体制を構築する。	成果については、指標をもとにした定量化が望ましいと記されており、その意味でPMの考え方と共通している。但し、使命を活動目的に咀嚼し、指標を選択する具体的方法が示されていない。
体制	組織のトップの理解と支援が必要。	同左。
Attribution	Attributionが主たる目的ではない。すなわち、プログラムと効果の因果関係を明らかにし、要因分析をすることはではなく、設定した目標に対する進捗監理が主たる目的である。	Attributionが主たる目的ではない。また、目標達成の証明も目的ではなく、むしろ変化しつづける顧客や外部環境と達成状況とのギャップを見出すことに注目する。

顧客の価値観に関するデータを新たに指標に組み込んだりする工夫がなければ、顧客概念が定着しているとは言い難いと考えられる。但し、顧客情報に対する関心自体は強く、顧客の視点から見て重視する成果を絞っている意識も見られる。また、本論でRBMの主要論文を提供しているOECD等国際機関の場合、事業の対象が国レベルで広範であるがゆえに、個々の住民の価値観まで注目することには物理的に限界があることも配慮すべきだろう²⁵。

「目標設定」

RBMは基本的に三段階のプロセスから構成される。まず、政策策定時あるいは事業形成時に設定した目標に基づいて、指標の目標値を設定する。次に、進捗状況を確認し変化を計測するために、モニタリングが行われる。最後に、目標値と進捗状況との比較分析を行うことで、当該事業の目標達成度を確認する。これは、事業を実施することによって対象が変化した結果、当初に設定した目標を達成することができるという仮説に基づいている。その意味で、RBMは過去に設定した「仮説検証型」をの評価を行っているといえる。

*自己評価手法*も同様の三段階のプロセスか ら構成される。しかし、外部環境と顧客の状 態は時事刻々と変化することを前提に、その 変化の状況を把握すること自体に価値を見出 す点において異なる。RBMは目標値と進捗 状況との比較分析の結果、両者に大きな乖離 がないことが望ましいとされた。しかし自己 評価手法の文脈では、組織が設定した使命や 目標に対し外部環境、成果や顧客変化が乖離 していること自体を重要な評価結果と見るの である。なぜならば、このギャップは組織の 課題であると同時に次の機会の可能性を示唆 しているからである。設定した目標の達成に こだわらず、むしろ環境や顧客の変化に応じ て新たな目標へと向かう組織のイノベーショ ンを促している。その意味において、*自己評 価手法*は「仮説刷新型」の評価といえる。

自己評価手法とRBMの相互補完性

RBM、*自己評価手法*は独自の思想に基づいて開発された手法であるが、いずれも長所、短所を抱えており、概して限界も多い。両者の相違点が明らかになったところで、相互にどう補完しうるのか考察したい。

「RBMが自己評価手法を補完しうる点」

自己評価手法では、改善とイノベーションを意識的に導入するツールとして評価のアウトラインが説明されている。しかし残念をがら、方法論としては具体性がない。使命を落った、目標というより具体的なレベルに落をし込んで表現することを求めているが、その方法は例示に留まっている。論理一貫性をには、PM やRBMで説明しているロジカルと思われる。

また、自己評価手法は、成果の測定方法として4つの指標をもとに定量調査を挙げているが、これも具体性に欠く。成果の定義のみによって、測定可能な指標を想定することは容易な作業ではなく、またデータ収集と分析にも評価技術が必要となる。この点においてもRBMは方法論を示していると考える。

自己評価手法は、5つの設問によって、実に平易で簡潔な表現で考え方を示した。しかし、指標の取り方、モニタリング方法、データ分析に関する技術や具体的な方法論は記されていない。RBMなどの技術や方法論を示したものとを併用することによって、自己評価手法はその有用性を高めることができるものと思われる。

「自己評価手法がRBMを補完しうる点」

RBMは多くの評価目的を有するがゆえに、評価結果をマネジメントにどう組み込んでゆくのかについて十分な解答を用意しきれていない。組織内の学習効果や意思決定支援をも視野に入れる一方、どのタイミングで誰がどのように評価情報を反映するのかを説明しきれていない。この点において、自己評価手法

では、イノベーションという目的のために何をどのように計画に組み込むのかを思考過程 に従って説明されている点が参考になる。

また、自己評価手法が目標達成に基づく成 果の確認というよりは、むしろ成果と環境変 化の間に起きたギャップとそこから得られる 機会に注目することを奨励している点が示唆 に富む。なぜなら、RBMが資源配分の効率 化を達成するために開発されたとはいえ、ド ラッカーの思想にも見られるように、今日の 組織における事業効率は有形資産の配分だけ では決定しないからである。つまり、情報の 蓄積や人材開発などの無形資産をモニタリン グの対象に入れることによって、はじめて事 業効率の改善を達成することができるはずだ からである。開発援助など公的機関において も、事業効果はこれらの無形資産に及ぼすプ ラスの影響や効果を含めて定義している。し かし、情報や人的資源は変化が顕著であり、 事業計画を策定した時点と評価を行った時点 では、その状態は異なる。評価を計画時に立 てた目的の達成度としてのみ捉えれば、たと え目的が達成されたとしても、現時点での事 業を取り巻く人々や社会状況にはそぐわない こともある。

したがって、RBMにも自己評価手法におけるイノベーションや自己学習という視点をより具体的に取り込む余地があると見るべきであろう。アカウンタビリティの確保とイノベーションの機会という2つの考え方は、成果と環境変化の間のギャップに対する捉え方の違いという二律背反する概念ではなく、ギャップの内容を機能的に分類したものとして理解するべきであろう。

5. 結語

ドラッカーの*自己評価手法*は、営利という 共通の業績判断基準を有しない民間非営利組 織のために開発された。業績判断基準に代え、 マネジメントを導入することによって問題解 決の方向性を示したのである。まず、民間非 営利組織が陥りがちな問題点として、自らの 善意と使命を信じるが故に対象者や支援者し、 言向をないがしろにするという点を指摘し、 二種類の顧客概念を取り入れた。そしてはているにするという指摘は、顧客の状態の改善にはないという指摘は、顧客の状態の改善にといる。民間非営利組織にとの原動けももは組織の存在意義かつ構成員の原動けももは組織の存在意義かつ構成員の原動けももいる。そのため、自己評価手法の問いかもまし、それは組織内で自己完結するのではなをし、それは組織内で自己完結するのではならが出る点が強調されるべきである。

科学的評価に始まり、様々な手法へと進化 を続けてきた評価であるが、これらの評価と ドラッカーの自己評価手法の間には大きく異 なる点があると思われる。それは、ドラッカ ーの*自己評価手法*は仮説とその検証にこだわ らず、むしろ過去に設定した仮説と時事刻々 と変化する対象との間のギャップに注目し、 使命と計画の見直しというかたちでイノベー ションを促している点である。RBMをはじ め従来型の評価の基本は、ロジカル・フレー ムワークやCausal Theoryに象徴されるよう に、事業実施によって対象が変化することを 想定し、事後にそれを確認してゆくことにあ った。すなわち、仮説の検証こそが評価の基 礎にあるとされてきたのである。しかし、 Simon (1976) が「人間は限定的にしか合理 的でない (Bounded Rationality)」と述べてい るように、情報を収集し諸条件を加味したと ころで、複雑な未来を予測することには常に 限界が付きまとう。評価が仮説検証型のロジ ックを基本にしている限り、この呪縛から逃 れることはできないと思われる。RBMのよ うに継続的にデータを収集することによって マネジメント改善を試みる評価手法において も、この点を克服できていない。現在、評価 者のみならず依頼主もが、評価結果が現行事 業や将来の事業形成に反映されにくいとの欲 求不満を抱く原因はこの点にあるのではない だろうか。ドラッカーの*自己評価手法*は、評 価の限界を示唆していると同時に、イノベー

ションという新たな領域の拡大可能性をも示しているかもしれない。

注釈

- 1 ここでの民間非営利組織とは、最大多数の利益を目的とし独立した意思決定機構を有する非政府(non-government)でかつ非営利(non-profit)の組織をいう。一般的に民間非営利組織はNGOやNPOという表現が使われることが多いが、非政府と非営利のその双方の特質を持つ組織として区別したい。
- 2 ドラッカーが提案した自己評価手法という表現は、1995年に邦語に翻訳書籍が紹介されてから、一般的に用いられるようになっている。本論では、様々な評価手法のうちドラッカーが提示した自己評価の手法を指す意味で、以下「自己評価手法」という表現を用いる。
- 3 静態的経済学から経済的発展への動態的 観察と、その中核としての技術革新を指 摘する経済学がこれに当たる。
- 4 米国におけるアウトソーシングの流行は 流動的知識ワーカーのメンタリティーを 表わしたものといえる。短期転職や起業 家などのような特定の組織に長期間帰属 しない「組織離れ現象」は、米国の大学 を卒業したばかりの若者たちの間でみら れることであり、欧州でも1980年頃から 同様の現象がみられるようになっている。
- 5 したがって、知識ワーカーは特定の組織に長期間帰属するよりは、短期に複数組織の間で転職を繰り返すか、独立して起業家となるというように、流動的な性質を有する。知識ワーカーは自分と共通の知識や関心をもつ人々とより密接な関係を保つのであって、組織との関係はそれほど重要としていない。
- 6 また、ドラッカーは『The New Realities』の中で民間非営利組織を「人間変革機関 (Hunan-Change Institute)」と呼び、その社会的役割の重要性を指摘している

- (Drucker 1989)。そこでは、学校が輩出するのはより知識を得た学生であり、病院が輩出するのは治癒した患者であるように、民間非営利組織は、その活動やサービスを通じて人々の資質や生活を向上させ人々を変革しているという。
- 7 「人間変革機関」については1989年、「市 民性創造機関」については1993年に発表 されているが、ドラッカーは民間非営利 組織の役割について継続的に思考を繰り 返しながら、その役割が、サービスを提 供することによって対象者の生活を改善 する側面と、民間非営利組織に寄付やボ ランティアとして参加する場を提供する 側面があることを徐々に気付いてゆく様 が伺える。
- 8 1990年、同氏は著書「Managing The Nonprofit Organization」を発表したが、研究者や実践者の間で、より明確なかたちでマネジメントの重要性が認識される契機になったといえる。経営学の権威として著明なドラッカーが、民間非営利組織のマネジメントの必要性を訴えたことは、その後の米国における民間非営利セクターの成長に多大な影響を与えたと言っても過言ではないだろう。
- 9 1990年には、民間非営利組織のための初の学会誌「NML: Nonprofit Management and Leadership」が米国、英国で刊行され、人材監理、資金開拓、マネジメントのみならず、ガバナンス、組織文化等、多岐にわたる分野での研究論文が発表されている。
- 10「最小有効多様性」とは、多様性をもたせることが過度でも過小であっても組織が崩壊する危険性があるため、最適解としての多様性をみきわめ調整する必要があるという意味である。組織過程とは、諸制約のもとで組織構造を変化させ権力分布を調整する過程のことを意味している。
- 11 組織現象の統合的コンティンジェンシー・モデル(図4)

- 12 環境(一般環境、タスク環境、組織間環境、創造環境)、コンテクスト(目標・戦略、規模、技術、資源)、組織構造(組織構造、風土)、個人属性(欲求、志気、価値、人格)、組織過程(リーダーシップ、意思決定、権力、紛争解消)、組織有効性(測定対象は、環境コントロール、目標達成、システム目標達成、システム間の団結)が挙げられている。
- 13 ドラッカーの評価は、マネジメント改善を通して組織および職員の能力向上を目指している。しかし、サービスの受益者はヒアリングの対象であり、参加は想定していない。その意味で、サービスの受益者を参加させエンパワーメントすることを目的とした参加型評価とは異なる点があるといえる。
- 14 PMにおけるアウトカムとは、プログラム の結果起こった出来事、状況、行動、態

- 度の変化のうち、顧客や市民にとって直接的に重要なものをいう。プログラム活動によって生み出された財やサービスを指すアウトプットとの区別はPMにおいて特に重要な意味を持ち、成果は意図されたものであるか、有益なものであるかは問わない。また、直接的な成果である中間成果(intermediate outcomes)と間接的な影響を受けた最終成果(end outcomes)も重要な区別である。(Hatry 1999, pp.13-15.256)
- 15 PMにおける効率性とは、インプットによって得られる生産物の総量に対するインプットの総量の割合と定義される。金銭的支出や労働時間などは典型的なインプットである。(Hatry 1999, pp.6,18-20)
- 16 PMにおける顧客とは、直接にサービスを 受ける市民またはその間接的な影響を受 ける市民および企業を指す。

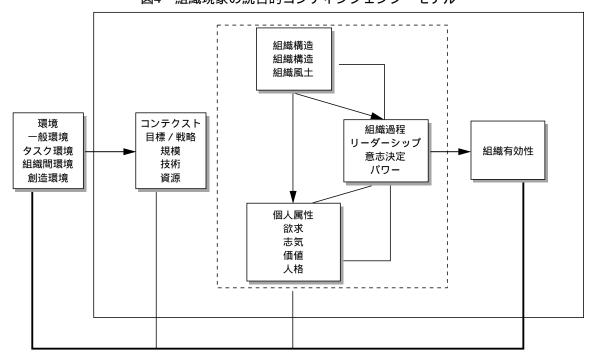


図4 組織現象の統合的コンティンジェンシーモデル

(出所)野中他1978, p.14

- 17 米国におけるPMの普及に大きな役割を担ったのが、すべての連邦行政機関や主要な事業にPMを義務付けた1993年政府業績成果法 (the Government Performance and Results Act of 1993)である。この時期を境に、イギリス、ニュージーランド、オーストラリア、カナダなど米国外でも、成果重視の業績測定に強い関心を寄せるようになってきている。
- 18 PMを成功させるためにはトップの主体的関与が不可欠である。十分な利用価値を持った評価が得られるには、プロセスの各段階における改善を含む評価工程に少なくとも3年程度の期間を要するからである。一貫したPMを行うための環境整備は実践的な評価の成否を左右することにもなりかねないほど、重要な課題であるといえる。
- 19 独立行政法人経済産業研究所が行った民間非営利組織法人アンケート調査結果報告(2002年7月10日)によると、全国の民間非営利組織の平均事業規模は年間1,860万円で4割強の民間非営利組織が500万円未満の規模であった。また、事務局スタッフ数は平均6.3人であったが、非常勤のスタッフが6割強を占めており、全体の4割が無給で働いている。このような状況を鑑みると、民間非営利組織において長期的に評価体制を行うには多くの課題を克服しなければならないと思われる。
- 20 犯罪の防止を評価するための代替的指標の例としては、再犯率や検挙率などが考えられる。これらの指標は統計学的に犯罪の防止と密接な相関関係にあることが証明されている。
- 21 ま た 、 DACに よ る 「 Results Based Management in the Development Cooperation Agencies: A Review of Experience」では、(i) ~ (iii) は戦略策定(Strategic Planning)の手法に該当するとしているが、PMの作業工程と戦略策定とは重複する部分こそあれ、内包する相関関係にないと考えられる。この点にお

- いてHatryも、将来に焦点を当てた戦略策定と違い、PMは判断技術を含まない手法として明確に区別されるべきであるとしている(Hatry 1999, pp.8-9)。
- 22 外部に対象を絞ることで簡潔化していた PMに対して、内部評価手法の欠如を補う フレームワークである点でBSC (Balanced Scorecard)の枠組みを評価する ことができる。組織の無形資産の成果を モニタリングすることは、他の業績測定 と並び、将来の成長に必要不可欠な視点 であるからである。そのため、戦略的フ ィードバックや仮説検証を通しての学習 改善と付加価値の発見を目指すという中 長期的な経営視点においてBSCの示唆が 伺える。しかし、小規模の組織へ一般適 用する場合には運用コストがかかる割に 効果が期待できない面もあり、また定量 的に測定しにくい事業活動の成果に対し て、どれだけインパクトを正確に把握す ることができるかにはまだ不確実な面が 多い。そのため、BSCはPMやRBMに代替 しうる評価手法とはいえず、むしろ必要 に応じてマネジメントを効果的に実践す るための経営管理ツールの特徴が色濃い。
- 23 ここでは、相違点を明示するために手法 に関する特段の事情を捨象している。特 にRBMは開発援助機関を自己評価手法 民間非営利組織を念頭に開発されてもた 経緯に鑑み、対象となるプログラムム ロジェクト、組織の規模、さらに比較 内の政治的事情から両者を単純に比較く。 また、民間非営利組織は自らの使命を また、民間非営利組織は自らの使命系 り、改変することができるが、政疫 関の場合には自らのみで使命を改変があ ことはできない点にも注意する必要があ る。
- 24 この点において、RBMは評価の準備段階から、組織トップの理解と支援策が必要であることを強調している(Hatry 1999, pp.27-28)。自己評価手法においてもこの点は共通する。

- 25 また、国レベル等、マクロレベルで成果 指標に注目する場合には、個々人の価値 観にまで及ぶことは現実的には困難が多 い。したがって、個々の顧客の変化をど こまで捉えるかの問題には、評価主体の 規模や立場が大きく影響するため、この 指標だけで顧客重視かどうかを判断する のは適当ではない。
- 26 事業形成当時に設定した目標にこだわらず、むしろ現状の環境変化にあわせて、目標も刷新することを意味する。事業を形成したときの目標は、事業というインプットの結果を受け達成するであろうという仮説に基づいている。現状変化に基づき目標を変えるということは、過去に設定した仮説に執着せず、むしろそれを刷新してゆくというイノベーションへの挑戦意識をも意味している

主要参考文献

- 上山信一(1998)『「行政評価」の時代 経営 と顧客の視点から 』 NTT出版
- 国際協力銀行(2003)『評価研修テキスト』
- 田中弥生(2002)「民間非営利セクターのインターメディアリ機能に関する研究」、博士学位論文(大阪大学)
- 田中弥生(1999)『NPO 幻想と現実 ~ そ れは本当に人々を幸福にしているのか~』 同友館
- 田中弥生 (1998)「NPOを成功させるドラッカーの五つの評価法」、『論座』、1998年7月号: 202-209、朝日新聞社
- 野中郁次郎・加護野忠男・小松陽一・奥村昭 博・坂下昭宣 (1978)『組織現象の理論と 測定』 千倉書房
- 野中郁次郎(1974)『組織と市場』、千倉書房 龍慶昭・佐々木亮(2000)『「政策評価」の理 論と手法』 多賀出版
- Center for Accountability and Performance, American Society for Public Administration、 谷口敏彦 訳、村岡政明 編、上山信一 監訳 (2001)『行政評価の世界標準モデル 戦略

- 計画と業績測定 』 東京法令出版
- Chen, H-T. (1990). *Theory Driven Evaluation*. Sage Publications.
- Drucker, P.F. 著、上田惇生 編訳 (2001) 『マネジメント [エッセンシャル版]』、ダイヤモンド社
- Drucker, P.F.・Stern, G.J. 編著、田中弥生 監訳 (2000)『非営利組織の成果重視マネジメント』、ダイヤモンド社Drucker, P.F.、林雄二郎 訳 (1969)『断絶の時代 ~来るべき知識社会の構想~』、ダイヤモンド社
- Drucker, P.F., & Stern, G.J. (1999). *The Drucker Foundation Self-Assessment Tool*. Jossey-Bass Publisher.
- Drucker, P.F. 編著、田中弥生訳 (1995) 『非 営利組織の*自己評価手法*』、ダイヤモンド 社
- Drucker, P.F.、上田惇生・田代正美・佐々木 実智男 訳 (1995)『未来への決断』、ダイ ヤモンド社
- Drucker, P.F. (1995). *Managing in the Great Change*. Truman Talley Books.
- Drucker, P.F. (1993a). *Post Capitalist Society*. Harper Business.
- Drucker, P.F. (1993b). Self Assessment Tool for Nonprofit Organization. Jossey-Bass Publisher.
- Drucker, P.F.、上田惇生・田代正美・佐々木 実智男 訳 (1993)『ポスト資本主義社会』、 ダイヤモンド社
- Drucker, P.F. (1992). *Managing the Nonprofit Organization*. Harper Business.
- Drucker, P.F.上田惇生・田代正美・佐々木実 智男 訳 (1992)『非営利組織の経営』、ダ イヤモンド社
- Drucker, P.F. (1989). *The New Realities*. Haper & Row Publishers.
- Drucker, P.F.、上田惇生・佐々木実智男 訳 (1989)『新しい現実』、ダイヤモンド社
- Harvard Business Review 編、DIAMONDハーバード・ビジネス・レビュー編集部 訳 (2001)『業績評価マネジメント』、ダイヤモンド社

- Hatry, H.P. (1999). *Performance Measurement: Getting Results*. The Urban Institute Press.
- IDA "Results Measurement System: Program and Proposal".
- Kaplan, R.S., and Norton, D.P. (1992). The Balanced Scorecard-Measures that Drive Performance, *Harvard Business Review*, January-February 1992, 70-79.
- Kotler, F. 著、井関利明 監訳 (1991)『非営利 組織のマーケティング戦略』、ダイヤモン ド社
- OECD, DAC Working Group on Aid Evaluation (2002). Managing for Development Results and Effectiveness.
- OECD, DAC Working Group on Aid Evaluation (2001). Result Based Management in the

- Development Co-Operation Agencies: A Review of Experience.
- Patton, M.Q. (1997). *Utilization-Focused Evaluation, The New Century Text, Edition 3*. Sage Publication, Inc.
- Rossi P.H., Freeman, H.E., and Lipsey, M.W. (1999). *Evaluation: A Systematic* Approach *Six edition*. Sage Publication.
- Simon, A. H. (1976). *Administrative Behavior*, 3rd Edition. The Free Press.
- United Way, (2003.July) http://national. unitedway.org/outcomes/initiatives/
- U.S. General Accounting Office (Apr. 1998).
 Performance Measurement and Evaluation Definitions and Relationships. GAO/GGD-9826.

(2003.7.17受理)

Drucker's Self-Assessment Tool for Nonprofit Organization -- How It Is Different from Evaluation, and How It Contributes to Evaluation --

Yayoi Tanaka

Daisuke Korenaga

The University of Tokyo yayoi-tanaka@ohriki.t.u-tokyo.ac.jp

Graduate School of Law, Keio University korenaga@1999.jukuin.keio.ac.jp

Abstract

Peter F. Drucker is one of the world famous authorities of Management. Although he was considered as a management expert of the for profit business sector, he has shown his interest in the private non-profit sector from more than 50 years ago. He predicted the growth of the nonprofit sector at the global level since 1980s. He also pointed out the fundamental problem of the nonprofit organization. It is the lack of the mechanism and consensus with respect to a bottom line of performance.

"Self-Assessment Tool for the Nonprofit Organization" was developed by Drucker Foundation as a tool of management improvement, in which Drucker guides to find out their problems and to think about innovating their business with simple five questions. This paper explains the outline of the self-assessment tool with Drucker's theory of nonprofit organization. And, since evaluation experts show their interest in how the result of performance measurement is integrated in management, we will try to identify the difference between the result based management and self-assessment tool simply. Drucker's assessment, which focuses on the innovation, always looks not at verifying attainment of the objectives but at comprehending the gap between the surrounding environment's changes and the organization's current achievement. This is the major difference of approaches between the RBM and self-assessment tool, and may tell us the limitations and possibilities of such evaluations.

Keywords

Drucker's self-assessment tool, result-based, management, nonprofit organization, innovation

【研究論文:依頼原稿】

戦略計画と一体となった評価 ~戦略計画を政策評価にどう活かすか~

龍 慶昭

佐々木 亮

城西大学 Bigliuryu@aol.com (財)国際開発センター sasa01@giganet.net

要約

日本における政策評価の導入と定着の過程の中で、アメリカでは実績評価(業績測定)と一対のものとして用いられることが多い「戦略計画」(Strategic Planning)の導入はあまり省みられて来なかった。そのことが逆に、日本において政策評価がスムーズに導入され定着することにつながったと言える一方で、一連の行政改革が目指した理念や目的を達成することに大きな貢献をして来なかったことにつながったと言えるだろう。今後行われる政策評価法の見直し作業においては、事業評価、実績評価、総合評価の3つの手法の横並びを見直し、戦略計画の策定と実績評価のセットでの実施を全省庁で全省的に導入することを求める一方、戦略計画の策定を通じて特定された個別施策の事前の検討と事後の検証のために、事業評価と総合評価が道具として補足的に用いられるという仕組みに変更することが検討されるべきであろう。

キーワード

戦略計画、経営戦略、業績測定、実績評価、パフォーマンス・メジャーメント

1. 現在の日本の政策評価の問題点 ~見落とされてきた戦略計画の 位置づけと視点~

2002年1月から「行政機関における政策の評価に関する法律」(以下、政策評価法)が施行された。この法律は、日本の行政機関へ政策評価の実施を義務づけることを定めている。この法律の制定による政策評価の導入は、1998年の中央省庁等改革基本法と軌を一にした動きであった。その法律のなかで、中央省庁等の改革の基本方針の一つとして政策評価

機能の強化が盛り込まれたのである(同法第四条六)。同法制定の契機となった「行政改革会議最終報告」における「行政改革に取り理念と目標~なぜ今われわれは行政改革に取り組まねばならないか~」では、21世紀型行政以下の仕組みや特性を上げている(行政改革会議 1997)。その中で見上げられたのが、総合性の確保と、が時と明である。「内外環境がはして、時と課題に即応し、国政全体と国際社会を見渡して、時と課題に応じていかなる価値を優先

するかを総合的、戦略的に判断し、大胆な価値判断と政策立案を行なうことが何より必要である。」と述べている。そして、そうした日本の行政改革に関する提言は、世界の行政改革の流れを的確に踏まえたものであったと言える。

中央省庁のスリム化・簡素化は、1998年に 中央省庁等改革基本法でまかりなりにも実行 に移された。「政策評価」という言葉は、上 述のとおり同法第四条で、法律の文言上初め て使われたが、その後、政策評価の手法や仕 組みについて様々な議論が行われ、まず2001 年1月に、「政策評価に関する標準的ガイドラ イン」(以下、政策評価ガイドライン)が公 表された(総務省2001)。同政策評価ガイド ラインでは、3つの基本的な評価の手法が整 理された。そして、2002年4月施行の政策評 価法で、全省庁に政策評価の実施が義務づけ られることになった。同法の施行に関しては、 政策評価ガイドラインで示された3つのうち のどれかの方法を用いれば、政策評価を実施 したことになると法案を起案した担当省庁は 説明したようだし、各省庁の政策評価担当部 署でもそう認識したようである。こうして、 省庁再編とともに、各省庁で政策評価法に基 づき、かつ政策評価ガイドラインに即した形 で政策評価が始まった。

政策評価ガイドラインで示された3つの方 式とは、ガイドライン自身の言葉を借りれば、 事業評価、実績評価、総合評価である(表1 を参照)。これらの用語は、評価研究におけ る一般的な用語の使い方とは違うが、複数の 研究者によって一般的に使われる用語との対 比が整理されている。その対比についてはす でに優れた解説があるので、ここで繰り返す ことはしない(古川・北大路 2001、古川 2002、山谷 2002)。ここで指摘されるべきは、 政策評価ガイドラインを一瞥して分かること が、政策評価、あるいは評価一般を非常に広 義に捉えているいろな手法を取り込んだとい うことである。評価研究(Evaluation Study) の対象とされてきた伝統的な評価とは極めて 厳密にはプログラム評価(Program Evaluation)

だけを指し、業績測定(Performance Measurement) は含まないという見解が有力 であり、しばしばプログラム評価と業績測定 を対比させたりその補完関係を検討する議論 がなされている (Hatry 1997, Wholey 2003)。 そもそも業績測定(Performance Measurement) という用語は、組織管理を主な対象とするマ ネジメント科学の分野から派生した用語であ り、評価研究の分野では、通常アウトカム・ モニタリング (Outcome Monitoring) と呼称 してきた (Hatry et al 1987)。また事業評価の 主な手法として取り上げられ、主に事前段階 で適用される費用 効率性分析は、経済学 の強い影響を受けつつ発展してきた政策分析 (Policy Analysis) で用いられてきた手法であ り、評価に含むことに抵抗を示す研究者もい るようである²。しかし一方では、クリティ カル・レビュー (Critical review) やポリシ ー・レビュー (Policy review) などと言われ る批評の類を含めなかったわけであり、ここ との間で一線が引かれたと見ることもでき る。したがって、ある程度手法が確立し、ア メリカの代表的な評価のテキストなどでも広 義に評価を捉えた場合に含まれる手法の範囲 をほぼ踏襲したと言えるとの見解がある (Wholey et al 1994、古川 2002)。

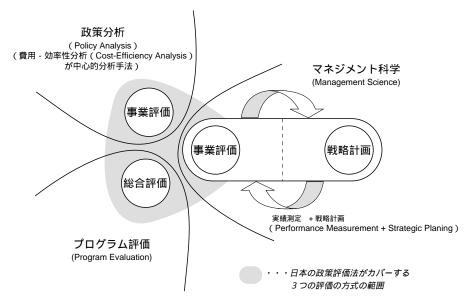
このうち本論文では、特に「実績評価」に 関連した諸事項について解説を試みる。実績 評価は、評価研究で言うところの「業績測定」 あるいば パフォーマンス・メジャーメント」 (Performance Measurement) とほぼ同一の方 法であるというのが多くの研究者の一致した 理解である。日本における法律制定に到る議 論の過程は必ずしも明らかではないが、少な くとも政策評価法の制定の動き自体が、アメ リカのGPRA(政府業績結果法、Government Performance and Results Act of 1993) に触発 され、それをある程度下敷きにしたものと理 解される。GPRAについては本論文の2.でさ らに詳しく解説するが、日本で政策評価の手 法の一つとして「実績評価」という名で取り 込む際に、GPRAで規定された業績測定 (Performance Measurement) の部分を主に参 考にしたと見てとれる。

ところで、GPRAの条文や解説(OMB 1993) によると、GPRAは、戦略計画 (Strategic Planning)と業績測定(Performance Measurement)の二つを柱としている、あるい は対として用いることを想定していると言え るのだが、日本の政策評価法では、このうち 業績測定のみが取り入れられたと解釈せざる を得ない。正確には、政策評価ガイドライン には、実績評価の前提として政策や施策を階 層化して樹型図上の体系図で表わすべきとい う記述があるが、これは戦略計画と同じもの を指しているのであろうか。GPRAが定めた 戦略計画に関する発想や仕組みについては次 節で若干の解説を試みるが、政策評価ガイド ラインが例示した体系図は戦略計画とは根本 的に違うものなのである。ちなみに、政策評 価法に規定された3つの手法と、政策分析、 プログラム評価、マネジメント科学の関係を 整理すると図1のように整理することができ るだろう。

日本の政策評価法が、GPRAで示された二 つの柱のうち、業績測定のみを取り入れたこ とに関しては、プラスに作用した面と、マイ ナスに作用した面があったと見ることができ る。プラスに作用した面とは、日本の行政組 織の仕組や運営の論理を大きく変えることな く導入できたということである。戦略計画と 業績測定は、コインと裏と表であるとも言え るが、そのうち、業績測定の部分のみ導入す ることによって、日本の行政組織の仕組みや 運営の論理を大きく変えることなく、つまり、 既存の組織構造と組織運営に追加的に導入す ることが可能だったわけである。したがって、 政策評価法の成立後、実績評価は、政策評価 の手法の一つとして中央省庁に幅広く普及す ることになった。

一方、マイナス面とは、上記のプラス面の全くの裏返しであるが、既存の行政組織の仕組みや運営の論理の抜本的な変革には結びつかなかったということである。GPRAを振り返れば、組織のミッション(使命)やビジョ

図 1 政策評価法がカバーする 3 つの評価の方式と各学問領域の関係



(出所)筆者作成

ン(将来像)をまず最初に明確に定めることや、策定された基本戦略にフィットするように組織構造を大幅に変更すること、そして成果に基づいた人事や予算管理を実行するということ、そして納税者や受益者の満足を最大の判断基準とするということ等が、その導入によって期待されたわけであるが、日本ではそこに示された抜本的な改革には結びつかなかったと言わざるを得ない。

その結果起きたことは何であったか。それ は、「実績評価」の名で導入された業績測定 のルーティン化、つまり日常業務化である。 そして、それは詰まるところ、政策評価法が 求めた政策評価全体のルーティン化である。 業績測定は導入されたが、それは年度末に指 標値を記入する毎年恒例の業務となり、指標 値の変化という情報に基づいて、組織構造の 変更、つまりある部を新しく作りだしたり、 ある部を廃止したりするということには結び つかなかった。また、人事や予算にも結びつ くことは少なく、また納税者や受益者の満足 度が高まったのかどうかも不明なままであ る。したがって、政策評価は、行政活動の一 部分に堕してしまい、省庁改革等基本等や政 策評価法の基礎となった行政改革会議最終提 言が掲げた理念である「戦略性の確保」につ ながらないばかりか、世界の行政改革の本質 的な動きにもついていっていない状況となっ ていると言わざるを得ない。

2. 戦略計画の理論的なバックグラウンド

(1) なぜ行政に戦略計画が必要とされるようになったのか

そもそもなぜアメリカを始めとして諸外国で、行政に戦略計画を導入するようになったのかについて以下、若干論じることとしたい。

「戦略」(Strategy)はもともと軍事用語であると言われるが、アメリカのパブリック・セクターを概観すると、自治体でまず戦略計

画の導入が進んだとされており、フロリダ州やテキサス州が先進事例としてよく紹介されている(Hatry et al 2001)。1980年代を通じて、アメリカの自治体で先行した動きはやがて連邦政府レベルへも及ぶこととなった。

戦略計画と業績測定の一体的運用が規定さ れているGPRAが制定される背景となった世 界的なパブリック・セクターの不信につい て、アメリカの研究者でGPRAの制定やその 導入にも関与したとされるホーリイは以下の ように解説している(Wholey 1997)。1980年 国内における規 中頃以降、アメリカでは、 制緩和と競争促進、 公共部門の規模縮小と 民間の参入拡大、 まったく新しい需要にも とづく新規産業の創出、 世界的な市場経済 化にともなう海外市場の拡大、等によって経 済成長が持続した結果、市場機能に関する信 頼がかつてなく高まっている。このことと表 裏一体であるが、現在、政府の存在理由と価 値に関する理解が浅くなっており、それに呼 応して政府に対する信頼も極めて低くなって いる。公共部門は、自らが提供するサービス の効果と効率性を改善し、納税者たる国民に 政府の存在理由を明確に説明する必要に迫ら れている。

こうしたパブリック・セクターを取り巻く 状況を改善するための処方せんは、その分野 の研究者から提案されることが期待されたは ずだし、また実際に適用されていくべきだっ たのだろう。しかし、実際には大きな改革の 提案は行政学の世界からはなかなか出て来な かった。こうした状況に対して、実際に提出 され、その後大きな影響を与えたのは、1992 に出版されたオズボーンとゲブラーの「政府 の改革:起業化精神を公共セクターに導入す るにはどうしたらいいか」(Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector) であろう。こ の本では行政組織が適用すべき10個の革新的 な原則を述べるとともに、具体的な様々な提 言がなされたが、極めて簡単に言うと、民間 の企業経営と政府の経営は本質的には何ら変 わるところはなく、民間企業経営で成功した

表1 政策評価の標準的ガイドラインで解説された3つの評価手法

	事業評価	実績評価	総合評価
基本的性格	事前の時点で評価を行い、 途中や事後の時点での検証 を行うことにより、行政活動の採否、選択等に資する 情報を提供することを主眼。	行政の幅広い分野において、 あらかじめ達成すべき目標 を設定し、それに対する実 績を測定しその達成度を評 価することにより、政策の 達成度合いについての情報 を提供することを主眼。	特定のテーマを設定し、様々な角度から掘り下で総合的に評価を行い、政との効果を明らかにするともに、問題点の解決に資する多様な情報を提供することを主眼。
対 象	事務事業が中心。必要に応じ、おおむね施策として捉えられる行政活動のまとまりについても対象(以下「事業等」。)	共通の目的を有する行政活動の一定のまとまり(おおむね施策程度のまとまりに相当。以下「施策等」。)を対象。各府省の主要な施策等に関し幅広く対象。	特定の行政課題に関連する行政活動のまとまり(おおむね政策(狭義)や施策ととらえられる行政活動のまとまりに相当。以下「政策・施策」。)を対象
評価の 時 点	事前の時点で評価し、途中 や事後の時点で検証。	あらかじめ目標を設定し、 定期的・継続的にその実績 を測定。 目標期間終了時に当該期間 全体における達成度を評価。	政策・施策の導入から一定 期間を経過した時点を中心。
評価の容	事めら析・ おいちが 予な推 の得的な 事内 び価充施。しいる かいらが 予な推 の得的な 事内 び価充施。しいれ分 いらが 予な推 の得的な 事内 び価充施。しいれ分 いらが 予な推 の得的な 事内 び価充施。しい。	主要なアース (政状体策因政を析政を時評緊次し社りさ国ニリ要従し図策況的・果策把。策検々価急の、会改れ民ー上請来てろのなか直つにそのに制み角に接いる原的が、策等では、策をに施関・握いるのは、一、では、大きに、大きに、大きなのでは、大きなのでは、大きなのでは、大きなが、大きなのがでは、大きなが、大きなが、大きなが、大きなが、大きなが、大きなが、大きなが、大きなが

(出所)総務省(2001)「政策評価に関する標準的ガイドライン(概要)」

仕組みは、政府経営に大胆に導入可能だとい う主張がなされた。そして著者たちは、クリ ントン政権のもとゴア副大統領が代表を務め たNational Performance Review(NPR)のアドバ イザーとして迎えられ、アメリカ政府全体に 彼らの主張を反映させることに成功した。 NPRは、Government Accounting Office (GAO) と協力して、1993年に成立したGPRAの導入 と普及を強力に推進した。したがって、オズ ボーンとゲブラーのReinventing Government という著書と、GPRAは直接の関係がないに も関わらず、目指した理想は極めて近似して いたため、1994年以降、同じ文脈で語られ、 理念としてReinventing Governmentという合 言葉が掲げられる一方、それを実現するため の具体的なツールとしてGPRAが紹介される ことになった。なお、同書に寄せてクリント ン大統領(当時)は、「アメリカの全ての被 選公務員 (every elected official) によってこ の本は読まれるべきだ。この本は、我々に (政府改革の)ブループリントを与えてくれ る」と述べている(Osborne et al 1993)。

上記で言及したGPRAについて簡単に解説 しておきたい。連邦政府レベルにおける戦略 計画と業績測定を一対の仕組みとして導入す ることを規定したGPRAは、1990年に起案さ れ、1992年に議会を通過し、1993年に発効し た。GPRAは、アメリカの全ての省庁に対し て戦略計画 (Strategic Plan) の策定を義務付 け、その上で明快な目標(つまり数値目標) を設定させ、その目標と実際に達成された毎 年の成果(当然これも数値的結果)を提示さ せるための仕組みを規定した法律である。こ れを通じて、プログラムの効果や効率性を改 善していくとともに、政府の存在理由と価値 を国民に対して示していくことが目的となっ ている。そして導入から今年(2003年)で10 年が経ち、その効果についても研究報告書が 出されるまでになっている(Hatry et al 2003)。

最初に軍事分野から始まり、民間企業の生き残りのために応用されて定着し、その後、公共組織の生き残りのために導入され徐々に定着している戦略計画の考え方は、アメリカ

の国内省庁だけではなく、GPRAの成立以降、他の先進国、国際機関、そして非営利組織などでも幅広く導入されるようになっていった(SIDA and UNDP 1997)。そして最近は学会組織やコミュニティ組織などでも採用されるようになってきている。

(2) 戦略計画の基本的な発想と構成要素の検討

本論文ではここまで、「戦略計画」という 言葉を特に定義しないで使用してきたが、そ もそも何を指して戦略計画と言うのであろう か。

すでに述べたとおり「戦略」(Strategy)はもともと軍事用語であるが、それを経営学において最初に明確に取り入れたのは、チャンドラー(Chandler)だと言われている。1962年に彼は「経営戦略と組織」(Business Strategy and Structure)を著した。そのなかで、チャンドラーは、戦略を「組織の基本的な長期目標や目的を決定し、これらの諸目標を遂行するために必要な行動を採択し、諸資源を割り当てること」と定義しており、これが戦略計画に関する最も初期の定義であるう。

この節では、上記の初期的な定義とその後 の普及の経緯も踏まえ、戦略計画の策定の手 順を解説した4種類の文書を事例として分析 することを通じて、戦略計画の定義と基本的 な構成要素を特定することを試みる。4種類 の文書とは、 中央政府レベルの戦略計画を 規定したGPRAの法律文、 自治体レベルの 戦略計画を解説したアメリカの国際シティ・ カウンティ経営協会のレポート、 特定セク ター(農業分野)の戦略計画の作成ガイドラ イン、そして 民間企業における戦略計画の 標準的な策定過程を解説したテキストであ る。

中央政府レベルの戦略計画を規定した GPRAの法律文(図2参照)

GPRAでは戦略計画を何と定義しているであるうか。GPRAの法律文(OMB 1993)に

は、特に戦略計画を定義した文言はないが、 戦略計画が含むべき項目が解説されており、 それらをつなぎあわせると次のようになると 筆者は理解している。「戦略計画は、(1)使命(ミッション)(2)ゴールと目標(Goals and objectives)(3)人材、資本、情報、その他の資源を使用しつつ、どのような過程を 通じてゴールと目標を達成するのかの説明、(4)(年間の)業績目標がどのようにゴール と目標の達成に貢献するのかの説明、(5)組織が管理し得ない主要な外部要因、他を含ま ねばならない。」

GPRAでは、次の三つの作業の段階が規定されている。(1)戦略計画の策定、(2)毎会計年度の当初に実施される年次実行計画の策定、(3)毎会計年度終了直後に実施される年次業績報告書の作成である。したがって業績測定は、この図中で矢印で示されたとおり、最後の(3)の段階の一部分の作業として実施されることになる。

自治体レベルの戦略計画を解説した国際シティ・カウンティ経営協会(ICMA)のレポート(図3参照)

アメリカのワシントンD.Cに所在する国際 シティ・カウンティ経営協会(ICMA)は、 1914年に設立されて以来、世界中の市町村の 首長・行政官・その支援スタッフに対して専 門的な研修や出版活動や支援サービスを実施 してきた非営利組織である。同経営協会が 発行した公共組織における戦略策定を解説し た文献の一つでは、戦略計画を以下のように 定義している (Gordon 1993)。「『戦略計画』 は、組織が将来起こる出来事にただ従うので はなく、組織が自らの運命をコントロールし ようという試みである。それは、将来の機会 (チャンス)について、自らが持つ現在と将 来の強み及び弱みを適用することによって、 成功を実現する一方、特定された問題を避け ていくということである。」

戦略計画の策定の流れは図3のように、(1)

図2 アメリカGPRAで規定された戦略計画と業績測定の仕組み

1 . 戦略計画の策定 (Strategic Plan)

組織として何を目指すのか、それを実現する長期目標等の以下の項目について特定して戦略計画とする。

- (1)組織のそもそもの社会的使命の特定
- (2)社会的使命を実現するための長期目標の決定
- (3)長期目標を達成するための方法・手段の特定
- (4)長期目標と毎年の目標の関係づけ
- (5)長期目標の達成過程に重大な影響を及ぼしかねない外部要因の特定

2 . 年次業績計画の策定 (Annual Performance Plan)

次に以下の項目で構成される各年の実行計画を立てる。

- (1)年間の活動目標及び数値目標の決定
- (2)目標数値の達成のために年間に投入されるべき各種資源と各種活動の決定
- (3)成果指標 (Performance Indicators)の選択
- (4)成果指標に表れた数値と、年度初めの目標数値を比較する仕組みの決定

3 . 年次業績報告書の作成 (Annual Performance Report)

年度が終了した時点で以下の項目で構成される報告書を作成する。

- (1)年間の活動結果を示す成果指標値の収集と一覧化
- (2)数値目標との比較と達成度合の判定
- (3)結論と提言(長期目標達成の有無、改善提言等)

(出所) Office of Management and Budget, The Executive Office of the President; Governmental Performance Results Act of 1993 (http://www.whitehouse.gov/omb/mgmt-gpra/gp;aw2m/html)

使命(ミッション)の確認、(2)環境分析 (外部環境分析、内部環境分析)(3)戦略計 画の策定(ゴール設定、目標設定、実行計画 策定、資源配分)の3段階から構成されると 解説されている。さらにその後、(4)戦略計 画の実行があり、その一部分である「進行中 の業績の測定と検討」(Monitor and assess ongoing performance)という名称の作業で、業 績測定が解説されている。前図と同様に、図 中に矢印で示した。

特定セクターの戦略計画の作成ガイドライン(タンザニアの農業分野)(図4参照)タンザニアはアフリカでも有数の援助受け入れ国である。同国の農業分野において、イギリス、デンマーク、アイルランド、ヨーロッパ連合(EU)世界銀行、FAO、日本など

が参加して、援助政策や事業の協調化が進められている。これらの援助国は、タンザニアの農業分野に対する援助の実施の事果レでも表して、中央政府だけではなく、にはないる。そして、農業セクターの戦略計画のガイドラインを対象とした戦略計画のガイドライを対象とした戦略計画のガイドライを対象とした戦略計画のが、北・西欧諸国や国際機関の協されて承認されたが、図4に示が、日国や国ので解説されたが、図4に示が、同域を表してもないである。なはである。なはである。なはではないではなく、保健セクターだけではなく、保健セクターだけではなく、保健セクターだけではなく、保健セクターだけではなく、保健セクターだけではなく、保健セクターだけではなく、保健セクターだけではなく、保健セクターだけではなく、保健セクターだけではなりした戦略計算とに作成している。

戦略計画に関する定義は、特に明示されていないが、そこに示された策定過程は、(1)

図3 アメリカの国際シティ・カウンティ経営協会 発行の文献で示された自治体向けの戦略計画と業績測定の仕組み

	7513 -	24100, 413. E 111 E E E E E E E E E E E E E E E E
1.	準備:使命の確認((Identify the mission)

- (1)法律、憲章、その他の情報源からの使命を特定する。
- (2)使命が存在しない場合には新たに策定する。

2. 環境分析 (Environmental Scan)

- (1)将来の特定 将来の「機会」と「課題」の特定
- (2)組織の診断 組織の「強み」と「弱み」の特定
- (3) SWOT マトリックスの作成と代替シナリオの検討

3. 戦略計画の策定(ゴール設定、目標設定、実行計画策定、資源配分)

- (1)組織が目指すべき複数のゴールについて合意する。
- (2)ゴールそれぞれに関する目標を立てて合意する。
- (3)目標を達成するための実行計画(Implementation Plan)を立てる。
- (4)人的資源、物理的資源、予算の配分の決定
- (5)管理のための仕組みやインセンティブの設定

4. 戦略計画の実行 (Implement the plan)

- (1)戦略の実行
- (2)管理の実施
- (3)業績の測定と検討 (Monitor and assess onging performance)
- (4)フィードバックと、必要に応じた実行計画の修正

(5. 時期計画の準備(Prepare for next planning cycle))

(出所) Gordon (1993) pp.3-6, pp81-83 を筆者が一部変更して掲載。

状況分析(外部分析、内部分析)(2)ビジョンの設定、(3)実行手段の特定、(4)村落レベルの計画策定、(5)(総合的な県レベルの)実行計画の策定、となっており、これが戦略計画を構成する要素であると言える。さらにその後、アウトプット・モニタリング・レポートの作成なる作業をすることになっており、この作成の際に、業績測定を実施することが想定されていると言える。前図と同様に、業績測定にあたる部分を、図4に矢印で示した。

民間企業における戦略計画の標準的な策定 過程

民間企業の戦略計画の策定と運用を解説したアメリカのスタンダードなテキストの一つ(Hunger and Wheelen 1993)では、以下のように戦略を定義している。「『戦略』とは、組織

の使命、将来像、及び目標(Mission, vision and objectives)をどのように実現するかを明示した総合的な計画のことであり、組織の比較優位を最大化しつつ比較劣位を最小化するものである。また戦略的マネジメントとは、外部環境分析、内部環境分析、戦略策定、組織戦略執行、業績測定、フィードバックを含む。」

戦略計画の策定の流れは図5に示したように、まず、環境分析(社会倫理の確認、外部環境分析、内部環境分析)を行なったあと、(1)状況分析(SWOT分析®)(2)使命(ミッション)の再確認、(3)全社戦略の策定、(4)部門別戦略の策定の4段階で行なうと解説している。さらにその後、戦略計画の実行前に、(5)人員配置(スタッフィング)(6)実行計画(アクション・プラン)の作成を行ない、実際に実行に移すことを解説している。

図4 特定セクターの戦略計画と業績測定の仕組みの一例 (タンザニアの「農業セクター開発戦略」における県農業開発計画)

1	十キャロム	ノギビ	۔ ا	レン	- 3	\	≠□中
	状況分	J 171		レン		/	'設定

- (1) 県農業の現状の確認
- (2)外部分析 機会(マーケットチャンス等)と脅威の特定
- (3)内部分析 強み(生産・加工・流通ポテンシャル等)と弱みの特定
- (4)ビジョンとキャッチフレーズの設定
- (5)実行手段の候補の選定

2. 村落計画の設定

- (1) 村落農業のビジョンの設定
- (2) 村落農業の計画・予算策定
- (3) 実行計画の策定他

3. オーナーシップの醸成

選定、統合、他計画との協調

- 4.参加型モニタリングと評価の計画策定
- 5. 県農業計画としての統合 計画書の完成と提出

6. プログレス・レポートの作成

(アウトプット・モニタリング・レポートを含む)

(出所) Government of Tanzania (2002)

そのあとで、業績の測定(Measuring Performance)が解説されている。業績測定の部分を、前図と同様に図中に矢印で示した。したがって、民間企業の解説でも、業績測定は、全体のプロセスの中の一部に過ぎないという位置づけであり、また実際の民間企業経営の実態からもそう理解するのが適切であるう。

(3) 戦略計画と実績評価の関係の再検証

戦略計画については、未だ定まった単一の手順はないと言わざるを得ない。しかし、前節で見た4つの戦略計画の流れに共通した事項を導き出すと、おおむね次の手順を取ることになると言えるだろう。本論文の結論の前に、簡単に一般的な手順を解説することにより、結論を理解するための助けとしたい。なお、戦略計画の策定の詳細な手順については別途、参考文献を参照されたい。。

言うまでもなく、前節の4つはあくまで事例に過ぎず、たとえ中央政府 / 地方政府、公共分野 / 民間分野、組織対象 / 分野対象というバランスが考慮されているとしても、厳密かつ体系的に選択された代表例というわけで

はない。これ以外の戦略計画の策定の仕方が存在する可能性は十分にあり得る。4つの中でも一定した手順があるとも言えない。例えば、ある事例では、ミッションの再定義や特定が、環境分析の前に来ている一方で、別の事例では、その順序が逆になっており、この点にもある特定の順番が定まっているわけではないことが表れている。したがって以下さい今回分析した事例から最大公約数的に導合分析が行われる際の何らかの参考となることをあらかじめお断りしておく。

1) 外部環境分析:成長機会と脅威の特定

最初に行われるであろう作業は、外部環境 分析である。外部環境は通常、3つの階層に 分けて認識するのが有効である。1)社会経 済環境、2)市場環境、3)タスク環境(自分 が日々触れる仕事環境)の分析である。

こうした外部環境分析を通じて、組織が生き延びるためのチャンスである「成長機会」と、落ちたら組織が破滅の危機にさらされる「脅威」が探知される。そして「脅威」に直面することをうまく避けながら、生き延びる

図5 民間分野における戦略計画と業績測定の仕組みの一例

1. 環境分析 (Scanning the Environment)

- (1)社会倫理の確認
- (2)外部環境の把握と分析
- (3)内部環境の把握と分析

2. 戦略計画の策定 (Strategy Formulation)

- (1) 状況分析 (SWOT 分析)
- (2)組織の使命(ミッション)の再確認
- (3)全社戦略 (Coorporate Strategy)の策定
- (4)部門別戦略の策定

3. 戦略計画の実行、業績測定、フィードバック(Strategy Implementation and Contr

- (1)人員配置(スタッフィング)と組織構造の変更
- (2) 実行計画(アクション・プラン)の策定と実行
- (3)業績の測定 (Measuring Performance)
- (4)適切なフィードバック (Proper Control)



(出所) Hunger (1993), pp.ix-xiii.を筆者が整理した。

ために「成長機会」を利用していくのである。 分析結果は、図6のSWOT分析図の縦軸(成 長機会、脅威)の欄に箇条書きでまとめるこ とになる。

2)内部要因分析:「強み」と「弱み」の特定 次に行われるであろう作業は、内部要因分 析である。自分が持つ資源を冷静に分析して、 ライバルよりも勝っている(優れている)と いう資源(「比較優位(強み)」)と、競合相 手よりも劣っている資源(「比較劣位(弱み)」)を識別するのである。なお、優れている、 のである。なお、優れている、 というのは競合相手と比較することによってしか分からない。資源には、人的 資源、物的資源、財務的資源、情報的資源な どがある。分析結果は、再び図6のSWOT分 析図の横軸(強み、弱み)の欄に箇条書きで まとめることになる。

3)組織のミッション(使命)の特定とビジョン(将来像)の決定

外部環境分析と内部要因分析が行われた後で、組織のミッション(使命)の特定を行なうことになる。ミッション(使命)は、時代の流れに沿った形で「組織が今後受け持ちたいと考えている社会的意義を内外に示す」ものである。そして、ミッションが示した社会

図6 SWOT (成長機会、脅威、比較優位、 比較劣位)分析図

内部	比較優位(強み)	比較劣位(弱み)
	•	•
外部	•	•
	•	•
成長機会		
•		

(出所)筆者作成

的意義に基づき、「一定の期間(例えば5年)後に、組織がどういう姿になっているかという将来像を描く」、これがビジョン(将来像)である。

4) SWOT分析図を作り、戦略目標の候補を 並べる。

「成長機会」と「脅威」を横軸に、「比較優位(強み)」と「比較劣位(弱み)」を横軸に据えると、4つのマスができる。それぞれを掛け合わせて図6のように並べる。

この図は次のように理解するのが一般的で ある。「成長機会」×「比較優位(強み)」 (図中の)のマスは、まさにチャンスその ものだ。このマスはどんどん大きくなって (成長して)いくし、しかも自分の組織が、 競合相手よりも優位に立つ資源を有してい る。ここで攻めない手はないので、このマス の基本は「攻め」の戦略であり、集中的に資 源を投入していく。この逆が、「脅威」× 「比較劣位(弱み)」(図中)のマスだ。こ のマスはどんどん縮小していくし、しかも自 分の組織の資源が競合相手の資源よりも劣っ ている。例えるなら、アリ地獄に引き込まれ るアリの中でも、他のアリに蹴られながら落 ちていくアリそのものだ。このマスからは、 とにかく早く資源を引き上げねばならない。 さて、残りのふたつのマスは微妙に判断が難 しくなる。「成長機会」×「比較劣位(弱み)」 (図中)のマスは、マス自体はどんどん大 きくなるが、自分の組織の資源が、競合する 組織よりも劣っている場合である。このマス では、成長の波に乗ることを利用して、自分 の組織の比較劣位を改善していく。最後が、 「脅威」×「比較優位(強み)」(図中)の マスだ。このマスはどんどん縮小していくが、 さりとて、自分の組織の資源が他の組織より も優位に立っている。アリ地獄に引き込まれ るアリのなかで、もっともヘリに近いアリが 自分だという場合である。もちろん撤退が基 本だが、ひょっとして他のアリは抱えていた 食物を残していくかも知れない。そして、自 分はその食物を拾う余力がある。したがって、

アリ地獄に落ちることを回避しながら、拾えるものを拾うのである。

5) ビジョンの達成に貢献する戦略目標の決定

SWOT分析図で得られた戦略の方向性に基づいて戦略体系図を作る。具体的には、ミッションとビジョンを一番上に書き込み、それを常に見ながら、ビジョンの実現に貢献する項目を「戦略目標」(Strategic Objectives)として、4つのマスの中から選ぶ。そして選択した項目を、下の図7のようにビジョンの下にぶら下げるように書き込む。

6) 戦略体系図の完成

ビジョン(将来像)の達成に貢献する戦略 目標が選択できたら、各戦略目標を最終成果 としてそれを達成するための施策や事業を書き込む。これで戦略体系図は完成する。なお、 戦略目標の達成のためにさらに複数の中間的 な戦略目標を並べる必要がある場合には、さらに枝分かれした体系図を作成することになる。

7)ビジョン(将来像)と戦略目標について 指標と数値目標を設定する。

ビジョン(将来像)とそれぞれの戦略目標について指標と数値目標を設定する。また、その下に羅列された中間的な戦略目標や施策についても指標を設定することが望ましいが絶対ではない。戦略目標の達成のために、戦略目標以下の部分は、適宜現場の判断でフレキシブルに変更されうるからである。

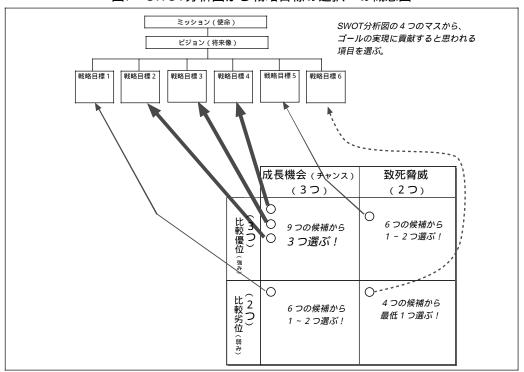


図7 SWOT分析図から戦略目標の選択への概念図

*図中の数字は、筆者達が読者の理解を助けるために記載したものであり、その数字が絶対というわけではまったくない。

(出所)龍(監修)/佐々木(著)(2003)

8)戦略目標ごとの実施計画(アクション・ プラン)を作る。

それぞれの戦略目標の達成ための行動計画となる実施計画(アクション・プラン)を作る。具体的には、上記6)で特定されている施策や事業に関する実施時期、概要、実施部署、予算などの内容で構成されることになる。なお、戦略目標ごとに明確に責任者を決めて、その責任者に大幅な人事権と予算権を与えることが理想的である。行動計画を別紙にまとめて、戦略体系図に添付すれば、戦略計画策定の作業は終了である。あとはすみやかに実施に移る。

以上が戦略計画の内容であるが、戦略計画 が実行に移されたらいよいよ実績評価(業績 測定)を実施することになる。

9)策定された戦略計画に示された指標を使用して実績評価を実施する。また、設定された数値目標に基づいて、戦略目標の達成度合を判断する。

戦略計画と実績評価の関係は、図8のような概念図で示されることがある。この図は両者が一対のものとして交互に、そして不断なく行われることを示していると理解することができる。アメリカの研究者ハーリー・ハト

リーはこの図を解説する際に、「戦略計画と 業績測定が一緒に使用されたとき、継続的な 営みとしての『成果のための行政』が実現す ることになるのだ。」と述べている (Hatry et al 2001, p.5)。

3. 日本での導入に向けた課題と提言

こうした戦略計画の策定は日本の行政機関 では行われているであろうか。筆者の知る限 りその実例はまだない。外部環境の分析から 機会と脅威を特定し、自身の組織を構成する 要素を冷静に分析して強みと弱みを特定し、 組織全体を統一するミッション(使命)やビ ジョン(将来像)を厳しく再検討する。そし て、機会が存在し、かつ自身の組織に強みが ある領域には、自身が持つ限られた資源を集 中的に投入する一方で、脅威に直面し、かつ 自身にこれと言って強みが存在しない場合に は、思い切って全面的に資源を引き揚げて撤 退する。まさに、「選択と集中」ということ であり、少し具体的に述べれば、「機会の選 択と資源の集中」ということである。なお、 「選択と集中」ということは、逆に言えば、 「選択と撤退」という非常に厳しい意思決定 も同時に行なうということである。それは現 象としては、ある特定の部署の廃止や、ある

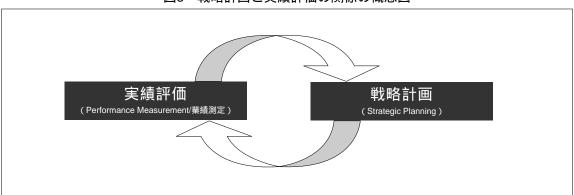


図8 戦略計画と実績評価の関係の概念図

(出所) Hatry et al. (2001), p.8

特定の人事ポストの廃止として表れることとなる。こうして戦略の方向性が決まれば、あとは実施計画を作り、資源の投入比率を決めて、実施に移していく。こうした作業を実施している日本の行政組織が存在して来たであるうか。日本の行政組織が行なってきた一般的な政策の立案や、各部署が上げてくる活動計画を合体させて年次計画とする作業とは、あまりに差が大きいと言わざるを得ないだろう。

以下、本論文での分析と議論を踏まえて、 今後の日本の政策評価の進め方についていく つかの提言を付記したい。

(1) 3つの政策評価手法の横並びの変更

現行の政策評価法では、政策評価ガイドラインに規定された3つの評価手法のうち、どれかを使用すれば政策評価を実施したことになると解釈できるが、これをやめて、次に立つに位置づけるべきである。まず、実績評価のもともとの考え方を忠実に踏まえて、戦高計画と実績評価をワン・セットで導入することを全省庁に義務づける。そして、戦略計画とを全省庁に義務づける。そりて、戦略計画によって設定された戦略目標の実現に貢献すると判断された特に評価を実施する価値があると判断された特

定の施策について事前段階での検証のために 「事業評価」を適用するとともに、同様に、 事後に評価を実施する価値があると判断され た施策について事後段階で「総合評価」を適 用するという仕組みである。つまり、組織全 体に適用され、組織全体をコントロールする 役割を担う戦略計画プラス実績評価に対し て、「事業評価」と「総合評価」が、選定さ れた個別の施策を評価するための道具として 仕えるという仕組みである。これを図示する と図9のようになる。このように仕組みを変 更して初めて、省庁等改革基本法や政策評価 が目指した「戦略性の確保」が実現されるの であろう。これに伴い法律の通称も現在の 「政策評価法」(Policy Evaluation Law)から、 より内容と目指すところをはっきりと明示す るために、「戦略計画・政策評価法」(Strategic Planning and Policy Evaluation Law) と変更さ れることが必要となっていくであろう。

(2) 成果に基づく人事・予算の本格的な導入

策定された戦略計画に基づいて、日本の行政組織も組織構造を大胆かつフレキシブルに変更することが必要だろう。つまり、「組織が戦略を作るのではなく、戦略が組織を作るのだ」ということへの発想転換である。また人事も成果に基づいて行われる必要が出てく

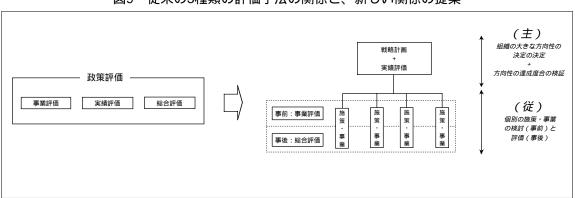


図9 従来の3種類の評価手法の関係と、新しい関係の提案

(出所)筆者作成

る。成果に基づくということは言い換えれば、 採用の種類ではなく、その一年、あるいはある決められた期間の間に達成することができた成果に基づいて昇進や降格を決めるという 民間分野では当たり前の原則を当たり前に実行するということである。さらに、予算についても成果が出た施策は拡大し、成果が出なかった施策からは予算を削るかあるいは施策自体を止めるということも重要であり、つまり予算の「選択と集中」を行うということである。

(3) 業績測定と評価の効果的な組み合わせ

アメリカの行政機関のひとつであるアメリ カ国際開発庁(USAID)では、1993年の GPRAの施行に伴って、全庁的に戦略計画と 業績測定を導入した。これに伴い、1994年に それまでの評価のあり方を根本的に見直し、 全ての施策で業績測定(実績評価)を実施す ることを決定した。そして、予想外にいい成 果が出た施策、あるいは逆に予想外に悪い成 果しか出なかった施策に関して、従来型のプ ログラム評価(総合評価とかなりの程度重な り合う評価)を適用して原因を探り改善提言 を得ることにした。ローコストで簡便な方法 と、高価ながら精緻な方法の組み合わせであ る。これは、次の考え方に基づいていると言 える。業績測定によって事業実施と成果発現 の因果関係が証明されるわけではないのは、 前述のアメリカの研究者ハトリーなどが以前 から指摘するとおりだが、逆に考えれば、予 想された成果がほぼ出た施策は、暗黙のうち に考えていた因果関係(例・予防注射を打て ばその病気の発生率が下がる)が、だいたい 証明されたと考え、深く考える必要はない。 むしろ、予想外の結果が出た施策こそ、何か 予想できなかった因果関係があったはずで、 それにどう対処すればいいのかの提言も得ら れるのではないか、という考えである。業績 測定 (実績評価)が導入され定着しつつある 日本においても、限られた評価予算の中で最 大限の効果をあげるために、今後このような 組み合わせが主流になっていくのが望ましいであろう。

なお、上記の提言(1)~(3)を実現するためには、それぞれ相当な研修や教育の実施が必要であることも指摘されねばならないだろう。

4. 結論

冒頭で解説した日本の政策評価法は、アメ リカの政府業績結果法(GPRA)を一部手本 にしたと見ることができることは既に述べ た。またその前に公表された総務省の政策評 価ガイドラインの中の実績評価も、GPRAが 求めた業績評価の手法を参考にしたと理解さ れる。アメリカで導入されたGPRAでは、戦 略計画 (Strategic Plan) の策定と業績測定 (Performance Measurement)をセットで導入 することが義務づけられている。ところが、 日本の上記法律や政策評価ガイドラインで は、そのうち業績測定の部分のみが導入され たり紹介されることとなった。日本では、従 来の政策や年次計画を、「戦略計画」になか ば強引に読み替えて、あるいは見立てて業績 測定(実績評価)のための指標設定の見取り 図として利用してきた。しかし、本来の戦略 計画はそれとは別物であり、戦略計画は業績 測定とワン・セットとして用いられて初めて 効果を発揮するものであろう。もし組織が生 き残りを考えるならば、評価の視点を「必要 性、効率性、有効性、公平性、優先性」など に求めるだけではなく、自らが策定した「生 き残り戦略」の達成度を測定すべきである。 「必要性、効率性、有効性、公平性、優先性」 は向上しましたが、生き残りには失敗しまし た、ということでは、その組織の設立や運営 費用を負担し続けてきた納税者は納得しない であろうし、組織トップの責任も問われる可 能性がある。戦略計画と業績測定(実績評価) をワン・セットで導入していくことが、今後 は必要であろう。

注記

- 1 アメリカの分類に従い、費用 便益分析 (Cost-Benefit Analysis)と費用 効果分析 (Cost-Effectiveness Analysis)を合わせた 大括りのカテゴリーとして費用-効率性分 析(Cost-Efficiency Analysis)という用語 を用いた。
- 2 2000年のEvaluator's Instituteのトレーニング・コースでの講師(Lipsay, M.)の発言。
- 3 この本の中で著者たちは、今後政府が導 入すべき以下の10の原則を上げている。 (1) 触媒的役割を担う政府(Catalytic Government (2) コミュニティに所有さ れた政府 (Community-Owned Government)、(3)競争する政府 (Competitive Government)(4)使命に推 進される政府 (Mission-Driven Government)、(5)結果重視の政府 (Results-Oriented Government)(6)顧客 重視の政府 (Customer-Driven Government)、(7)企業心を持つ政府 (Enterprising Government) (8) 先回りし て対処する政府(Anticipating Government)、(9)分権化された政府 (Decentralized Government)(10)市場指 向の政府 (Market-Oriented Government) である。なお、このベストセラーの Appendixは、Appendix A.がAlternative Service Delivery Options、Appendix B.が The Art of Performance Measurementとなっ ており、これら10の原則の実現を担保す るための必須の方法としてパフォーマン ス・メジャーメント(業績測定)を位置 づけて解説している。なお、この本の脱 稿は1991年、ハードカバー版の出版が 1992年、ペーパーバック版の出版が1993 年と理解される。
- 4 さらに最近では、戦略計画に対する批判的な意見も出てきている。例えば、ミンツバーグは、トップダウン的な戦略のアンチテーゼとして、「創発戦略」(Emergent Strategy)を提唱しているが、

- 彼によると創発とは、個人の自由で自発的な行動が、自然に全体の秩序や行動を生み出すことを指しており(Mitzgerg, 1989)日本の状況にとって示唆に富む。また、戦略計画はある時点の状況下での組織の行動パターンを将来に向かって決定するものであるが、現実には、組織のの権力の相対関係や組織経営者の直感的な判断などが関係する組織運営プロセスの中で変容していくものであることから、つきにロジカル・インクリメンタル戦略(Logical Incremental Strategy)などの新しい主張がなされて来ている。
- 5 ~ の記号は筆者たちが独自に追加。
- 6 より詳しくは同経営協会のウェブサイト を参照されたい(http://www2.icma.org)。
- 7 同書は、カリフォルニア大学ビジネス・ スクール(リバーサイド校)で長年、使 用されている戦略計画のテキストでもあ る。
- 8 SWOT: 強み (Strength)、弱み (Weakness) 機会 (Opportunity) 脅威 (Threat) の頭文字の組み合わせ。
- 9 例えば、龍・佐々木(2002) 龍(1997)。

参考文献

- 上山信一(2002)『日本の行政評価 総括と 展望 』第一法規出版
- 大住荘四郎(2002)『パブリック・マネジメント 戦略行政への理論と実践』、日本評論社
- 古川俊一(2002)『異種の評価レジームをどう使うか:理論の発展と実務的統合の可能性』、日本評価学会第3回全国大会発表論文
- 古川俊一・北大路信郷 (2001) 『公共部門評価の理論と実際』 日本加除出版
- 山谷清志 (2002)「わが国の政策評価 1996 年から2002年までのレビュー」、『日本評価 研究』2(2)
- 山谷清志 (1997) 『政策評価の理論とその展 開』、晃洋書房

- 龍慶昭(監修)・佐々木亮(著)(2003)『政 策評価トレーニング・ブック 7つの論争 と7つの提案』。多賀出版
- 龍慶昭・佐々木亮(2002)『戦略策定の理論 と技法~公共・非営利組織の戦略マネジメ ントのために~』、多賀出版
- 龍慶昭・佐々木亮 (2000) 『「政策評価」の理論と技法』。 多賀出版
- 龍慶昭(1997)『経営管理』、創成社
- 行政改革会議(1997)『行政改革会議 最終 報告』
- 総務省(2001)『政策評価に関する標準的ガイドライン』
- Bryson, J.M.(1995). Strategic Planning for Public and Nonprofit Organizations. San Francisco, CA: Jossey-Bass Publishers.
- Chadler, A.(1962). *Strategy and Structure*. Cambridge, MA: The MIT Press.
- Gordon, G.L. (1993). Strategic Planning for Local Government. Washington, D.C.: International City / County Management Association (ICMA).
- Government of Tanzania (2002). Guidelines for District Agricultural Development Plans (DADP), Dar es Salaam, Tanzania.
- Hatry, H.P., Morley, E., Rossman, S.B., and Wholey, J.S. (2003). How Federal Program Use Outcome Information: Opportunities for Federal Managers. Washington, D.C.: National Academy of Public Administration (NAPA).
- Hatry, H.P., Liner, B., and others (2001). *Making Results-Based State Government Work*. Washington, D.C.: Urban Institute.
- Hatry, H.P. (1999). *Performance Measurement:* Getting Results. Washington, D.C.:Urban Institute Press.
- Hatry, H.P. (1997) Where the Rubber Meets the Road. In Katerin E. Newcomer (Eds.), Using Performance Measurement to Improve Public and Nonprofit Programs, New Directions for Evaluation 75 (Fall 1997), San Francisco, CA: Jossey-Bass Publishers.

- Hatry, H.P., Louis Blair, Donald Fisk, Wayne Kimmel (1987). *Program Analysis for State and Local Governments*. Washington, D.C.: Urban Institute.
- Hunger, J.D., and Wheelen, T.L.(1993). *Strategic Management (4th ed.)*. Reading, Massachusetts: Addision-Wesley Publishing Company.
- Mintzberg, H (1994). *Rise and Fall of Strategic Planning*. New York, NY: The Free Press
- Mintzberg, H.(1989). *Mintzberg on Management*. New York, NY: The Free Press
- Osborne, D. and Gaebler, T. (1993). Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector. Reading, Massachusetts: Addision-Wesley Publishing Company.
- PUMA, OECD (1994). Performance Management in Government: Performance Measurement and Result-Oriented Management. Paris: OECD.
- SIDA and UNDP (1997). Measuring and Management Results: Lessons for Development Cooperation. New York, NY: UNDP.
- Wholey J.S.(2003). Improving Performance and Accountability: Responding to Emerging Management Challenges Results-Oriented Management. In Stewart I. D and Michael S.(Eds.), Evaluating Social Programs and Problems: Visions for The New Millenium Ch.3. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wholey, J.S.(1997). Trend in Performance Measurement: Challenges for Evaluators. Evaluation for 21st Century. Thousand Oak, CA: Sage Publication.
- Wholey, J.S., Hatry, H.P., and Newcommer, K. (1994). *Handbook of Practical Program Evaluation*. Washington, D.C.: Jossey-Bass Publishers, San Francisco, CA.

(2003.7.2受理)

Evaluation as an Integral Part of Strategic Planning and Management Process -- How Strategic Planning Can Contribute to the Current Exercise of Policy Evaluation Activities in Japan --

Yoshiaki Ryu

Ryo Sasaki

Josai University Bigliuryu@aol.com

International Development Center of Japan sasa01@giganet.net

Abstract

In the process of introducing evaluation activities into Japanese public sector, 'strategic planning', which is generally recognized as counterpart of performance measurement in the United States, has not been seriously examined and considered. This fact has made it very smoothly to introduce evaluation activities in Japan, but also this has led the result that evaluation activities have not well contributed to realization of the idea and the goals sought in the trials of the public sector reform in Japan. Based on this situation, the followings should be seriously examined in revision of the current policy evaluation law, which is expected to be conducted soon. Firstly, strategic planning and performance measurement should be introduced as 'one set' into all governmental ministries and agencies as obligation. Secondly, Japanese term of 'project evaluation', which is almost similar to ex-ante cost-benefit analysis in the United States, and Japanese term of 'integrated evaluation', which is almost the same as program evaluation in the United States, should be employed in order to assess and evaluate programs and projects scrutinized and selected as a result of strategic planning.

Keywords

strategic planning, performance measurement, result-oriented management, outcome monitoring, performance monitoring,

【研究論文】

「市民参加型」二次評価の試み JICAの「インドネシア国スラウェシ貧困対策支援村落 開発計画」プロジェクト評価を事例として

森田 智

高松 香奈

三好 皓一

東京大学大学院農学生命科学研究科 aa07181@mail.ecc.u-tokyo.ac.jp 東京大学大学院新領域創成科学研究科

立命館アジア太平洋大学

要約

本稿は、日本の公益事業の中でもJICAプロジェクトを事例に、国民が行政活動に参画する手段としての「市民参加型」二次評価の新たな枠組みを提示した。本手法の手順は、1)インターネット等を利用した主に評価報告書からなる一次資料収集、2)情報公開法等に基づいた二次資料収集、3)PTMの策定、4)DAC評価5項目による二次評価の実施からなる。

本手法では、国内にいたまま既存資料を用いた低コストでの事業評価実施が可能となり、また評価者が自由な立場や視点から主体的に評価を実施することを可能にする。さらに、二次評価結果の公表により当該事業の評価内容が深められる。

こうした二次評価の概念は「評価」の歴史においても新しく、評価のもつ可能性が広がるという点でその意義は大きい。さらに、NPO等が公益事業を対象とした二次評価を積極的に行うことで、行政の政策形成における市民の発言力の拡大、および行政事業の透明性の向上と効率化が期待される。

キーワード

二次評価、市民参加型、プログラム・セオリー・マトリックス(PTM)、情報開示、DAC評価5項目

1. はじめに

近年の日本における新公共管理の考え方に基づいた行政改革の展開により、行政活動をサービスと捉え国民をその顧客とみなした「国民本位の行政」の考え方が広まりつつある。そのような中で、行政側はどのように自らの事業に関する国民への説明責任を果た

し、また国民側はどのように行政活動に積極 的に参画していくべきかについて関心が高ま っている。

行政への民間の自発的関与という考え方は、米国において比較的早くから存在した。 民主主義型の政治制度においては、政策形成・決定の過程が明瞭かつ透明性が高くオー プンであるべきという認識から、立法や行政 から独立した立場において、公共行政・政策 の決定に関与し影響を及ぼすといった役割を 果たす民間非営利のシンクタンクが現在、既 にいくつも存在する(アーバン・インスティ テゥート 2002)。ここでみられるのは、多様 な意見が民主主義の発展ないし市民社会の成 熟にとって不可欠であるという共通認識であ る。

日本でも近年、公益事業に関わる事業者の マネジメントや事業改善を支援する目的で、 評価システム開発や評価を通じた政策提言を 行っている「評価みえ」のように、行政と NPOとの協働を積極的に進める動きがみられ る。さらに、2001年には日本の行政において 政策評価(行政評価)制度が導入されるなど、 公共事業における評価の役割がこれまで以上 に重要視される中、納税者である一般市民が 主体となり行政活動を評価するというような 枠組みは未だ存在せず、行政事業の透明性や 効率化という観点からも、市民の立場や視点 からの評価の有効性が期待される。

本稿では、行政活動としてのODA事業の 国民に対する説明責任の遂行に寄与すること を目的とし、日本の行政の中でも「評価」に 関して経験の蓄積があるとされるJICAのプ ロジェクトを対象に実施した評価調査を基 に、一般市民が行政活動に能動的に参画し、 行政側の今後の政策形成に影響を与え得る一 つの手段として「市民参加型」二次評価の新 たな枠組みを提示する。JICA事業の多くは 開発途上国・地域で行われるという性格上、 その内容は一般市民にとって分かりにくいも のであり、たとえ興味があっても、ほとんど の一般市民には事業視察の機会がなく、また 仮に機会があっても、現地訪問のコスト負担 が大きくなる。さらに、プロジェクトを視察 できたとしても、それが必ずしもプロジェク ト評価の実践にはつながらない。そのような 状況をふまえた上で、上述した「市民参加」 という観点から、一般市民が現地へ赴いて調 査を行わずとも、日本国内にいたまま収集し た資料を用いてJICA事業を評価できる二次 評価の意義は大きい。また、このような評価 を主体的に実施することによって、評価者で ある一般市民が行政活動への関心を高めると ともに、建設的かつ具体的な提案を持って行 政へより積極的に関与することが可能とな る。

ODA事業一般に関しては、ODA大綱や中 期政策において述べられているように、援助 実施におけるNGOとの連携の重要性が高ま る中、JICAと外務省は独自に定期協議会を 立ち上げ、それぞれ「開発パートナー事業」 や「NGO事業補助金」等によりNGOとの連 携が進められているが、これらのODA事業 全体に占める割合はまだ大きいとはいえな い。NGOと援助実施機関のより一層の連携 強化が求められ、NGOを含めた市民社会の ODA事業への関わり方が模索される中で、 本研究での「市民参加型」二次評価の手法は 新たな方向性を示し得る。

この手法では、情報収集の対象が主に既存 資料となるため、評価実施のコストを低く抑 えることが可能となり、従来ならば実施機関 以外の外部者が評価を行う際に問題となって いた高コストの問題が解決される。また、こ の手法は、評価者の幅広い興味に対応し得る といった利点、すなわち、評価者が自らの興 味に基づいて、外部から制約を受けない自由 な立場や視点から主体的に評価を実施できる といった、既存の評価手法とは異なる新しい 可能性を有する。さらに、この手法は、一般 市民が行政側に対し、事業に関する情報開示 を請求できる制度が近年整備されたことで、 NPOやNGOにおいてもその潜在的利用度が より高まることが期待される。

本稿の構成は、次のように整理される。ま ず、第2節では本研究で実施した「市民参加 型」二次評価における枠組みとして情報収集 の手順を説明し、第3節では対象案件に関す る政策体系とそれに基づいた二次評価結果を 提示する。第4節では当該案件の意義および 本二次評価手法の可能性と限界点について考 察し、第5節では結論と今後の方向性を述べ る。尚、第3節での評価調査は、東京大学大 学院新領域創成科学研究科での講義「国際開 発協力論 」内の演習として実施されたものである。

2. 「市民参加型」二次評価の枠組み

本研究では、次に示す手順において二次評価実施のためのプロジェクト選定と情報収集を行った。

(1) 調査対象地域・分野の絞り込み

まずは二次評価の対象となるプロジェクト に関してであるが、JICAのホームページ上 において、1998年度以降の個別案件の評価調 査結果の要約等をまとめた事業評価報告書 や、2000年からは終了後の案件に関する国 別・特定テーマ別の事後評価報告書等が公開 されており、これらの資料からは案件名、国 名、協力期間、要請背景、目標、活動内容等、 各種プロジェクトの概要とその評価結果を見 ることができる。さらに、「地球規模問題」 としてジェンダー・WID、貧困、環境、教育 等、全9つの分野において、それぞれ複数の プロジェクト概要が紹介されている。これら の資料を利用することで、自ら興味をもつ国 や地域、または分野やテーマ等に基づき、調 査対象としてのプロジェクトの絞り込みが可 能となる。本研究では、日本のODA供与が 重点的に実施されている「アジア地域」、 MDGsの第一目標に掲げられている「貧困撲 滅」、近年の参加型手法を取り入れた「社会 開発」等のキーワードに着目し、プロジェク トの規模や実施時期等、さらに同ホームペー ジ上での情報検索結果も参考とした上で、ま ずいくつかの案件に焦点を絞った。

(2) プロジェクト選定と一次資料の収集

次に、これらの案件から最終的に調査対象 プロジェクトを選定する際には、調査可能性 という観点から、個々の案件に関してどの程 度情報収集が可能かということが重要とな る。JICAが作成したプロジェクトの各種報告書に関しては一般に、JICA図書館や国会図書館においてアクセスできる。JICA図書館では2002年12月の時点において、各種調査、技術協力プロジェクト、無償資金協力等、JICA事業の報告書を含むJICA刊行物が約38,300冊、さらにJICA調査団が海外で収集した開発途上国の政府刊行物、図書、調査関系が約16,700冊所蔵されており²、またインターネット上での目録検索システムも利用できる。さらに、同図書館が所蔵する資料は館内において閲覧および複写が可能である。国会図書館では、JICA図書館の目録検索システムを用いて

- ・特定プロジェクトに関する全評価報告書、 関連調査報告書、その他関連資料の入手可 能性
- ・政策体系の策定(後述)を念頭とした、隣接地域・類似分野における他の協力案件の存在
- ・JICA事業全体の中でみた際の特定プロジェクトの意義と重要性(本研究では特に、パイロット・プロジェクトの特性としての他の事業への適用可能性)

といった基準により、最終的に「インドネシ ア国スラウェシ貧困対策支援村落開発計画」 を選定した。同案件はインドネシア国内務省 村落開発庁および南スラウェシ州村落開発庁 を相手国実施機関として、1997年~2002年に 実施された。同国では国民の貧富の差ならび に地域格差の拡大が次第に顕著になりつつあ ることを背景に、経済発展がイ国内でも相対 的に遅れている東部インドネシアの南スラウ ェシ州タカラール県において、貧困対策事業 を含む住民参加型村落開発事業の立案形成・ 運営能力の強化を目的とした技術協力が行わ れた。プロジェクト目標は「南スラウェシ州 において適用可能な参加型村落開発モデル (以下、タカラール・モデル)が開発される」 上位目標は「スラウェシにおいてタカラー

ル・モデルが普及・応用される」が設定され た。プロジェクト・サイトには、モデル構築 のためのラボ村落として同県内の4村落が選 定された。タカラール県人口は22万人である のに対し、対象4村落の全人口は約12,300人 である。

同案件に関して留意すべき点は、村落住民 の自主的・主体的な村落開発計画の立案形 成・運営等の能力向上を図るという見地か ら、ドナーから村落民への支援が直接行われ ないという主旨のもとでプロジェクトが実施 されたことである(国際協力事業団社会開発 協力部 2001)。

尚、表1にJICA図書館より入手した資料の 一覧(一次資料)を示す。この表をみても分 かるように、一次資料の多くは通常、特定プ ロジェクトに関する評価報告書となり、この 既存評価資料が二次評価に必要な情報源の中 核となる。

(3) 文献調査と二次資料の収集

これまでに得られた一次資料を用いた文献 調査を行うことで、プロジェクトの全体像が 把握できる。この一次資料においては、当該 案件の実施に関連して行われた現地調査の結 果からの情報や、対象地域の社会・経済状況

に関する統計データ等がよく用いられている が、こうした現地調査に関しては通常、独立 した別の報告書としてまとめられ、さらに派 遣専門家による報告書等と共に、JICAの担 当事業部において保管されていることが多 い。このように、一次資料からさらに必要な 資料 (二次資料)をリスト・アップして入手 することで、当該案件に関して、対象地域で の実際の状況の詳細な把握が可能となる。

二次資料を入手するには、一次資料の中の プロジェクト関連報告書からJICAの担当事 業部および職員を割り出し、資料提供依頼の ための連絡をとる。尚、「独立行政法人等の 保有する情報の公開に関する法律」が2002年 10月より施行されたため、所定の手続により 当該案件の関連資料等に関する開示請求を行 うことでも資料入手が可能である。これらの 資料収集においてかかる費用は開示請求手数 料と、法人文書の閲覧や複写による交付を受 ける際の開示実施手数料となるが、全て実費 の範囲内である。今回の調査では、JICA社 会開発協力部社会開発協力第一課の当該案件 担当職員に協力を得ている。表2に入手した 資料のリストを示す。

表1 一次資料リスト

国際協力事業団社会開発協力部「インドネシア共和国スラウェシ貧困対策支援村落開発計画」

- ・事前調査団報告書(平成8年5月)
- ・実施協議調査団報告書(長期調査員報告書)(平成8年12月)
- ・事前打ち合わせ調査団報告書(平成10年5月)
- ・巡回指導調査団報告書(平成12年4月)
- ・終了時評価報告書(平成13年10月)
- ・98年度研究会報告書(平成12年2月)

龍谷大学「インドネシア・スラウェシ貧困対策支援村落開発計画プロジェクトと連携した社会統計分 析」(平成10年度 国際協力事業団委託業務完了報告書)(平成11年3月)

国際協力事業団国際協力総合研修所

- ・「ソーシャルキャピタルと国際協力「事例分析編]」(平成14年8月)
- ・「インドネシア・スラウェシ貧困対策支援村落開発プロジェクトと連携した地域社会開発手法の研 究報告書」(平成9年12月)

表2 二次資料リスト

- ・国際協力事業団社会開発協力部「インドネシア・スラウェシ貧困対策支援村落開発計画平成9年度 リーダー会議個別打ち合わせ資料No.2」(平成10年2月)
- ・国際協力事業団社会開発協力部「社会経済調査()報告書概要版」(平成10年8月)
- ・柿沼潤「インドネシア・スラウェシ貧困対策支援村落開発計画WID分野短期派遣専門家業務報告書」 (平成12年4月)
- ・原口孝子「短期専門家活動報告書」(平成11年3月)
- ・福島州馬「ニーズ調査報告書」(平成11年6月)
- ・渡邊敦「南スラウェシ村落経済支援構想のために ウサハの創意と多様性からの出発 (村落経済インタラクション調査最終報告書)」(平成11年9月)

3.「市民参加型」二次評価の実施

本研究では、収集した資料に基づき、次に示すような手順において実際に当該案件に関する二次評価を実施した。

(1) 評価の中心視点と評価設問の設定

まずは、二次評価を実施するにあたり、評価を通じて何を明らかにしたいのか、評価設問を設定する必要がある。本研究では、村落開発というプロジェクトの性格上、対象地域の住民側の立場からの公平・公正性の観点を評価の中心視点に据え、

- 1. 当該案件の対象地域における村落民の 真のニーズが案件形成および実施に反 映され、村落内の社会・経済格差に配 慮した支援が実施されていたか?
- 2. 当該案件において構築されたモデルの 有効性が保証され、またモデル普及が 実現するようにプロジェクトが適切に デザインされていたか?

という設問を設けた。

(2) 政策体系の策定

次に、インドネシアの発展における当該案件の意義を再確認する必要性から、プログラム・セオリー・マトリックス(PTM)を用

いた政策体系を策定し、その中でのプロジェ クトの位置づけを捉え直すことが重要となる (三好2002)。PTMとは、個々のプロジェク トがどのような政策あるいはプログラムの下 で行われているかについて、その因果関係を 明確化する機能をもつツールであり、通常は 最終成果、中間成果、結果、活動、投入の5 項目および前3項目の指標からなる。投入か ら結果までが行政のコントロールが及ぶ範囲 であり、中間成果と最終成果は、それぞれ行 政活動によるターゲットグループの変化、社 会の変化でありアウトカムに該当する(三好 他2003)。本来ならば、こうした政策体系は 案件実施前に策定されているべきものである が、JICAにおいて国別での事業実施体制が 強化され始めたのは近年のことであり、また 当該案件の実施開始時には政策体系の概念自 体が日本の開発援助関係者の間でもまだ一般 的なものではなかったことから、本研究では 独自に、二次評価で用いることを目的とした 当該案件に関する政策体系を想定した。

その際に留意すべき点は、対象国における 国家開発計画および戦略、対象国からの援助 要請内容、対象案件における目的や活動内容、 そして評価の中心視点等である。本研究においては、これらの点を全てふまえて表3の政 策体系を策定し、まず最終成果として国・地域の貧困緩和を設定した。ここでは、イ国の 行政区を細分化すると、州、県、郡、そして 最終的には村の単位になるため、表3では便 宜上、県レベルで区切り、イ国がA~Z県か

ら構成され、各県においてa~z村が存在する と仮定した際、イ国における貧困緩和とは、 イ国を構成する各A~Z県における貧困緩和、 さらに県を構成する各a~z村の貧困緩和に他 ならない、という認識に基づいている。

また、当該案件では、4モデル村落におけ る住民参加型村落開発モデル構築のためのパ イロット事業の実施に加え、イ国内の他地域 へのモデル普及のための情報整備とマニュア ル作成が主に実施されたが、表3の政策体系

では便宜上、パイロット事業に焦点を当てて おり、表3の下部でみられる4モデル村落以外 を対象としたモデル普及事業に関しては、因 果関係をより明確化するため、別の政策体系 (表4)として表した。図1~図4は、プロジェ クト実施前からモデル機能想定時までの各段 階における利害関係者間での資源フローの概 念図(全て筆者作成)であり、これらを用い て上述した2つの政策体系を説明する。尚、 ここでの「資源」には知識やノウハウなどの

表3 当該案件の政策体系(パイロット事業)

最終成果	(国・地域)	中間成果	結果	(指標)	活動
インドネシア	タカラ (A) 県		一	・増加した住民組織数 ・組織参加する住民 の割合 ・行政側へ機関を対した。 ・行政側へ関連を対した。 ・一では、一部では、一部では、一部では、一部では、一部では、一部では、一部では、一	(a ドナー 現地NGO) ・NGOスタッフへの研修実施と ソフト面での技術協力 ・ 現地NGO 村落住民) ・NGOスタッフによる村落住民の 組織化支援および住民組織の 運営支援 ・ 地方政府/NGO 村落住民) ・ 産業特性に応じた技術協力 ・ 社会的格差に配慮した貧困層の村落民へのエンパワメント (ドナー) ・ 各種社会調査実施 (c ドナー 地方政府) ・ 支援システム事業(SISDUK)の 企画運営、マニュアル整備と 人材養成 ・ 参加型計画手法(PLSD)の
	b~ d木 <u>パイロット事業</u>	住民組織の自主 運営体制の確立 <u>モデルb~d</u>			
	(A県) e~ z木				
	他県(B~Z県)a~ における貧困 z木 緩和				 <u>モデル普及事業</u>
	投入予定額(当初): 4億円(専門家関連経費は除く) 実績:機材供与2,700万円、ローカルコスト負担9,800万円				

⁽注)結果の指標のうちイタリック体(*)のものは関連データが存在せず、また活動のうち イタリック体(*)のものは実際に実施されていない

(出所)筆者作成

ソフト面も含まれる。

パイロット事業実施前

まず、タカラール県(A県)の任意のモデル村落(a村)を一応想定した当該プロジェクト実施前の地域社会においては、図1で示すように、地方政府と現地NGOによる村落住民への資源の流れが小規模(レベル1)であり、イ国中央政府から地方政府への資源配分も十分でない。村落住民から地方政府への資源のたりが、期待されるフロー(事業案件の提出)である。尚、地方政府から現地NGOへの資源の流れはないものとここでは便宜的に仮定する。

パイロット事業実施とモデル構築

次に、ドナー(JICA)により4つの各モデル村落(a~d村)においてそれぞれパイロット事業が開始されると、表3で示すように、「タカラール県の貧困緩和」という最終成果のもと、「住民組織の自主運営体制の確立」という中間成果を達成すべく、 住民の事業形成・実施能力を養成し、 住民に対し社

会・経済状況に配慮した支援を実施し、 地方政府による住民主体の事業への支援制度を確立するという、「結果」としての行政サービスを提供するために、数々の活動を関連である。これらの活動に関係者間での資源フローという観味がら整理してあり、イタリック体で示してものであるものは、実際には行われなかったが行われるがりは全て実施された活動である。また、結果の項目における指標に関して、イタリセン体で示してあるものは関連データが存在しないものである。

図2がパイロット事業における資源フローの概念図であり、図中の矢印とアルファベットは表3の活動の各項目と対応している。例えば、ドナーから地方政府への支援 (c) と現地NGOへの支援 (a) が実施されたことで、地方政府と現地NGOから村落住民への働きかけと資源の流れ(それぞれd とb)が増大し(レベル2) その結果として村落住民の活動が活発となり、地方政府への事業案件の提出数が増大した。

表4 当該案件の政策体系(モデル普及事業)

最終成果国	最終成果(地域)	中間成果	結果	(指標)	活動
インドネシア	県村	住民組織が	モデルを	* 広報活動のための	<u>(c</u> <u>ドナー 地方政府</u>)
において	において	持続的に	導入・普及	セミナーの実施数	・イ国の地方行政レベルにおける
貧困を緩和	貧困を緩和	自主運営	する	* 広報活動の予算額	モデルの広報活動
する	する	できる体制		*人材養成のための	・モデル普及のための人材養成
	(A県A~D村	を確立する		ワークショップの	(<u>d</u> 地方政府 村落住民)
	を除いた全ての			実施数	* 住民への広報活動
	村落が対象)			*人材養成のための	(<u>e</u> <u>ドナー 中央政府</u>)
				活動の予算額	* イ国の法政面の環境整備への
				* 関わっているNGO	支援
				およびスタッフの数	* イ国の政府レベルにおける
				* 地方政府の関係者	モデルの広報活動
				間での資料・教材の	・モデル普及のための人材養成
				普及率	(<u>f</u> <u>地方政府 現地NGO</u>)
				* 住民事業への支援	*NGOスタッフへの研修実施と
				を義務づける法律・	ソフト面での技術協力
				政令の数	(<u>ドナー</u>)
				! ! !	・パイロット事業での経験に基づい
				1 ! !	たモデル構築と資料・教材作成

(注)結果の指標のうちイタリック体(*)のものは関連データが存在せず、また活動のうちイタリック体(*)の ものは実際に実施されていない

(出所)筆者作成

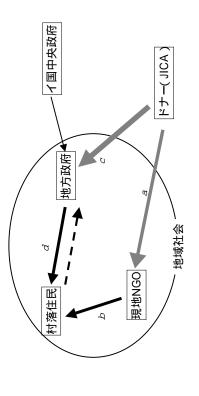


図2 パイロット事業における関係者間の 資源フローの概念図

パイロット事業実施前の関係者間の

<u>×</u>

資源フローの概念図

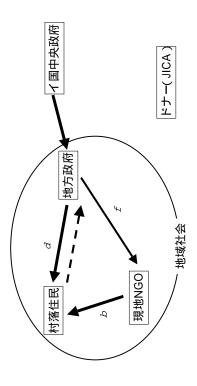
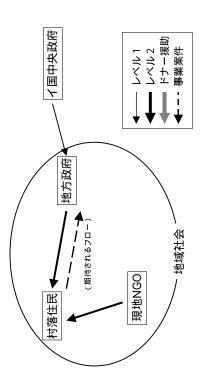


図4 モデル機能想定時における関係者間の 資源フローの概念図



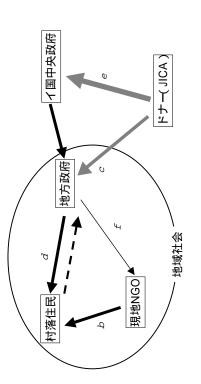


図3 モデル普及・導入時における関係者間の 資源フローの概念図

これらの活動に基づき、タカラール県(A県)での4モデル村落(a~d村)におけるパイロット事業実施の経験(モデルa~d)を全て合わせることで、タカラール・モデルが構築される。この「モデル」とは、

ある事業を実施する際に設定すべき活動 の枠組み、およびその中で従うべき事項 を含む模範的体系

とここでは定義する。タカラール・モデルには、表3で示すようなパイロット事業で実施された活動が、体系的に整理された上で全て含まれる。この中で特に重要なのは、ドナーにより村落の現況把握を目的とした各種調査が実施されていること、地方政府により支援システムの法制化がなされたこと、現地NGOが、村落住民に対する支援において非常に重要な役割を果たしていること、といえよう。

モデル普及事業実施

構築されたタカラール・モデルが有効であ るという前提のもと、表3の政策体系で示す ように、今度はタカラール県内の他村(A県 e~z村)およびB~Z県の全村において同モ デルを普及するためのプロジェクトがドナー により開始されることとなる。表4が、任意 の村(県村)を対象とした政策体系とな っているが、表3と同一の最終成果のもとで 同一の中間成果を達成すべく、「モデルを導 入・普及する」という行政サービスを提供す るための活動を実施することが想定される。 これらの活動も同様に利害関係者間での資源 フローという観点から整理してあり、またイ タリック体で示したものは、プロジェクト活 動として行われるべきであったと筆者らが判 断したものである。結果の指標に関してもイ タリック体で示してあるものは関連データが 存在しないものである。尚、モデル構築の資 料・教材作成は主にドナーが担当している。

図3が、 県 村を対象としたモデル普及 事業における資源フローの概念図であり、表

4の活動の各項目と対応している。ここでは ドナーが、特にイ国中央政府に対する支援 (e)の実施を強化するため、対象村落を管轄 する地方政府へのフロー(c)は減少する (レベル1)こととなるが、その分中央政府か ら地方政府への資源配分が増大することが期 待される。しかしながら、このフローはプロ ジェクト活動の範囲外であることに留意しな ければならない。また、モデル普及において はドナーから現地NGOへの支援(a)がなく なるため、代わりに地方政府から現地NGO への支援 (f) が適切なレベルで行われる。 つまり、図2と図3を比較した場合、2つの政 策体系における活動内容が全体的に異なるた め、図2におけるドナーからの*aとc*のフロー において図3では変化が見られるものの、モ デル普及事業実施によりパイロット事業の場 合と同じ効果が出ることが期待される。換言 すると、村落住民に対するbとdのフローが同 じままであることが、モデルが導入されたこ との証明といえよう。

モデル機能想定時

タカラール・モデルが4モデル村落以外の 全村に導入・普及された後には、図4で示す ように、ドナーからの支援 (cとe) なしでモ デルが機能することが想定される。この場合、 図3と図4を比較すると分かるように、イ国中 央政府からは地方政府への資源配分が継続的 になされることが前提となっており、また図 3における村落住民への支援 (bとd)、さらに 地方政府から地元NGOへの支援 (f) がその まま継続している。さらに、図1と図4を比較 すると、政策体系に基づいた場合の当該案件 の実施前後において、中央政府から地方政府 への資源配分が増大し、それが地域社会内で の各利害関係者間における資源フローの増大 となって現れているなど、地域社会内での諸 活動がより活発になり、その結果として村落 住民の活動も活発になることで、行政側への 事業提案数が増大していることが分かる。

(3) プロジェクト実績とDAC評価5項目に よる評価

次に、本研究において策定した政策体系と プロジェクト実績、および設定した評価の中 心視点と評価設問に基づき、既存のJICAプ ロジェクト評価手法で用いられているDAC 評価5項目により、対象プロジェクトの二次 評価を行う。実際の作業においては、平成13 年9月に刊行された「JICA事業評価ガイドラ イン」を参考とした。尚、同ガイドラインは JICAのホームページにおいても入手可能で ある。表5において政策体系とDAC評価5項 目との関係を示す。評価項目は妥当性、イン パクト、有効性、効率性、自立発展性の5項 目、評価領域は網目状になっている部分が該 当する(三好 2002)。 プロジェクト実績に関 しては、評価項目の該当部において同時に示 す。

妥当性

妥当性をみる主な観点は、相手国の政府レ ベルおよび本案件の対象地域における村落住 民レベルでのニーズ、および日本側の援助政 策との整合性とした。

第一に、策定した二つの政策体系に共通す る事項として、インドネシア政府は1994年に 公表した第6次国家開発5ヵ年計画の中で「平 等と貧困軽減 を中心目標として掲げており、 また経済危機後の1998年には、主に貧困層を 対象としたソーシャル・セーフティーネッ ト・プログラムを開始するなど、貧困緩和に 重点的に取り組んでいる(国際協力事業団 2002)。一方、日本のODA大綱や中期政策に おいても貧困問題への対応が重点課題の一つ として挙げられている。さらに、モデル普及 事業に関して、インドネシアでは1999年に地 方行政法および中央地方財政均等法の二つの 法律が成立するなど、地方分権化が進められ てきており、それに対応した形で、日本政 府はインドネシアに対し、日イ間の首脳合意 に基づいて地方分権化の分野での政策アドバ イスを実施している(国際協力事業団 2002)。 本案件はこうした流れに沿ったものと理解さ れるため、インドネシア国の国家開発計画や 日本のインドネシアに対する援助政策との整 合性が全体的にとれているといえよう。しか しながら、モデル普及事業に関しては、イ国 の法政面の環境整備への支援がより考慮され るべきであったように思われる(後述)。

第二に、インドネシア政府は第6次国家開 発5ヵ年計画の中で、国家開発のプロセスに おいて女性の地位向上を目指す旨を明記して おり(国際協力事業団企画部 1997) また日 本のODA大綱や中期政策ではジェンダーが 取り上げられ、開発への女性の積極的参加や 受益の確保など、女性支援の視点の重要性が 述べられている。これらとの関連において、 特にパイロット事業に関しては、報告書等の 記述に基づく限り、ジェンダーを意識した活 動が行われていたことが伺えるものの、事前 調査、プロジェクト実施、評価の全段階にお いてジェンダーが配慮されていたとはいえ ず、インドネシア国側のニーズおよび日本側

表5	政策体系。	とDAC評値	価5項目。	との関係

	最終目標	中間目標	成 果	活 動	投 入
妥 当 性					
インパクト					
有 効 性					
効 率 性					
自立発展性					

(出所)三好(2002)

の援助政策との整合性に欠けるように思われる。

第三に、上記の事項に関連して、パイロット事業では4モデル村落内の住民のニーズ調査が実施されたが、調査手法の性質や統計データの入手性という要因等により、住民の中でも特に貧困層や女性のニーズを正確に捉えることが非常に困難であったようだ。これは、「貧困」をどのように捉え、援助の対象平であるべきかという問題であるが、公平・ットを絞った活動の実施や、ジェンダーへの配慮という視点を活動に十分反映させるなど、会・経済状況により配慮した支援を実施する。

有効性

有効性をみる主な観点は、中間成果の達成度、および結果の中間成果への寄与度とした。第一に、パイロット事業では、表3の結果やでみられるように組織や制度等のハード面を重視した支援が行われてきたが、特に結果に関して、ソフト面、すなわち村民が従事する産業を考慮した支援は積極的に行われていなかった。村落における住民組織が持続的に自主運営できる体制を確立するためには、事業活動から継続的に便益を得ることが

重要となり、そのため産業の発展が鍵となる。 4モデル村落を選定した際にはモデル構築が 念頭に置かれたため、表6で示すように村落 ごとに産業が異なるよう考慮されていたこと を考えると、表中の「開発支援の視角」でも みられるように、産業特性に応じた技術支援 の意義は大きい。そのような観点から、村落 住民に対し彼らが従事する産業を考慮した支 援が行われるべきであったように思われる。

第二に、パイロット事業での結果 に関連して、4モデル村落を管轄するタカラール県での行政において、制度面でさまざまな変化がみられたが、それらが村落の住民組織のキャパシティー・ビルディングに及ぼす影響はあくまで間接的なものであり、また長期的なものであり、また長期的なり、の結果と中間成果との明確な関係を示すったは容易ではないといえよう。さらにこのことは、タカラール・モデル自体の有効性にも関係してくる(後述)。

第三に、モデル普及事業に関しては、本来ならばその有効性がある程度確認された段階で、他村において徐々に導入していくことが望ましく、例えばタカラール県内では早いうちからの普及事業への着手が可能であったと考えられるが、実際には4モデル村落以外への普及が少なくともプロジェクト実施期間内

20 年10 7711日の工産来に開ル文版の抗治			
村落名	主産業		地域内での開発支援の視角(概要)
Lassang	低地農村	農業	・国策中心から、個々の農家経営の所得向上への農業政策への 転換
Barugaya	高地農村	農業	・農業生産の多様化と、農民を主体と見なした環境作り(優良 種苗と肥料・農薬の供給)
Bonto	 	農業	・行政による投入財供給管理体制・耕作指導の廃止
Kassi 産業村	家内手工業	・品質向上による付加価値の増加	
		漁業(舟製作 を含む)	・自主的協同組合組織の実現への支援 ・資源保全が根本的問題
Bonto Loe 沿岸村	沿岸村	養殖業	・より安定的な技術の開発(小規模) ・経営状態の安定(大規模)
		工業・製造業	・品質向上による付加価値の増加
		農業	(上記参照)

表6 4モデル村落の主産業と開発支援の視角

(出所)渡邊(1999)を基に筆者作成

には実現しなかったため、モデル普及という 見地からのプロジェクトの有効性は非常に低 いといわざるを得ない。

効率性

効率性をみる主な観点は、本案件における 結果の産出までの資源活用の有効性とした。 表7は、終了時評価の時点でのパイロット事 業の結果に関する関連指標と測定されたデー 夕を示す。

第一に、パイロット事業の結果 に関して は、各指標の関連データがそれぞれ存在し、 数字上ではある一定の成果がみられることか ら、その効率性は高いといえよう。結果 に 関しては、設定された指標のうち支援システ ムの職員数や予算額に関するデータがないも のの、同システムの法制化の実現や行政の住 民参加型事業への支援状況において、ある一 定の成果がみられることから、その効率性は 高いといえよう。

第二に、パイロット事業の結果に関して、 設定した指標の関連データが一つしか存在せ ず、そのデータに関しても、住民組織へ参加 する女性の数が増えた一方で、全体に占める 割合は減ったことを示すのみであり、女性に 対する支援が結果となって現れたことは評価

できるものの、このことから全体的な状況を 判断するのは容易でない。しかしながら、表 8で示すように、プロジェクト開始後2年間で 村落住民開発ニーズ調査や女性経済活動調査 を含む9件の調査が行われているが、本研究 において入手した資料から判断すると、村落 住民の真のニーズの把握という観点からは、 これらの調査結果の案件実施における活用状 況が十分でなかったように見受けられ、特に 住民に対し産業を考慮した技術支援が行われ なかったことを考えても、社会・経済状況に 配慮した支援の実施ということに関して、効 率性はそれほど高くないと思われる。

第三に、モデル普及事業に関しては、資 料・教材作成に重点が置かれ、実際にモデル 導入・普及を促進するための、イ国の法政面 の環境整備への支援が行われなかったことか ら、効率性は低いといわざるを得ない。設定 した指標に関しても、関連データは存在しな L1.

インパクト

インパクトをみる主な観点は、村落レベル と国レベルそれぞれにおける、予測的見地か らの最終成果の達成度とした。

第一に、パイロット事業を想定した村落レ

結果	指標	データ		
	住民組織数	56組織(内、女性集団29組織)		
	サンプル集団20組織の参加住民数の変化 (内、女性の数)	354名 1054名:増加率300% (163名 354名:増加率220%)		
	組織会合の頻度	週1回 - 月1回		
	行政側に提出された住民参加型事業数	170件(内、女性集団91件)		
	実施された事業数	40事業(内、女性集団13事業)		
	サンプル集団20組織の女性参加割合の変化	46% 34%		
	政策形成のための開催協議数	1999~2000年に5回		
	地方行政による住民提案審査件数と認可数	審査数170件、認可数40件(43%認可)		
	県支援システム法制化	県知事令2本制定(1999、2000年各1本) (他に県条例1本、県知事令5本が制定予定)		

(出所)国際協力事業団社会開発協力部(2001)を基に筆者作成

ベルでは、村落住民の組織への加入により住民の所得向上の可能性が見込まれる一方で、特に貧困層の住民が受益すると保証された訳ではなく、そのため村落内での貧困層の割合が減少するかどうかは予測不可

能である。また、本案件実施の影響により、 村落内において資源配分の状況に変化が起こり、所得の格差が増大する等といった負の効果が現れる可能性も完全には否定できないように思われる。

第二に、国レベルでは、モデル普及事業によりタカラール・モデルが4モデル村落以外の多くの村に普及しなければ、本案件が国民全体の所得向上には寄与しないため、本案件におけるモデル普及の現状を考慮すると、インパクトは低いといわざるを得ない。

自立発展性

自立発展性をみる主な観点は、行政側および村落住民側の両者における、キャパシティー・ビルディングと貧困緩和に向けた意識の

変化とした。

第一に、パイロット事業における村民の意識の変化に関しては、案件実施中および終了後を通じた、村落組織と住民に対するエンパワメントの視点からの継続的な支援が一貫して重要であり、特に終了後における行政から現地NGOへの資源フローが開始または増加することで、NGOから住民への支援が継続されることが期待されるが、現時点では予測不可能である。

第二に、パイロット事業における行政側の 意識の変化に関して、タカラール県ではタカ ラール・モデルの採用が条例化され、案件実 施により確立された支援体制を保持している ことに行政側は積極的であるため、自立発展 性は高いように思われる。一方で、モデル 及事業においては、本来ならば案件終了後に おけるイ国中央政府から地方政府への継続的 な資源配分が望まれるが、タカラール県を除 いて実際にはモデル普及が実現しなかったた め、自立発展性は低いといえよう。

表8 各種調査の目的と調査結果の活用状況

年度	調査名	調査目的	活用状況
	社会経済	村落住民の生活水準と経済多様性メ	村落開発政策及び政策意思決定への適用
	統計分析	カニズムに影響を及ぼす要因析出	方法を検討(実際の状況は不明)
1007	ラボ村落選定	タカラール県内での4ラボ村落の選定	本調査結果に基づき、4ラボ村落を選定
1997	ラボ村落社会	4ラボ村落内での村民の生活環境の詳	事業における基本情報、事業評価の指標
年度	経済	細な把握	として活用(実際の状況は不明)
	州・県開発	村落開発関連行政機関どうしの関係、	プロジェクト実施のための基礎データと
	資源	協同組合等の存在の有無の明確化	して、インドネシアの行政機構の把握
	ラボ村落社会 経済調査()	4ラボ村落内の各種組織に関する情報 収集	プロジェクト実施の基礎資料として活用
	村落経済イン	村落住民の外部資源へのアクセスお	産業の特徴把握とプロジェクトへの活用
	タラクション	よび流通の把握	(実際の状況は不明)
1998	女性経済活動	4ラボ村落内で経済活動に携わる女性	4ラボ村落内の女性へのエンパワメント
年度	調査(WID)	の生活状況把握	事業形成への活用(実際の状況は不明)
	州・県別職員 研修ニーズ	(記載なし)	(記載なし)
	村落住民開発 ニーズ	4ラボ村落内の村民のニーズ把握 (PRA手法の使用)	村民の生活状況の把握、支援ターゲット特定とニーズ把握(実際の状況は不明)

(出所)国際協力事業団社会開発協力部(2000b)を基に筆者作成

4. 考察

以上の評価結果をふまえ、本案件の意義を 再確認する目的のもと、そのプロジェクト・ デザインに関して、モデルの有効性および何 故モデルが普及されなかったのかという観点 から、政策体系との関連における妥当性を考 察し、さらに本研究における二次評価手法の 可能性と限界について述べる。

(1) プロジェクト・デザインに関する考察

パイロット事業とタカラール・モデルの有効性 策定した政策体系において、パイロット事 業はタカラール・モデルを構築するための事 業と位置付けられ、その普及に関しては同モ デルの有効性が前提とされているが、同モデ ルの有効性自体はいくつかの想定に基づいて いる。

まず、モデル村落の選定方法に関してである が、村へのアクセスの良さ、潜在力の高い産業 の存在、人口流動性の低さ等の基準が設けら れていたため、これらの条件に合致しない村 においては同モデルの適用可能性が低くなる おそれがある。しかしながら、他村へのモデル 普及においてこのことは考慮されていない。

次に、パイロット事業では、ドナーから村 落住民への直接の支援は行わないという前提 のもと、ドナーによる村落の状況把握のため の各種調査に基づき、地方政府や現地NGO による住民支援の実施が含まれている。また、 同モデルにおけるNGOの役割は非常に重要 であり、欠かすことは実質上できない。その ため、他村でのモデル普及を念頭に考えた際 には、同モデルと同じ役割を果たし得る NGOが存在すること、さらにこのNGOと特 に管轄の行政側が対象村の社会・経済状況を 把握していることが想定されており、場合に よってはドナーが実施したのと同様の調査を 実施する必要があろう。

これらの想定が正しくなければ、それだけ で同モデルの有効性に問題が生じる可能性が ある。さらに、想定が正しい場合でも、同モ

デルの有効性を厳密に判断するためには、実 際には中・長期的な視点での検証が必要とな ろう。当該案件では、実施期間5年のうち、 最初の2年が準備、次の2年が事業実施、最後 の1年が終了に向けての準備期間と位置づけ られているが(国際協力事業団社会開発協力 部1998a)、モデル普及を念頭に置いた場合に は、早い時点でその有効性を見極めなければ ならないというジレンマが存在するため、最 初の2年の準備期間が短縮されるのが望まし かったように思われる。

モデル普及事業

同モデルがイ国における他の行政区へ導 入・普及されるためには、ドナーからイ国中 央政府、特に地方行政を管轄する内務省への 働きかけが必要であったように思われる。 同モデルが真に有効であるという前提に基づ くならば、イ国政府が主体となり、法的に義 務づけるという形でモデル普及事業を全国に 展開することが望ましく、そのためにはドナ ーからイ国政府への広報活動等の強い働きか けが必要であり、またドナー側としてもイ国 における法政面での基礎的環境整備への支援 等が可能であったように思われる。このこと は、図3において、ドナーからイ国政府への フロー (e) はプロジェクト範囲内の活動で あるのに対し、イ国政府から地方政府へのフ ローはプロジェクト外のものとなっているこ とからも分かる。

しかしながら、実際にはそのような活動は 実施されておらず、これが本案件におけるプ ロジェクト・デザインの「最大の課題」(滝 村2002)であったといえよう。

また、同モデルの有効性との関連において、 モデル普及の際には各村落において同モデル の内容を個別に修正して導入することが当然 期待される。

(2) 本手法の可能性と限界点

ODA事業の透明性向上 JICA事業を対象とした外部評価は、JICA

が費用負担を行なうJICA事業の一環として、 事業の透明性および評価の質と中立性の確保 を目的とし、これまで主に外部の機関や有識 者により実施されてきたが、その手法として、 既存資料を用いた文献調査とプロジェクト・ サイト訪問による現地調査を組み合わせて実 施するという形が一般的であった。これは、 評価主体者がそもそもプロジェクト実施に関 わっていないため、現地へ行き情報収集する ことで、既存の資料では不十分な情報量を補 うといった認識に基づいているためといえよ う。しかしながら、ODA事業に全く関係の ない外部者は、これまで内部に保存されてき た既存資料にさえアクセスしにくく、ある事 業に関して十分な情報が得るためには現地へ 行くことが必須であった。

行政事業の透明性という観点から市民による評価の意義と重要性が高まる中、本研究における「市民参加型」二次評価の手法は、行政側からの情報公開に基づき、市民が行政活動としてのJICA事業を低コストで評価できる新たな枠組みを提示しており、特に国際協力に関連しているNGOやNPOが積極的にこの手法を用いてJICA事業の二次評価を実施することで、日本のODA事業の透明性が大きく向上することが期待される。

その一方で、当初の(一次)評価者と二次 評価者が同じ問題意識や視点を共有しない場 合もあり、二次評価結果の内容が当初の(一 次)評価結果とは異なったものとなる可能性 も存在する。その際には、二次評価結果が時 として不合理な事業批判に偏ったものとなり 得ることも否定できない。ここでは、そうし た危険性を排除するためにも、二次評価者は その評価において信頼性や妥当性といった側 面を重視し説得力のある批判を行うことが求 められる。

情報収集に関する制約

本手法を用いた調査は国内で収集できる既 存資料のみに基づいているため、その内容は 収集した資料の量と質に大きく左右される。 行政側からの情報公開が適切に行われたと仮 定しても、例えば必要とする情報が文書化さ れていなかったり、もともと存在しなかった りする等の理由で情報の絶対量が不足してい る場合、あるいは資料内容の科学的客観性や 信頼性等、情報の質の面において問題がある と判断された場合は、本手法の有効性は薄れ てしまうであろう。実際に本案件においても、 一次資料の文献調査に基づきリスト・アップ した資料に関して、保管状況との関連からそ の全てが入手できず、二次資料の収集状況が 当初の期待に沿わなかった。さらに、パイロ ット事業の結果 ~ それぞれにおける投入 の区別が不可能であったことや、4モデル村 落における住民所得のデータが欠如していた ことによる評価実施への影響(特に効率性) さらに案件終了時に4モデル村落において村 落開発ニーズ調査と同様の調査が実施されな かったため、対象村落内の変化が明確にみえ ない等の問題点があった。

これらの問題点を解決するためには、プロジェクト・デザインを適切に行うこともきった。派遣専門家など関係者が事業実施において得られた知見等を適宜適切に記録することが求められる。また、それらの情報の再利用を図るためには、行政側が既存のの両利用を図るためには、行政側が既存のの情報開示請求に対し迅速に対応できるような形で、あるいは市民が随時容易にアクセス可能な形で、資料管理の体制が整備されること

が期待される。

他方、報告書等がもともと存在せず情報の 絶対量が少ない場合には、それらを基に実施 された当初の(一次)評価結果の信頼性にも 問題が生じ得るであろうし、また既存のプロ ジェクト評価手法を用いて適切な評価を実施 するために、どれだけの情報量が必要かとい うことに関する共通認識や基準等もこれまで なかったように思われる。今後、さまざまな 形での二次評価が実施され、多くの事例が報 告されることで、これらの事項に関する議論 がより活発になることが望まれる。

5. 結論

既存評価資料等を活用した二次評価は、他 の公益事業においても有効であり、また以下 に示すような多くの潜在的機能を有してい る。まず、事業主である行政側に対して、納 税者としての市民への情報開示の責任を明確 にし、行政事業に関する説明責任の遂行を促 す。また、サービスの受益者である市民に対 して、これまで基本的に受け身の姿勢で関わ ってきた行政活動に積極的に参画する一つの 手段を提示することで、行政事業の顧客とし ての市民の存在感が増すようになる。さらに、 二次評価の実施により、外部者としての市民 の公正かつ厳格な目が行政事業に向けられ、 またその評価結果の公表を通じて、行政の政 策形成における市民の発言権が拡大され得 る。これらの相乗効果により、行政事業の透 明性の向上と効率化がもたらされ、健全な行 政活動の発展につながることが期待される。

こうした二次評価の概念は「評価」の歴史 をみても比較的新しいものであり、これまで 実施機関がほとんど独占的に実施してきた 「評価」に対する認識を変え、市民の手に委 ねることで「評価」のもつ可能性が広がると いう意味においても意義が大きいといえる。 以上の理由から、特にNGOやNPOにおいて 公益事業を対象とした二次評価が積極的に行 われ、多くの実践例が報告されることが強く

望まれる。さらに今後、二次評価の分野にお いてさまざまな手法が開発され、それと同時 に評価結果公表のための場や機会がこれまで 以上に増えることで二次評価の経験が蓄積さ れ、その結果として日本における「評価」へ の人々の関心が高まり、「評価」文化の形成 および育成につながることが期待される。

謝辞

本稿は、東京大学大学院新領域創成科学研 究科での講義「国際開発協力論 」内の演習 として、他のメンバー(高田友美氏、野村綾 子氏、李京南氏)と共に実施したグループ調 査によるところが大きい。調査を進めるにあ たり、国際協力事業団社会開発協力部社会開 発協力第一課の長谷川敏久氏には資料提供に 関する便宜を図って頂いた。また、本稿の査 読者からは有益なご指摘を頂いた。ここに記 して深く感謝いたします。

注記

- 1 「評価みえ」ホームページ http://www. hyouka.org
- 2 http://www.jica.go.jp/library/shozo.html (2003年3月17日アクセス)
- 3 国立国会図書館オンライン蔵書目録Web-OPACにおいて、著者名に「国際協力事 業団」を入力し検索した結果より(2003 年3月17日アクセス)。国会図書館ホーム ページhttp://www.ndl.go.jp
- 4 JICAの技術協力プロジェクトにおいては 通常、事前調査、事前打合せ、巡回指導、 実施協議、終了時評価の計5冊の報告書が ある。
- 5 詳細に関しては国際協力事業団(2000) を参照されたい。
- 6 ドナーがイ国政府へモデル導入・普及の ための支援を行う際には、本来ならばま ずタカラール・モデルと、戦後の日本に おいて採用された農村開発のモデルや JICAがこれまでに実施した事業において

構築されたモデル等を比較検討し、その中から最適なモデルが選択されるべきであるう。

参考文献

- アーバン・インスティテゥート (2002) 『資料 政策評価機関と政策評価プロジェクト』
- 柿沼潤(2000)『インドネシア・スラウェシ 貧困対策支援村落開発計画WID分野短期派 遣専門家業務報告書』
- 国際協力事業団 (1999)『インドネシア国別 援助研究会報告書(第3次)』
 - (2000)『第4次インドネシア国別援助研 究会報告書』
 - (2002)『国際協力事業団年報2002』国際協力出版会
- 国際協力事業団企画部(1997)『国別WID情報整備調査 インドネシア 』
- 国際協力事業団国際協力総合研修所(1997) 『インドネシア・スラウェシ貧困対策支援 村落開発プロジェクトと連携した地域社会 開発手法の研究報告書』
- 国際協力事業団社会開発協力部(1996a)『インドネシア共和国スラウェシ貧困対策支援 村落開発計画事前調査団報告書』
 - (1996b)『インドネシア共和国スラウェシ貧困対策支援村落開発計画実施協議調査団報告書(長期調査員報告書)』
 - (1998a)『インドネシア共和国スラウェシ貧困対策支援村落開発計画事前打ち合わせ調査団報告書』
 - (1998b)『インドネシア・スラウェシ貧困対策支援村落開発計画平成9年度リーダー会議個別打ち合わせ資料No.2』
 - (1998c)『社会経済調査()報告書概要版』
 - (2000a)『インドネシア共和国スラウェシ貧困対策支援村落開発計画巡回指導調査 団報告書』
 - (2000b)『インドネシア共和国スラウェシ貧困対策支援村落開発計画98年度研究会

報告書』

- (2001)『インドネシア共和国スラウェシ貧困対策支援村落開発計画終了時評価報告書』
- 滝村卓司(2002)「第1章 地域社会開発とソーシャル・キャピタル 1.地域社会開発における開発課題とソーシャル・キャピタル 国際協力事業団
- 国際協力総合研修所『ソーシャルキャピタル と国際協力[事例分析編]』、17-30
- 田中由美子(2002)「ジェンダー評価」、田中 由美子・大沢真理・伊藤るり編著『開発と ジェンダー』、国際協力出版会、59-71
- 原口孝子(1999)『短期専門家活動報告書』
- 福島州馬(1999)『ニーズ調査報告書』
- 三好皓一(2002)「プログラム・セオリー・マトリックスの活用について 評価における分析の有用性を高めるために 」、『日本評価研究』2(1):11-27
- 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):65-79
- 三好皓一・平田慈花・和田知代・中澤哉・喜 多悦子(2003)「プログラム評価の試み フィリピン感染症対策分野の評価を事例と して、『日本評価研究』3(1):43-56
- 龍谷大学(1999)『インドネシア・スラウェ シ貧困対策支援村落開発計画プロジェクト と連携した社会統計分析(平成10年度国際 協力事業団委託業務完了報告書)』
- 渡邊敦(1999)『南スラウェシ村落経済支援 構想のために ウサハの創意と多様性から の出発 (村落経済インタラクション調査 最終報告書)』
- The World Bank (2001), Integrating a Gender Dimension into Monitoring & Evaluation of Rural Development Projects. Washington, D.C.: The World Bank.

(2003.7.2受理)

A Study on Participatory Secondary Evaluation for the Public: The Case of JICA's Evaluation on "Poverty Reduction and Rural Development Project in Sulawesi, Indonesia"

Satoshi Morita

Graduate School of Agricultural and Life Sciences, the University of Tokyo aa07181@mail.ecc.u-tokyo.ac.jp

Kana Takamatsu

Graduate School of Frontier Sciences, the University of Tokyo

Koichi Miyoshi Ritsumeikan Asia Pacific University

Abstract

This article proposes a new method of secondary evaluation as a means for the public to participate in governmental activities, using a JICA project as a case study. In this method, various procedures are followed, such as: 1) collecting, using the internet and libraries, primary data mainly composed of evaluation reports; 2) collecting secondary data based on literature review and the law concerning access to information held by administrative organs; 3) formulating the Program Theory Matrix; and 4) conducting secondary evaluation using the DAC five criteria.

This method makes it possible for the public to carry out evaluation on governmental activities from their own standpoint and viewpoint, using existing materials at low costs without implementing research at the project site. Thus, secondary evaluation enhances the contents of primary evaluation.

The concept of secondary evaluation is relatively new in the history of evaluation, and is meaningful in the sense that the potential of evaluation is widened. Through non-profit organizations actively conducting secondary evaluations on activities of public utilities, public's voice will be raised in formulating policies of the government and transparency and efficiency of governmental activities will be increased.

Keywords

secondary evaluation, public's participation, PTM (Program Theory Matrix), disclosure of information, DAC Five Criteria

【研究論文】

Incidences of Public Education Spending in Developing Countries

Takako Yuki

Research Center for Advanced Science and Technology, University of Tokyo yuki@ip.rcast.u-tokyo.ac.jp

Abstract

This paper quantitatively analyzes the incidence of public education spending on the poor in developing countries by using the results of standard benefit-incidence studies and additional national data. Although there is considerable variation across studies, it is found that public education spending generally does not favor the poor, but it does favor them at a lower level of education. The poorest quintile (20% of the population), on average, receives 16.3% of total public education spending while the richest quintile receives 25.9%. In primary education, the poorest quintile receives 22.4% of public spending but only 5.5% in higher education. The cross-country analysis of variations in the incidence of public education spending implies that increased spending on education will be associated with increased share for the poor given that it is not devoted to spending on higher education.

Keywords

benfit-incidence analysis, meta-analysis, education, public expenditure, developing countries

1. Introduction

Many people would agree that education is one of the most important basic human rights and is central to equitable growth and poverty reduction. Equity, as well as efficiency, is an important reason for the government financing and provision of education services. The recent movement toward a more comprehensive and prioritized development strategy and the associated donors' assistance, for example a poverty reduction strategy, is expected to

strengthen macro-level policy and financial commitments to educational goals in developing countries. Meanwhile, it will be further demanded by wider stakeholders to articulate and evaluate the rationale for public funds allocated to education. Certainly, it is important to articulate equity in education financing.

For measuring equity in education, four distinguished approaches can be seen in the literature. To be brief, studies aim to analyze differences in the following: (i) access to education and the quality of access; (ii) the

accumulation of human capital (e.g., educational attainment and return to education) as gains associated with access to education; (iii) the appropriation of public resources, in other words public subsidies associated with access to education; or (iv) private contributions such as direct costs to gain access and tax contributions. In any approach, decisions on equity would be based on norms and facts concerning distribution across various units (e.g., individuals and households) and/or the groups (e.g., by gender and income) for a period (e.g., current or lifetime differences) in a society.

The third approach, occasionally combined with the fourth approach, is called "benefitincidence analysis." Such analysis in the education sector began with a study of higher education financing in the US (Hansen and Wesbrod 1969), and for developing countries, Jallade (1974). Jallade's is the first study to analyze the benefit-incidence in detail, followed by Meerman (1979) and Selowsky (1979). Recently, a fairly simple standard benefitincidence analysis has become more available for developing countries. The analysis is expected to provide a measure for the effectiveness of public spending regarding the transfer of current benefits to the poor, as defined mostly by household consumption or income.

This paper aims to quantitatively analyze the incidence of public education spending on the poor in developing countries, using the results of 43 standard benefit-incidence studies from 34 countries. More specifically, it will assess whether and to what extent public education spending is effectively distributed to the poor and how the degree of equity in distribution varies across studies. Furthermore, with additional national data, it will consider what may explain the variation in the studies, with a focus on interand intra-sectoral allocation policies, and whether increased public education spending may increase the share for the poor. This cross-

country analysis may help policy makers establish a benchmark from which to evaluate the current state and future impact of policy changes with respect to targeting public spending on the poor.

This paper is organized as follows. Chapter 2 elaborates a little further the methodology of a standard benefit-incidence analysis and describes the previous studies reviewed in this paper. Chapter 3 presents the results of the review and Chapter 4 concludes the paper.

2. Measuring the benefit-incidence and the studies used in this review

2.1 Measuring the benefit-incidence and the interpretation for policy implications

Broadly speaking, the empirical benefit-incidence literature may be categorized into two types. *Firstly*, a standard benefit-incidence analysis, which is widely used, combines the cost of providing public services with information on their use in order to generate distributions of the benefit of government spending.² This approach is also called expenditure-incidence analysis. *Secondly*, a fiscal- or budget-incidence analysis sums both the tax- and benefit-incidence estimates [e.g., Jallade (1974); Fields (1975)]. This approach is appropriate particularly in a society that has an education earmarked tax system or/and an income tax system strongly linked to workers' education.³

Recently, the standard benefit-incidence analysis has become more available for developing countries. The analysis is expected to provide a measure for the effectiveness of public spending regarding the transfer of current benefits to the poor. Although norms on equity between the poor and non-poor vary across countries and societies, studies tend to analyze resource distributions across groups classified by

a welfare indicator related to household income and consumption levels, such as consumption deciles and quintiles.4 Representative formats that present the educational subsidy imputed to each of the classified groups include: subsidy per capita terms, as a share in the total subsidy (to be compared to share of population or share of school-aged children), and as a percentage of household consumption/income.5 For example, Castro-Leal et al. (1999) found in their review of studies in nine African countries that the poorest 20 percent of the population (quintile) gains less than 20 percent of the subsidy, and significantly less in most cases. In addition, the studies often address the different distributional role of public education spending between a lower and higher level of education. Despite some differences in measurements and decision criteria on equity, as suggested in Selden and Wasylenko (1992), the studies generally confirm the arguments of Psacharopoulos (1977) that expenditure on primary education is pro-poor, either as a percentage of household income or in absolute terms, while expenditure on higher education levels tends to be pro-rich, but the exact point in the education system at which the rich become favored differs among countries.

The benefit-incidence results typically support the following policies for improving the equity: (i) increasing public resources to education while aligning sub-sectoral budget allocations, often towards primary education; and (ii) introducing cost-recovery and/or private provision at a higher level of education. In other words, benefit-incidence studies try to foresee what would happen on a distribution pattern of public spending if governments change the level and pattern of budget allocations to the education sector.

Broadly speaking, the following approaches may be taken to provide some thoughts or estimates of changes in the benefit-incidence over time. The first approach is to

simply take an existing distribution pattern of public spending (or pattern of use of subsidized services) as given and generate a simple analysis by assuming that the pattern does not change significantly and that the observed-incidence of current spending would hold also for any additional spending (Demery 2000, p.20). The second is to examine changes in the benefitincidences for two or more points in time within a country that might have arisen from changes in public spending [e.g., van de Walle (1992) for Indonesia, World Bank (1994a) for Columbia]. The third is to estimate "marginal" gains (gains from a change) that may be distributed differently from the average incidence in a group using cross-sectional data in a country [e.g., Lanjouw and Ravallion (1999)6]. In addition, a crosscountry analysis of benefit-incidence results and the variation, which is the aim of this paper, may be helpful to draw policy insights for improving the targeting public spending to the poor.

2.2 Description of studies used in this paper and framework for comparison

I prepared a dataset that includes empirical results with respect to estimated benefit-incidences across income (consumption) quintiles measured as the quintile's share of the total subsidies for the education sector as a whole and for each level of education. A standard formula can be written as follows:

$$X_{j} = \sum_{i=1}^{n} \frac{S_{i}}{E_{i}} E_{ij} / \sum_{i=1}^{n} S_{i} *100$$
 (1)

where X_j is the benefit-incidence (incidence of public spending) for the j^{th} quintile, S_i is total public spending on the i^{th} level (often primary, secondary, or higher education, n=3), E_i is enrollment at the i^{th} level of education and E_{ij} is enrollment at the i^{th} level of education among the j^{th} quintile. The sum of the benefit for the j^{th} quintile at all levels of education divided by the

total public education spending provides the benefit-incidence for the jth quintile for education as a whole. The benefit-incidence ranges from 0 to 100 percent.

In finding studies, I specifically rely on the two previous reviews, Selden and Wasylekno (1992) and Castrol-Leal, et. al (1999), and the World Development Source.7 Most studies use data from the 1990s and a few use data from the 1970s and 1980s. I tried to incorporate information on how studies deal with measurement issues that analysts could face in estimating benefits and imputing the benefits to beneficiaries. These issues are related to the following questions: whether unit subsidy has regional/geographical variation at a level of education or variation by program (e.g., vocational vs. general education); whether revenues from cost-recovery are netted out of public unit subsidies; how investment expenditure is included as part of unit subsidy or excluded; and how benefit for those enrolled in private provision of education is estimated (e.g., no benefit as there is no government subsidy to private school).

The issues involved in imputing benefits to beneficiaries include: whether the unit of analysis is household or individual; whether the analysis includes all population in grouping to quintiles or disaggregated population (e.g., by gender, urbanrural area, and region); and whether a classifier for grouping to quintiles is per capita household consumption (expenditure/income), total household consumption, or per adult equivalent consumption. While some studies provide comprehensive answers to these issues, most do not provide adequate information. This may be partly due to a lack of data and limited time for the studies.

Table 1 provides selective information on the dataset based on 46 studies in 34 developing countries. When a study provides a few distinguished sets of results through addressing measurement issues, the dataset includes the results separately. The table provides information on the country and year of data, the mythological type, the unit of analysis (whether it is individual or household), and the incidence of public education spending for the poorest quintile for education as a whole and by level of education. The sample countries have relatively regional variations, nine countries in Africa, two countries in Central Asia, six in East Europe, eight in East Asia and the Pacific, seven in Latin America, and two in the Middle East and North Africa. Overall. Table 1 shows that the distribution of public education spending is less likely to be targeted to the poor. In many cases in the table, the poorest quintile does not gain the larger share in the spending than their share in the total population (of households or individuals), i.e., 20 percent. Yet, the table also shows that the degree of nonpoor varies among studies.

This paper will further compare their results through quantitatively analyzing: (i) variation across studies and (ii) factors that may explain the variation in consideration of typical policy recommendations in the previous studies. To analyze these factors, this paper will focus on the relationship of the incidence of spending on the poor to the level of public education spending and a sub-sectoral allocation of the spending:

$$X_i = {}_0 + {}_1SPEND + {}_2SHAREH + Z + \mu_i$$
 (2)

where SPEND is public education spending as a percent of Gross National Income (GNI) and SHAREH is higher education share of total education recurrent spending. Z denotes indicators as proxy for other country's socioeconomic factors such as per capita GDP. I use SHAREH as a most available proxy of subsectoral allocations.⁹ The national aggregated data for the independent variables are added to the dataset corresponding to the country and year for each of the studies surveyed. The data mainly comes from the World Bank (2001) and

UNESCO (2002).10

3. The results of standard benefit-incidence studies in developing countries

3.1 The incidence of public education spending across the population

Table 2 summarizes a meta-analysis of the incidences of public spending across quintiles for education as a whole and for each level of education (primary, secondary and higher education). It shows the means of the incidence, the standard deviations, the minimum and maximum incidences, and the number of cases used for this analysis together with the numbers of the corresponding countries and studies. The table confirms that the distribution pattern of public education spending is less likely to be propoor, as a poorer quintile tends to receive a lower share of spending. On average, the poorest quintile receives 16.3% of the total education spending (with a standard deviation of 5.9 percentage points) and the richest quintile receives 25.9% (with a standard deviation of 9.3 percentage points).

Table 2 also confirms that the distribution pattern is greatly different by level of education. Spending on primary education is likely to be pro-poor, while spending on secondary and higher education is likely to be pro-rich. In primary education, the quintile's share is above 20% in the bottom three quintiles and it becomes below 20% in the top two quintiles. The richest quintile receives the least share, 14.8% of public spending on primary education with a standard deviation of 4.7 percentage points. On the other hand, the distribution pattern of higher education spending is largely skewed towards the rich. The poorest quintile receives only 5.5% of spending on higher education (with a standard deviation of 4.6 percentage points) and even the bottom two

quintiles (i.e., the poor 40%) account for only 15%. More than a half of the spending goes to the richest quintile. In secondary education, although the distribution pattern is not pro-poor, the distribution pattern across quintiles is quite different from higher education. While the poorest quintile receives only 11.9% of the spending (with a standard deviation of 7.4 percentage points), wide variance exists and the second and third bottom quintiles receive relatively fare shares, 18.5% and 19.3%, respectively.

Table 2 also suggests that the unit of analysis would affect the results. The magnitude of inequality becomes slightly smaller for studies that analyzed the distribution across household quintiles rather than individual quintiles. For example, the mean of the poorest household quintile's share in total spending is 18.2% and the richest is 23.5%. On the other hand, the poorest individual quintile's share is 15.2% and the richest is 27.3%. This is partially because poor households tend to have more household members (especially younger ones) than the rich.

3.2 Cross-country analysis of factors that may explain the incidence for the poor

Although the distribution of public spending on education as a whole is likely to be non-poor, it appears that wide variance exists among the sample studies with respect to the degree of inequality. In some countries public education spending is strongly non-poor (e.g., 5% of the poor quintile's share in Guinea, 1994), while most countries have a modestly non-poor or neutral distribution pattern and the pattern is even pro-poor in a few countries (e.g., 30% in South Africa, 1993). Although measurement issues, e.g., unit of analysis, might explain some of the differences, other factors could be further considered. As equation (2) illustrates, the level of public education spending and sub-sectoral

Table 1 Poorest Quintile's Share in Public Education Spending in Selective Studies

rable i Poc			Snare					n Selective Studies
Country	Year	Method	Unit -			e's share in* lecondary 1		- Source
Albania	of 1996	id 1	I		27.0	7.2		World Bank (2000b)
Armenia	1996	1	H	7.0				World Bank (2000a)
Bangladesh	1994	1	Н	13.8	19.4	6.0		World Bank (1996a)
Bulgaria	1995	1			17.7	8.1		Pradhan and Zant (1998)
Bulgaria	1997	1	•••	20.0	21.1	15.9		World Bank (1999a)
Colombia	1974	1	 H	19.8	32.1	16.8		Selowsky (1979)
Colombia	1992	1	Н	23.0	39.0	21.0		World Bank (1994)
Cote d'Ivoire	1995	1	I	14.0	19.0	5.2		Demery (2000)
Cote d'Ivoire	1995	2	H	19.4	28.8	11.2		ditto
Cote d'Ivoire	1995	3 M	I	16.2	22.8	6.8		ditto
Cote d'Ivoire	1995	4 F	I	8.9	14.0	4.2		ditto
Dominican Rep.	1996	1		22.7	28.8	17.6		World Bank (2000e)
Ecuador	1998	1	I	11.0	24.0	9.0		World Bank (2000d)
Ghana	1989	1	I	17.1	21.2	16.8		Demery, et.al. (1995)
Ghana	1989	2 M	I	17.1	22.3	18.7		ditto
Ghana	1989	3 F	I	16.8	19.8	13.8		ditto
Ghana	1992	1	I	16.4	21.8	14.9		ditto
Ghana	1992	2 M	I	17.7	22.7	17.0		ditto
Ghana	1992	3 F	I	14.6	20.7	11.8		ditto
Ghana	1992	4 C	I	16.7	17.4	18.6		ditto
Ghana	1998	1	I		20.4	18.8		Canagarajah and Ye (2001)
Guinea	1994	1		5.0	11.0	4.0		Castro-Leal, et. al. (1999)
India	1994	1	I		14.2			Lanjouw & Ravallion (1999)
Indonesia	1987	1	I	21.9	24.5	21.4		Van de Walle (1992)
Indonesia	1987	2 U	I	23.4	25.0	28.8		ditto
Indonesia	1987	3 R	I	22.9	23.7	23.3		ditto
Indonesia	1989	1	I	15.0	22.0	5.0		World Bank (1993)
Indonesia	1998	1	I		24.8	12.7		Lanjow, Peter et. al (2001)
Jamaica	1992	1	H	18.0	24.0			World Bank (2000a)
Jordan	1996	1						Hammer (1999)
Kazakhstan	1996	1	Н	8.0				World Bank (2000a)
Kenya	1992	1	I	17.0	21.8	6.4		World Bank (1995a)
Kenya	1992	2 R	Ī		20.8	6.4		ditto
Kenya	1992	3 U	Ì		28.1	10.6		ditto
Kyrgyz	1993	1	H	14.0				World Bank (2000a)
Lao	1993	1	I	11.6	18.4	6.6	0.0	World Bank (1995b)
Macedonia, FYR	1996	1	Ī	18.5	25.4	12.6		World Bank (1999b)
Madagascar	1994	1	Ī	8.3	17.2	2.0		World Bank (1996b)
Madagascar	1994	2 V	Ī	8.2	16.8	1.9	0.0	ditto
Madagascar	1994	3 MV		8.0	17.0	3.0	0.0	ditto
Madagascar	1994	4 FV	Ī	8.0	16.0	1.0		ditto
Madagascar	1994	5 UV	Ī	9.2	21.5	5.9	0.0	ditto
Madagascar	1994	6 RV	Ì	11.5	16.5	1.9		ditto
Malawi	1991	1	Ī	10.0	15.0	7.0	3.0	Castro-Leal (1996a)
Malawi	1995	1	Ī	16.0	20.0	9.0		Castro-Leal (1996a)
Malaysia	1974	1	Н	21.9	28.2	19.0		Meerman (1979)
Malaysia	1989	1	Н	26.0	35.0	31.0	10.0	Hammer, et.al. (1995)
Moldova	1999	1	Н	26.0				mil 1 (0000)
Moldova	1999	2 U	Н	24.3				ditto
Moldova	1999	3 R	Н	28.2				ditto
Morocco	1999	1	I	12.0	21.1	7.3		World Bank (2001a)
Nepal	1999	1	Н	11.0				World Bank (2000a)
Nicaragua	1990	1	Н	9.0	•••	•••		World Bank (2000a)
Pakistan	1993	1	Н	14.0	•••	•••		
Panama	1991	1	Н	20.0	•••	•••		*** 11D 1 (0000)
Peru	1997	1	н	15.0	•••	•••	•••	World Bank (2000a)
Peru	1994	1	П	20.6	28.4	15.5	6.2	World Bank (2001b)
ı cı u	1 77 /	1	1	∠0.0	∠0.4	13.3	0.2	world Dalik (20010)

Table 1 (Continued)											
Romania	1994	1		I	24.0	27.0	18.3	7.8 World Bank (1997)			
Romania	1997	1		I	24.0	31.0	26.0	10.0 Tesliuc, et. al (2000)			
South Africa	1993	1		Η	30.0			Castro-Leal (1996b)			
South Africa	1993	2	V	Η	21.0	26.9	18.3	10.6 Castro-Leal (1996b)			
South Africa	1993	3		I	19.9	25.8	18.8	6.1 Castro-Leal (1996b)			
South Africa	1993	4	V	I	14.1	18.9	11.5	6.1 Castro-Leal (1996b)			
Tanzania	1994	1		Η	13.0	20.0	7.6	0.0 World Bank (1999c)			
Uganda	1992	1			13.0	19.0	4.0	6.0 Castro-Leal, et. al. (1999)			
Vietnam	1993	1		I	12.0	19.9	4.7	0.0 World Bank (2000c)			
Vietnam	1998	1		I		26.1	8.6	0.6 World Bank (2000c)			

Notes: Method-id: 1, Method most closes to standard among available results in a study; Other numbers are explained by the following: M, the sample includes males only; F, females, U, urban population; R, rural population; C, classifier is per adult equivalent consumption; V, variation in unit spending (e.g. variation by region).

Table 2 Summar Statistics of Incidence of Public Education Sendin Across Quintiles

	Maan	Standard	Minimum	Maximum	No. of	No. of	No. of
	Mean	deviation	Minimum	Maximum	obervations	studies	countries
Education as a whole							
Quintile 1	16.3	5.9	5.0	30.0	58	38	31
Quintile 2	18.8	3.4	12.0	28.5	55	35	28
Quintile 3	18.8	2.6	13.9	23.0	55	35	28
Quintile 4	20.3	2.8	15.0	27.0	55	35	28
Quintile 5	25.9	9.3	11.2	46.0	58	38	31
Primary							
Quintile 1	22.4	5.4	11.0	39.0	53	34	24
Quintile 2	22.8	2.4	17.7	28.0	50	31	21
Quintile 3	21.0	1.7	16.5	25.4	50	31	21
Quintile 4	19.1	2.8	11.0	24.2	50	31	21
Quintile 5	14.8	4.7	4.0	27.8	53	34	24
Secondary							
Quintile 1	11.9	7.4	1.0	31.0	52	33	23
Quintile 2	18.5	5.8	8.2	36.7	49	30	20
Quintile 3	19.3	4.2	10.6	26.6	49	30	20
Quintile 4	22.6	5.2	11.5	33.6	49	30	20
Quintile 5	28.1	11.5	10.0	50.3	52	33	23
Higher							
Quintile 1	5.5	4.6	0.0	17.4	46	33	24
Quintile 2	9.1	6.5	0.0	27.1	42	29	20
Quintile 3	14.3	7.7	0.0	32.4	42	29	20
Quintile 4	20.2	6.6	7.0	33.0	42	29	20
Quintile 5	51.9	20.3	21.0	100.0	45	32	23
Education as a whole by	y unit anal	ysis					
Unit analysis is individ	dual						
Quintile 1	15.2	4.9	8.0	24.0	33	18	14
Quintile 5	27.3	8.9	14.0	43.0	33	18	14
Unit analysis is housel	nold						
Quintile 1	18.2	6.7	7.0	30.0	21	18	16
Quintile 5	23.5	9.4	11.2	46.0	21	18	16
Source: Author's calcul-	ation main	r the data fro		anamaa din Te	1.1. 1		

Source: Author's calculation using the data from source presented in Table 1.

Unit: H, Household is unit of analysis; I, Individual is unit of analysis. (...) Information is not available.

^{*} For some countries, primary education refers to basic education. For Indonesia and Vietnam, the data for secondary are the arithmetic averages of junior and secondary education.

allocation and the other country's socioeconomic factors might explain differences in the degree of inequality or equality.

To assess the extent of their relations, I regress the incidence of public education spending for the poorest quintile on the variables: SPEND and SHAREH with some specifications including the other two variables: a constant GDP per capita in US dollars (GDPPC) and the adult illiteracy rate as a percentage of people aged 15 and above (ILLITERATE). The regression results are reported in Table 3. The results are derived through using the Huber-White robust standard error method. To assess heteroskedasticity, I performed the Cook-Weisberg test. The test result suggests that there is no evidence of heteroskedasticity.

The specification (1) of the table shows that the coefficient upon SPEND is positive (1.563) and significant at the 1% level. A one-percentage point increase in SPEND increases the incidence among the poorest quintile by about 1.6 percentage points. However, not only the level of spending but also the sub-sectoral allocation

appears to be important as expected (specification 2). The coefficient upon SHAREH is negative (-0.238) and significant at the 5% level¹¹ while the coefficient upon SPEND remains positive and significant at the 1% level.¹²

The specifications (3) and (4) show the results become robust after controlling for GDP per capita¹³ and illiteracy rate. The coefficients upon SPEND and SHAREH remain positive and negative, respectively, and both statistically significant. This result does not change even if either of these two variables is dropped from regression (specifications 5 and 6), suggesting that multicollinearity between the two is not a serious problem. In the specification (3), GDP per capita is positively associated with the incidence for the poorest (0.002) at the 5% level of significance. In the specification (4), the coefficient upon GDPPC remains positive but not significant while the coefficient upon ILLITERATE is negative (-0.102) and significant at the 5% level.14

Table 3 also provides regression results for a reduced size of sample studies by keeping only

Table 3 Ordinary Least Square Regressions, with Robust Standard Errors, a Dependent Variable is the Poorest Quintile's Share

				Specif	<u>fication</u>			_
	(1)	(2)	(3)	(4)	(5)	(6)	$(3)^{b/}$	$(4)^{b/}$
Constant	9.292 ***	13.684 ***	13.294 ***	20.305 ***	26.694 ***	12.683 ***	* 11.766 **	* 17.273 ***
	(6.850)	(5.130)	(5.600)	(4.490)	(7.210)	(4.820)	(4.620)	(3.390)
SPEND c/	1.563 ***	1.465 ***	1.174 ***	0.920 ***		1.098 **	1.128 **	* 1.089 ***
	(6.340)	(5.650)	(4.220)	(3.090)		(4.420)	(3.050)	(3.670)
SHAREH ^c	/	-0.238 **	-0.256 ***	-0.354 ***	-0.409 ***		-0.235 *	* -0.338 **
		(-2.240)	(-2.850)	(-3.350)	(-3.400)		(-2.100)	(-2.670)
GDPPC			0.002 **	0.001	0.001	0.001	* 0.003 ***	* 0.003 ***
			(2.480)	(1.340)	(1.320)	(1.730)	(7.110)	(4.320)
ILLITER				-0.102 **	-0.140 ***	-0.081 **	k	-0.075
				(-2.290)	(-2.940)	(-2.080)		(-1.470)
N	55	50	50	46	49	51	32	28
R^2	0.340	0.401	0.493	0.610	0.447	0.5158	0.540	0.655

Notes: *** Statistically significant at the 1% level, ** Significant at the 5% level, * Significant at the 10% level. a/ Huber-White robust t-statistics in parentheses. b/ For a reduced sample size by keeping only one result for each study if a study has multiple results with different methods. c/ The pairwise correlation coefficient between SPEND and SHAREH for the full sample is -0.24.

one set for each study in order to eliminate a possible bias from country- and year-specific factors of studies with multiple results. Overall, the regression results for the reduced sample are consistent with the findings for the full sample. In all specifications, the coefficient upon SPEND is positive and significant at the 1% level. The coefficient upon SHAREH is negative in all specifications, but the significance is less when compared to the regression results for the full sample. Specification (4) shows the coefficient upon SPEND is 1.089 at the 1% of significance level and the coefficient upon SHAREH is -0.338 at the 5% of significance level.

The results for both full and reduced samples suggest that the incidence for the poor would be higher when a country allocates more public resources to education but the incidence would not be much higher when a country allocates a larger share of the resources to higher education. For example, even if a country has a higher level of public education spending by 1 percentage point of GNI than the other country with the same economic and literacy levels, but the country also has a higher share of higher education by 6.2 percentage points, the difference in the incidence for the poor would be almost zero. That is no advantage in equity despite the high financial commitment to education as a whole.

In other words, the results imply that reallocating public education resources from higher education to other levels of education might be a policy measure to improve equity when GDP per capita and the level of public education spending are held constant. For countries where total education spending is rather small, it might be more appropriate to consider an increase in the overall spending while providing the highest priority to lower levels of education. This would result in a relatively smaller share of higher education without a reduction in the existing level of spending on higher education. ¹⁵

3.3 Changes in the incidence for the poor over time in a country

Obviously, the cross-country analysis above does not observe various factors that might also explain the differences in the benefitincidence for the poor. Some factors might be related to country-specific socioeconomic contexts, while others might be related to education sector policies (e.g., cost-recovery policy) and beneficiaries' response. It may be useful to compare changes over time in a country for which the results are available for multiple years, in order to ensure reliability of the crosscountry regression analysis and get more clues on other factors that might explain a higher incidence for the poor. Thus, I tried to examine whether changes in the benefit-incidence favor the poorest quintile (poor 20%) by focusing on ten cases of medium- or long-term changes in six countries among the studies presented in Table 1. When the poor increases their share in spending over time, the change is considered to be propoor (if not, non-poor). Although the data are very limited, it is found that variations still exist across cases and there are a few more cases of pro-poor changes. As the information is not sufficient for any systematic review (e.g., panel regressions), two country cases are explained descriptively below.

The case of Malawi shows a strongly propoor change, which is an increase of 6 percentage points in the poor's share in public spending on education as a whole between 1990/1991 (10%) and 1994/1995 (16%). In particular, the improvement was seen in primary education. In Malawi, total public spending on education increased from 3.2% of GNI in 1991 to 5.4% in 1995 and the share of higher education declined slightly from 22.3% to 20.5%. And primary education increased the share in total recurrent education expenditures from 46.1% to 58.8%. The government also abolished primary school

fees during the period (Castro-Leal, 1996). Households responded positively to the policy changes, i.e., increased public resources and decreased private cost in primary education. The gross enrollment ratio jumped by 44.6 percentage points.

The case of Ghana shows mixed changes over 10 years. The change in the benefit-incidence from 1989 to 1992 was modestly non-poor for education as a whole (-0.7 percentage point from 17.1% to 16.4%), reflecting non-poor changes in secondary and higher education despite a pro-poor change in primary education. However, the changes became strongly pro-poor in secondary and higher education from 1992 to 1998 and modestly non-poor in primary education.

In Ghana, total public spending on education increased from 3.5% of GNI in 1989 to 4.5% in 1992, but slightly decreased to 4.2% in 1998, which is still higher than the level in 1989. However, the share of primary education in the total education budget declined from 45% in 1990 to 41% in 1998, while the shares of secondary education (junior and senior secondary) and university remained stable and the share of teacher training college increased (Canagarajah and Ye, 2001). Although primary education is free, in practice, all schools collect obligatory contributions from students. Between 1992 and 1998, the share of private enrollment in primary school remained at the same level, while private secondary school enrollment has increased substantially from one to seven percent. As a part of the education reforms initiated by the government in the late 1980s, the length of preuniversity schooling was shortened from 17 to 12 years. This might have helped poor households keep their children in the education system to be eligible for universities. University education remained almost free from tuition in the 1990s (Girdwood, 1999).

4. Summary

This paper has attempted to quantitatively analyze the incidence of public education spending on the poor in developing countries by using the results of standard benefit-incidence studies and additional national data. Although there is considerable variation across studies, it is confirmed that public education spending generally does not favor the poor but it favors them at a lower level of education. The poorest quintile (20% of the population), on average, receives 16.3% of total public education spending while the richest quintile receives 25.9%. In primary education, the poorest quintile receives 22.4% of public spending but only 5.5% in higher education.

The cross-country regression analysis suggests that the incidence of public education spending on the poor is higher when a country allocates more public resources to education (as a percentage of GNI), but the incidence is not much higher when a country allocates a larger share of the spending to higher education. With caution on country specificities, this finding implies that increased spending on education will be associated with increased share for the poor given that it is not devoted to spending on higher education.

In addition, a descriptive analysis of changes in the incidence over time in two countries implies that policies regarding out-of-pocket education spending and private provision will be also among important factors determining the incidence on the poor. With improved information availability, future research is expected to examine the determinants of pro-poor distribution of public spending in a more comprehensive way. It is also expected to provide more policy insights on budgetary allocation with a sub-sector and budget execution process.

Despite the limited size and policy information of the reviewed studies, the findings

of this paper still provide a benchmark on which to assess equity in public education spending with respect to targeting to the poor. When the benefit-incidence is estimated and monitored appropriately over time in a country (e.g., preand post-poverty reduction strategy) and is possibly compared with other countries, it can become a more useful indicator of evaluating the country's commitment to poverty reduction. It may enable policy makers to evaluate the impacts of policy changes such as an increase in public education budget and the share allocated to primary education.

Such information, complemented by other dimensions of equity and efficiency in public education spending, is growing in demand by international donor communities that are moving toward a more coherent and coordinated approach in their assistance to developing countries. Even if some donors continue a more traditional approach, that is, project-type rather than sector-wide or basket-funding support, they are also expected to look at macro and sectoral policies and budgetary issues in a recipient country. This is required to rationalize their specific projects in broad policy contexts and to establish coherent goals.

Acknowledgement

I am very grateful for advice on draft of this paper from Hiromitsu Muta and Yasuyuki Sawada. I also thank for information and advice from authors of the surveyed literature, especially Lionel Demery, Dominique van de Walle, and Diane Steele. Anonymous referees provided helpful comments. All errors remain my own.

Notes

- 1 This is a slightly modified categorization on the basis of Mingat and Tan (1986).
- 2 Benefits for recipients of direct government

- payments (or indirect beneficiaries) such as salary for teachers can be analyzed in a so-called "money flow" approach which has a different viewpoint in assessing distributional impacts of governmental expenditures (De Wulf, 1975).
- 3 In addition, van de Walle (1998) suggests that a number of recent benefit-incidence studies can be interpreted as moving in a new approach combining a standard benefitincidence analysis with a behavioral approach. The "behavioral approach" addresses the important limitations of a standard benefit-incidence analysis in assessing the impact of public spending on living standards. It aims to adopt the method of benefit valuation on the basis of individual preferences rather than a single utility function and to measure consumer surplus, benefits relative to a counterfactual world without government [e.g., Gertler and Glewwe (1990,1992) Lopex-Acevedo and Salinas (2000)]. However, no empirical benefit-incidence study with the new approach seems to be yet available for the education sector.
- 4 In some cases, studies aim to assess distribution across groups by other types of classifiers that are related to ethnicities and religions. For details of methodological steps, see Demery (2000).
- 5 The Lorenz curve (curve of concentration) and the Gini-coefficient are also used.
- 6 The study analyzed primary education only, not education as a whole.
- 7 http://www-wds.worldbank.org (January 2002, retrieved). I also checked other databases such as ERIC and ECONLIT.
- 8 In addition, some other issues include: how public spending on items for which users are not easily identified (e.g., scholarship abroad) is included in unit subsidy or excluded and how studies control potential biases in the

- household survey data that cause common problems for benefit incidence (e.g., underestimate of university enrollments). Most studies do not provide the information on these issues.
- 9 Some other indicators are considered but not used in the model, for example the higher education share of total education spending (both recurrent and investment spending) and the private share of financing (or enrollments) in higher education. This is because these indicators are less available in international databases.
- 10 See Table A.1 that presents data added to a dataset of selective benefit-incidence studies presented in Table 1. Table A.1 also shows that the average public education spending of sample studies is 3.9% of GNI, close to an average for lower-middle income countries (4.1%) and a little higher than that for low income countries (3.3%). This may imply that there is no clear evidence on biased sample selection in terms of a priority given to education in budget allocation of public funds.
- 11 Due to data limitations, I could not calculate the disaggregated share for primary and secondary education in total recurrent spending. When SHAREH is replaced with the aggregated share for primary and secondary education (SHAREPS), the coefficient upon SHAREPS becomes also negative but insignificant. This suggests further studies required to see the effect of sub-sectoral allocations with much improved data.
- 12 On the other hand, when regressions are run for the incidence on the richest quintile, the coefficient upon SPEND is significantly negative and the coefficient upon SHAREH is positive.
- 13 If GDPPC is replaced with the natural log of GDP per capita (LGDP), the coefficient upon

- LGDP remains positive (3.307) and significant at the 1% level.
- 14 If ILLITERATE is replaced with the gross enrollment ratio for primary, secondary, or higher education, the coefficient upon any of these variables is not significant at all.
- 15 The lower share of higher education might also reflect cost-recovery policy in sample countries, whether public higher education institutions are free from tuition fees. As information on tuition is available only in some of the sample studies, the relation with the incidence is not analyzed here.

References

- Canagarajah, Sudharshan and Xiao Ye. (2001). Public health and education spending in Ghana in 1992-98: issues of equity and efficiency. World Bank.
- Castro-Leal, Florencia, Julia Dayton, Lionel Demery, and Kalpana Mehra. (1999). Public social spending in Africa: do the poor benefit. *The World Bank Research Observer*, 14 (1), 49-72.
- Castro-Leal, Florencia. (1996a). Who benefits from public education spending in Malawi? results from the recent education reform. World Bank Discussion Paper, 350.
- Castro-Leal, Florencia. (1996b). Poverty and inequality in the distribution of public education spending in South Africa. World Bank.
- De Wulf, Luc. (1975). Fiscal incidence studies in developing countries. *International Monetary Fund Staff Papers*, 22, 61-131.
- Demery, Lionel. (2000). Benefit incidence: a practitioner's guide. (http://www1.worldbank. org/publicsector/pe/benefitabstract.htm)
- Demery, Lionel, Shiyan Chao, Rene Bernier and Kalpana Mehra. (1995). The incidence of social spending in Ghana. World Bank.
- Fields, Gary. (1975). Higher education and

income distribution in a less developed country. *Oxford Economic Papers*, 27 (2), 245-259.

- Gertler, Paul and Paul Glewwe. (1990). Willingness to pay for education in developing countries: evidence from rural Peru. *Journal of Public Economics*, 42 (3), 251-275.
- Gertler, Paul and Paul Glewwe. (1992). The willingness to pay for education for daughters in contrast to sons: evidence from rural Peru. *World Bank Economic Review*, 6 (1), 171-188.
- Girdwood, Alison. (1999). Tertiary education policy in Ghana: an assessment 1988-1998. World Bank.
- Hammer, Jeff. (1999). Education chapter for the Hashemite Kingdom of Jordan public sector review. World Bank (Mimeo).
- Hammer, Jeffrey S, Ijaz Nabi, and James A. Cercone. (1995). Distributional effects of social sector expenditures in Malaysia, 1974-89. In D. van de Walle & K. Nead (Eds.), *Public Spending and the Poor-Theory and Evidence*. Baltimore, MD: Johns Hopkins University Press, 154-186.
- Hansen, W.L. and Weisbroad, B.A. (1969). Benefits, Costs and Finance of Public Higher Education. Chicago: Markham.
- Jallade, Jean-Pierre. (1974). Public Expenditures on Education and Income Distribution in Colombia. Baltimore, MD: Johns Hopkins University Press.
- Lanjouw, Peter, and Martin Ravallion. (1999). Benefit incidence, public spending reforms, and the timing of program capture. *World Bank Economic Review*, 13 (2), 257-73.
- Lanjow, Peter, Menno Pradhan, Faida Saadah, Haneen Sayed, Robert Sparrow. (2001). Poverty, education and health in Indonesia: who benefits from public spending. World Bank.
- Lopex-Acevedo, Gladys and Angel Salinas. (2000). Marginal willingness to pay for education and the determinants of enrollment

- in Mexico. World Bank.
- Meerman, Jacob. (1979). Public Expenditures in Malaysia: Who Benefits and Why? New York: Oxford University Press.
- Mingat, A. and J.P. Tan. (1986). Who profits from the public funding of education? a comparison by world regions. *Comparative Education Review*, 30 (2), 260-270.
- Pradhan and Zant. (1998). Incidence of public education and health care expenditure in Bulgaria. Background paper prepared for the Bulgaria Poverty Assessment.
- Psacharopoulos, George. (1977). The perverse effects of public subsidization of education, or how equal is free education. *Comparative Education Review*, 21(1), 69-90.
- Selden, Thomas M., and Michael J. Wasylenko. (1992). Benefit incidence analysis in developing countries. World Bank.
- Selowsky, M. (1979). Who Benefits from Government Expenditure? A Case Study of Colombia. New York: Oxford University Press.
- Tesliuc, Cornelia, and Lucian Pop. (2000). Poverty, inequality and social protection. In C. Ruhl & D. Daianu (Eds.), *Economic Transition in Romania: Past, Present and Future* (Proceedings of the Conference Romania 2000: Ten Years of Transition, October 1999).
- Tibi, Claude, Sue E. Berryman and Michael Peleah. (2001). Moldova education financing study. (Mimeo).
- Van de Walle, Dominique. (1992). The distribution of the benefits from social services in Indonesia, 1978-87. World Bank.
- Van de Walle, Dominique. (1998). Assessing the welfare impacts of public spending. *World Development*, 26 (3), 365-379.
- World Bank. (1993). Indonesia: Public Expenditures, Prices and the Poor. Report No. 11293.
- World Bank. (1994). Colombia Poverty Assessment Report. Report No. 12673.

- World Bank. (1995a). *Kenya Poverty Assessment*. Report No. 13152.
- World Bank. (1995b). Lao PDR Social Development Assessment and Strategy. Report No. 13992.
- World Bank. (1996a). Bangladesh Public Expenditure Review. Report No. 15905.
- World Bank. (1996b). *Madagascar Poverty Assessment*. Report No.14044.
- World Bank. (1997). *Romania Poverty and Social Policy*. Report No. 16462.
- World Bank. (1999a). *Bulgaria Poverty during the Transition*. Report No. 18411.
- World Bank. (1999b). Former Yugoslav Republic of Macedonia Focusing on the Poor. Report No. 19411.
- World Bank. (1999c). *Tanzania: Social Sector Review*. A World Bank Country Study.

- World Bank. (2000a). World Development Report.
- World Bank. (2000b). *Making Transitional Work* for Everyone. Poverty and Inequality in Europe and Central Asia.
- World Bank. (2000c). Viet Nam Managing Public Resources Better: Public Expenditure Review 2000. Report No. 21021.
- World Bank. (2000d). *Ecuador: Crisis, Poverty and Social Services*. Report No.19920.
- World Bank. (2000e). *Dominican Republic:* Social and Structural Policy Review. Report No.20192.
- World Bank. (2001a). *Kingdom of Morocco Poverty Update Morocco*. Report No. 21506.
- World Bank. (2001b). *Peruvian Education at a Crossroads Challenges and Opportunities for the 21st Century*. Report No. 22357.

(2003.6.30受理)

Table A.1 Data Added to Selective Benefit-Incidence Studies Presented in Table 1

Ta		ata Added t	o Selectiv	e Benefit-	Incidenc	e Studies	Presented i	n Table	1			
Country	Year a/				Data							
		Public	Sub-sectoral share in recurrent		GDP pc b/	Illiteracy	GER	GER	GER			
		education	educ	education spending		P-	rate, adult c/	basic	second	higher		
		spending	Pre-	Higher	Other		,					
			higher	C								
		0/ COM	0/	0/	0/	TIOO	0.7	0./	0/	0/		
Albania	1996	% of GNI 3.1	% 84.5	10.3	5.2	US\$ 828	17.9	107.2	37.5	12.0		
Armenia	1996	2.0	78.8	13.2	3.2 8.0	810	2.0	87.4	37.3 89.6	12.0		
Bangladesh	1996	2.0	88.6	7.9	3.5	307	62.7	07.4	89.0	12.2		
Bulgaria	1994	4.7	75.1	15.8	9.1	1511	2.3	88.9	72.2	35.4		
Bulgaria	1993	3.2	73.1	18.0	8.2	1317	1.9	98.9	76.8	41.2		
Colombia	1974	2.2	73.6	16.0	0.2	1613	1.9	106.4	36.1	6.8		
Colombia	1974	3.5	56.5	14.8	28.7		19.4	100.4	56.1	14.6		
Cote d'Ivoire	1995	5.3	83.6	16.4	0.0	720	59.6	68.9	23.0	14.0		
Dominican Rep.	1996	2.0	62.0	13.0	24.9	1608	18.0	93.9	51.6	22.9		
Ecuador	1998	3.5	74.4	21.3	4.3	1560	9.4	126.7	31.0	22.9		
Ghana	1989	3.5	65.4	12.1	22.5		43.0	73.0	38.7	1.4		
Ghana	1992	4.5	63.5	11.0	25.5		38.8	79.6	36.4	1.4		
Ghana	1998	4.2	05.5	11.0	25.5	401	30.9	77.0	30.4	1,7		
Guinea	1994	1.9	64.7	26.1	9.2	550	30.7	46.3	12.0	1.3		
India	1994	3.5	65.5	14.4	20.0	359	47.5	100.6	48.8	5.8		
Indonesia	1987		52.1	19.9	27.9	649	23.1	116.4	48.2	7.0		
Indonesia	1989	0.9	52.1	19.9	27.9	727	21.2	116.4	45.6	8.5		
Indonesia	1998	1.4	73.5	24.4	2.1	975	14.3	112.7	55.7	11.3		
Jamaica	1992	4.7	66.1	26.3	7.6	1794	17.0	100.7	65.3	6.8		
Jordan	1996	7.9	64.4	33.0	2.5	1608	12.7	70.2	56.3	17.9		
Kazakhstan	1996	4.7	71.5	13.4	15.1	1285	12.7	97.8	86.5	33.3		
Kenya	1992	6.7	73.6	17.7	8.7	337	26.4	91.7	27.3	1.6		
Kyrgyz	1993	4.3	77.6	9.5	12.9	968		112.1	89.8	13.3		
Lao	1993	1.9	85.7	3.9	10.4	349	60.2	108.9	26.2	1.5		
Macedonia, FYR	1996	5.1	78.0	22.0	0.0	1275		99.1	62.9	19.5		
Madagascar	1994	2.2	61.6	24.6	13.8	241	38.5	88.6	15.5	2.3		
Malawi	1991	3.2	58.7	22.3	19.0	153	46.8	78.9	8.9	0.6		
Malawi	1995	5.4	67.7	20.5	11.8	146	44.1	133.5	16.5	0.6		
Malaysia	1974	6.0				1739	36.0	94.5	45.7			
Malaysia	1989	5.7	72.3	14.9	12.8	2933	19.9	94.1	57.8	6.7		
Moldova	1999	10.6	77.4	13.3	9.2	617	1.3	97.4	80.5	26.5		
Morocco	1999	5.0	83.5	16.5	0.0	1359	52.0	86.0	39.1	11.1		
Nepal	1996	3.1	70.1	17.9	11.9	212	63.0	113.0	42.3	4.8		
Nicaragua	1993	3.5				419	34.1	102.3	46.9	8.9		
Pakistan	1991	2.6	75.9	14.2	9.9	460	62.8	65.0	25.6	3.0		
Panama	1997	5.1	50.9	26.1	23.0	3123	8.8	105.5	68.5	31.5		
Peru	1994	3.9	56.4	16.0	27.5	2132	12.5	120.6	69.3	26.8		
Peru	1997	2.9	56.4	16.0	27.5	2405	11.2	122.8	72.5	25.8		
Romania	1994	3.0	68.7	15.9	15.4	1310	2.5	94.6	77.8	19.7		
Romania	1997	3.6	66.5	16.0	17.5	1373	2.2	103.5	78.4	22.5		
South Africa	1993	7.1	81.4	15.0	3.6	3788	17.4	129.2	85.6	15.7		
Tanzania	1994	5.0				176	31.1	67.5	5.3	0.4		
Uganda	1992	2.0				256	41.6	73.5	11.8	1.3		
Vietnam	1993	2.7	69.0	22.0	9.0	241	8.5	110.8	35.5	2.2		
Vietnam	1998	3.0	69.0	22.0	9.0	330	7.1	113.5	56.8	6.9		
Arithmetic Average		3.9	69.7	17.4	13.0	1063	26.4	97.8	49.6	12.3		

Arithmetic Average 3.9 69.7 17.4 13.0 1063 26.4 97.8 49.6 12.

Notes: (...) Not available. a/ For some studies where the data for the referring year are not exactly available, the data for the most closet year (within three years) are used. b/ GDP pc, Gross domestic produce per capita (constant 1995 US\$); c/ Illiteracy rate, adult total (% of people ages 15 and above). Sources: World Development Indicator 2001. UNESCO statistical database (on line, 2002).

【研究ノート】

The Tao of Evaluation

Craig Russon

W. K. Kellogg Foundation Craig.Russon@wkkf.org

Abstract

In his treatise on strategy entitled, A Book of Five Rings, famed samurai, Miyamoto Musashi, puts forward a nine-step method for learning Kendo or "the Way of the sword." Musashi's words also speak to those of us who practice the evaluation profession. The same nine-step method can be applied for learning "the Way of evaluation."

Keywords

evaluation, tao, Japan, samurai, strategy

1. The Tao of Evaluation

"I have been many years training in the Way of strategy, called Ni Ten Ichi Ryu, and now I think I will explain it in writing for the first time." It is with these words that Miyamoto Musashi began his treatise on strategy entitled, *A Book of Five Rings*. Born in 1584, Musashi was destined to become one of Japan's most famous warriors or samurai (Harris, 1974).

Samurai were easily distinguishable from the other classes of sixteenth century Japanese society (i.e., farmers, artisans, and merchants) by the two swords that they wore in their belts. The longer of the two was carried out of doors only, while the shorter was worn at all times. Dueling and tests of arms were common, with both real and practice swords. These took place within fencing halls (Dojo), before shrines, behind castle walls, and even in the streets. Duels were fought to the death or until one of the contestants was disabled.

By the age of 30, Musashi had fought and won more than 60 tests of arms by killing his opponents (Harris, 1974). Satisfied that he was invincible, Musashi turned to formulating his philosophy of "the Way of the sword" or Kendo (Harris, 1974). Over the next 30 years, Musashi refined his philosophy. A few weeks before his death in 1645, he committed his philosophy to paper.

A Book of Five Rings consists of the following sections: Ground, Water, Fire, Tradition (Wind), and Void. Ground contains an explanation of strategy and provides the context

160 Craig Russon

for the other four books. Water contains instruction on the fundamentals of long sword fighting such as stance, gaze, grip, footwork, and attitudes. Fire contains an exposition of some 28 strategies for long sword fighting. (Many contemporary Japanese and American businessmen use these to guide their business practices.) Wind provides a critique of other schools of warfare and other traditions of strategy. Void instructs students to practice the Way until "the clouds of bewilderment clear away."

Musashi's words also speak to those of us who practice the evaluation profession. In Ground, he prescribes a nine-step method for learning his strategy. These nine steps, listed below, also apply to those who seek to learn "the Way of evaluation."

- 1. Do not think dishonestly.
- 2. The Way is in training.
- 3. Become acquainted with every art.
- 4. Know the Ways of all professions.
- 5. Distinguish between gain and loss in worldly matters.
- 6. Develop intuitive judgment and understanding for everything.
- 7. Perceive those things which cannot be seen.
- 8. Pay attention to even trifles.
- 9. Do nothing which is of no use. (pg. 49)

2. Do Not Think Dishonestly

In an effort to promote honesty, many regional and national evaluation organizations around the world have put forward guidelines for ethical conduct. The American Evaluation Association's (AEA) *Guiding Principles* (1994) is an early attempt to promote ethical behavior. The document contains five principles: (a)

Systematic Inquiry; (b) Competence; (c) Integrity/Honesty; (d) Respect for People; and, (e) Responsibilities for General and Public Welfare. The Integrity/Honesty principle reads, "Evaluators ensure the honesty and integrity of the entire evaluation process." This is followed by seven statements that explain how the principle should influence every step of the evaluation process. The statements deal with contract negotiations, changes to the contract, conflict of interest, misrepresentations, misleading conclusions, and financial sources.

The Canadian Evaluation Association (CES) adopted its Guidelines for Ethical Conduct in 1996. The document contains three guidelines:
(a) Competence; (b) Integrity; and (c) Accountability. The Integrity guideline reads, "Evaluators are to act with integrity in their relationships with all stakeholders." This is followed by four statements that operationalize the concept. These statements deal with accurate representation, conflict of interest, cultural sensitivity, and contract negotiations.

A very extensive effort to promote ethics was made by the Australasian Evaluation Society (AES). In 1997, AES endorsed its Guidelines for the Ethical Conduct of Evaluations. The preface of the document is devoted to the background, purpose, scope, audience, relationship to other guidelines, approach, and terminology. The document contains five principles that deal with critical phases of the evaluation process: (a) Commissioning and preparing for an evaluation; (b) Conducting an evaluation; and (c) Reporting the results of an evaluation. Each principle has several guidelines associated with it. Guidelines deal with contract negotiations, changes to the contract, potential risks, competence, conflict of interest, conducting one's self in an honorable manner, etc. Altogether there are 22 guidelines.

Even newly formed regional and national evaluation organizations have addressed this issue. The International Network of Program and Evaluation Specialists (Georgia, Moldova, Russia, and Ukraine) has adopted a set of five Program Evaluation Principles: (a) Program evaluation can be based only on accurate and reliable information; (b) Evaluation should be competent; (c) Program evaluators guarantee honesty and transparency throughout evaluation; (d) Program evaluators respect safety and dignity of people with whom they cooperate in the course of their professional activity; and (e) Evaluators have professional obligations determined by public interests and public welfare. The honesty and transparency principle is followed by three statements dealing with contract negotiations, communication with clients, and conflict of interests.

(While Musashi advocated for honesty and integrity, there are aspects of the ethical codes explained above that he did not practice. For example, it probably could not be said that Musashi respected the safety and dignity of the people with whom he competed during the course of his professional activity.)

3. The Way is in Training

Evaluation is generally acknowledged as an emerging profession with an expanding body of knowledge and skills that require special training (Mertens, 1994). In general, there are three types of evaluation training: short, medium, and long-term. An example of short-term training would be the informal, non-degree, professional development sessions that are often offered prior to conferences sponsored by regional and national evaluation organizations.

An example of medium-term training

would be the World Bank's International Programme for Development Evaluation Training (IPDET). In 2001, the Bank partnered with Carleton University in Canada to offer about 160 hours of intensive training in development evaluation. Eighty hours of core training focused on basic evaluation knowledge/skills and development evaluation issues. Afterwards, another 80 hours of electives were offered. In 2003, the World Bank and Carleton University will be offering their third IPDET session.

An example of long-term training would be degree programs offered by colleges and universities. A directory of such programs was published in 1994 (Altschuld, Engle, Cullen, Kim, and Macce). The authors are in the process of updating this directory, with special emphasis on degree programs at colleges and universities outside the United States of America. The World Bank has also developed an on-line directory of evaluation training programs. The URL is http://wbln0018.worldbank.org/oed/evalcat.nsf? OpenDatabase These are both valuable information resources.

Patton (2002) believes that training can be made more effective through ongoing "reflective practice." Evaluators should take what they learn in training and put into practice. Then, they should follow-up on their practice and reflect on the results. This reflection should be the basis for improving the evaluators' future work. Of course, this process could be done vicariously by reflecting upon the practice of other evaluators. This important way of deepening evaluator competence is very consistent with Musashi's own self-training.

Some regional and national evaluation organizations have undertaken efforts to link training with systems of licensure, certification, and/or credentialization. For example, Altschuld

162 Craig Russon

(1999) described the Canadian Evaluation Society's system that requires evaluators to successfully participate in four daylong workshops in order to ensure that credentialed evaluators have been trained at a basic level in key evaluation areas. Worthen (1999), pointed out the serious challenges that must be overcome. They include

- Determining the basic approach certification should take
- Reaching agreement on a definition of evaluation and core competencies
- Constructing a professionally and legally defensible system
- · Garnering support for the process

4. Become Acquainted with Every Art

The author of this paper served for a time as editor of Western Michigan University, The Evaluation Center's Occasional Papers Series. During this period, Daniel Stufflebeam (1999) published an Occasional Paper entitled, Foundational Models of 21st Century Program Evaluation. In his publication, Stufflebeam critically appraised 22 evaluation approaches in order to determine which were worthy of continued application and further development. Twenty of the approaches were found to have legitimate value for evaluating programs. Stufflebeam (1999) divided these approaches into three categories: questions/methods-oriented approaches, improvement/accountability approaches, and social agenda/advocacy approaches. The approaches are listed by category below, along with the briefest of descriptions and a reference for further information. Evaluators following the injunction of Musashi to become acquainted with every art should seek to become familiar with as many of these approaches as possible.

Questions/Methods-Oriented Evaluation Approaches

- Objectives-Based Studies-seeks to determine if a program's objectives have been achieved (Madaus & Stufflebeam, 1988)
- Accountability (particularly payment by results studies)-used to ascertain responsibility for good and bad outcomes (Lessinger, 1970)
- Objective Testing Programs-assesses student achievement as compared to norms and/or standards (Lindquist, 1951)
- Outcome Monitoring/Value-Added Assessment-recurrent objective testing in order to assess trends and partial out effects (Sanders & Horn, 1994)
- Performance Testing-effort of off-set limitations of objective testing using authentic assessment of performance (Baker, O'Neil, and Linn, 1993)
- Experimental Studies-compares subjects randomly assigned into groups that receive different interventions to establish causality (Campbell & Stanley, 1966)
- Management Information Systems-supplies managers with information to guide programs (Deming, 1986)
- Benefit-Cost Analysis Approach-factors program costs discounted over time with accrued benefits (Tsang, 1997)
- Clarification Hearing-uses a judicial model to put a program on trial (Wolf, 1975)
- Case Study Evaluations-a focused, in-depth description and inductive analysis of a program (Stake, 1995)
- Criticism and Connoisseurship-expert appraisal akin to art criticism and wine connoisseurship (Eisner, 1983)
- Program Theory-Based Evaluation-uses a program's theory as basis of evaluation.
 Operationalized through logic-modeling (Chen, 1990)

 Mixed Methods Studies-eclectic mix of approaches (National Science Foundation, 1997)

Improvement/Accountability-Oriented Evaluation Approaches

- Decision/Accountability-Oriented Studiesprovides managers information for making decisions that promote cost-effective services (Stufflebeam, 1967)
- Consumer-Oriented Studies-uses evaluation to determine worth and merit of products and services offered to consumers (Scriven, 1967)
- Accreditation/Certification Approachinstitutions are studied for approval to perform specified functions

Social Agenda-Directed (Advocacy) Approaches

- Client-Centered Studies (or Responsive Evaluation)-evaluators interact continuously to respond to clients' evaluation needs (Stake, 1983)
- Constructivist Evaluation-the evaluator weaves together the stakeholders' various constructions of reality (Guba and Lincoln, 1989)
- Deliberative Democratic Evaluation-applies the principles of democracy to evaluation (House and Howe, 1998)
- Utilization-Focused Evaluation-users of evaluation are identified and a variety of methods employed to ensure that evaluation meets their needs (Patton, 1997)

5. Know the Ways of All Professions

There are two basic approaches to professional knowledge. One can know much about a single profession or one can know a little about many professions. Based upon this injunction, Musashi appears to favor the latter. This issue has been debated in the evaluation profession as well. Worthen and Sanders (1984) question whether evaluators should be trained as generalists, whose knowledge and skills are applicable to a broad range of professions. Or is it better to train them as specialists, with knowledge and skills in a single profession and less extensive evaluation training? To answer the question, they recommend taking six factors into account:

- Difficulty and uniqueness of the content-if the content is neither difficult nor unique, then a specialist may not be necessary.
- Reference groups and impartiality-the training and experience of specialists may influence them to exclude important stakeholder groups.
- Evaluation roles and tasks-do the job requirements call for the knowledge and skills of an evaluation generalist or a professional specialist?
- Scope of work-does the evaluator's scope of work require that he or she deal with a variety of professional specializations?
- Implication for training-is it possible to provide the specialist with the training necessary to do a quality evaluation?
- Professional status and rewards-the more time specialists spend on evaluation, the less they are able to devote to their profession. Evaluators who limit their work to one profession will have low career security.

Based upon the above factors, Worthen and Sanders (1984) conclude that it is often better for evaluators to be trained as generalists, who know a little about many professions, rather than as specialists, who know much about one profession.

164 Craig Russon

6. Distinguish Between Gain and Loss in Worldly Matters

Musashi believed that'worldly matters' should not cloud the intuitive judgment of the swordsman, which is always the single source of his strength; for that, he should in daily practical dealings resort to simple calculation of gain and loss. Evaluators must do the exactly the same thing. We refer to this in technical terms as controlling bias. Bias is systematic error introduced into evaluation by selecting or encouraging one outcome or answer over others. According to Scriven (1975), there may be economic or social incentives that can cloud the judgment of evaluators. He offers four methodological approaches to controlling bias:

- Standardization or routinization of qualitative aspects of the procedures
- Upgrading the training procedures for evaluators, especially in the qualitative dimension
- Focus on the organizational and economic sources of bias external to the evaluator
- Take a team approach to generating alternative plans which can then be comparatively evaluated

In addition to these four approaches, Scriven (1975) also recommends the use of metaevaluation as a way of controlling bias. Metaevaluation is simply an evaluation of an evaluation. Often times the code of ethics discussed in a previous section of this paper or the standards to be discussed in a later section serve as the basis for such metaevaluation.

7. Develop Intuitive Judgment

The life of a samurai like Musashi depended in large measure on the judgments that

he made. Likewise, a fundamental part of the evaluator's job involves making judgments. Unlike the samurai, however, there is some difference of opinion among evaluators over who should make the judgments. Scriven (1974) believes that the evaluator is the best qualified to judge. Stake (1973), on the other hand, believes that the evaluator's role is to process the judgments of others. Musashi's injunction suggests that he would have been Scriven's ideal type of evaluator.

The process of judging consists of three steps: (1) determining the criteria for making judgments, (2) collecting necessary data, (3) applying the criteria to the data. For Musashi, the criterion for making judgments was clear-cut: whether or not he could kill his opponent. It should also be noted that Musashi's judgments had a high degree of finality, so in a sense he could be considered as a specialist in summative evaluation. For many contemporary evaluators the criteria are not so clear cut nor their judgments as final.

Evaluators typically make judgments of two kinds: merit and worth (Western Michigan University, The Evaluation Center, 2002). Merit is defined as "the excellence of an object as assessed by its intrinsic qualities or performance." Worth is defined as "the value of an object in relationship to a purpose." For example, an expensive silk kimono might have a lot of merit-but little worth if one is trying to stay warm outdoors during the winter in northern Hokkaido.

8. Perceive Those Things Which Cannot be Seen/Develop an Understanding for Everything

The injunction to perceive those things

which cannot be seen is influenced by one's worldview. The traditional scientific worldview holds that there is a single, objective, tangible reality that one can see, know, and measure. The alternative worldview holds that there exist many, unseen, socially constructed realities (Guba and Lincoln, 1989). Evaluation is an iterative process of analysis leading to the emergence of a joint construction of that which is being evaluated.

The evaluator's worldview will tend to influence the approach (see the section on knowing every art) that he or she adopts. For example, an evaluator who believes the traditional scientific worldview would tend to use approaches such as Accountability, Objective Testing Programs, Outcome Monitoring, Performance Testing, Experimental Studies, and Cost-Benefit Analysis. An evaluator who believes in the alternative worldview would tend to use approaches such as Case Study, Criticism and Connoisseurship, Client-Centered Studies, and Constructivist Evaluation.

According to Stake (1979), the products of evaluations using traditional scientific worldview are different from those of using the alternative worldview. The product of the former is **explanation** that emphasizes formalistic generalizations from sample to population. The product of the latter is **understanding** with emphasis on particulars and experiential knowing. Stake counsels us not to confuse explanation and understanding.

9. Pay Attention Even to Trifles

The injunction to pay attention even to trifles, for the evaluator, is an exhortation to produce quality work. One way that evaluators ensure the quality of their work is by conforming to professional standards. One of the most enduring and broadly accepted sets of evaluation standards is *The Program Evaluation Standards* (Joint Committee for Educational Evaluation, 1994). *The Standards* contain 30 principles that can help improve evaluation of programs, projects, and materials. The principles fall under four categories: Utility, Feasibility, Propriety, and Accuracy.

The Joint Committee has always maintained that *The Standards* are uniquely American and may not be appropriate for use in other countries without adaptation. Several regional and national evaluation organizations from around the world have undertaken projects to create standards suitable for use in their own cultural contexts. Many think that standards could provide a model to guide practice.

In an Occasional Paper entitled, *The Program Evaluation Standards in International Settings*, Widmer, Landert & Bachmann (2000) put forward the evaluation standards recommended by the Swiss Evaluation Society (SEVAL). The SEVAL standards are, in part, based on *The Program Evaluation Standards*. This was an historic first. These were the first evaluation standards to be developed in a country outside of the USA.

Patel & Russon (2000) described a 1998 evaluation workshop of the United Nations Children's Fund (UNICEF) held in Nairobi that initiated a discussion of the appropriateness of *The Program Evaluation Standards* for use in African cultures. During the discussion, it appeared that, in some cases, *The Standards* were culturally inappropriate or misleading. The results of the discussion were presented at the Inaugural Conference of the African Evaluation Association (AfrEA) as a draft document.

166 Craig Russon

In December, 2000, the board of the Australasian Evaluation Society voted to authorize the development of Australian and New Zealand standards of evaluation (Fraser, 2001). A special committee is developing the standards for endorsement by the membership. These standards may, ultimately, be registered with the standards authorities of both countries (Fraser, 2001).

10. Do Nothing Which is of No Use

Utilization is widely seen as the sine qua non of evaluation. It is often said that it is better not to conduct an evaluation at all, than to conduct an evaluation that is not used. This is because an evaluation that is not used wastes money that could have been devoted to programming. The professional literature historically has recognized three main types of evaluation utilization: instrumental, conceptual, and symbolic (Kirkhart, 2000).

Instrumental utilization occurs when major stakeholders immediately implement the recommendations from an evaluation report. Conceptual utilization occurs when an evaluation report serves to alter the way that stakeholders think about the object of the evaluation. Symbolic utilization occurs when the evaluation is used for political purposes, for example, to muster support for the continuation of a program.

Utilization is such an important matter that an entire approach has been devoted to it. Michael Quinn Patton's Utilization-Focused Evaluation is a widely practiced evaluation approach. According to Patton, "Utilization-focused program evaluation . . . is evaluation done for and with specific, intended primary users for specific, intended uses" (p. 23). Patton's writings have recently been translated into Japanese and published by G. Pam

Communications (2000).

11. Conclusion

To conclude this article, I wish to paraphrase Musashi. It is important to start a profession in evaluation by setting these broad principles in your heart. With your spirit settled, you must accumulate practice in evaluation day by day, year by year. As you grow in the art and the science of evaluation, eventually you will come to think of things in a wide sense and, taking the void as the Way, you will see the Way as Void. "In the void is virtue, and no evil. Wisdom has existence, principle has existence, the Way has existence, spirit is nothingness" (p. 95). Study this well.

Notes

- This article represents the views of the author and not necessarily those of the W. K. Kellogg Foundation.
- The author wishes to thank Michael Quinn Patton and Blaine Worthen for their valuable feedback on an early draft of this manuscript.

References

American Evaluation Association. (1994). AEA guiding principles for evaluators. http://www.eval.org/Publications/publications. html#Guiding%20Prin

Altschuld, J. (1999). The case for a voluntary system for credentialing evaluators. *American Journal of Evaluation*, 20(3), 507-517.

Altschuld, J., Engle, M., Cullen, C., Kim, I., & Macce, B. (1994). The 1994 Directory of Evaluation Training Programs. In J. Altschuld and M. Engle (Eds.), The preparation of professional evaluators: Issues, perspectives,

- and programs. *New Directions in Evaluation*, 62, 71-94.
- Australasian Evaluation Association. (1998). Guidelines for the ethical conduct of evaluations. Curtin, Australia: AES.
- Baker, E., O'Neil, H., & Linn, R. (1993). Policy and validity prospects for performance-based assessment. *American Psychologist*, 48, 1210-1218.
- Campbell, D. & Stanley, J. (1966). *Experimental* and quasi-experimental designs for research. Boston, MA: Houghton Mifflin.
- Canadian Evaluation Association. (1996). CES guidelines for ethical conduct. http://www.evaluationcanada.ca/site.cgi?sectio n=5&ssection=4&_lang=an
- Chen, H. (1990). *Theory driven evaluations*. Newbury Park, CA: Sage.
- Deming, W. (1986). Out of the crisis. Cambridge, MA: Center for Advanced Engineering Study, MIT.
- Eisner, E. W. (1983) Educational connoisseurship and criticism: Their form and functions in educational evaluation. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), Evaluation models. Boston: Kluwer-Nijhoff.
- Fraser, D. (2001). Development of AES standards. *Evaluation Journal of Australasia* 1(1), 59.
- Guba, E. G., & Lincoln, Y. S. (1989). Fourth generation evaluation. Newbury Park, CA: Sage.
- Harris, V. (1974). Translators introduction. In M. Musashi's *A book of five rings: The classic guide to strategy*. The Overlook Press.
- House, E. R. & Howe, K. R. (1998). *Deliberative* democratic evaluation in practice. Boulder: University of Colorado.
- International Program Evaluation Network.

 Program evaluation principles.

 http://ipen21.org/ipen/en/default.html
- Joint Committee on Stands for Educational Evaluation. (1981). *Standards for evaluations*

- of educational programs, projects, and materials. New York: McGraw-Hill.
- Kirkhart, K. (2000). Reconceptualizing evaluation use: An integrated theory of influence. In V. Caracelli and H. Preskill (Eds.), The expanding scope of evaluation use. *New Directions in Evaluation*, 88, 5-24.
- Lessinger, L. (1970). Every kid a winner: Accountability in education. New York: Simon and Schuster.
- Linquist, E. (Ed.) (1951). *Educational measurement*. Washington D. C.: American Council on Education.
- Madaus, G. & Stufflebeam, D. (1988). Educational evaluation: The classical writings of Ralph W. Tyler. Boston: Kluwer.
- Mertens, D. (1994). Training evaluators: Unique skills and knowledge. In J. Altschuld and M. Engle (Eds.), The preparation of professional evaluators: Issues, perspectives, and programs. *New Directions in Evaluation*, 62, 17-28.
- Musashi, M. (1974). A book of five rings: *The classic guide to strategy*. The Overlook Press.
- National Science Foundation. (1993). Userfriendly handbook for project evaluation: Science, mathematics, engineering and technology education. NSF 93-152. Arlington, VA: Author.
- Patel, M., & Russon, C. (2000). Appropriateness of the Program Evaluation Standards for use in African cultures. Monitoring and Evaluation Capacity Development in Africa: Selected Proceedings From a Seminar and Workshop Organized by the Development Bank of Southern Africa, the African Development Bank, and the World Bank. Johannesburg, South Africa.
- Patton, M. Q. (2002). Personal communication.
- Patton, M. Q. (1997). *Utilization-focused* evaluation: The new century text (3rd ed.). Newbury Park, CA: Sage.
- Patton, M. Q. (2000). *Utilization-focused* evaluation: The new century text (3rd ed.).

168 Craig Russon

Japan: G. Pam Communications.

- Sanders, W. & Horn, S. (1994). The Tennessee value-added assessment system (TVAAS): Mixed model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8(3) 299-311.
- Scriven, M. (1975). Evaluation bias and its control. *Occasional Papers Series # 4*. Kalamazoo, MI: Western Michigan University, The Evaluation Center.
- Scriven, M. (1974). *The exact role of value judgments in science*. Berkeley, CA: University of California (mimes).
- Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Curriculum evaluation*. *AERA Monograph Series on Curriculum Evaluation* (Vol. 1). Chicago: Rand McNally.
- Stake, R. (1973). The countenance of educational evaluation. In B. Worthen & J. Sanders, *Educational evaluation: Theory and practice*. Belmont, CA: Wadsworth.
- Stake, R. (1979). Naturalistic study of such things as science teaching. Paper delivered at the May 17 meeting of the North Carolina Association for Research in Education.
- Stake, R. (1983). Program Evaluation, particularly responsive evaluation. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), Evaluation models, pp. 287-310. Buston: Kluwer-Nijhoff.
- Stake, R. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

- Stufflebeam, D. (1967, June). The use of and abuse of evaluation in Title III. *Theory Into Practice*, 6, 126-33.
- Stufflebeam, D. (1999). Foundational models for 21st century program evaluation. *Occasional Papers Series # 16*. Kalamazoo, MI: Western Michigan University, The Evaluation Center.
- Tsang, M. (1997). Cost analysis for improved *educational* policymaking and evaluation. *Educational Evaluation and Policy Analysis*, 19(4), 318-324.
- Western Michigan University, The Evaluation Center. (2002). Glossary of evaluation. http://ec.wmich.edu/glossary/
- Widmer, T., Landert, C. & Bachmann, N. (2000).
 The Swiss evaluation standards. In C. Russon (Ed.) The Program Evaluation Standards in international settings. Occasional Papers Series # 17. Kalamazoo, MI: Western Michigan University, The Evaluation Center.
- Wolf, R. L. (1975, November). Trial by jury: A new evaluation method. *Phi Delta Kappan*, *3*(57), 185-187.
- Worthen, B. (1999). Critical challenges confronting certification of evaluators. *American Journal of Evaluation*, 20(3), 533-556.
- Worthen, B. & Sanders, J. (1984). Content specialization and educational evaluation: A necessary marriage. *Occasional Papers Series* # 14. Kalamazoo, MI: Western Michigan University, The Evaluation Center.

(2003.7.7受理)

【研究ノート】

社会指標による政策効果測定に関する一考察 時系列指標を比較する方法

小野 達也

富士通総研 onotatsu@fri.fujitsu.com

要約

時系列の社会指標を互いに直接比較するためには、何らかの方法により個々の指標を共通の尺度に変換することが必要となる。近年の業績測定・プログラム評価においては社会指標が多く用いられており、時系列の評価指標の推移を解釈しまた比較することは重要な意味を持つ。

旧経済企画庁の社会指標の系譜に連なるものであり、「構造改革の成果を評価する」という踏み込んだ目標を掲げて試算された内閣府の「暮らしの改革指数」における変換の方法の問題点を明らかにした上で、業績測定・政策評価においても活用できる変換の方法を提案する。

キーワード

社会指標、プログラム評価、業績測定、時系列指標

はじめに

行政評価においては様々な社会指標'が用いられる。プログラム評価program evaluationにおいて施策・事業のアウトカムを測定する指標として、あるいはインパクトを評価するためのデータとして個々の社会指標が広く用いられて来たし、ここ十数年の間に多くの国において取り組まれるようになった業績測定performance measurementではアウトカムを測定する指標として多くの社会指標が用いられている。

一方、社会統計の一分野としての社会指標 の目的に関しては社会状態のモニタリングや 市民への報告に徹するのか、それとも公共政策へ反映させるのかについて常に議論があった。とりわけ社会指標を政策効果の評価に社会指標を政策効果の評価に開発された頃よりあった。するないである。この関係性を前提とした試みは、失敗に終担なで説明しようとするものである。このを説したされる。しかしその原因は、関係把握が充される。のではなく、このモデルが純すぎたことにあるといえる(Noll 1996)。

近年は欧州連合(EU)やカナダ、オーストラリアなどにおいて総合的な社会指標の体系を構築する試みが再び活発になっている。

日本評価学会『日本評価研究』第3巻第2号、2003年、pp.169-178

170 小野 達也

このような最近の試みにおいては、社会指標は実態把握と課題設定・政策選択・政策評価という一連の過程を通じて公共政策に反映すべきであると再認識されてプログラム評価・業績測定との連携も論じられている(Beck et al 2001, Finn 1998, AIHW 2001, Land 2000)。

さて、複数の時系列指標を比較することは、 近年急速に導入が進む行政評価において、時 系列で整備される評価指標を解釈するうえで 必要な手順である。また複数の時系列指標を 総合化してひとつの指標とする際には、加重 平均あるいは単純平均を計算する前に、個々 の指標を互いに比較可能な状態にすることは 本来必要な手順である²。

本稿は、時系列指標が比較可能になるように変換する方法について、最近発表された「暮らしの改革指数」など公的に作成されている指標における問題点を指摘し、実務的に有効な新しい変換方法を提案するものである。

1. 時系列指標を比較する方法について

一般に複数の時系列指標を比較する場合には、最新時点の値を比べるほかに、過去から最新時点までの推移を比べることになる。そのためには通常指標の値そのものの推移と、値の変化率(連続する2時点間の変化率の推移)をまたは基準時点に対する変化率の推移)をあわせて用いるのが基本である(Morley et al 2001, p. 65)。基準時点に対する変化率は、基準年の水準を100とする指数の形で表すことも多い。

しかし、同一の指標を複数の地域間で比較するような場合を除き、異なる指標同士を精密に比較しようとすれば、各指標の過去から現在までの変化を同じ観点から解釈することが必要であり、各指標のとりうる値がどのような分布になるのか、実際に生じた変化がその分布に対して相対的にどれだけの大きさか、という点に関する配慮が求められる。す

なわち、指標の値の変化が確率的な変動の範囲を超えているか、ある期間の変化のスピードやパターンが別の期間と比べてどう変化しているか、などについて明らかにし、指標同士を共通の尺度で表現することが求められる。

近年の社会統計としての社会指標には、次に述べる「暮らしの改革指数」など、この目的のために個々の指標に「変換」を施している例がある。ただし、今のところこの変換の方法について標準となる方法はないとされる(国民生活審議会 1985, p. 17, Beck et al 2001, p. 32, Sharpe 1999, p. 49)。社会指標におけるこのような変換は、総合指標をより科学的に作成するために個々の指標のもつ尺度を揃えることが目的であるが、既に述べた通り、総合指標の作成の有無に関わらず行政評価等において時系列指標を精密に比較するために必要な手順であるということができる。。

なお、このような「変換」について、欧米 の文献では尺度(scale)を揃えるという意味 でscaling (Sharpe 1999)、同じscaleへの conversion (Industry Canada 1997) あるいは 単にstandardization (Schetagne et al 2001)な どの用語が使われている。一方、日本の国民 生活審議会・旧経済企画庁から現在の内閣府 にいたる社会指標の取組みにおいては「標準 化」という語が使われている(国民生活審議 会 1985, 経済企画庁 1997, 国民生活審議会 2002)。また後述の旧経済企画庁・内閣府に おける景気動向指数のコンポジット・インデ ックスの作成においては「基準化」という語 をほぼ同様の意味で用いている(内閣府経済 社会総合研究所 2003)。本稿では分野により 異なる意味で用いられる「標準化」「スケー リング」などではなく、「変換」という語を 用いた。

2. 「暮らしの改革指数」における 方法の問題点

(1) 暮らしの改革指数における変換の方法

総合指数を含む社会指標として、日本で唯 一継続的に作成されてきたのが、国民生活審 議会・旧経済企画庁の「社会指標(Social Indicators,略称SI)」(1974~)「新社会指標 (New Social Indicators, 略称NSI)」(1986~) 「新国民生活指標(People's Life indicators,略称 PLI)」(1992~1999)である(国民生活審議 会 1985、経済企画庁 1997)。 しばしの中断を 経て、内閣府は昨年12月末に「暮らしの改革 指数」という新しい社会指標を1990年から 2001年まで試算し公表した(国民生活審議会 2002)。この指数は、現在進められている構 造改革の成果を国民生活の視点から評価しよ うとするもので、公共政策・政策評価との関 係を意識した社会指標である。今後、毎年末 に前年分を公表する予定とされる。

指数は構造改革の目指す社会を「住みやすい社会」「働きやすい社会」など9つの側面に分けて評価している。その手順は、それぞれの側面ごとに暮らしの改革の進展度を代表すると考えられる指標(個別指標)を抽出し、総合化して側面別の指数を算出するというものである。

この指数では、個別指標を総合化する前に、 指標により異なる変動の幅を調整する目的 で、各指標を変換する。その変換の方法は、 指数作成の対象期間である1990年から2001年 に至る12年間の各年の変化率の絶対値の平均 が1となるように各年の変化率を変換した上 で、その変換後の相対変化率を用いて、基準 年(1990年)の水準100から出発して、各年 の基準化指数を順次計算していくというもの である。

変換の具体的手順は次の通り。なお、この 個別指標を変換する手順はかつての新国民生 活指標の時系列指数作成におけるのと同様で ある⁴。 手順1:対称変化率の算出

・ケース1:指標が通常の指数や現実の レベルそのものの場合

$$C_{i(t)} = \frac{D_{i(t)} - D_{i(t-1)}}{\left(\frac{D_{i(t)} + D_{i(t-1)}}{2}\right)} = 100$$

 $D_{i(i)}$:個別指標,i:指標番号,t:時点

C:(:): 対称変化率

・ケース2:指標が構成比等の場合、または0値や負値をとる場合

$$C_{i(t)} = D_{i(t)} - D_{i(t-1)}$$

手順2:平均絶対変化率(変化率の 絶対値の平均 タ(A_i)の算出

$$A_i = \frac{\sum_{t=2}^{N} \left| C_{i(t)} \right|}{N-1}$$

N: 時点数 (今回はN=12)

手順3:相対変化率(平均絶対変化率 A_i で基準化(相対化)した個々の変化率 $J(B_{iii})$ の算出

$$B_{i(t)} = \frac{C_{i(t)}}{A_i}$$

手順4:基準化指数(基準年次を100とした指数) (S_{ii}) の算出

基準年次のSimを100とし、次式により算出。 今回はSi(1990)=100。指標値上昇が構造改革 として望ましくないものは、次で求めた指数 を200から引いた値を標準化指数とする。

・ケース1:

$$S_{i(t)} = S_{i(t-1)} \times \frac{100 + \frac{B_{i(t)}}{2}}{100 - \frac{B_{i(t)}}{2}}$$

(2時点の指数の対称変化率が Bi(t)となる。)

172

小野 達也

・ケース2:

$$S_{i(t)} = S_{i(t-1)} + B_{i(t)}$$

(2) 現行の方法の問題点

上の手順1~4による変換方法には、次のような問題がある。

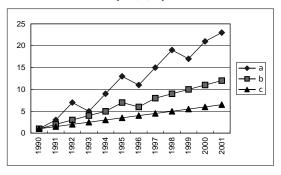
実数値を読み取れない

基準年次を100として各年の変化率を積み上げる形となっているため、実数値の上昇・下降は反映するが、手順2,3の変換により変換後の指標値の高低が大幅に変化してしまう。図1に示すような指標a,b,cの場合、大小関係は全く逆転してしまう。。

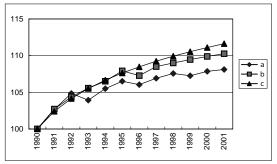
上昇・下降の度合いを比較できない

全期間の指標値を対象に相対変化率を計算するため、指標ごとの変化率の変化パターンは観察できるが、指標の上昇・下降の状況を

図1 問題点 の模擬例 (変換前)



(変換後)



(出所)何れも筆者作成

指標間で比較することができない。図2に示したような指標a,b,cの場合、近年の上昇・下降の状況は変換後の指標値の個々の推移パターンに反映されるが、指標間で比較することはできない。

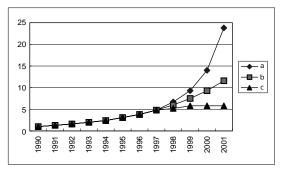
変化率の符号の反転の影響が大き過ぎる

手順2により、変化率の絶対値の平均を用いて振幅の調整を行うため、期首と期末の間の変化率は同じであっても期間内に変化率の符号の反転があるか否かにより、変換後の指標値が大きく異なることとなる。

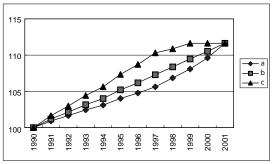
図3に示す指標a,b,cの場合、変化率の変動が少ないa、変動があっても符号の反転がないbに対し、符号の反転があるcの変換後の指標値には大きな違いが生ずる。

図4には暮らしの改革指数の「働きやすい 社会」側面に含まれている「裁量労働制採用 企業割合」「大学発ベンチャー数」「完全週休 2日制採用企業割合」「失業率」(失業率は値

図2 問題点 の模擬例 (変換前)



(変換後)

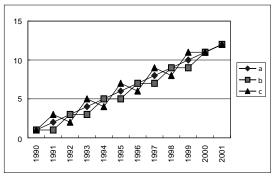


(出所)何れも筆者作成

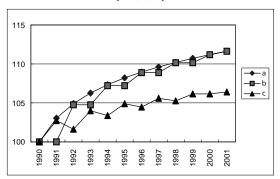
が小さい方が望ましいため大小関係を逆転して指標化)「公共職業安定紹介所の就職件数(対有効求職者千人比)という5個の個別指標を例として取り上げ、その変換前(基準年次の1990年が100になるように表示した)と変換後を示した¹⁰。

指標作成対象の期間中変化率の符号が反転しなかった失業率が、期間中一度以上変化率の符号が反転した他の指標に比べて変換後の変化幅が大きくなるなど、変換後の指標値は互いに直接比べることはできず、個別の指標ごとに変化率の推移のパターンをみることしかできない11。従ってこれら変換後の指標を平均して得られる総合化指数は明確な意味をもたないことになる。

図3 問題点 の模擬例 (変換前)



(変換後)



(出所)何れも筆者作成

3. 望ましい変換方法とその表示に ついて

上に指摘した問題点を解決する変換方法の1つとして、以下の方法が考えられる。即ち、実数値の推移と変化率の推移は別々に観察することとし、期間を区切ってそれぞれの振幅の調整をする変換のための因子を計算する。この因子は平均絶対変化率ではなく、標準偏差とする。

具体的な計算手順は次の通り。

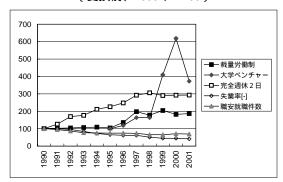
手順1:対称変化率の計算

暮らしの改革指数と同様。

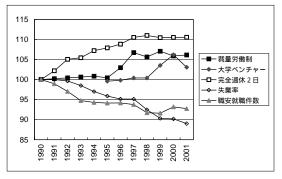
手順2:振幅調整のための因子の計算

全期間ではない特定の期間について、個別 指標の対称変化率の平均μ×、標準偏差××

図4 「働きやすい社会」の5指標 (変換前、1990年=100)



(変換後)



(出所)国民生活審議会2002に掲載のデータをもとに筆者作成

174 小野 達也

及び実数値の平均 μ, ˇ、標準偏差 ¸ ˇを計算する。

手順3:対称変化率・実数値のZ得点(通称偏差値)を計算する。

全期間にわたって、各年の対称変化率及び 実数値のz得点(実数値の平均からの偏差を 標準偏差で除したもの)を計算し、z得点を 用いて平均50、標準偏差10の Z 得点を計算す る。

$$egin{aligned} z_{i(t)}^X &= rac{C_{i(t)} - \mu_i^X}{X} & C_{i(t)} : 対称変化率 \ & z_{i(t)}^Y &= rac{D_{i(t)} - \mu_i^Y}{Y} & D_{i(t)} : 実数値 \ & Z_{i(t)}^X &= 50 + 10 imes z_{i(t)}^X \ & Z_{i(t)}^Y &= 50 + 10 imes z_{i(t)}^Y \end{aligned}$$

このように計算すれば、各指標の実数値及び変化率の推移を継続的に直接比較することができる。したがって、必要となれば両者別に単純平均や加重平均による総合化を行うための前提を満たすことにもなる。

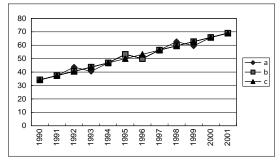
先に指摘した問題点 ~ について、同じ模擬例を変換した結果を以下に示す。手順2において調整因子の計算対象は1990~2000年としている。これはこの期間の水準を基準として以後の期間(今回のデータでは2001年のみ)を基準化することになる¹²。

まず問題点 の実数値の比較について同じ 模擬例を変換した結果を図5に示す。Z得点 化によって位置と振幅を調整したことによ り、3指標とも同じような推移をしているこ とがわかる。また90年代の平均50に対して順 調に上昇していることも読み取れる。

続いて問題点 の上昇・下降の変化の比較について、同じ模擬例を変換した結果を図6に示す。2000~01年の上昇率の違い、及び90年代の上昇率の平均50との差を直接読み取ることができる¹³。

図5 新方法による実数値の変換結果

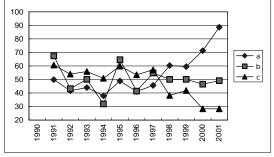
(問題点 - 図1の模擬例と同じデータ)



(出所)筆者作成

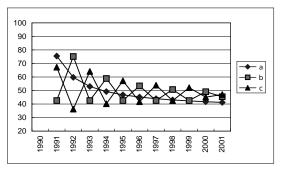
図6 新方法による対称変化率の変換結果

(問題点 - 図2の模擬例と同じデータ)



(出所)筆者作成

図7 新方法による対称変化率の変換結果 (問題点 - 図3の模擬例と同じデータ)



(出所)筆者作成

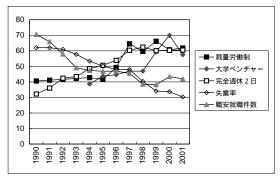
問題点 で指摘した変化率の符号反転の影響は、調整のための因子として平均絶対変化

率ではなく、標準偏差を用いることで解決する¹⁴。図7に問題点 と同じ模擬例を変換した結果を示す。振幅の違いはあるものの変化率自体は同じような傾向で推移していることがわかる。

次に、暮らしの改革指数の「働きやすい社会」側面の5指標について、新しい変換を施した結果を図8及び図9に示す。

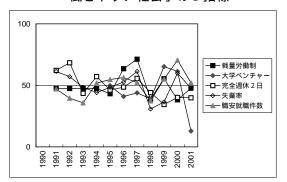
図8では、各指標が2000から01年にかけて 上昇したか下降したか、90年代の平均と比べ てどうか、個々の指標の実数値の推移を読む とともに、水準・振幅の違いを調整した上で 指標同士を直接比較することができる。

図8 新方法による実数値の変換結果 「働きやすい社会」の5指標



(出所)国民生活審議会2002より筆者作成

図9 新方法による対称変化率の変換結果 「働きやすい社会」の5指標



(出所)国民生活審議会2002より筆者作成

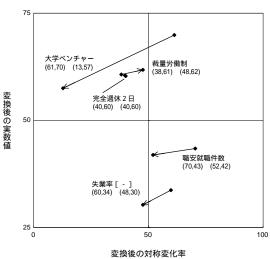
図9では、各指標の変化率の変化について、 図8と同様個々の指標の推移を読むとともに 水準・振幅の違いを調整した上で指標同士を 直接比較することができる。

最後に、変換後の実数値を縦軸に、変換後の対称変化率を横軸にとり、これら5指標の2000~01年の推移を散布図上にプロットしたものを図10に示す。実数値と対称変化率の推移を同時に比較することができる。

謝辞

本論文の初稿に対して、査読者から数々の 有益なご指摘をいただきました。深く感謝し ます。

図10 「働きやすい社会」の5指標の推移 (新方法による変換後、2000~01年)



(出所)国民生活審議会2002より筆者作成

176 小野 達也

注記

- 1 「社会指標」は指標の体系を意味すると同時に、1つ1つの指標も社会指標と呼ばれる。
- 2 このような手順を踏まずに総合指標を算出すれば、確率的な変動であれ、意味のある水準の変化であれ、各指標の変動が、総合化の段階で明示的に付与するウェイトとは別に、事実上のウェイトとして作用することとなる。
- 3 この変換による比較は、もとの個別指標のそれぞれを解釈・評価することが目的であり、変数間の関係を多変量解析や時系列モデルなどの統計的手法を用いることによって、データ全体を要約したり景気指標などの合成指標をつくることとは目的が異なる。
- 4 手順1~4の計算式自体は、米国のNational Bureau of Economic Researchによって景気 観測のために考案されたコンポジット・ インデックスにおける個別指標の変換方 法と基本的に同一である。異なる点は、 手順3のあとで採用系列全体の平均変化率 を求め、それから手順4によって出発点を 100とする指数を作成することである。な お、コンポジット・インデックスは景気 変動の大きさやテンポを測定することを 目的とし、一般の社会指標とは目的が異 なる。日本の景気動向指数のコンポジッ ト・インデックスもかつてはこの手順で 作成されていた(刈屋1986、竹内編1989) が、その後旧経済企画庁における検討を 経て、手順2,3(注記15参照)及び手順4 の前の採用系列全体を合成する手続きが 別のものとなっている(内閣府経済社会 総合研究所 2003)。
- 5 時点を逆転した変化率も絶対値は等しい という対称性をもつ変化率。例えば、100 から200への変化率も200から100への変化 率も絶対値はともに66.7%。
- 6 国民生活審議会 (2002) では「標準化因 子」と呼んでいる。

- 7 国民生活審議会 (2002) では「標準化変 化率」と呼んでいる。
- 8 国民生活審議会(2002)では「変化率標準化指数」と呼んでいる。
- 9 手順1と手順4においては、3指標ともケース1として計算した。ケース2として計算 しても全く同じ問題を指摘できる。続く 問題点 及び についても同様。
- 10「大学発ベンチャー数」「失業率」はケース2、その他の3指標はケース1として取り 扱われている。
- 11 今回の報告書(国民生活審議会 2002)では、変換後の指標の推移について、問題点 ~ に起因すると考えられるミスリーディングな記述が散見される。
- 12 前2内閣の続けた景気優先路線から構造改 革路線への転換を打ち出した小泉内閣の 誕生が2001年4月であり、構造改革の成果 を2000年以前と2001年以後を比べること で評価する想定とした。
- 13 暮らしの改革指数に用いられている変換 方法を始めに提案した国民生活審議会の 調査委員会の中間報告(国民生活審議会 1985)及びそれを受けた当初の新社会指 標(NSI)においては、実は指数作成対 象の全期間ではなく、昭和45~55年の10 年間を対象に変化率を調整し、55年を基 準として、その後の変化率をそれ以前の 平均変化率と比べる形となっていた。
- 14 注記5で述べた旧経済企画庁・内閣府の作成する景気動向指数のコンポジット・インデックスにおいても、現在では暮らしの改革指数と同じ手順2,3ではなく、平均変化率の平均、標準偏差を用いてz得点化(平均からの偏差を標準偏差で除すしている(内閣府経済社会総合研究所2003)。また、カナダ政府が最近開発したPersonal Security Index (Schetagne et al 2001)では、総合指標の計算の過程で、個々の指標の基準年に対する成長の量として、基準年次以前の一定の期間の指標値の平均・標準偏差からz得点を求め、それを基準年の指標値に加算する方法を

とっている。

参考文献

- 刈屋武昭 (1986) 『計量経済分析の考え方と 実際』、東洋経済新報社
- 経済企画庁(1997)『新国民生活指標PLI (People's Life Indicators)』
- 国民生活審議会(1985)『NSI (New Social Indicators) 国民生活指標』、第10次国民生活審議会総合政策部会調査委員会中間報告
- 国民生活審議会(2002)『暮らし指数検討委員会報告書』 第18次国民生活審議会総合 企画部会暮らし指数検討委員会報告書
- 竹内啓編(1989)『統計学辞典』東洋経済新 報社
- 内閣府経済社会総合研究所 (2003) 『景気動 向指数の利用の手引き』http://www. esri.cao.go.jp/jp/stat/di/di3.html
- Australian Institute of Health and Welfare (2001). Needs, interventions, outcomes: measurement frameworks. In *Australia's Welfare 2001*. Canberra.
- Beck, W.; Keizer, M.; van der Maesen, L.J.G.; & Phillips, D. (2001). General Paper. A paper presented at the 1st Plenary Meeting of the Network 'Indicators Social Quality'. Amsterdam Oct. 2001.

- Finn, P.J. (1998). Social Indicators for the Strategic Evaluation of Major Social Programs. Ottawa: Human Resources Development Canada.
- Industry Canada (1997). Measuring Sustainable Development: Review of Current Practice. Industry Canada Occasional Paper No.17. Ottawa.
- Land, K. (2000). Social Indicators. In Borgatta, E.F. & Montgomery(eds.), Encyclopedia of Sociology, Revised Edition. New York: Macmillan.
- Morley, E.; Bryant, S.P. & Hatry, H.P. (2001). *Comparative Performance Measurement*. Washington, D. C.: The Urban Institute Press.
- Noll, H.-H. (1996). Social Indicators and Social Reporting: The International Experience. A paper presented at the Symposium on Measuring Well-being and Social Indicators held by Canadian Council on Social Development. Ottawa Oct. 1996.
- Schetagne, S.; Jackson, A. & Harman, S. (2001). Gaining Ground: The Personal Security Index, 2001. Ottawa: Canadian Council on Social Development.
- Sharpe, A. (1999). A Survey of Indicators of Economic and Social Well-being. Ottawa: Canadian Policy Research Networks.

(2003.7.23受理)

178 小野 達也

A Study on Outcome Measurement with Social Indicators -- A Method for the Comparison of Time-series Indicators --

Tatsuya Ono

Fujitsu Research Institute onotatsu@fri.fujitsu.com

Abstract

When developing social indicators, a method is needed to transform each variable into units on a common scale so that they can be compared and combined. Recently, this issue has become common to performance measurement and program evaluation, because they use so many outcome measures many of which are a kind of social indicators. In order to analyze the outcome of programs, it is often necessary to compare the indicators in a common framework.

We review the method for converting the indicators used in the new social indicators developed by the Cabinet Office of the Government of Japan to evaluate the outcome of structural reform in Japan, and propose another method of conversion which are applicable to performance measurement and program evaluation.

Keywords

social indicators, program evaluation, performance measurement, time-series indicators

【研究ノート】

地方自治体における行政評価の実際と新しい評価者モデル エンパワメント型行政評価者

西出 順郎

福井県庁/地域構想センター Junronishide@aol.com

要約

本稿の目的は、行政評価の現状を基に行政評価における新しい評価者モデルを構築し、その 養成の必要性を問うことにある。地方自治体において、行政評価は行政改革の切り札として導 入された。行政改革には三つのステージとして、制度改革、マネジメント改革、行動規範改革 がある。後者二つの改革において、行政評価はその力を発揮する。しかしながら、現実には著 しい成果は上がっていない。マネジメント改革と行動規範改革、そのいずれのステージにおい ても障壁は立ちはだかっている。その克服のためには、行政評価を牽引し、組織力を向上させ る評価者、すなわちエンパワメント型の行政評価者モデルを構築し、そして、そのような新型 評価者養成の必要性を地方自治体間に浸透させなければならない。エンパワメント型行政評価 者は、(狭義の)エバリュエーターだけでなく、メンター、ファシリテーター、そしてプロデューサーとしての役割を駆使し、マネジメント改革、行動規範改革を実現させる。すなわち、行 政評価はエンパワメント型行政評価者によって再生されるのである。今、エンパワメント型行 政評価者モデルの構築とその養成は、喫緊の課題である。

キーワード

行政評価、行政改革、エンパワメント型行政評価者

はじめに

バブル経済崩壊後、地方自治体を取り巻く 環境は激変した。財政状況の悪化、多様化す る社会ニーズ、高齢社会の到来、そして度重 なる不祥事・・・。このような状況の下、地方 自治体は、深刻かつ複雑な社会環境に対応し 新たなる規範を確立するため、変革に向け始 動した。行政改革である。行政評価は、この 改革の切り札として、1990年代後半から地方 自治体間で急速に普及した。

現在、行政評価はほぼ全国の都道府県や政令指定都市で導入されている(三菱総合研究所 2002)。しかしながら、現実には行政改革が遅々として進まず、掛け声だけで終わっている自治体も決して少なくない。行政評価は未だ発展途上の段階にあり、実際の行政運営の中で様々な問題に直面しているのだ。その

日本評価学会『日本評価研究』第3巻第2号、2003年、pp.179-187

180 西出 順郎

克服には評価そのものを見直すだけではなく、行政評価を牽引する人材、いわゆる行政 評価のための評価者が新たに必要とされている。

本稿では、まず行政改革における行政評価の位置づけを整理し、行政評価の目的と必要性を明確にする。そして、その実像を捉えながら、行政評価における新しい評価者像を考察する。

なお、本稿でいう行政及び地方自治体とは、 中央政府ではなく、地方政府の行政(執行) 機関を想定していることを付記する。

1. 行政改革の三つのステージ

行政改革、これは、地方分権の進展そして厳しい行財政状況の中、行政を抜本的に改革することである。効率的な行政運営、質の高いサービスの提供、住民やNPOとの協働など、その領域は多岐にわたるが、行政改革を成功させるには、階層化された三つのステージで改革が実現しなければならない。すなわち、制度改革、マネジメント改革、そして行動規範改革である。

第一の制度改革は、法律・法令に基づき、 行政の活動に対する正当性の付与もしくは制限によって実現される、最上層の改革である。 これは立法によってコントロールされる改革 であり、国から地方への権限委譲はその最ものである。第二のマネジメント改革は、 法律・法令に基づく事務・事業及び地方独自 の事務・事業を運営するシステム、いわゆる マネジメントにおける改革である。これは回る 域の意思決定に基づく組織自身によるものである。第三の行動規範改革は、最下層に位置 し、組織のマネジメントを実際に掌る職員の ための改革である。

制度改革による行政改革の限界はすでに指摘されている(上山 2002)。また、市町村合併のような、自主的に自らの存立を問うものは除き、制度上の改革を地方自治体自らが実現できる環境には未だ至っていない。すなわ

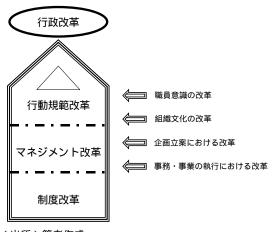
ち地方自治体にとって行政改革とは、マネジメント改革と行動規範改革を指すものであり、前者は計画策定、予算編成などの企画立案における改革もしくは事務・事業の執行における改革である。後者は職員意識や組織文化の改革である(図1)。

2. 行政評価の定義

行政改革を推進する具体的な方策、まさしくそれが行政評価である。行政評価の先駆的事例として紹介される三重県の事務事業評価は1995年に導入された。しかし、行政評価が実際に脚光を浴び出したのは、行政改革が声高に叫ばれていた1997年後半からであり、この原動力は、専門紙「地方行政」による米国の行政評価を紹介した長期連載特集にあった。

行政評価の明確な定義はない。「業績測定 (Performance Measurement) 又は包括的に各種の評価の類型を代表する概念」(古川・北大路 2001)、「業績測定とプログラム評価(Program Evaluation)の和集合」(島田他1999)、「数値目標の導入、民間手法の活用」(上山1998)、「業績測定」(上山2001)など

図1 行政改革における三つの改革ステージ



(出所)筆者作成

様々であり、実務者間で急速かつ強力に支持された結果、行政評価はその明確な定義なし に普及したのが実状である。

厳格な定義への探求は容易ではないが、行 政評価の定義にあたっては、普及した当時の 社会背景を考慮したい。すなわち、行政評価 は、ハコモノ行政や巨額累積赤字だけではな くカラ出張問題など、行政のモラルにまで及 んだ批判に応える、改革の切り札であったこ とである。既存の事業評価という用語が行政 評価に自然と駆逐されてしまったのもそのよ うな背景がある。行政評価は、個別事業のア カウンタビリティを強化もしくはその存廃を 判断するミクロ的な評価の適用を超越した、 非効率の象徴たるマートンのいう官僚制の逆 機能にマクロ的大転換をもたらす社会装置と して期待されたのである。すなわち、行政自 らのパラダイムシフトを強烈に志向したもの であった。

また、評価手法の用語による定義だけでは、 その意味する幅を狭めてしまう。よって行政 評価を数少ない分析・評価手法の包括的意味 合いとすることは避けたい。業績測定、プロ グラム評価のほかに、事業別のバランスシー トや行政コスト計算書も行政評価として活用 されている。また、費用便益分析(CBA [Cost Benefit Analysis]) や費用有効性分析 (CEA [Cost Effectiveness Analysis]) などの学 術的手法、更には業績測定においてもABC (Activity-Based Costing), BSC (Balanced Score Card) などの応用型モデルがある。改 革のツールとして活用され得る分析・評価手 法は、行政評価の一類型とすべきである。し たがって、行政評価とは、「行政におけるマ ネジメントの機能強化や職員の意識改革な ど、行政改革に活用される定性もしくは定量 的な分析・評価手法の包括的な概念」である と位置付けたい。

このような定義においては、行政評価の客体は非常に幅広い。実際には、政策体系に沿って政策・施策・事務事業を評価客体とすることが多いが、事務事業をより細分化した活動を客体とする場合もある。ABCはその一

例である。その一方で事業直接ではなく、BSCのように部局などの組織単位でその運営全体を客体とする評価もある。また、評価の実施時期についても同様で、評価目的に応じその時期は決定する。CBAやメタ評価によって政策・事業のインパクトを予測するのなら事前評価、業績測定やBSCでモニタリングをするなら期中評価となる。プログラム評価によって政策・事業の有効性を実際に検証する場合には、それらの完了時となろう。

3. 行政改革における行政評価の活用

行政評価はどのように活用されるのか。ここではマネジメント改革と行動規範改革のプロセスを通じて整理する。パットン (Patton 1997) は、評価結果の使い道と評価プロセスの使い道を定義し、評価目的によって、評価の活かし方を区別している。行政改革においても、改革のステージが異なれば、評価の活かし方は異なる。マネジメント改革の場合は評価結果が改革への端緒となり、その一方、行動規範改革では、評価プロセスと関わることによって無意識のうちに改革がなされていく。

(1) マネジメント改革と行政評価

マネジメントのプロセスにおいて評価結果 を活用する主なものは、合理的判断を追求す る計画の策定や予算の編成、そして効率性・ 有効性を追求する事務・事業の改善である。 いずれにおいても行政評価は論理的な意思決 定を導くための有効な支援ツールである。

第一に計画の策定。概して行政が策定した 計画は、美辞麗句が多く、具体的目標が曖昧 である。また、各部局の策定案をそのまま綴 じた、いわばホッチキス的計画策定である。 そこで行政評価は、計画との連動を図りその 中に具体的な数値目標を設定させる。また、 目標の達成過程を演繹的に説明するロジック モデルを構築し、計画の合理性を確保すると 182 西出 順郎

ともに各部局間の連携関係を明らかにする。 このような取組みは戦略的計画策定 (Strategic Planning)と呼ばれている。いわば 評価結果を合理的にフィードバックするため の基盤作りである。

第二に予算の編成。これは評価結果に即して合理的な予算配分を目指すものである。一般的に、事務事業評価やABC、事業別のバランスシートや行政コスト計算書がそのために用いられる。公共事業と称される大規模な事業の実施にあたっては、メタ評価やCBAもしくはCEAなどの妥当性の高い事前評価が実施される。

第三に事務・事業の改善。ここでは実施事務・事業の効率性や有効性向上のために評価結果を活用する。業績測定の場合、有効性の改善にはアウトカム志向、効率性の改善にはアウトプット志向の業績指標が設定される。評価結果は、時系列比較や他団体比較(ベンチマーキング)、設置目標との比較分析などに用いられ、事務・事業の改善すべきポイントを探る。プログラム評価は、事業とそのインパクトの因果関係をより深く追求する場合に導入される。

その他、評価結果は政治家や利害関係者、メディアなどに対する政治的なマネジメント・ツールとして活用される。政治的マネジメント(Political Management)とは、外部関係者の理解・支援を得て業務遂行を果たすことを指す(Moore 2000)。すなわち、行政とその外部の相互理解を促進すべく、評価結果が共通言語的役割を担うのである。

このように、評価結果の活用は広い範囲に 及んでおり、今後は、定員管理や人事マネジ メントで活用されることが予想される。

(2) 行動規範改革と行政評価

職員意識と組織文化の改革。マネジメントにおける意思決定はツールではなく、あくまで人間が行なう。マネジメントを理想的に機能させるには、顧客満足重視、事業の成果重視など時流を掴んだ行動規範に対して、職員

自身が共鳴しなければならない。行政評価は、 その一連のプロセスを通じ、変革へのインセ ンティブを職員ひいては組織全体に与え、職 員意識と組織文化の改革を引き起こす。すな わち、職員は評価プロセスの中で顧客満足の 必要性や成果志向の重要性を自問自答する。 職員間ではその答えを探すべくコミュニケー ションが生まれ、新たな規範に対する共有意 識が芽生える。そして、その反復と継続によ って新たな価値観が職員の心に浸透するので ある。更には行政評価で得た業績にスポット ライトを当て、それを職員間で共有すること で組織の新しい価値観は統一される。行動規 範改革と行政評価の密接な関係は、三重県の 事務事業評価システムが導入当初は組織体質 改善のためのツール(中村 1999)であった ことからも窺える。現在、福岡市では、 「DNAどんたく」と称し、改善運動例の発表 会が毎年開催されている(福岡市ホームペー ジ)。まさしく組織文化の改革である。

4. 行政評価の実際

行政改革のツールとして導入された行政評価だが、現実にその力を十分発揮させるには克服すべき課題は多い。ここでは実際の行政評価に内在する問題点を、マネジメント改革及び行動規範改革の視点から考察する。

(1) マネジメント改革としての問題点

第一に、評価技術における精度の低さである。例えば、業績測定では、実施事業している。例えば、業績測定では、実施事業している。のもれず、その難とそのインパクトの因果関係を顕しまである場合が多い。その結果、アウトカム指標をしたが設定は避け、アウトプット指標もしくは避け、アウトカム指標が設定される。仮っている。である。 一夕収集がごく限られた地域のみの場合、測定結果は妥当性に欠く。更に測定結果が適正でも、他団体の情報が入手困難ならば ベンチマーキングは不可能である。プログラム評価においては、その分析手法が専門的であるため、学術的調査の経験がない多くの職員にとって、外部によるプログラム評価の導入でさえ心理的抵抗がある。

(2) 行動規範改革としての問題点

第一に、自治体職員間に潜む事務(いわゆる事務・事業)に対する主体性の欠如である。これは、自治体事務の多くが条例ではなく、法律又は法令に依拠することに起因する。確かに機関委任事務は廃止され、法定受託事務と自治事務に生まれ変わった。しかし、いずれの事務であれ、その運用の多くは国がいる。このように自らの裁量権の放棄が自らの事務に対する主体性の欠如をもたらし、計画の形骸化、事務・事業のインパクトに対する無管任感、そして行政評価への無関心さを増幅さてある。

第二に、システム偏重による行政評価の機能不全である。行政評価はマネジメント機能を強化するツールであり、システムそのものではない。予算編成や事務・事業の執行など、各マネジメントに即して評価は活用され、その運用については各マネジメントの所掌セクションの判断を尊重すべきである。しかしな

がら、多くの自治体では、評価が活用される 舞台演出も考慮されず、官房系セクションが、 行政評価の全庁的なシステム構築、いわゆる 評価の体系化と評価シートの様式作りに終始 している。その表層的な取組みは、官房系セクション以外の職員から評価の活用意識を削 ぐばかりか、評価に対する関心さえも摘み取ってしまう。シート作成というペーパーワーク、信頼性の低い分析・評価を、ここでも助 長することになる。

5. 新しい評価者モデルの構築

概して行政評価は内部職員が実施する。複数の評価段階を踏む場合、担当職員に限らず、課長もしくは部局長級の管理職も評価者となる。外部評価者の必要性が問われても未だ普及しないのは、学術的に質の高い判断重視型評価ではなく、行政改革のための改善重視型評価として行政評価が期待されているからである。よって業務に携わる職員全員が評価者であり、また、そのような職員意識なくして改革は前進しない。

発展途上の行政評価において、エンパワメント型行政評価者には、評価の技術だけでなく、評価の普及・形成・活用を確実に遂行する力量が要請される。すなわち、エバリュエ

184 西出 順郎

ーター(ここでは、単なる評価技術に熟練した評価者を指す。)として評価技術を習得させるだけではなく、メンター(mentor)ファシリテーター(facilitator)、そしてプロデューサー(producer)としての役割が求められているのである。換言すれば、行政評価の真髄を普及する(mention)、その実施を促進する(facilitate)、その価値を創造する(produce)役割である。

ここでクイン他(1996)による定義を用いて、メンター、ファシリテーター、プロデューサーをエンパワメント型行政評価者に援用する。

メンターは学習コーチである。部下が個人的目標や業務上の目標を達成するための道筋を作り上げることができるよう、補助する役割を有する(Quinn他 1996)。これをエンパワメント型行政評価者として捉えると、そのメンター的な役割とは、職員とのフィードバックを絶えず繰り返しながら行政評価の意義や目的を説き、その普及に努めることにある。

ファシリテーターは、チームの高い業績を

作り上げ、それを維持し、チーム自体を発展させるよう支援する役割を有する(同上)。エンパワメント型行政評価者のファシリテーター的な役割とは、権威的ではなく、協力的な役割を演じながら、権限や意思決定をチームに任せ、行政評価を推進する。そして、評価が有機的に形成されるよう導くことにある。

プロデューサーは、部下がやる気を出し、 業績を上げるため、部下のやる気を刺激する 要因は何かを考える役割を有する(同上)。 そのプロデューサー的な役割とは、行政評価 の実績を高く評価することで、評価に対する 姿勢及び価値観を明確にする。そして評価者 (職員)のやる気を引き出すと同時に評価の 実効性を向上させることにある。

このようにエンパワメント型行政評価者は、エバリュエーター、メンター、ファシリテーター、プロデューサーという四つの役割を好循環させながら、行政評価の普及・習得・形成・活用プロセスの機能強化を図り、行政評価を実践的な改革のツールに変貌させ

 メンター
 エバリュエーター

 普及
 習得

 行政評価
 形成

 ファシリテーター
 プロデューサー

 (出所) 筆者作成

図2 エンパワメント型行政評価者モデル

るのである(図2)。

6. エンパワメント型行政評価者に 求められる技能

エンパワメント型行政評価者にとって、行 政評価独自の知識に限らず一般的な分析・評価手法やコミュニケーション・スキルが必ら となるのは言うまでもない。しかしながらら 政内部にこのような新タイプの人材をエンがらら 動きなどほぼ皆無である。したがってるため ワメント型行政評価者を具現化させるため ワメント型行政評価者を具現化させるるの は、そこに内在すべき技能を明確にすっ がある。ここでは、メンター、エバリューと がある。ファシリテーター、プロデューサー して、それぞれに不可欠な技能について鳥瞰 する。

エンパワメント型行政評価者は、メンター として高いコミュニケーション能力を駆使 し、行政評価の意義や目的の普及に努めなけ ればならない。よって行政評価に関する幅広 い知識が求められる。行政評価の沿革、日本 だけではなく先進国の歴史も紐解かねばなる まい。特に、PPBS (Planning and Programming Budget System) † ZBB (Zero-Based Budgeting) の理念とその失敗、プログラム評価の降盛か ら衰退、GPRA (Government Performance and Results Act)の誕生など、米国における評価 の歴史は、今後、日本の行政評価が歩むべき 方向性を示唆している。また、行政評価の必 要性を論じるには、CBAのような学術的手 法の限界を反芻し、まずは業績測定などの実 践的手法の普及に理解を求めるべきである う。

エバリュエーターに必要なものは、多様な評価技術を実践できる能力である。プログラム評価では、回帰分析などの定量分析やフォーカスグループなどの定性分析手法が必須である。業績測定においては、ベンチマーキングやABC、BSCなどの応用手法が挙げられる。また、バランスシートや行政コスト計算書などの財務の視点に立った手法も軽視してはな

らない。評価技術の向上は、即、評価の信頼性につながる。質の高い評価結果は予算編成時の政治的圧力に対峙するコミュニケーション・ツールになり得るだろう。

ファシリテーターにとっては、各マネジメ ントにおける評価の執行を支援し、評価の実 用化を推進する能力である。例えば以下のこ とが挙げられよう。第一に、ロジックモデル を強化させること。これによって誰もが事業 インパクトを論理的に共有できる。第二に、 中間アウトカムの業績指標を活用させるこ と。最終アウトカムの測定が困難な場合でも 業績指標を設定できる。第三に、評価結果を 一般市民に理解しやすいよう加工させるこ と。政治的なマネジメントには平易なレポー トが必要である。最後に、それぞれのマネジ メントに適合した評価手法を提示すること。 これによって評価の使い道が明確になる。関 係職員との連携を図り、このような助言や支 援を行なうことで、行政評価の形成にその力 を如何なく発揮する。

プロデューサーには、行政評価の実効性を 高める戦略的能力が求められる。評価結果の 活用を確実に改革へつなげるため、そして、 行政評価の市民権を確立するため、硬軟交え た方策を展開しなければならない。例えば官 房系セクション主導の評価システムの見直し である。各マネジメント所掌セクションが独 自に評価プランを設計するよう軌道修正し、 管理統制的な評価体質を一掃するのである。 更には、評価事例の発表会や褒賞制度の設置 による、評価及び改革の意識と自らの事務に 対する主体性の醸成である。中央政府への依 存体質や補助金等に潜むモラルハザードの排 除、その他、無謬性・表層的完全主義の一掃 など、エンパワメント型行政評価者はプロデ ューサーとして、これら可視化できない難問 に挑戦することになる。

以上のように、エンパワメント型行政評価者に求められる技能は多岐にわたる。メンター、エバリュエーター、ファシリテーター、プロデューサー、それぞれの力をタイミングよく発揮し改革を推進するによってエンパワ

186 西出 順郎

メント型行政評価者は屹立するのである。

7. エンパワメント型行政評価者の 養成

エンパワメント型行政評価者に要求される 技能、それは極めて多様である。では、その 養成は実際に可能なのだろうか。これに対す る解答はそう簡単には見出せないだろう。し かし、エンパワメント型行政評価者は決して 絵空事であってはならない。そこで最後に、 エンパワメント型行政評価者の養成に向けて の基本的な方向性を提示する。

行政評価を導入する際、地方自治体の多く は職員研修を実施している。しかし、知識・ スキル・能力、それらすべてを網羅した研修 など無きに等しく、一人当たりの研修時間も 決して長くない。エンパワメント型行政評価 者として、多様な役割を循環させながら成果 を導き出すためには、集中的にすべての技能 を摂取する専門プログラムの開発が求められ る。但し、それを各々の地方自治体が開発す ることは、決して効率的ではなく、したがっ て第三者機関による質の高いプログラム開発 が強く期待される。例えば、米国には "Evaluators' Institute "と呼ばれる評価者養成 の専門機関がある。年に二回、米国内で評価 研修セミナーを一週間にわたって開催する。 その短期間にどこまでの実効性を見出せるか 疑問ではあるが、このような第三者機関プロ グラムが日本でも定期的に開催され、多くの 地方自治体職員が参加することになれば、エ ンパワメント型行政評価者の養成に確実なー 歩を踏み出すことになる。

詳細なカリキュラムについては割愛するが、米国専門機関のような評価手法を重視したカリキュラム構成では不十分である。評価の歴史や技術だけでなく、アクティブ・リスニングやプレゼンテーションを要としたコミュニケーション・スキル、そして行政評価の牽引者として不可欠なリーダーシップなど、人的要素を内包したカリキュラムが基本的な

枠組みとなろう。

おわりに

現下の状況において、行政評価は大きな障 壁に直面している。一時の勢いが衰退したの は、評価に対する確たる目的意識もなく、や みくもに導入した結果であり、導入自体が自 己目的化するに至ったからである。今ある障 壁を乗り越えるためには、行政評価自体が成 長を遂げなければならない。そして、実用重 視の行政評価へと進化し、行政改革を担う起 爆剤として、復活しなければならない。エン パワメント型行政評価者は、そのためのキャ スティングボードを握る存在であり、改革プ ロセスを支える重要なプレーヤーである。地 方自治体にとって、新しい評価者モデルの構 築と一人でも多くのエンパワメント型行政評 価者を養成することは、喫緊の課題の一つで ある。

参考文献

上山信一(1998)『行政評価の時代』、NTT出 版

(2000)『行政の経営改革』、第一法規出 版株式会社

上山信一監訳・監修(2001)『行政評価の世界標準モデル』、東京法令出版社(Broom, C., Jackson, M., Harris, J., and Vogelsang-Coombs, V.(1998). Performance Measurement: Concepts and Techniques 3rd Edition. Washington D.C.: The Center for Accountability and Performance, American Society for Public Administration.)

島田晴雄・三菱総合研究所政策研究部(1999) 『行政評価』 東洋経済新報社

中村征之(1999)『三重が、燃えている』、公 人の友社

福岡市ホームページ http://www.city.fukuoka. ip/

古川俊一・北大路信郷(2001)『公共部門評

価の理論と実際』、日本加除出版株式会社 三菱総合研究所(2002)「地方自治体におけ る行政評価への取組みに関する実態調査」、 http://www.gyousei.gr.jp/

Moore, H. M. (2000). Creating Public Value: Strategic Management Government 5th Edition. Cambridge, Massachusetts: Harvard University Press.

Patton, M.Q. (1997). Utilization-Focused

Evaluation 3rd Edition. Thousand Oaks: Sage Publications. (大森彌監修、山本泰・長尾眞文編(2001)『実用重視の事業評価入門』、清水弘文堂書房)

Quinn, R. E., Faerman, S.R., Thompson, M.P., and McGrath, M.R. (1996). *Becoming Master Manager 2nd edition*. New York: John Wiley & Sons.

(2003.7.1受理)

A New Evaluator Model for Local Government Reform -- An Empowerment-oriented Evaluator for Local Government Reform --

Junro Nishide

Fukui Prefectural Government / the center for Citizenship and Social Capital (CSC)
Junronishide@aol.com

Abstract

The purpose of this paper is to present a model for a new role of evaluators for local government reform based on the current practice of its evaluation. Evaluation was introduced as a highlight for local government reform. Local government reform has the three stages: system reform, management reform, and norm reform. At the stages of management and norm reforms, evaluation must be utilized, however, not a remarkable outcome has been seen and obstacles exist for the evaluation to be fully utilized. In order to overcome the obstacles, the author argues that local government foster an Empowerment-oriented Evaluator for local government reform, who would lead local government reform and improve organizational capacity. The evaluator utilizes the roles of mentor, facilitator and producer as well as evaluator, and realizes management and norm reforms. The evaluator may revitalize government reform. Creating Empowerment-oriented Evaluator Model and training of Empowerment-oriented Evaluator are one of the most pressing challenges.

Keywords

local government reform, evaluation, empowerment-oriented evaluator

188 陸路 正昭

【研究ノート】

住民意識の分析による水循環の施策と評価について

陸路 下昭

千葉経済大学短期大学部 m.mutsuro@dream.com

要約

治水の時代、利水の時代を経て水環境の時代になった(高橋 1988)といわれており、健全な水循環のあり方を考えることが急務である。経済的な観点だけでなく、環境に配慮した行政のあり方が模索されている。また、地方自治体の行政においては、地方分権や行政改革の流れと共に、住民とのパートナーシップも求められている。

岐阜県では、水循環健全化大綱を作成するとともに、アンケートを実施して、住民の水循環に対する意識を確認しようとしている。水循環への施策を行政が一方的に決めるのでなく、住民の意向を十分に反映し、さらにその評価を受けていこうとする試みである。

この研究では、このアンケート結果を活用して、設問同士の関係や自由記述の分析を通じて、 水循環への住民意識を掘り下げていく中で、問題解決の展開を組み立てながら、施策の評価に ついて考察していく。

キーワード

水循環の健全化、住民意識、アンケート、問題解決、水循環の施策と評価

1. はじめに

水に関する関係省(環境省、国土交通省、厚生労働省、農林水産省、経済産業省)による「健全な水循環系構築に関する関係省庁連絡会議」によれば、健全な水循環系とは「流域を中心とした一連の水の流れの過程において、人間社会の営みと環境の保全に果たす水の機能が、適切なバランスの下に、ともに確保されている状態」と定義している(国土交通省水資源部 2002)。

岐阜県における水循環健全化大綱において は、水循環が「海洋、陸域からの蒸発、降 雨・降雪、流出・流下という過程を繰り返しつつ、気象、地形の改変等の重要な現象を支配している。」と述べた上で、水循環系が「物質や熱エネルギーを移動させつつ、人間はもとよりあらゆる生命活動をささえている」一方で、「人間の活動は水循環系に影響を与えてきた」ことを述べている。

この大綱が示すとおり、「近年の都市化の 進展、土地利用状況等の急激な変化により、 常時の河川流量の減少、保水能力の減少によ る洪水流量の増大、逆に、渇水被害の頻発と 被害ポテンシャルの増大、産業排水、生活排 水等による一部水域における水質改善の未達 成、化学物質等による新たな水質汚染問題の 発生、治水安全度が低い地域での居住者の増加、人々に安らぎを与える水辺環境の減少な ど水に関わる多くの、そして重大な問題」が、 人間による活動に関連しており、水循環を健 全化していくための方策を明らかにしていく ことが重要であると考えられる。

本研究は、水の循環を成り立たせるための 岐阜県の水循環健全化大綱におけるアンケー ト結果によって住民の意識を明らかにしなが ら、水循環健全化のための施策と評価の手が かりにすることを目的として行う。

本研究に関連する従来の研究は、次の3つに分類できると考えられる。

1つ目は、住民の意識や行動等に関する研究である(原田 2000)。また民間企業のマーケティング分野では、顧客の意識を分析するための自由記述の分析も行われている(大隈 2002)。

2つ目は、自治体の政策や施策に関する研究であり、理論的な考察(小野 2002)や特定の自治体における考察(鎌田 2002)もある。

3つ目は、水循環等に関するものであり、 技術的な視点や環境面での考察がある。

さらに、この分類の組合せとして位置づけられる研究もあり、住民の意識や行動を地域づくりや施策に結びつける研究(馬場 2002)(佐藤 2001)(岩松 2001),水循環等に関する住民意識や行動の研究(村瀬 2002)さらに、水循環等に関する施策の研究(伊藤他2002)などがあげられる。

本研究においては、これらの3つを組み合わせ、水循環に関する住民意識を自治体の施策と評価に結びつけることを意図している。 具体的には、水循環に関する住民意識の分析のためにアンケートを活用し、選択肢と自由記述の回答を組み合わせた分析から施策と評価の具体的な例示をしていくものである。

2. アンケートの概要

(1) アンケートの内容

アンケートは、岐阜県において、水循環に 対する住民の意識を分析するために、

「県土の根幹をなす水循環系を望ましい姿に改良し、将来にわたって維持していくために、水循環健全化大綱を策定しております。この大綱に対するご意見を伺い、皆様と共に、豊かな岐阜県を創りたいと考えています。」という趣旨で、岐阜県建設管理局河川課がモニターに対して実施したものである。

調査期日は、平成12年9月であり、アンケート対象者は、県政モニターの対象者800名 (一般推薦400名、一般公募200名、電子公募200名)に対して、郵送法による実施で、アンケート回収数 612名(回収率76.5%)である。

質問項目は、以下の8問である。

質問1. 年代と性別(プロフィール)

質問2. 水循環健全化大綱の趣旨への意見(賛成、 どちらかといえば賛成、 どちらかといえば賛成、 どちらとまいえかい どちらかとい

どちらともいえない、 どちらかといえば反対、 反対、 その他)(選択式) 質問3. 反対理由(で反対の番号選択者のみ)(記述式)

質問4. 10問題のうち3つの問題の選択 およびその他の問題点の記述(手入れ が行き届かない森林の増加、流域の開 発・都市化の拡大、農地の近代化、

ダムの建設、 河岸、水路のコンクリート化、 地下水位低下、湧水枯渇、地盤沈下、 河川常時水量の減少、 渇水の頻発、被害ポテンシャル増大、 河川等の水質悪化、地下水汚染、 地球温暖化の影響)(選択式と記述式)

質問5. 大綱についての共感できる目標の選択(良質な水を安定的に供給できる水循環系、 異常気象時などにおいても県民の生命、生活を守ることのできる水循環系、 県民に潤いを与え、自然と

190 陸路 正昭

の共生を可能とする水循環系、 水を活 かした産業、観光、文化等諸活動を活性 化する水循環系)(選択式、個数は自由) 質問6. 9つの施策から、3つ選択および 意見の記述(水を貯める、 水を浸み 込ませる、 水のむだをなくす、 水に 親しむ、 水をきれいにする、 災害に 自然と共に生きる、 備える、 水を調 水を活かす)(選択式と記述式) べる. 質問7. その他の取り組むべき推進方策 (記述式)

質問8. 大綱全体についての意見(記述式)

なお、質問4の原文の「課題」を、ここでは、「問題点」と変えた。後述の通り、質問4、5、6を問題解決の展開として組み合わせて分析するためである。

対象者の年齢、性別および居住地に関して は、無作為抽出を施していないので、県民の 意見全体を反映しているわけではないが、モ ニターの性格上、比較的、高い問題意識を有 する住民の回答が得られており、今後の水循 環に関する行政のあり方や行政と住民の関係 のあり方への提言を得ることができると判断 できる。

回答者の性別は、男性280名、女性332名であり、それぞれの構成比は45.8%と54.2%であった。母集団としての県全体の人口における構成比は男性48.5%、女性51.5%である。年代は、10代5名、20代73名、30代116名、40代123名、50代99名、60代以上196名であり、それぞれの構成比は0.8%、11.9%、19.0%、20.1%、16.2%、32.0%である。

母集団としての県全体では、10才未満が9.9%、10代11.5%、20代13.2%、30代12.4%、40代13.0%、50代15.4%、60代以上が24.6%である。60歳以上がともに最大の構成比を持っている点が共通している。

(2) アンケートの集計結果

集計結果の概要は、

(1) 賛否については、「賛成である」、「どちらかといえば賛成である」をあわせて、88.9%の高率に達する。 「どちらかといえば反対である」、および「反対している」は、合わせて1.3%であるが、これらは趣旨に関するものであることと、モニターという立場の前向きと考えられる住民が対象であるからと推測される。各設問への自由記述の部分にも注意を払っていく必要があろう。

(2)問題点についての回答の構成比は、 多い順に、「手入れが行き届かない森林」(21.6%)、「水質悪化・地下水汚染」(21.1%)、「水路のコンクリート化」(12.3%)、以下順に「地球温暖化の影響」、「地下水位低下」、「流域の開発」であった。

(3)目標についての回答の構成比は、「良質な水の安定的供給」が30.2%であり、「自然との共生」が27.5%、「県民の生命、生活を守る」が、26.3%、を占め、「産業、観光、文化の活性化」が比率的には低い。すべてが目標として取り上げられる必要があるものであるが、優先順位ということでの結果が現れている、と解釈できそうである。 (4)施策に関しては、上位3項目の構成

比が、「水をきれいにする」(20.6%) 「自然とともに生きる」(19.1%) 「水のむだをなくす」(17.6%)であり、 次いで、「災害に備える」が続き、 「水を貯める」、「水を活かす」と 「水を浸み込ませる」が続いている。 (5)自中記述の回答者は「趣旨への意

(5)自由記述の回答者は、「趣旨への意見」と「問題点への意見」は選択肢の他に、という問いかけなので記入者がそれぞれ、27名、31名と少ないと思われる。「施策への意見」はそれらの2倍強に当たる50名が答えている。「その他の推進方策」と「全体への意見」は、それぞれ、168名、269名と多くの回答が得られており、単に選択肢の回答では示せない考え

方も含まれている。

(6) 問題点について、3個答えた人は581 人、2個が16人、1個が12人、0個が3人で あった。処理の関係ですべての回答は同 じ重みで扱うことにする。目標の選択と 施策の選択についても同様とする。目標 については、4個答えた人が149人、3個 答えた人が157人、2個答えた人が230人、 1個答えた人が67人、0個が9人であった。 (8)施策については、3個が579人、2個が 24人、1個が6人、0個が6人であった。 無回答の人についての他の設問への回答 について調べると、例えば施策に関して の選択肢が空欄の人は、施策についての 自由回答が見られない傾向が高く、この 選択肢だけでは、不足ということではな いと判断できる。ただし、これらの人々 のうち全体意見や推進方策についての記 述を記入した人々が多く見られた。

3. アンケートによる住民意識の分析を活用した施策評価

図1は、アンケートの各設問の内容を示す 図である。本研究の目的は、1章で触れたと おりであるが、そのための住民意識の分析方 法として、第3章では、(1)問題点 目標 施策(図1の真ん中の枠内の3つの設問)の組 合せを決め、(2)組合せとのつながりから、 回答者の自由記述の特徴的な単語を取り出 し、これらをもとに(3)施策評価の例示を 試みる。

なお、図2は、本研究における3章、4章全体の分析のステップであり、さらに、第4章では、(1)回答者の自由記述の単語を分析をした上で、(2)さらに回答者を分類して、(3)施策評価推進の手がかりを得ようと試みていく。

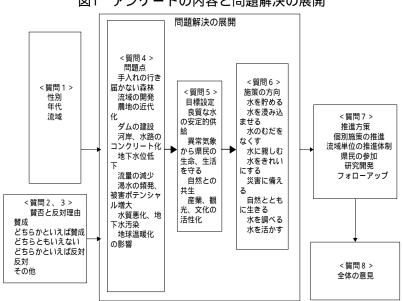


図1 アンケートの内容と問題解決の展開

(出所)岐阜県のアンケート項目より筆者作成

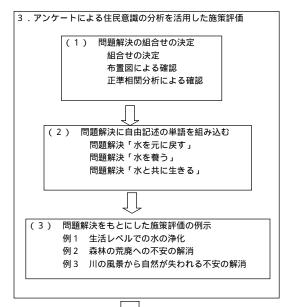
192 陸路 正昭

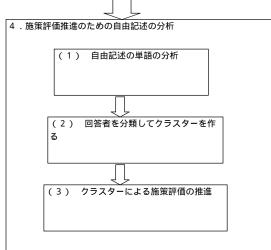
(1) 問題解決の組合せの決定

組合せの決定

アンケートにおける問題点 目標 施策の 3つの設問において、意味のあるつながりを 見出していくことによって、それぞれの問題 を解決していく手がかりを得ていく。そのた

図2 分析のステップ





(出所)筆者作成

めに、まず、各設問間クロス表を作成して分析を進める。

問題点の各選択肢と目標の各選択肢の組合せ、および目標の各選択肢と施策の選択肢の組合せの比率の高いものに注目する。特に、ここでは、設問ごとのもともとの回答人数の上位3項目に注目していくことにする。人数の多い順に問題点では、、、であり、目標では、、、、施策では、、「であり、信では」であり、結びした。更に意味内容を把握しながら、結びった。更に意味内容を把握しながら、結びった。更に意味内容を把握しながら、結びった。更に意味内容を把握しながら、結びった。更に意味内容を把握しながら、結びった。更に意味内容を把握しながら、結びった。更に意味内容を把握しながら、結びった。更に意味内容を把握しながら、結びった。

「水質」 「良質な水」 「水をきれいにする」

「森林」 「県民の生活」 「水のむだをなくす」

「コンクリート化」 「自然との共生」 「自然とともに生きる」

の通り、線を結ぶことができる。これらの組み合わせにおける上記の意味のクロス表における比率の積はそれぞれ、が0.49、

が0.36、 が0.45である。 この比率の積の大きさは、3つの設問における選択肢の組み合わせ360個(10.×4×9)の うち、それぞれ、6位、78位、14位である。

布置図による確認

各選択肢の位置を数量化 類によって定め、相互の位置関係を確認してみる。数量化類とは、質的データを量的データに変換する方法をとることによって、データに隠れている構造を求めるものである。ここでは、_問題点、4目標、9施策をすべて同レベルの変数として扱いながら、相互の位置関係によって分類できるか(Brian S. Evenitt and Graham Dunn 2001)を考えてみる。なお、4つのクラスターはこの布置図の座標をもとに、ユークリッド距離を用いて、合併後の距離を重心距離によって計算したものである。

2次元の布置で示すと図3の通りである。横軸は、正の方向が水を調査、負の方向が水の 浸透、縦軸は、正の方向が流量、負の方向が 水への親近性を示していると考えられる。累積寄与率よりもとのデータの16.4%の変動を説明したものに過ぎないが、上述の問題点目標 施策の選択肢に関して、 「水質」

「良質な水」 「水をきれいにする」、 「森林」 「県民の生活」 「水のむ だをなくす」、 「コンクリート化」 「自然との共生」 「自然とともに生きる」 の3つの組み合わせとも原点付近ではあるが、 相互に近い位置にプロットされていることが わかる。

4つのクラスターのうち、クラスター1が最 も多くの選択肢を含み、それぞれの選択肢の 回答数も多く、上述の3つの問題解決に関す る選択肢をすべて含んでいる。問題解決1の 問題点 目標 施策 と問題解決2の問 題点 目標 施策 および問題解決3の 問題点 目標 施策 であり、全体を通 して生活レベルにおける水との自然なかかわり、と解釈できるであろう。クラスター2は問題点と施策を含み、水への親近性と考えられる。クラスター3は、施策とともに、問題点を含み、水の状況なども含めた大きな観点からの調査ととらえられる。クラスター4は施策、と問題点を含み、水のフローとストックを関連させており、水の流量と解釈できるであろう。

上述の通り、この分析においては、10問題点、4目標、9施策の3つの変数群は同レベルで扱ったが、目標に関してはすべてクラスター1に含まれている。目標の各選択肢の選択比率が高いためと考えられる。問題点と施策に関しては、水循環においてそれぞれの選択肢がどういう位置づけとして回答者にとらえられているかが、各クラスターに反映してい

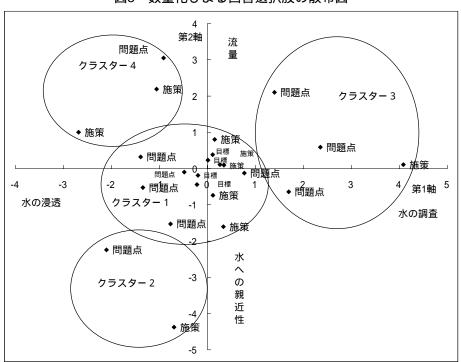


図3 数量化びよる回答選択肢の散布図

(出所)筆者作成

194 陸路 正昭

ると考えられる。

正準相関分析による確認

ここで、問題点 目標 施策のつながりを 正準相関分析によって、とらえてみる。この 分析法は、問題点の10変数、目標の4変数、 施策の9変数のそれぞれの3つの群における線 形結合による合成変数間の相関係数を最大に するように、合成変数をあらわす元の変数の 係数を決める手法である。最大の相関係数を 正準相関係数といい、このときの合成変数を 示す各変数の係数の大きさによって、2組の 変数間の関係を解析するものである。

まず、質問4の問題点と質問5の目標をそれぞれ変数群とする正準相関分析の結果は、

第1正準相関係数は0.30、第2正準相関は0.17と大きくはないが、サンプル数が多く、Bartlettの検定(Barbara G. Tabachnick and Linda S.Fidell 2000)により、第1正準相関係数、第2正準相関係数とも、ほぼ有意であるといえる。また、元データが質的変数(0と1)であるが、ここでは、計算上の支障はないものと判断した。第1正準変数は、第1変数群(問題点)のほぼ全変数と、第2変数群(目標)のうちの「良質な水」の関係を示している。第2正準変数は、第1変数群の「地球温暖化」、「河川の常時流量」と第2変数群の「県

上述の3つの問題解決の組合せをそのまま示すものではないが、第1正準相関によって、組合せの前半が表現されていると解釈できる。

民の生活」の関係を示している。

次に、質問5の目標と質問6の施策をそれぞれ第2変数群、第3変数群とする正準相関分析では、第1正準相関係数は0.29、第2正準相関は0.21と大きくはないが、Bartlettの検定により、第1正準相関係数、第2正準相関係数とも有意であるといえる。

第1正準変数は、第2変数群(目標)の「県民の生活」、「良質な水」と第3変数群(施策)の「災害に備える」、「水を貯める」、さらに、「水のむだをなくす」の関係を示している。第2正準変数は、第2変数群

の 「産業の活性化」と第3変数群の 「水 の活用」の関係を示している。

上述の3つの問題解決の組合せをそのまま、 示唆しているわけではないが、第1正準相関 によって、組合せ の後半が、かす かに、表現されていると解釈できる。

ここまでで確認できた3つの問題解決の組合せの意味内容から、それぞれ に「水を元に戻す」 に「水を養う」 に「水とともに生きる」という名前をつけることにする。

(2) 問題解決に自由記述の単語を組み込む

ここで、3つの問題解決の組合せごとに、 それぞれ該当する回答者の施策(質問6)と推 進方策(質問7)での自由記述を組み込む。

問題解決「水を元に戻す(このパターンの該当者192人の回答におけ る、施策(質問6)と推進方策(質問7)の自 由記述から検索できるKW(キーワード、以下 同)で頻度の高いものは、「水」が22回、「川」、 「必要」が9回、「人」、「大切」、「利用」が7回、 「行政」、「場所」、「生活排水」が5回、「雨水」 「家庭」、「気持ち」、「子供」が4回、「ゴミ」、 「下水道」が3回等である。「水」、「川」から 検索できる意見として、「工業の排水の浄化、 ゴミ等の川への投げ捨て(キャンプ残飯、釣 り具の針等)色々問題がある。それらは規制 があるはずと思うが野放し状態に近い。それ らの取り締まりを強化して、不法行為は認め ないという姿勢も必要である。」、「水辺環境 汚染の防止(四輪車の乗り入れ禁止、河川で の洗車禁止)」、などが挙げられる。これらと 記述内容から、生活レベルでの水の浄化が求 められていることがうかがえる。さらに、そ れをサポートする行政による規制を住民が求 めていることが推測できる。

問題解決「水を養う()」

該当者133人における施策(質問6)と推進 方策(質問7)の自由記述から検索できるKWで 頻度の高いものは、「水」が16回、「川」が7 回、「雨水」が5回、「自然」が5回、「コンク リート」が4回、「認識」が4回、「県政」が3 回、「行政」が3回等である。「水」から検索 できる意見として、「山林を針葉樹ばかりに せず根を張り保水力の大きい広葉樹林を作 る。河川流域に遊歩道や公園等を作れば川に ゴミを捨てる人も少なくなるのではないでし ょうか。」、「深い山にはいって、きれいな水 が力強く流れ、落ちてくる姿、それも絶え間 なくです。この自然の力を生かすこと、国民 が関心を持つよう情報を発すること、自然の 恵に対する(生かされていること、又こわさ を知らなさすぎます)知識がないというか無 関心であること、これらに目を向けてもらい たい。」、「都市部に降った雨水の有効活用。」 などが挙げられる。これらと記述内容から、 森林の荒廃に不安を感じている人が多いこと がうかがえる。都会に住んでいると気づきに くいことであり、正しい情報を伝えていく責 任が行政に求められるであろう。

問題解決「水と共に生きる()」

該当者110人における、施策(質問6)と推 進方策(質問7)の自由記述から検索できるKW で頻度の高いものは、「水」が16回、「川」が 9回、「自然」が7回、「きれいな」、「必要」が 6回、「コンクリート」、「子供」、「大切」が5 回、「ダム」、「県民」、「生活」が4回、「行政」 「水路」が3回等である。「水」や「コンクリ ート」から検索できる意見として、「水は予 想を超える大きな力を持ちます。時には恵み の力を時には凶器となって私達をおそれさせ ます。土は水によって運ばれ今の国土を築き ました。土は必ず崩れるものなのではないで しょうか?無理にコンクリートで川の両側を 固めても必ず崩されるものです。私は自然の 形又は、それに近い形をとどめられるような メンテナンスが必要だと思います。」、「水路 を集中させず、分散させることにより、豪雨 時の流水負担を軽減させると共に、生活用水 路としての機能を果たせる工夫(融雪、植物 への水かけなど)」などがあげられる。これ らと記述内容から、自然が失われつつあることへの不安が読みとれる。護岸工事などについても、岐阜県で進めている民間との共同研究による新しい工法なども含め、流域の住民の意向を反映した河川の維持が求められているといえよう。

(3) 問題解決をもとにした施策評価の例示

ここまでの分析結果をもとに、「住民参画の行政評価」(上山 2002)が行われるための検討をする。今日の自治体においては、「新しい行政評価システム」が求められているが、ここでは、今回のアンケートの分析をもとに、施策評価について、アンケートによる住民の声、岐阜県での既存の施策と評価指標、特に定量化した²指標、これらの、

を結びつけるために測定していくべき効用 としてのアウトカム指標[®]、の3点から検討を 加えていく。

具体的には、上述(2)の3つの問題解決から、施策における課題として、例1「生活レベルでの水の浄化(武田 2001)、例2「森林の荒廃(レスター・ブラウン 2002)への不安の解消、例3「川の風景から自然が失われる不安の解消(環境省 2002)」の3項目をとりあげ、そのための解決策として、それぞれ「行政による規制」、「正しい情報を伝えていく」(William James 1992)、「護岸工事のあり方を変える。」を設定し、上記、、を検討したものが表1である。

4. 施策評価推進のための自由記述 の分析

本章での分析のステップは、前述の図2の 通りであり、自由記述の分析によって、施策 評価を推進するための手がかりを得るもので ある。 196 陸路 正昭

表1 施策評価の3事例

事例	例1「生活レベルでの水の浄化」を 課題として、「行政による規制」を 解決策とする	例2「森林の荒廃への不安の解消」 を課題として、「正しい情報を伝え ていく」を解決策とする	例3「自然が失われることへの不安 の解消」を課題として、「護岸工事 のあり方を変える」を解決策とする
アンケートによ る住民の声	「各個人に出来ること。また、すべきことについて具体的にアピールする必要性有り。例:家庭における洗剤の種類の選定、流しのゴミ受け袋の使用普及・排水の利用・洗車の方	貯める」活山、森林の保全をもっと 上位に配置し、重点的な施策を。木 曽檜や貴重なブナ林の保全を。」	「家庭汚染が川にどのように影響するかは、目に見える川が身近にあってこそ理解できるものと思う。コンクリートでかくされた排水路ばかりでなく、本物の小川を多く作るべきだと思う。」
	法etc」 「日常生活の水の無駄遣いに対する 教育・宣伝等」などの声があげられ る。	「森林の手入れだと思います。 県から間伐促進を奨励していただきもっと山林保有者の意識向上を図っていただくようお願いします。」 の声などがあげられる。	「家のそばのくつわの川もホタルが飛びますが、ところどころ護岸がコンクリート化され、ホタルが少なくなっています。護岸工事も工夫が必要だと思います。」などがあげられる。
	岐阜県における「施策評価表」の 「健康によい大気・水・土壌環境作	「施策評価表」の「快適ネット(自 然環境、生活環境、文化環境)」に ついては、H12地域情報文化育成事 業の県実施機関及び市町村数7機 関+14市町村と測定されている。	「施策評価表」の「豊かな自然との ふれあい対策の推進」において、 H12自然公園等の面積18.4%に対し て、目標H17県土面積の19%が掲げ られており、H12自然ふれあい観察 会の開催日数・参加者数32回・931 人が測定されている。
岐阜県の既存の 施策と評価指標、 特に定量化した指標(注記2)	リの推進」について、H12公共用水域水質環境基準達成率92.8%と目標100%が掲げられている。 「下水道化の推進」について、H12下水道普及率63.2%とH22到達目標92%が掲げられている。	「人と自然の共生の森林づくりの推進」においては、H12広葉樹人工林面積1067ha、育成複層林面積9146ha、間伐面積10000haと算定されている。	「貴重な動植物の保全対策の推進」 において、H12県自然環境保全地域・緑地環境保全地域・緑地環境保全地域面積31地域・ 3283haが算定されている。
		「森林を支える人づくり」において、 H11森林技術者数1687人に対して、 H13目標2100人が掲げられている。 H12森林愛護隊登録者数598人に対してH17目標745人等が設定されている。	「自然環境と調和した公共事業の推進」において、H12河川整備計画設定数12プロック 目標達成率100%にたいして、H15目標14プロックが掲げられており、H13自然工法管理士認定数168人 目標達成率100%に対して、H18目標5000人が掲げられている。
上記の 、 を 結びつけるために 測定していくべき 効用としてのアウ トカム指標	例えば、「車所有者の洗車の方法で 川の浄化に反するような行為をして いる人の比率の低下」や「生活レベ ルでの水の浄化への普及活動が十分 であると感じている人の比率の向 上」などが考えられるであろう。	例えば、「あなたの周りの住民に森林の情報が正しく伝わっていると考える人の比率の向上」や「森林の荒廃の歯止めによって不安が減少したと感じる人の比率の向上」などが考えられる。	例えば、「川で遊んだことのある児童の世帯比率の向上」や「最近、外出した際に、河川の自然維持に配慮されていると見受けられるような場所を見かけた人の比率の向上」などが考えられるであろう。

(出所)筆者作成

(1) 自由記述の単語の分析

まず、アンケートの質問8「全体の意見」における自由記述を分析する。この問は、全体のことに触れており、記述量も最も多い。現段階では自由記述の文章の分析法は確立しているわけではなく、選択肢の回答などの分析と異なり、困難さの伴うものである。ここでは、水循環のための施策の手がかりを得るために、文章を分かち書きした単語の出現頻度をもとに分析を試みる。

問8への自由記述の回答は、269人分、9588 語とカウントできた。この中には、記号や、 句読点、助詞なども含まれ、繰り返し出現す る単語も数えられている。異なる単語の数は 2122単語である。

さらに、助詞や接続詞など(を、に、へ、及び、そして等)を削除した上で、頻度が2以上の単語に絞り、さらに、同じ意味の言葉を統一する(たとえば、「いい」を「良い」に、「お願いしたい」を「お願いします」に、「とっても」を「とても」に、「どちら」を

「どれ」などに変換する)ために語を置き換え、ひとまとめにした。この結果、全単語数が3696語で、実質の異なる単語数は559語となり、これらを今回の分析対象とした。50回以上の頻度の単語は、思う125回、水119回、する71回、ある64回、自然55回、いる51回であった。

(2) 回答者を分類してクラスターを作る

縦軸にサンプル269人、横軸に異なり単語559のマトリックスをつくり、サンプルと単語にスコアを与える対応分析をする(Brian S. Everitt and Graham Dunn 2001)。このマトリックスは、非常に疎である傾向があり(大隈2000)、多くの成分をとっても、累積寄与率があがらず、2軸までの累積寄与率が2.1%であるので、サンプルを6クラスターに分けてみる。

クラスター数については、その数を変化させたときの比較をするとともに、クラスター内の分散を計算して、クラスターのまとまり具合を考慮した。ここでは、特に、6前後のクラスターについて、解釈のしやすさから判断することによって決めた。

また、各クラスターのサンプル人数はサンプル1から順に、226人、14人、13人、7人、2人、1人である。合計人数は、分析の対象とする単語の頻度を2以上としたので、263人となっている。1人でクラスターを形成するものについては、除外する考え方もあるが、意味内容から、そのまま1つのクラスターとして残した。

ここでは、各クラスターに含まれるサンプル(回答者)の使用頻度の高い単語を検定統計量の順に並べ、それぞれの上位構成要素から判断できる名前をつけることにする。

クラスター1 = 「自然」(含まれる単語は「水」、「川」、「自然」、「思う」、「大切」、「水循環健全化大綱」、「いう」、「考える」、「人」など)

クラスター2=「他県との連携」(含まれる 単語は、「全国」、「岐阜県」、「水質」、「他県」 「知識」、「スペシャリスト」、「ノウハウ」、「願って」、「岐阜県民」、「汲み上げる」、「提供」、「地域」、「供給」、「情報」など)

クラスター3 = 「汚染」(含まれる単語は、「農薬」、「いない」、「生活排水」、「汚染」、「どうして」、「減少」、「豪雨」、「全然」、「される」、「下水道」、「基準」、「わからない」、「早急」、「必要」、「災害」、「実行」、「10年」など)

クラスター4=「治水・植生」(含まれる単語は、「山林」、「杉」、「近年」、「何とか」、「根」、「土砂災害」、「洪水」、「檜」、「ありました」など)

クラスター5=「目標達成の意識」(含まれる単語は、「高め」、「金」、「頑張って」、「必要性」、「聞く」、「保全」、「目標達成」、「いきたい」、「意識」、「一人」、「認識」、「もっと」、「10年」、「1つ」、「1つ」、「21世紀」、「PR」など)

クラスター6 = 「参加と要望」(「参加」、 「要望」、「森林」など)としてみる。

(3) クラスターによる施策評価の推進

問8の自由記述の回答者269人のデータについて、6つのクラスターへの所属と質問6の9施策への選択結果をクロス表にしたものから、対応分析を実施することによって同じ布置図にプロットすることができる。この図4における質問8全体意見の6つのクラスターとの位置関係から、質問6の9つの施策を実施する際の考え方のよりどころを得ることが可能になる。ここでは、さらに、上述の岐阜の施策評価表の中から関連するものを取り上げることによって、施策評価の推進の手がかりとしていく。

この分析における成分1、成分2の寄与率は、 それぞれ、46.3%、24.7%であり、累積寄与 率は71.0%である。

図4上のC1・・・C6とは、それぞれ、クラスター1・・・クラスター6を示し、同図の・・・ は、それぞれ、施策・・・施策を示すものとする。

198 陸路 正昭

成分1を横軸に、成分2を縦軸にして、横軸 は正の方向が自然な水、負の方向が水の影響 と対応、縦軸の正の方向が他県とのつながり、 負の方向が実現への働きかけと考えられる。

図4上で、原点から近い順に見ていく。

クラスター1 = 「自然」は原点に近く、これと接近しているのが施策の 「水をきれいにする」 「自然とともに生きる」 「水のむだを無くす」である。これらは、上述の3つの問題解決に組み込まれている施策であり、「自然」との関係を大切にする中で実施されるべきであるといえる。また、このクラスターに含まれる単語に「水循環健全化大綱」や「人」があり、「自然」との関係を大切にしながら、水循環健全化大綱に取り組むべきであることが示唆されている。

さらに、施策の 「水に親しむ」と 「水を浸み込ませる」も原点付近にあり、「自然」に目を向けながら、3つの問題解決を展開する際に、これらの施策も総合的に取り組むべ

きであると考えられる。岐阜県の施策評価表の「快適ネット(自然環境、生活環境、文化 環境)」は、こうした観点で取り組むべきで あろう。

次に、クラスター2 = 「他県との連携」は、施策 「水を調べる」に近い位置にある。水に関するデータ等が、県内だけでなく、広域的に共有されるべきことを示唆していると解釈できる。上述の岐阜県の施策評価表の「他県・市町村との連携(地方分権)」は、こうした水の問題と結びつけて実施されるべきであろう。

クラスター4 = 「治水、植生」は、施策「水を活かす」と近い位置にある。森林のあり方が川の保全や洪水の問題とともに、川の活用と結びつけられており、その位置から、クラスター3 = 「汚染」とも関連づけられることが示唆されている。このクラスター3は岐阜県の施策評価表の「安全ネット」と関連づけて実施されるべきであろう。

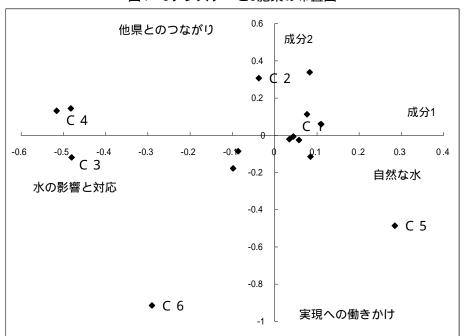


図4 6クラスターと9施策の布置図

(出所)筆者作成

なお、同じく、第3象限にあるのが、施策 「水を貯める」と施策 「災害に備える」 であるが、これらは、距離的には原点に近く、 「自然」と関連づけてとらえるべきであろう。

クラスター5 = 「目標達成の意識」、とクラスター6 = 「参加・要望」については、特に近い位置にある施策はないが、水循環を健全化する際に、考え方や行動面で欠かせない項目であることを示唆している。岐阜県の施第一番の「県民総参加の仕組みづくり」や「県民の意向の把握と世論の喚起」あるいは「協働事業・パートナーシップの構築」などはクラスター6に関連するであろう。これらは、水循環を正していくための条件が行政へ投げかけられているものであるが、パートナーシップの新しいあり方を求めているともいえよう(世古 1999)。

5. まとめ

岐阜県においては、8つの流域を有しており、水の循環に関する住民の意識は高いものがあった。今回の研究においては、特に、アンケートの自由記述の部分を活用した分析を試みたが、恣意的になる傾向があるところであり、今後、より客観的な指標に基づいた分析方法を確立していく必要がある。

本研究から指摘できることは、次の3項目である。

1)アンケートの選択肢の回答から意味のある問題解決の展開を考え出すことが可能である。

2)自由記述の文章を組み込むことによって、それぞれの問題解決に適する施策とその評価についての方法を探っていく糸口が得られる。

3)アンケートの選択肢と自由記述にあらわれた住民の意識をもとに、施策の優先順位や評価のための指標などを得ることによって、水循環健全化のための問題解決を進めていくことができる。

さらに、今後の課題として、

1)施策をどのように関連づけ、体系化していくか。施策相互の優先順位をつけながら、施策全体の整合性をつけていくことが求められる(地方行政活性化委員会 1994)。そのためには、水循環全体のビジョンの組織内における位置づけの明確化が不可欠であり、さらに、その達成状況が評価されるシステムを作る必要があろう。

2)どのようにすれば、計画段階だけでなく、評価の段階においても、住民の参加を促進していくことが可能か。アンケートの対象者のような比較的、高い意識の住民をベースにしていくことによって県民主体の行政の展開を試みる必要がある」。

の2点があげられる。

謝辞

最後になりましたが、本研究を支援していただいた岐阜県建設管理局の河川課の皆様に厚く御礼申し上げます。

注記

- 1 島田(1999)は「新しい行政評価システム」の特徴として、「事業の効果を測る、住民の主観的意識・効用を測る、客観的測定(数量化)、既存資源の活用、住民視点からの整理、行政視点と住民視点の統合、行政活動と住民満足の関係分析、行政評価指標の体系、行政判断に客観的な材料を提供」の9項目を挙げており、「新しい行政システムのイメージ」として、「行政活動のサイクルと新しい行政評価システム」を示しながら、9つの項目を説明している。
- 2 岐阜県の「県政の総合的基本方針」に基づき設定された評価の施策体系における、 岐阜県経営管理部行政管理課行政評価担当による施策評価システム(H12年度導入)より引用した。H13においては評価対象施策152施策、評価対象事業3530事業とされている。原文には施策の通し番号

200 陸路 正昭

が振られているが、ここでは省略した。

- 3 島田(1999)は、評価指標の対象による 分類として、活動量を測るインプット指標とアウトプット指標、成果を測るアウトカム指標をあげ、さらにアウトカム指標を測る指標と効用(満足を測る指標に分けて説る。より計で表現の指標を住民調査により計であると当れるものに分けでよれば、評価の中心調でありにより計測されるアウトカム指標の活用は今後の課題であると考えられる。
- 4 大隅(1994)によるハイブリッド法などがあり、ここでは、結合水準値の計算などを参考にした。クラスターの数を決めるルールは必ずしも明確に存在するわけではない。
- 5 大隈(2000)は、出現単語を頻度分布に よって検定したり、序列化することを示 している。
- 6 例えば、今回のアンケートの声を参考にして、施策間の優先順位や施策同志のつながりをどうつけていくか、検討すべきである。また、岐阜県の施策評価表の「健康によい大気・水・土壌環境づくりの推進」と「他県・市町村との連携」などをどう連動させるかなどの検討が必要である。また、組織間の連携(例えば、河海にを担当している行政管理課と農山村整備、大田の連携と農山村整備、大田の連携がある。とに、河川課と農山村整備、大田の連携があることに、河川課と農山村を開政策課、さらに、河川課と農山村整備政策課の連携がなども求められることになるう。
- 7 PDSサイクルの展開に住民を着実に組み 込んでいくために、住民の声を吸収して、 双方向にコミュニケーションを展開して いくことである。例えば、アンケートの 声に対して行政がどこまで対応している のか、あるいは率直に対応できない部分 はどのあたりなのかを明確に住民に示し ていくとともに、それらに対する住民の

評価を得ていくべきであろう。この際に、 岐阜県の施策評価表の「便利ネット(各 種生活サービス)」(H12県民情報ネット ワークアクセス件数266万件、H12予算編 成プロセス公開アクセス件数3853件など) も工夫次第では活用できるであろう。

参考文献

- 伊藤達雄他(2002)「流域論からみたローカルガバナンスに関する考察 矢作川を例として 」、『環境情報科学』、31(4):43・49岩松文代他(2001)「山村集落の活性化に関する合意形成と住民リーダー」、『日本林学会誌』、83(4):307・314
- 上山信一(2002)『日本の行政評価』、第一法 規
- 大隅昇(2002)「インターネット調査の適用 可能性と限界 データ科学の視点からの考 察」(大隅昇)「ブランドからの連想言語 の構造」、『行動計量学』、29(1):20.44
- 大隈昇(2000)「調査における自由回答データの解析」、『統計数理』48(2):339·376 大隈昇(1994)『記述的多変量解析法』、日科
- 小野達也(2002)「地方自治体の行政評価システムの課題と展望」、『日本評価研究』 2 (1): 29.37
- 鎌田徳幸(2002)「岩手県の政策評価と外部 評価委員会」、『日本評価研究』、2(2): 37.46
- 環境省編 (2002)『平成14年版環境白書』、ぎょうせい
- 国土交通省水資源部(2002)『日本の水資源』 財務省印刷局
- 佐藤慶他(2001)「参加型地域づくりへの一般住民の参加の質に関する調査研究」、『農村計画論文集』、3:241.246
- 島田春雄(1999)『行政評価』、東洋経済 世古一穂(1999)『市民参加のデザイン』、ぎょうせい
- 高橋裕(1988)『都市と水』 岩波新書 武田育郎(2001)『水と水質環境の基礎知識』

オーム社出版局

- 田辺智子(2003)「プログラム評価の手法外観」、『政策研究』、16(5): 22-7
- 地方行政活性化委員会 (1994) 『職員参加の目標による行政運営』、産能大学
- 日本水環境学会(1999)『日本の水環境行政』 ぎょうせい.
- 馬場仁志 (2002)「市民と行政との連携による仙台地域水循環再構築について」、『第6 回水資源に関する論文集』、343.348
- 原田昌幸(2000)「地球環境問題に対する住 民意識と意識啓発手法に関する研究」、『衛 生工学会論文集』、77:58.63
- 村瀬勝彦 (2002)「水資源計画における合意 形成プロセス」、『第6回水資源に関する論 文集』 31.36

- 森邦恵(2003)「北海道の公共投資に関する 一考察」。『日本評価研究』 3(1):77-88
- 龍慶昭、佐々木亮 (2000)『「政策評価」の理 論と技法』多賀出版
- レスター・ブラウン (2002)『エコ・エコノ ミー (「環境としての」人間と経済)』、家 の光協会
- Barbara G. Tabachnick and Linda S.Fidell (2000). *Using Multivariate Statistics 4th edition*. Boston: Allyn & Bacon.
- Brian S. Everitt and Graham Dunn (2001). Applied Multivariate Data Analysis. New York: Arnold Publishers.
- William James (1992). *Water, Development and the Environment*. Chelsea, Michigan: Lewis Publishers.

(2003.7.1受理)

What Measures Should Be Taken for the Water Cycle Based on the Consciousness of Residents?

Masaaki Mutsuro

Chiba-Keizai College m.mutsuro@dream.com

Abstract

With the development of decentralization and administrative reform, local self-governing bodies are being asked to enter into a partnership with residents. The administration must fully consider the environment, not just economics.

The purpose of this research is to determine measures for the water cycle by assembling a solution through a questionnaire analysis of prefecture Gifu on consciousness of the water cycle among residents.

Keywords

healthy water cycle, consciousness of residents, questionnaire, problem solving, measures for the water cycle

【実践・調査報告】

インドネシアにおけるジェンダー統計分野への協力の評価

浜野 敏子

国際協力事業団 hamano6@attglobal.net

要約

ジェンダー統計は、ジェンダー視点に立った政策策定、計画立案の基礎として欠かせないものである。

インドネシアでは、ジェンダー統計の不整備がジェンダー視点に立った政策策定の促進を妨げていることが認識された。そこで、国際協力事業団(JICA)はインドネシア政府に対するジェンダー統計システム整備の協力(以下ジェンダー統計プロジェクト)を2000年2月から2002年3月に実施した。

ジェンダー統計プロジェクトの上位目標は、ジェンダー統計が政策策定や計画立案に活用されることであり、直接の目標は、1)ジェンダー統計利用者と生産者の連携促進 2)利用者の能力向上 3)生産者の能力向上 4)ジェンダー統計の入手可能性の改善とした。実際の協力は、ニーズアセスメント、ジェンダー統計研修、ジェンダー統計集の生産・出版、ジェンダー統計集の普及、ジェンダー統計集のレビュウを通じて行われた。

本稿では、まず直接目標にそってその成果と実施プロセスを確認し、さらに提示された統計の政策策定への活用可能性を検討した。そして、このプロジェクトがジェンダー統計を活用した政策策定の促進に貢献したかどうかについての評価を行った。その結果、プラスの外部要因の上に政策策定に必要なジェンダー統計生産の具体的体制あるいはモデルを形成したことが認められた。

キーワード

ジェンダー統計、政策策定、連携

1. はじめに

社会における女性と男性の状況、その関係性を示す統計情報は、社会におけるジェンダー偏見を可視化し、ジェンダー平等と公平をもたらす効果的な政策策定の基礎となる(Beck 1999, p. 7)。

1995年の第四回世界女性会議で採択された 北京行動綱領の戦略目標「H. 3企画と評価の ためのジェンダー別データと情報の作成と配 布」では、さまざまなジェンダー統計に関す る重要事項が表明されている。たとえば、統 計を性別に収集・編集・分析・提示するこ と、ジェンダー統計プログラム強化のための 職員を配置すること、ジェンダーに関する統計出版部物を定期的に発行すること、統計生産者と利用者が政府統計システムの妥当性と必要なジェンダー課題を取り上げているかを定期的に検討すること、政策立案や事業実施においてジェンダーに敏感なデータをより活用すること、などである。(北京行動綱領1995, Q/CONE,177/L5)。

ジェンダー格差とジェンダー課題についての体系的な調査を可能にする性別データとしてのジェンダー統計(Vaidayanathan 1999)を生産するための体制は、多くの発展途上国でまだ整備されていない。統計情報なしに政策が策定され、またモニタリングや評価のための指標も確立していないのが現状である。その理由は、利用者と生産者の連携の欠如、利用者のジェンダー別統計に対する低い意識、生産者のジェンダー課題についての無理解、統計の質や入手可能性の低さの問題などが指摘されている。(Headman 1996, pp. 9-10)

インドネシアでもジェンダー統計の不整備がジェンダー視点に立った政策策定の促進を妨げていることが認識された。そこで、JICA専門家としてインドネシアの女性のエンパワーメント担当国務大臣府(以下、女性のエンパワーメント省)にジェンダー主流化支援アドバイザーとして1999年から2002年にかけて派遣されていた筆者は、女性のエンパワーメント省と中央統計局とともにジェンダー統計整備を2000年に開始した。

本稿では、インドネシアにおけるジェンダー統計整備へのJICAの協力(以下、「ジェンダー統計プロジェクト」)の成果と実施プロセスを確認し、さらに提示された統計の政策策定への活用可能性について検討したうえで、このプロジェクトがジェンダー視点に立った政策策定の促進に貢献したかどうかを評価する。

2. インドネシアの国内本部機構 と女性政策

女性の地位向上のための国内本部機構である女性のエンパワーメント省の任務は2000年に公布された大統領令」によると、女性のエンパワーメントおよび子供の福祉と保護のための国家政策策定、関連省庁へのジェンダー視点に立った政策策定、事業計画とそのモニタリング・評価への助言、広報啓発、大統領への評価報告と提言、と定められている。

インドネシアの女性政策は、表1に見られるように1970年代から1990年代にかけて女性への福祉中心のプロジェクトアプローチであったが、1990年代後半から国家開発の全ての分野²における政策策定、計画の立案と実施、モニタリング・評価へジェンダー視点を組み込むアプローチへその焦点は移行している。1999年に作成された国家政策大綱、2000年2004年の国家開発計画、2000年に公布されたジェンダー主流化に関する大統領告示³において、そのことが表明されている。国際的には、インドネシアは1984年に「女性に対する全ての差別撤廃条約」、1995年には「北京行動綱領」を採択している。

上記のようなアプローチの移行にともない、図1のように女性のエンパワーメント省はセクター別(教育、労働、保健、文化)から機能別体制(政策・情報、ジェンダー主流化、女性の生活向上、子供の福祉、地域参加)へ再編成された。新体制のもとで、ジェンダー統計は副大臣Iの政策・情報の部署で扱われることとなった。

3. インドネシアのジェンダー統計 の動向

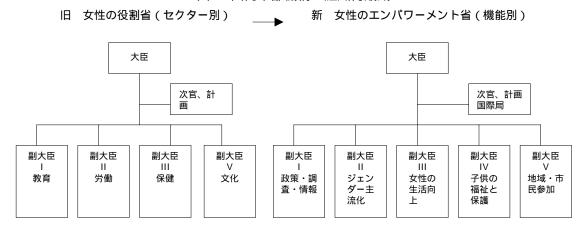
1980年代後半から1990年代にかけて、国際的にジェンダー統計への理解が促進し、ジェンダー統計集の作成・発行が各国で進んだ(伊藤 2002,p.3)。それを追うようにインドネ

表1 インドネシアの国内政策と女性政策、および国際的動向

1	ンドネシア国内政策	インド	ネシアの女性・ジェンダー政策		国際的動向
1945 1966	インドネシア憲法成立 スハルト政権樹立	1945	憲法に男女平等が明記		
1974	婚姻法成立	1974	婚姻法に前提としての男女 の役割分担明記	1975 1979	メキシコ世界女性会議 国連女子差別撤廃条約採択
		1978	女性の役割担当大臣府設立 女の福祉、WID中心の政策	1980	「国連女性の10年」世界会議(コペンハーゲン)
		1984	「女性差別撤廃条約」批准	1985	「国連女性の10年」世界会議
		1985	「雇用機会における差別禁		(ナイロビ)
1998	スハルト政権崩壊、民		止条約」批准	1995	世界女性会議(北京)
	主化促進	1995	「北京行動綱領」を採択		「北京宣言および行動綱領」
1999	ワヒド民主政権樹立				主要戦略としてのジェンダ
1999	国家政策大綱制定	1999	女性のエンパワーメント担		ー主流化アプローチ
			当大臣府に改称 (GAD中心		
			の政策へ転換)		
2000	国家5ヵ年開発計画開	1999			
	始		女性のエンパワーメントが 明記		
		2000	国家 5 ヵ年開発計画におい		
			てジェンダー視点にたった		
			施策と事業の強調	2000	国連特別総会「女性2000年
2001	地方分権化が施行	2000			会議」
			大統領告示が公布		

WID:Women in Development (出所)筆者作成(2002) GAD: Gender and Development

図1 国内本部機構の組織再編成



(出所) インドネシア女性の役割担当国務大臣府ならびに女性のエンパワーメント担当国務大臣府組織図

シアでは、1990年代後半からUNIFEMの中央 統計局へのジェンダー統計支援が始まった。 1997年と2000年には一般利用者向けのジェン ダー統計小冊子が刊行され、2001年には国レ ベルの研究者や計画立案者向けのジェンダー 統計集が刊行された。その他、教育省では教 育分野における男女別統計を1990年代後半に 刊行した。また、ジェンダー統計について NGOへの研修なども行われた。このように ジェンダー統計について、さまざまな努力が 行われていたが、それらは主に生産者主体の ものであり、利用者との連携を基礎にしたジ ェンダー統計システムは構築されていなかっ た。そして、生産されたジェンダー統計が利 用者に十分に普及されておらず、したがって 政策策定あるいは計画立案にも活用されてい なかった。

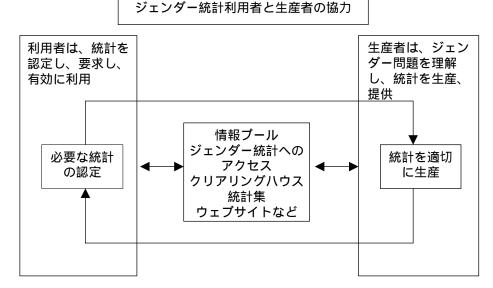
4. ジェンダー統計プロジェクトの 概要

ジェンダー統計プロジェクトは、2000年2

月から2002年3月まで行われた。活動主体は

女性のエンパワーメント省と中央統計局であ る。上位目標は、ジェンダー統計が政策策定、 計画立案、モニタリング・評価に活用される こととし、直接の目標は、(1)統計利用者と 生産者の連携促進、(2)利用者の能力向上、 (3) 生産者の能力向上、(4) ジェンダー統計 の入手可能性の改善(普及の拡大) とした。 これら4つの目標は、図2のジェンダー統計 システムの枠組みをもとに設定された。枠組 みは、1) 左ボックスの利用者が必要な統計 を確定し、右ボックスの生産者に要求する、 2) 生産者はジェンダー課題を理解し、要求 された統計を適切な方法で生産する、3)そ の生産された統計を中ボックスの情報プール (クリアリングハウス、ジェンダー統計集、 ウェブサイトなど)をとおして幅広い利用者

図2 インドネシアのジェンダー統計システムの枠組み



(出所)大竹美登利作成(2000) (Surbakti, Soedarti(2000)より) に提供する、4)これら一連のシステム全体を上ボックスの統計利用者と生産者の連携体制によって管理運営・調整する、というものである。

図3はこのプロジェクトで実施された一連の活動である。その支柱となったのは「連携」と「参加」であった。まずジェンダー統計利用者の代表である女性のエンパワーメント省と生産者の代表である中央統計局が中心となって中央レベルでジェンダー統計システム全体を管理運営する運営委員会と日常的に実践してゆくテクニカルワーキンググループを作った。初年度には中央レベルのセクター省庁職員もの能力向上を目的としたジェンダー統計ワークショップの企画、運営を行った。

次年度にはこの経験を基礎に、12州®を対象にした州レベルでのジェンダー統計ワークショップを企画・運営した。このワークショップには各州から統計局、女性研究センター、女性局(あるいは部)の職員が参加した。ワークショップの前半は、ジェンダー統計に関する基礎知識の講義を行い、後半は各州チームがそれぞれの州のジェンダー統計プ

ロファイル (人口動態、教育、保健、世帯・家族、経済、政治・行政・司法、環境、女性に対する犯罪に関するジェンダー統計を収集・編集したもの)原案を作成した。各州のジェンダー統計プロファイルの監修には、中央統計局あるいは女性のエンパワーメント省の上級職員があたり、州チームとの協議を繰り返した。

中央レベル 問題把握とニーズアセスメント 運営委員会とテクニカルワーキンググループの設立 中央セクター省庁職員へのジェンダー統計の研修(12州) 州レベル 州別ジェンダー統計プロファイル作成ワークショップ 州別ジェンダー統計プロファイルの発行・配布(12州) ジェンダー統計プロファイル普及セミナー開催(12州) ジェンダー統計プロファイルのレビュウ(2州)

図3 ジェンダー統計システムプロジェクトの活動の流れ

(出所)筆者作成(2002)



図4 バリ州ジェンダー統計プロファイルの レビュウの結果。によるものである。 表紙

者による統計を用いたその州のジェンダー課 題についての発表、参加者との質疑応答と協 議、そして最後に今後の課題と提言によって セミナーは締められた。

以上のように活動の各段階において異なる 組織、あるいは中央政府と地方政府が連携し、 各関係者が積極的にそれぞれの役割を果た し、プロジェクトが展開された。

5. プロジェクトの成果と実施プロ セスの確認

プロジェクトの成果と実施プロセスで何が 起きたかを確認するために、当初設定された 4つの目標に照らしてプロジェクト以前と以 後の状況の変化を概観したものが表2である。 プロジェクト以前の状況については2000年2 月に行ったニーズアセスメントでの結果を用 い、以後については各州で行われたジェンダ ー統計普及セミナー報告書®と2003年3月の終 了時に行ったジェンダー統計プロファイルの

プロジェクト第1の目標であるジェンダー 統計の利用者と生産者の連携については、中 央レベルでは利用者の代表である女性のエン パワーメント省と生産者の代表である中央統 計局で構成された運営委員会とテクニカルワ ーキンググループがワークショップの企画・ 運営からジェンダー統計プロファイルの出 版・普及までをその連携体制のもとで行っ た。また州レベルでは、州統計局、女性研究 センター、州女性局(あるいは部)の3者が チームを組み、ジェンダー統計プロファイル を共同執筆し、ジェンダー統計プロファイル 普及セミナーを共同開催した。このように具 体的な活動を通して連携は強化・促進してい った。また、ジェンダー統計プロファイル普 及セミナーでは、これまでほとんどジェンダ 一統計に接したことのなかった地方議員、州 政府職員、ジャーナリスト、あるいはジェン ダー統計やジェンダー課題に関心の深い研究 者やNGO職員、そしてジェンダー統計の活 用を促進しようとする女性のエンパワーメン ト省や中央統計局など様々な立場の利用者と 生産者が対話の機会を持った。その結果、ジ ェンダー統計プロファイルに欠損しているデ ータやその出所、統計の定義や収集方法の問 題点、質的情報の必要性、ジェンダー統計プ ロファイルの作成過程におけるNGOや女性 組織の参加の重要性が協議された。さらに、 参加者からジェンダー統計改善のための定期 的な協議が提案されるなど、積極的な姿勢が みられた。

第2の目標であるジェンダー統計利用者の 能力向上については、プロジェクトの責任者 (活動主体)でもある女性のエンパワーメン ト省がジェンダー統計の重要性を一層認識 し、プロジェクト終了後に独自でジェンダー 統計プロファイル生産のための予算措置と人 員配置を行い、中央統計局との連携のもとで 新たに12州のプロファイルを完成させた¹⁰。 女性のエンパワーメント省のこのような積極 的な態度を動機づけた主な理由は、ジェンダ 統計プロファイルという具体的な道具とし

表2 ジェンダー統計整備プロジェクトの以前と以後の比較

目標	プロジェクト開始前	プロジェクト開始後
1)利用者と 生産者の連携 体制の強化	個人レベルでの協力はあったが、組織 としての連携体制はなかった。 幅広い利用者と生産者の対話の機会は なかった。	中央レベルでは、女性のエンパワーメント省と中央統計局の連携体制が確立した。 州レベルでは、州統計局、女性研究センター、州女性局の連携が促進した。 幅広い利用者と生産者(地方議員、中央政府職員、州政府職員、研究者、NGO職員、ジャーナリストなど)の対話が促進した。
2)ジェンダー統計利用者の能力向上	必要なジェンダー統計の認定は行なわれていなかった。 女性のエンパワーメント省のジェンダー統計へのコミットメントは高くなかった。 一般の政策策定者や計画立案者のジェンダー統計についての関心や要求は低く、したがってそれを活用することはほとんどなかった。	必要な統計データの確定作業を開始した。 女性のエンパワーメント省はジェンダー統計プロファイル生産のための予算措置と人員配置を確立し、独自でプロジェクト終了後に12州のプロファイルを完成させた。 ジェンダー統計プロファイル普及セミナー参加者からのアンケート回答のうち93%がジェンダー統計の活用の意思を表明した。
3)ジェンダー統計生産者の能力向上	中央統計局のジェンダー配慮は行われていたが、主流ではなかった。セクター省庁のジェンダー統計生産は、教育省以外は非常に少なかった。男女別に収集しているが集計していない、あるいは集計しているが公表していない統計があった。 概念や収集方法の見直しが必要な統計があった。	中央統計局の上級職員のほとんどがジェンダー問題を理解し、統計におけるジェンダー配慮の重要性を認識した。中央統計局は包括的情報システムに男女別項目を挿入した。 州レベルの既存統計の再集計や再分類により新たなジェンダー統計を生産した。
4) ジェンダ・統計入手可能性の改善	ジェンダー統計が分散している、あるいは公表されていないため、入手が困難であった。 刊行されたジェンダー統計集は全国レベルの小冊子や教育分野などに限定されていた。 刊行物の普及が十分ではなかった。	ジェンダー統計情報のインベントリーが生産された。 12州の州レベルジェンダー統計プロファイルが刊行され、計7500冊が配布された。- 県レベルでのジェンダー統計プロファイルが作成された。ホームページやCD-Romによるジェンダー統計の入手が可能になった。

(出所)筆者作成(2002)

ての有効性、そしてジェンダー統計をとおしてジェンダー課題が明確に認識できたという利用者からの肯定的な反応である。たとえば、全州知事が集まる会議で女性のエンパワーメント省大臣が12州のジェンダー統計プロファ

イルを用いてジェンダー課題とそれに対する 政策の必要性を提言したところ、プロファイ ルをまだ作成していない州から自分達の州で もこのプロファイルを作成したいという要望 が急増した。また、女性のエンパワーメント 省副大臣は、州や県の政策策定者や計画立案 者がジェンダー統計をそれぞれの地域開発計 画に利用していきたいという声を聞き、ジェ ンダー統計はジェンダー主流化への具体的な 出発点になると認識するに至った。実際にジ ェンダー統計プロファイル普及セミナーの参 加者へ行ったアンケートでは、回答のあった 539人のうち93%がジェンダー統計プロファ イルを今後の仕事に利用していきたいと答え ている。また12州のいずれのセミナーでも、 州政府や県政府職員から県レベルのジェンダ 統計プロファイルの必要性が指摘され、女 性のエンパワーメント省にその作成支援が要 請された。ランプン州のある県ではジェンダ 統計プロファイルを独自の予算で生産し、 また東ジャワ州では州政府が予算を割り当て 全県の政府と協力してジェンダー統計プロフ ァイルを作成中である1。

第3の目標であるジェンダー統計の生産者 の能力向上については、以前から中央統計局 ではUNIFEMの支援やある上級職員の努力に よりジェンダー配慮は進んでいたが、このプ ロジェクトを通して局のほぼ全ての上級職員 がジェンダー統計プロファイルの編集を担当 することになり、それぞれがジェンダー課題 を理解するようになった。そして、中央統計 局が開発した統計情報システム に各統計デ ータの段階別(収集、集計、出版)男女別項 目を導入した。このことにより中央統計局へ の統計の報告を義務付けられているセクター 省庁では全ての統計を男女別に集計しなけれ ばならなくなる。たとえば村レベルで男女別 に収集されているが県、州、中央へ上がって いく過程で男女混合に集計されてしまう保健 省の統計(死亡率、傷病率、外来患者数、入 院患者数、乳幼児の栄養状態など)や社会福 祉省の統計(孤児数、ホームレス数、障害者 数、薬物中毒者数など) などそれらに当ては まる12。

また、ジェンダー統計プロファイル作成に あたって州レベルでジェンダー視点にたった 既存統計の再集計や再分類を行ない、新たな 統計や統計表を提示した。たとえば男女就業 者の教育レベルと平均賃金のクロス表、産業分野別や職業別の男女就業者比、政党別男女議員数、行政職員や行政管理職の男女比、司法職の男女従事者比、避妊方法別の男女の利用割合などである。

第4の目標であるジェンダー統計の入手可 能性の改善については、表2にあるようにプ ロジェクト以前に刊行されたジェンダー統計 刊行物は全国レベルの一般向け小冊子、また は教育分野のものなど、その利用の範囲が限 られていた。今回作成された州別ジェンダー 統計プロファイルは、インドネシアの地方分 権化政策13に伴い重要性が増した地方レベル での主要政策分野の統計(人口動態、教育、 保健医療、世帯・家族、経済、政治・行政・ 司法、環境、女性に対する犯罪)を包括的に 収集・編集した。計7500冊の各12州のジェン ダー統計プロファイル14を各関係者に配布し た。また、これらジェンダー統計プロファイ ルを配布するだけでなく、活用を促進するた めの1日セミナーを各12州で開催し、1州あた リ100人から80人、計1060人以上の利用者が 参加した。これによって、州レベルの利用者 のジェンダー入手可能性は高まった。

以上のようにプロジェクト前後の比較をとおして、1)各レベルでの生産者と利用者の連携が強化・促進したこと、2)利用者のジェンダー統計への関心が高まり、重要性が認識され、ジェンダー配慮した政策策定や計画立案を促す態度が生じたこと、3)生産者のジェンダーへの理解、知識が向上し、既存統計のジェンダー配慮が促進したこと、4)ジェンダー統計プロファイルの刊行・配布・普及によってジェンダー統計への入手可能性が改善したこと、が確認できた。

6. 提示された統計の政策策定への 活用可能性

ジェンダー統計プロファイルで提示された 統計がジェンダー視点に立った政策策定に活 用され得るかどうか、つまりジェンダー統計 210 浜野 敏子

としての妥当性があるかどうかについて、検討する。政策策定に活用するためには、統計がジェンダー課題を明示し、ジェンダー・バイアスやジェンダー格差の発生の原因や結果についての情報を提示する必要がある(伊藤2001、p. 54)。このような観点から、ジェンダー統計プロファイルのレビュウから得た利用者の意見をもとに以下の3例を取り上げる。

インドネシアでは、女性は男性に比べてその労働の価値が適切に評価されていないという問題が認識されている(Indonesia Country Report 2001: p. 10)。西ジャワ州では、女性の経済的労働力率は30.3%と男性の68.9%より低い。反面、失業率では女性は9.4%と男性の7.4%より高い。図5から女性労働力をと関けの7.4%より高い。図5から女性労働力をと関けの7.4%より高い。図5から女性労働力を大きの後減少と対しまされている。都市と農村におけるのポークがあることが示されている。都市の場合、女性は20-24歳に約45%のピークがありその後減少し横ばいに向かうが、農村におりその後減少し横ばいに向かうが、農村においては年齢とともに徐々に上昇し40-49歳で、大は年齢とともに徐々に上昇し40-49歳で、大は年齢とともに徐々に上昇し40-49歳で、大は年齢とともに徐々に上昇し40-49歳で、大は年齢とともに徐々に上昇し40-49歳で、大は年齢とともに徐々に上昇し40-49歳で、大は年齢とともに徐々に上昇し40-49歳で、大は年齢とともに徐々に上昇し40-49歳で、大は年齢とともに徐々に上昇し40-49歳で、大は年齢とともに徐々によりで、大きないる。

は困難であり、特別な技術を習得していない これら女性の多くはインフォーマルセクター に参入する。農村では、既婚女性にとって家 事と家計を支える経済活動の二重労働は当た り前であり、雇用機会が少ない農村では就業 の場はやはりインフォーマルセクターであ る。このインフォーマルセクターへの労働力 率を表3でみると、男性53.0%、女性53.8%と 大差はないが、このうち無報酬の家族従業者 率では女性が26.6%、男性が3.8%と差がある。 このようにインフォーマルセクター内にも就 業上の地位において男女格差がある。また、 フォーマルセクターにおいても表4に示され るように、男性労働者の平均賃金100に対し て女性は57.4(初等教育レベル)と低い。こ の賃金格差を引き起こす原因ともなる男女の 教育レベルを表5からみると、学校に行った ことのない女性の割合は農村では13.9%と男 性の5.8%より高い。これらの統計データか ら女性労働者における低い教育、未熟練労働、 不安定な就業上の地位、低賃金・無報酬とい う関連が示唆される。

このような状況はすでに国レベルで認識されており、女性雇用労働者の労働条件の改善、

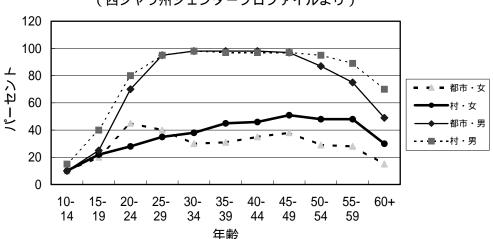


図5 男女別・年齢別・都市・農村別の労働参加率、2000年 (西ジャワ州ジェンダープロファイルより)

出所: BPS Propinsi Jawa Barat (2000)

訳:筆者

(1) (3/1) 1/3/1/06/9)								
就業セクター・地位	農村		都	市	計			
が来でクラー・地位	女(%)	男(%)	女(%)	男(%)	女(%)	男(%)		
フォーマルセクター(計)	60.8	59.4	34.8	36.9	46.2	47.0		
雇い主	0.9	2.7	0.4	1.6	0.6	2.1		
雇用者	59.9	56.7	34.4	35.3	45.6	44.9		
インフォ - マルセクター(計)	39.2	40.6	65.2	63.1	53.8	53.0		
従業員なし自営業	17.6	27.4	15.9	31.4	16.7	29.6		
従業員あり自営業	7.1	10.7	13.1	26.9	10.5	19.6		
無報酬従業員	14.4	2.5	36.2	4.8	26.6	3.8		

100.0

100.0

表3 男女別・就業セクター・都市・農村別の労働力率、2000年 (西ジャワ州ジェンダープロファイルより)

計 (出所) BPS Propinsi Jawa Barat (2000)

訳:筆者

注: 筆者 フォーマルセクターは登録されている事業主とその雇用者 インフォーマルセクターは上記以外のすべての就業者

100.0

表4 男女別・教育レベル別、一ヶ月あた平均賃金2000年(西ジャワ州ジェンダープロファイル)

教育レベル	女(ルピア)	男(ルピア)	賃金比
未就学	33,147	57,706	57.4
小学校中退	36,472	70,047	52.1
小学校卒業	53,190	91,138	58.4
中学校卒業	226,863	149,574	151.7
高校卒業	344,230	304,363	113.1
専門学校	313,068	326,597	95.9
資格取得学校	436,705	475,124	91.9
短大	510,432	558,278	91.4
大学	590,254	762,547	77.4
平均	134,301	163,676	82.1

(注)賃金比は男性を100とした場合の女性の比率 (出所)BPS Propinsi Jawa Barat (2000)

訳:筆者

適正賃金の設定、ILO条約の順守などの目標が国家政策として設定されている(国家開発計画2000、p. 30-31)。しかし、インフォーマルセクターにおける女性労働の社会経済開発における貢献、その正当な報酬を確保する法的整備、労働条件改善のための施策については、ほとんど関心が払われてこなかった。地方分権化にともない、西ジャワ州が中央に率先してこれらの課題について、独自の政策策

定や法律を制定することが可能である。 (The World's Women 2000,p. 122)。

100.0

100.0

100.0

第2の例は、女性世帯主世帯の農業生産に ついてである。農村の女性世帯主世帯の多く は貧困層に属し、一般に貧困削減プログラム の受益者とみなされている。しかし、表6の ジャンビ州のデータは、女性世帯主世帯が男 性世帯主世帯に比べ、生産性の高い農業を行 っていることを示している。ジャンビ州では 女性世帯主の管理土地面積と耕作面積は男性 世帯主の半分しかないが、ヘクタール当たり の生産額はほぼ同じであり、さらにヘクター ル当たりの収益は男性世帯主世帯の2.846.135 ルピアに比べ女性世帯主世帯は3.112.837ルピ アと高い。このような傾向は、他の州でも同 様に見られた。この州では女性世帯主世帯の 47.2%が2人以下の家族数であり、限られた 労働力の中で何がその生産性の高さの要因と なっているのかについて、農機具や農薬・肥 料の使い方、生産作物の種類、収穫方法、マ ーケティングなどについてさらに分析する必 要がある。インドネシアでは零細農民の生産 性向上が農業分野の主要課題の一つでもあり (国家開発計画2000、p. 32) この統計に基づ いてジャンビ州での効率的な小規模農業につ いての政策策定、あるいは農業普及や研修計 212 浜野 敏子

表5	男女別・教育レベル別・都市・農村別	10歳以上の就学率、	2000年
	(西ジャワ州ジェンダープロ・	ファイルより)	

物会し がリ	都市			農村		
教育レベル	女(%)	男(%)	計(%)	女(%)	男(%)	計(%)
未就学	5.3	1.9	3.6	13.9	5.8	9.3
小学校	19.2	15.6	17.4	30.9	29.8	30.3
中退小学校卒業	33.0	30.7	31.8	43.1	45.2	44.2
中学校卒業	18.0	19.3	18.6	8.7	11.0	9.9
高校卒業	20.1	26.4	23.3	3.8	7.3	5.6
短大・大学卒業	4.4	6.1	5.3	0.6	0.9	0.7
	100.0	100.0	100.0	100.0	100.0	100.0

(出所)BPS(2000)

訳:筆者

表6 男女世帯主世帯別 平均管理土地面積と耕作面積および所得、1999年 (ジャンビ州統計プロファイルより)

項目	女性世帯主世帯	男性世帯主世帯
管理土地面積(平方m)	3,381	6,449
耕作面積(平方m)	2,249	5,189
耕作率(パーセント)	66.2	80.7
ヘクタール当たりの生産額(ルピア)	3,779,797	3,332,165
ヘクタール当たりの収益ルピア)	3,112,837	2,846,135

(出所)1998・1999米と他の農作物の費用構造調査

訳:筆者

画立案へのヒントとして活用することが可能 性である。(Filling the Gap 1999, p. 5)。

第3の例は、家族計画への男性参加についてである。表7は南スラウェシ州の家族計画プログラムで利用されている避妊方法をジー視点から再分類したものである。農村では避妊の96.5%は女性が利用するる方法、2.9%が伝統的方法、わずか0.6%が男性がはは2.8人と減少し、家族計画の主要方針はは2.8人と減少し、家族計画の主要方針はによりに含まれた。現在、全国を出りである。現在、全国では2.8人の国家政策に含まれた。現在、全国では2.3%の男性参加を8%まで引き上げのスの男性参加を8%まで引き上げで家目標が設定されている(国家開発計画 2000、中. 121)が、南スラウェシ州でも独自の数値

目標を設定し、モニタリングしていくことが 可能である。

避妊方法	都市(%)	農村(%)
南スラウェシ		
男性が利用する方法	1.5	0.6
(精管切除、コンドーム)		
女性が利用する方法	96.3	96.5
(インプラント、注射、IUD、ピル) 伝統的方法	2.2	2.9
12,11/07/37/21	2.2	2.9
インドネシア 男性が利用する方法	1.7	1.0
(精管切除、コンドーム)	1.7	1.0
女性が利用する方法	96.5	97.1
(インプラント、注射、IUD,ピル)		
伝統的方法	1.8	1.9

表7 都市・農村別 避妊方法の利用率、2000年 (南スラウェシ州ジェンダープロファイルより)

(出所) BPS (2000)

訳:筆者

7. プロジェクトの全体評価と今後の課題

このプロジェクトが、ジェンダー統計を活用した政策策定の促進に貢献したかどうか、 という視点から全体を評価する。

地方分権化とジェンダー主流化という国家 政策が進められるなか、州レベルにおいてジ ェンダー視点に立った政策策定に活用可能な 統計を生産・提示出来たことは、妥当性の面 から評価される。さらに実践活動を通して向 上した統計生産者と利用者のジェンダー統計 に関する能力と両者の連携体制の確立はジェ ンダー統計生産の基盤形成としての有効性が ある。そしてプロジェクト終了後に活動主体 であった女性のエンパワーメント省と中央統 計局が継続して他州のジェンダー統計プロフ ァイルを生産・出版したことはオーナーシッ プの向上と制度上と財政上の自立発展性とい う面で評価される。このプロジェクトの具体 的成果は州レベルのジェンダー統計生産と普 及であったが、実際に活動を行った主体は国 レベルであり、この経験は国レベルでのジェ ンダー統計生産にも適用されること、また州 に続いて県レベルでもジェンダー統計集作 成・出版の動きが活発化したことはプロジェクトの効果(インパクト)として、評価できる。外部要因として、インドネシア全体にはけるジェンダー主流化の大きな流れ,それに伴う女性のエンパワーメント省におけるもは高いでも、その組織再編成、そして中央統計局を紹心による地方の権化による地方政府の政策による地方の権限拡大がプロジェクトの効果発見に必要なジェンダー統計生産の具体的体制あるいはモデルを形成することが出来たといえる。

このように、ジェンダー統計プロジェクトはジェンダー視点にたった政策策定に向けて、多極的な面から貢献した。しかし、まだ多くの課題が残っている。ジェンダー統計の改善にむけては、セクター省庁での男女別統計の収集・集計・公表の徹底、統計の概念や収集方法の見直し、無報酬労働や女性に対する暴力など欠損情報の収集、NGOを含めた幅広い情報源からのデータの収集、より多くの利用者の入手可能性をあげるためにウェブサイトでのデータの公開¹⁵(伊藤2001、p. 12)やCD-ROMの作成・配布、利用者の統計利用

の能力向上、生産者のジェンダー意識の向上、 そして利用者と生産者の連携制度の強化など が、具体的にあげられる。

さらに、ジェンダー統計の改善と同時にジ ェンダー統計を実際に政策策定へ活用する具 体的取り組みが重要である。インドネシアで はカナダ国際援助庁(CIDA)の支援を得て、 ジェンダー視点にたった政策評価の方法論が 開発された。それは、認知された政策課題に ついてジェンダー情報(統計と質的情報)を 用いて分析し、政策を評価し、修正された政 策目標を達成するための指標を設定し、政策 の実施計画を作成するものである (Rosalin2001)。これまで、教育や農業分野 などでこの方法論の実践が試みられたが、ジ ェンダー統計の入手困難が一つの障害にな り、なかなか一般には普及していかなかった。 ジェンダー統計の収集・集計の基盤が整い、 必要とされる基本的なジェンダー統計が提示 され、利用者の関心も高くなった現段階で、 その方法論は有効に活かされるはずである。

ジェンダー統計を政策策定に活用していく ことは、同時にジェンダー統計そのものの改 善のステップでもある。ジェンダー統計は実 際に利用されて初めてその効果が評価され、 修正点が指摘され、不足している情報が確定 される。現在ある統計が完璧でなくても可能 な限り利用してゆくことで、ジェンダー統計 と政策立案の両方の能力は高められてゆくは ずである。その実証が次の課題である。

注記

- 1 政令に相当し、法的拘束力を持つ。大統 領が公布。
- 2 国家開発計画においては、司法、経済、 宗教、教育、社会・文化、天然資源・環境分 野においてジェンダー配慮プログラムが 設定されている。
- 3 政令・通達の一種で、法的拘束力は持た ない。大統領が公布。
- 4 労働省、教育省、農業省、保健省、司法 省、協同組合省、大統領府、家族計画庁、

- 商業省、工業省からの情報システムある いは計画局の職員、計25人。
- 5 12州は、スマトラ島のブンクル州、ジャンビ州、ランプン州、ジャワ島の西ジャワ州、中部ジャワ州、東部ジャワ州、バリ島のバリ州、カリマンタン島の南カリマンタン州、スラウェシ島の南スラウェシ州、中部スラウェシ州、ヌサテンガラ島の西ヌサテンガラ州、東ヌサテンガラ州。
- 6 中央レベルでセミナーのガイドラインを 作成し、州はそれに基づいて企画・開催 したので、各州のセミナーの形式は同じ である。
- 7 2000年2月から2ヶ月間にわたってJICA短期専門家と現地コンサルタントが中心となって行った。JICA短期専門家は主にジェンダー統計システムの枠組みの開発とそれに伴うシステム上の問題点の把握と現地コンサルタントは主に中央統計局とともなりター省庁にある既存統計について表しての生産者、内容、収集時期・方法・ベル(収集、集計、出版、ウェブサイト公開といるといるによいでは、「インドネシアにおけるジェンダー配慮した統計システムの評価報告書」として2000年に出版した。
- 8 12州で作成された各州のセミナー報告書を一つにまとめたもので、内容は各州のセミナー全体の要旨、質疑応答の内容、ジェンダー統計の利用や今後の要望についての参加者からのアンケート回答などである。2002年2月に出版した。
- 9 全12州のジェンダー統計普及セミナーが 終了後、今後刊行予定のジェンダー統計 プロファイルの改良を目的に州の利用者 によるグループディスカッション、中央 レベルで関係機関からの聞き取りを2002 年3月に行い、その利点や改善点などにつ いて意見を収集した。
- 10 女性のエンパワーメント省が独自に作成した12州は、北スマトラ州、西スマトラ

- 州、南スマトラ州、リアウ州、ジャカルタ特別市、ジョクジャカルタ特別市、中部カリマンタン、東カリマンタン州、西カリマンタン州、中部スラウェシ州、北スラウェシ州、イリアンジャヤ州である。これらのジェンダープロファイルはプロジェクト終了後の2002年4月に完成した。
- 11 2002年12月時点における状況。
- 12 プロジェクト終了時においては、システムが稼動し始めた段階であり、まだ各セクターが十分体制を整備していないので、成果は確認していない。
- 13 1998年に崩壊したスハルト独裁体制以後 の新政権において、民主化促進、トップ ダウン型開発からボトムアップによる開 発促進というパラダイムの転換が起こり、 この改革の中で1995年の地方行政法と中 央地方財政均衡法が制定され、2001年か ら地方分権化が施行された。地方政府が 独自性を発揮すること、オーナーシップ を高めることで行政能力の向上を図るこ とを目的として、中央集権の機能縮小、 県政府への権限委譲、州と県の関係の並 列化、地方議会の権限強化による行政府 とのチェックアンドバランス、そして財 政的にもそれまで中央政府による一方的 な地方の資源収入の収奪から地方の天然 資源からの収入の中央と地方の分配や中 央から地方政府の財源の移転についての 均衡を図っていくことが規定された。ジ ェンダー平等政策においても、主要方針 は中央で策定されるが、個々の政策策定 や法律制定については各地方自治体の自 主性が尊重され、決定する権限が拡大し た。
- 14 人口の多い西ジャワ州、中部ジャワ州、 東部ジャワ州は各1000冊、残りの州は各 500冊を印刷した。
- 15 すでに全国レベルの性別統計データはホームページに載っている。

参考文献

- 伊藤陽一(1994)『女性と統計 ジェンダー 統計論序説 』、梓出版社
- インドネシア共和国(2000)、バクティア ル・アラム日本語訳監修『国家開発計画 (PROPENAS) 2000 2004年』、JICAイン ドネシア事務所
- ジェンダー統計研究グループ(2002)「ジェンダー統計関係論文等(日本)集成 No.1:第4回世界女性会議前後間まで」、 法政大学日本統計研究所
- 法制大学日本統計研究所(2001)、「ECE地域 のおけるジェンダー統計ウェブサイトの構 築 関係報告書の翻訳と論評」、統計 研究参考資料: No.75
- BPS (2000). *Indicator Kesejahteraan Rakyat* 1999 (The People's Welfare Indicators 1999), Jakarta
- BPS (1999). Struktur Ongkos Usaha Tani Padi Palawija 1998/1999 (The 1998/1999 Rice and Second Crop Cost Structure), Jakarta
- BPS Propinsi Jawa Barat (2000). *Gambaran Keaddan Sosial Ekonomi Masyarkat Jawa Barat 1999* (The Socio-Economic Situation in the Communities of West Jawa Province 1999), Bandung.
- Beck, T. (1999). Using Gender-Sensitive Indicators. Commonwealth Secretariat, London.
- Headman, B., Perucci, F. and Sundstrom, P. (1996). Engendering Statistics. Statistics Sweden.
- Rosalin, L.N., Surbakti, S., Rahardjo, Y., Carriere, E. and Heroe, H. (2001). *Gender Analysis Pathway*. BAPPENAS. CIDA and WSPII, Jakarta.
- Statistic Division Economic and Social Department, Women and Population Division Sustainable Development Department, FAO (1999). Filling the data gap. Rome.
- Surbakti, ASoedarti (2000). LARORAN STUDI: Evaluasi Kebutuhan Sistem Statistik yang

Responsif Jender di Indonesia (Needs Assessment on the Gender Responsiveness of the Statistical System in Indonesia), Jakarta.

The Office of the State Ministry of Women's Empowerment Republic of Indonesia (2001). Indonesia Country Report the Implementation of All Form's of Discrimination against Women During 1995-2000. Jakarta.

United Nations (2000). The World's Women 2000

Trends and Statistics. New York.
Vaidyanathan, K. E. (1999). Gender Statistics:
Present situation and future perspectives,
Bulletin of the International Statistical Institute

Bulletin of the International Statistical Institute Finland 1999 (インターネットでアクセスしたが、巻や頁の番号は不明).

(2003.7.7受理)

Evaluation on Cooperation in the Area of Gender Statistics in Indonesia

Toshiko Hamano

Japan International Cooperation Agency Hamano6@attglobal.net

Abstract

Gender Statistics are basis for gender-responsive policy formulation and planning.

In Indonesia lack of appropriate gender-responsive statistics system hindered the promotion of the formulation of gender-responsive policies. In this situation the Japan International Co-operation Agency (JICA) started assisting the Government of Indonesia in developing the functional gender-responsive statistics system (Gender Statistics Project) in February 2000 and continued until March 2002.

The overall goal of the Gender Statistics Project was to utilize gender statistics to formulate policies and plan, and the direct objectives were 1) to promote cooperation between gender statistics producers and users 2) to improve capacities of producers 3) to improve capacities of users 4) to increase users' access to gender statistics. Under the Project the following activities took place; Needs assessment, Trainings, Production and publication of gender statistics booklets, Dissemination of gender statistics, Review of the published gender statistics booklets.

In this paper first the outputs of the project were reviewed, and then the possibility of the utilization of the presented statistics was examined from the gender-responsive policy formulation point of view. Finally it was evaluated whether this Project contributed to promote the gender-responsive policy formulation. It was recognized that the Project made the concrete system of gender statistics necessary for formulating gender-responsive policies based on the positive external factors.

Keywords

gender statistics, policy formulation, cooperation

研修委員会

活動方針

研修委員会では、「評価の普及および評価に関する人材育成に寄与すること」を目的として、 各種セミナーの開催、研修プログラムの開発、教材の作成、講師の派遣、その他人材育成に関 わる活動の実施支援を行うことを基本方針としている。その方針のもと、今年度は以下のよう な活動を展開している。

活動状況

「評価インターン出前サービス」に係るパイロット・プロジェクト

学会の学生会員の中から「実際に評価を経験してみたい」と考えている希望者を募り、2~4 名程度の学生評価インターンチームを編成して、「評価を外部に委託したい」と考えている地方自治体やNGOで、実際に評価を体験させていただくという企画である。学生会員にとっては評価を体験する機会として、委託者である自治体やNGOにとっては、外部の評価者(インターン)による評価を低コストで受け入れる機会となる。また、学会として評価インターンチームによる評価の品質を一定レベルに保つ意味から、各チームに学会の理事がアドバイザーとして1名加わる仕組みとする。研修委員会では、評価インターン希望者が希望する評価対象テーマ・対象機関と、評価インターンの受け入れを希望する自治体・NGOをマッチングし、調整業務を行う。

今年度は初めての試みであることから、「パイロット・プロジェクト」として試行的に実施し、 その結果を踏まえ、来年度からの本格的な事業としての取り組みをにらみ、今後の実施方針・ 実施方法を再検討することを考えている。

現在、学生会員4名の希望があり、研修委員会で関係自治体に打診中である。

評価研修プログラム認証制度創設に係るパイロット・プロジェクトの実施

評価の普及と評価活動の質の改善にとって、その実施に携わる人材の育成は不可欠であるが、そのひとつの手段として、官民諸団体が実施する評価研修活動を対象とする認証制度を創設することにより、質の高い研修の普及を図ることができると考えている。研修内容、方法、教材、講師に関する事前査定と研修のインパクトの事後評価および受講生自身のプログラム評価の検討をもとに、一定の水準を充たしていると判断される研修プログラムを事後的に認証する形で、評価人材育成を間接的に支援する試みである。

制度創設にあたり、まずパイロット・プロジェクトとして、7月28日~31日の4日間、広島県立教育センターの主催により実施される「広島県学校評価研修講座」の協力を得、同研修プログラムを対象とした認証作業を行うこととなり、現在、その作業中である。審査方法については、研修員会で「認証審査基準」を作成し、本パイロット・プロジェクトのために立ち上げた「学校評価審査委員会」(牟田博光委員長)に、認証に関わる審査を依頼している。

今後は、パイロット・プロジェクトの結果を踏まえ、その実施方法及び審査方法を再検討するとともに、当面は学校評価の他にODA及び行政分野における評価研修への適用を推進したいと考えている。

研修委員会共同委員長 長尾 眞文源 由理子

国際交流委員会

7月15~16日世界銀行にて同銀行主催の「第5回開発評価国際会議」が開催されました。本会 議の本年の主題は「開発効果の評価:その課題と今後の展望」であり、世界銀行専務理事によ る挨拶に始まり、毎日午前午後それぞれ二つのセッションが設定され、主要援助国ODA評価局 長、国際通貨基金を含めた国際開発金融援助機関の評価局長、開発評価専門家、NGO評価担当 者が世界各地域から総数250名ほど参加しました。各セッションではこれら参加者代表による開 発評価に関する基調講演とパネル討議がなされ、最終セッションでは世界銀行評価局長による 討議総括報告がなされました。小生は、日本評価学会副会長と国際開発評価学会副会長の資格 で、第一日午後の「開発援助機関の活動評価」セッション議長をつとめましたが、いずれのセ ッションでも開発評価関係者から率直な課題指摘・分析と明確な方向性を示した素晴らしい基 調報告と白熱したパネル討論が展開されました。特に世界における開発評価の主要な潮流と開 発評価手法・体制に関する議論からは参加者がお互いに多くを学んだという印象を強くしまし た。しかし、開発途上国の観点から見た主要援助国と主要国際機関の開発評価活動に対する評 価に関する議論が必ずしも十分でなかったという小生の指摘に対して、会場全員から賛同拍手 があがったことを考慮すると、世界銀行主催の第6回の開発評価国際会議の方向が決まったよう です。なお、わが国はこの点では主要援助国に比べると進んでおり、日本外務省は、開発途上 国の開発評価活動の発展とその能力育成を目指して、既に昨年と一昨年にアジアの途上国の評 価担当者と国際機関の評価担当者を招いて国際会議を開催して世界各国から高い評価を得てい ることは周知の事実であり、今年の「ODA白書」にも明確に記載されているごとくです。それ にしても、このような主要な開発評価国際会議への日本人専門家の参加が、東京の日本政府、 国際協力事業団、国際協力銀行から一人も無かったこと大変寂しく思いました。しかし駐ワシ ントン日本大使館、現地駐在のJICAから一人ずつ参加なされたことはせめてもの幸いでした。 日本は受信は積極的だが、世界への知的発信は中々少ないという一般的指摘は、小生の議長・ 討論者としての参加と大使館勤務経済協力担当官からの討議への有意義な参加を除いては、こ こでも見受けられたといってよいでしょう。本国際会議後、ワシントン在住の開発問題・開発 協力課題に関心を持っている援助関係者、専門家、大学院生が毎週集まって討議している「ワ シントン開発フォーラム」で世銀主催の本会合の討議内容についての小生の意見を開陳する 機会を与えられましたので、それについての詳細は同研究会のwebsite(http://www. developmentforum.org/)を見ることをご推薦いたします。なお、世界銀行国際会議後の17日には 世界銀行にて、「国際開発評価学会第4回役員会」が開催されましたので、これについては国際 交流委員会での討議を踏まえて次回に報告させていただきます。

国際交流委員長 廣野 良吉

平成15年8月25日 日本評価学会 第4回全国大会 実行委員会

日本評価学会 第4回全国大会のお知らせ

日本評価学会第4回全国大会を以下の要領で開催いたします。会員の皆様には奮ってご参加いただきますようご案内申し上げます。ご参加の方はお手数ですが別添申込用紙で10月10日(金)までにhyoka4@apu.ac.jpにお申込ください。

記

日時:2003年11月1日(土),2日(日)

【受付開始時間:9:15】

場所:立命館アジア太平洋大学(APU)

F棟(下記のAPUのHPアドレスよりキャンパスマップをご確認ください)

プログラム:次頁をご参照ください。(詳細なプログラム内容につきましては完成次第別途学会

ホームページに掲載いたしますので、そちらをご覧ください。不都合等ございましたら実行委員事務局hyoka4@apu.ac.jpまでご照会頂きたく宜しくお願いします。)

大会参加費:正会員 4000円、学生会員 2000円

非会員(一般) 6000円、非会員(学生) 3000円

懇親会費: 正会員 5000円、学生会員 2500円

お願い:

大会参加費、懇親会費は当日受付でお支払いください。

学生の方は学生証を提示して下さい。

キャンパス内には、生協食堂と生協ショップがあります。両日とも11:30~14:00が営業時間となっておりますので、昼食はそちらでとられることをお勧めします。

宿泊先は各自でご手配をお願いいたします。なお、大会用に下記の旅行代理店でフライト・宿泊の斡旋を行なっておりますのでご利用ください。

日本旅行大分支店 大会用HP

http://www3.nta.co.jp/convention/hyouka 0.htm

また、会場、懇親会場にはクロークがございませんので御了承下さい。

会場までの交通手段と大学キャンパスの地図は以下のHPをご照会ください。

立命館アジア太平洋大学(APU) URL: http://www.apu.ac.jp/index j.html

(APUについて 大学へのアクセス、キャンパスマップ)

日本旅行大分支店 大会用HP URL: http://www3.nta.co.jp/convention/hyouka_0.htm

日本評価学会(JES) URL: http://www.idcj.or.jp/JES/

日本評価学会 第4回全国大会プログラム

各セッションの座長、コメンテーター、報告者名、報告演題についてはプログラム委員会で確定後、大会参加者に通知する予定です。また、学会ホームページにも掲載する予定ですのでご参照下さい。

第1日:2003年11月1日(土)

	会場1	会場2	会場3	会場4
午前の部 10:00~11:45	セッション1 経営と評価	セッション2 インドネシア / 教育評価	セッション3 国際機関 / 開発政策	企画セッション ジェンダーの主流化 とインパクト評価
午後の部 12:45~15:05	セッション4 政策評価 / 行政評価	セッション5 教育 / 評価手法	セッション6 ODA評価 / 評価のあり方	セッション7 評価研修
会員総会 15:15~16:15				
共通論題 16:25~18:15	シンポジウム「評価文化の構築を目指して」			
懇親会 18:30~20:00				

第2日:2003年11月2日(日)

	会場1	会場2	会場3	会場4
午前の部 9:15~11:30	企画セッション 経営システム としての評価	企画セッション JBIC円借款の有効性	企画セッション 政策プログラム 評価の手法開発	セッション8 NPO / 参加型評価
午後の部 12:30~14:50	セッション9 地方自治体の評価	セッション10 教育評価	企画セッション 保健セクターに影響 を与えるプロジェク トのインパクト評価 及び、その手法の考察	セッション11 プロジェクト / プロジェクトの評価
午後の部 16:00~16:45	企画セッション 市民参加型二次評価 の可能性について	セッション12 開発評価 / 成果重視評価		

備考:

各セッションには基本的に2名のコメンテーターがつきます。

各セッションにおける時間配分は原則として報告20分、コメント5分、質疑応答10分の計35分です。

以上

日本評価学会規約

2000年9月25日設立総会承認

第1章 総則

(名称)

第1条 本会は、日本評価学会と称し、英語名は、The Japan Evaluation Society (略称JES) とする。

(所在地)

- 第2条 本会の本部は、東京都江東区に置く。ただし、その他の地に支部を置くことができる。 (目的)
- 第3条 本会は、評価に関する研究及び応用を促進し、会員相互及び関連学・協会との情報交換を図るとともに、この分野の学問の進歩発展及び評価に携わる人材の育成を通じ、評価活動の向上と評価の普及を目的とする。

(事業)

- 第4条 本会は、前条の目的を達成するために次の事業を行う。
 - 1)評価に関する学術研究会、講演会、国際シンポジウムなどの開催
 - 2)評価に関する研究助成、報奨
 - 3)評価に関する調査及び研究並びに研修
 - 4) 学会機関誌及びその他の刊行物の発行等の普及啓発活動
 - 5) その他、本会の目的を達成するために必要な事業

第2章 会員

(会員の種別及び入会)

- 第5条 本会の会員の種別は、次の通りとする。総会での議決権は、正会員のみが持つものと する。
 - 1)正 会 員 本会の目的に賛同して入会した個人。
 - 2) 学生会員 本会の目的に賛同して入会した、原則として大学以上の学生で、学生会員を希望する者。尚学生会員は、卒業と同時に正会員となることができる。
 - 3) 賛助会員 本会の目的に賛同し、その事業を後援する者。
 - 4)名誉会員 本会に功労のあった者及び広く評価分野に関連ある分野における学識経験者で理事会の推薦に基づき総会の承認を経た者。
- 2 本会の会員になろうとする者は、入会申込書を会長に提出し、理事会の承認を得なければならない。また、名誉会員として推薦された者は、入会の手続きを要せず、本人の承諾をもって会員となるものとする。

(会費)

- 第6条 会員は、総会の定めるところにより、会費を負担しなければならない。 (退会)
- 第7条 会員は、退会しようとするときは、事前にその旨を書面をもって会長に届け出なければならない。
- 2 会員は、次の各号のいずれかに該当するときは、本会を退会したものとみなす。
 - 1)正会員又は学生会員にあっては、会費を2年以上滞納した場合

- 2)本会の名誉を毀損し又は設立の趣旨に反する行為を行なったことにより、総会において、除名すべきものと認められた場合
- 3) 前各号に掲げるほか会員たる資格を喪失した場合
- 3 本会は、会員がその資格を喪失しても、既に納入した会費その他の拠出金品は返還しない。

第3章 役員等

(役員の種別)

- 第8条 本会には、次の役員を置く。
 - 1) 理事 20名から25名(うち会長1名、副会長若干名)
 - 2) 監事 2名

(選任)

- 第9条 理事及び監事は、総会において、正会員の中から選任する。
- 2 会長及び副会長は、理事会において、理事の互選により定める。
- 3 理事及び監事は、相互に兼ねることができない。

(職務)

- 第10条 理事は理事会を構成し、会務の執行を決定する。
- 2 会長は、本会を代表し、会務を統括する。
- 3 副会長は、会長を補佐し、会長に事故あるとき又は欠けたときは、会長があらかじめ指名した順序で、その職務を代行する。
- 4 監事は、会務の執行及び会計を監査する。

(任期)

- 第11条 役員の任期は、2年とする。ただし、連続3期までの再任を妨げない。
- 2 補欠又は増員により就任した役員の任期は、前項本文の規定にかかわらず、前任者又は現任者の残任期間とする。
- 3 役員は、任期満了の場合においても、後任者が就任するまで、その職務を行なわなければならない。

(幹事)

- 第12条 本会に幹事若干名を置く。
- 2 幹事は、会長が理事会の同意を得て任命する。
- 3 幹事は、共同して会務の執行を補佐する。
- 4 幹事の任期については、第11条第1項の規定を準用する。

(顧問)

- 第13条 本会に顧問若干名を置くことができる。
- 2 顧問は、会長が理事会の同意を得て委嘱する。
- 3 顧問は、本会の運営に関し、会長の諮問に答え、又は意見を述べることができる。
- 4 顧問の任期については、第11条第1項の規定を準用する。

第4章 総会

(招集)

- 第14条 会長は、毎年1回通常総会を招集する。
- 2 理事会が必要と認めた場合、臨時総会を招集することができる。

(開催及び議決)

第15条 総会は、正会員総数の3分の1以上の出席が無ければ開催することができない。

- 2 総会の議事は、この規約に別に定めるもののほか、出席正会員の過半数の同意でこれを 決し、可否同数の場合は、議長の決するところによる。
- 3 総会の議長は、会長をもってこれにあてる。

(書面表決等)

- 第16条 やむを得ない理由のため、総会に出席できない正会員は、書面又は代理人をもって表 決権を行使することができる。
- 2 前項の代理人は、代表権を証する書面を総会毎に議長に提出しなければならない。
- 第1項の場合において、正会員は、表決内容等について、総会の議長に一任することができる。
- 4 第1項及び前項の規定により、表決権を行使する場合は、当該正会員は総会に出席した ものとみなす。

第5章 理事会

(構成及び機能)

第17条 本会に理事会を置く。

- 2 理事会は、理事をもって構成する。ただし、監事、顧問、幹事ならびに第30条第2項に 定める事務局長は、理事会に出席し意見を述べることができる。
- 3 理事会は、この規約に別に定めるもののほか、次の事項を議決する。
 - 1)総会の議決した事項の執行に関すること
 - 2)総会に付議すべき事項
 - 3) その他総会の議決を要しない会務の執行に関すること

(開催及び招集)

第18条 定例理事会は、毎年3回開催する。

- 2 前項にかかわらず会長が必要と認めた場合には臨時理事会を開催することができる。
- 3 理事会は会長が招集する。
- 4 会長は、緊急に理事会を招集する必要がある場合において、やむを得ない事情によりこれを開催できないときには、理事の承諾を得て、書面により議決を得ることができる。この場合、理事会は開催されたものとみなす。

(議決等)

第19条 理事会は、理事現在数の3分の1以上の出席がなければ、開会することができない。

- 2 理事会の議決は、この規約に別に定めるもののほか、出席理事の過半数の同意でこれを 決し、可否同数のときは、議長の決するところによる。
- 3 理事会の議長は、会長をもってこれにあてる。

(書面表決等)

- 第20条 やむを得ない理由のため、理事会に出席できない理事は、書面又は代理人をもって表 決権を行使することができる。
- 2 前項の場合において、理事は、表決内容等について理事会の議長に一任することができる。
- 3 前2項の規定により表決権を行使する場合は、当該理事は理事会に出席したものとみな す。

第6章 資産及び会計

(資産の構成及び管理)

第21条 本学会の資産は、次に掲げるものをもって構成する。

- 1)財産目録記載の財産
- 2) 会費
- 3)寄付金品
- 4) 資産から生ずる収入
- 5) 事業に伴う収入
- 6) その他の収入
- 2 本会の資産は、会長が管理し、その方法は理事会の議決による。

(経費の支弁)

第22条 本会の経費は、資産をもって支弁する。

(事業計画、収支予算、事業報告及び収支決算)

- 第23条 本会の事業計画書及び収支予算書は、会長が作成し、理事会の議決を経た後、毎事業年度の開始前に総会の議決を得なければならない。ただし、やむを得ない事情により、当該事業年度開始前に総会を開催できない場合にあっては、理事会の議決によることを妨げない。この場合、当該事業年度の開始の日から90日以内に総会の議決を得るものとする。
- 2 会長は、前項の事業計画書及び収支予算書を変更しようとするときは、理事会の議決を 得なければならない。
- 3 本会の事業報告書及び収支決算書は、会長が毎事業年度終了後遅滞なくこれを作成し、 監事の監査、理事会の議決を経た後、当該事業年度終了後90日以内に総会の承認を得な ければならない。

(特別会計)

- 第24条 本会は、事業の遂行上必要がある場合には、理事会の議決を経て、特別会計を設ける ことができる。
- 2 前項の特別会計は、前条の収支予算及び収支決算に計上しなければならない。

(会計年度)

第25条 本会の会計年度は、毎年10月1日に始まり、翌年9月末日に終わる。

第7章 規約の変更

(規約の変更)

第26条 この規約は、総会において出席正会員の3分の2以上の議決を経なければ変更すること はできない。

第8章 雑則

(年次大会)

第27条 本会は、毎年1回年次大会を開催する。

2 年次大会の開催は、大会実行委員長がこれを指揮統率する。

(委員会及び分科会)

- 第28条 本会は、事業の円滑な遂行を図るため、委員会及び分科会を設けることができる。
- 2 委員会及び分科会は、その目的とする事項について、調査及び研究し、または審議する。
- 3 委員会及び分科会の組織、構成及び運営その他必要な事項は、理事会の議決を経て、別

に定める。

(支部)

第29条 本会は、事業の円滑な遂行を図るため、支部を置くことができる。

- 2 支部には、理事会の同意を得て会長が委嘱する支部長を置く。
- 3 支部の位置、組織、運営その他必要な事項は、理事会の議決を経て、別に定める。 (事務局)

第30条 本会の事務処理のため、事務局を置く。

- 2 事務局には、理事会の同意を得て会長が委嘱する事務局長を置く。
- 3 事務局の場所、組織、職員及びその他必要な事項は理事会の議決を経て、別に定める。

附則(2000年9月25日)

- 1.この規約は、本学会の設立の日(以下「設立日」という。)から施行する。
- 2.設立発起人は、第5条の規定にかかわらず、本学会設立当初の会員とする。設立日までに入会の申し込みを行なった者も同様とする。
- 3. 本学会の設立初年度の年会費は、第6条の規定にかかわらず、設立総会の定めるところによる。
- 4. 本学会の設立当初の役員は、第9条第1項及び第2項の規定にかかわらず、設立総会の定めるところとし、その任期は、第11条第1項本文の規定にかかわらず、2001/2002年度の事業報告及び収支決算の承認を審議する総会の日までとする。
- 5. 本学会の設立初年度の事業計画及び収支予算は、第23条第1項の規定にかかわらず、設立総会の定めるところによる。
- 6. 本学会の設立当初の事業年度は、第25条の規定にかかわらず、設立日から2001年9月30日までとする。

日本評価学会誌刊行規定

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15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、英文又は和文とする。
- 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ーの投稿についても、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。

(抜刷の配布)

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

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。
- 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、 研究論文、 総説、 研究ノート、 実践・調査報告、 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、必要なリライトを経た後に、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。 研究論文20,000字、 総説15,000字、 研究 ノート15,000字、 実践・調査報告20,000字、 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 X 43行 X 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

〒135-0047 東京都江東区富岡2 - 9 - 11 京福ビル財団法人国際開発センター内 日本評価学会事務局

TEL: 03-3630-8031 FAX: 03-3630-8095

E-mail: yabuta.m@idcj.or.jp

『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)

(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:	-		
Couroc.			
Note:	rable 1	Number	of Accidents in the State of
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'。)とする。英文の場合は、(...')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)おとしで記述する。

編集後記

第3巻第2号は、「評価手法の現状と課題」をテーマとして特集を組みました。全部で7つの論文・研究ノートを掲載ことができました。今後の評価の議論に寄与することを期待します。次号第4巻第1号では「ジェンダーの主流化とインパクト評価」をテーマとして特集を組みます。

また、投稿も増え、本号では7つの論文、研究ノート、実践・調査報告を掲載しました。11月 には全国大会があります。大会の結果を受け多くの投稿があることを期待いたします。

編集委員長 三好 皓一(文責)

『日本評価研究』第3巻第2号 2003年9月29日

編集・発行 日本評価学会

〒135-0047 東京都江東区富岡2-9-11京福ビル

財団法人国際開発センター

TEL: 03-3630-8031 FAX: 03-3630-8095

印 株式会社三好デジタルプリンティング

© 日本評価学会

The Japanese Journal of Evaluation Studies

Vol. 3, No. 2, September 2003

CONTENTS

Special Issue:

Evaluation Methods: Current Situation and Discussion

Editor's Note Ryo Sasaki

Using Policy Evaluation to Support Decisionmaking and Improve Program Performance

Joseph S. Wholey (Abstract: Hiraki Tanaka)

Towards Social Policy-making Based on Systematic Reviews:

Challenges of the Campbell Collaboration

Hiroshi Tsutomi

Toward Constructing More Suitable Program Theory for Japan's Evaluation

-- Focusing on International Cooperation Evaluation and Policy Evaluation --

Koichi Miyoshi, Satoshi Morita, Yoshio Aizawa

Utilization-Focused Evaluation: Theory and Issues Masafumi Nagao

Characteristics of Empowerment Evaluation and Possibility of Its Application

-- Focusing on Theory of Empowerment Evaluation by Dr. Fetterman -- Yuriko Minamoto

Drucker's Self-Assessment Tool for Nonprofit Organization

-- How It Is Different from Evaluation, and How It Contributes to Evaluation --

Yayoi Tanaka, Daisuke Korenaga

Evaluation as an Integral Part of Strategic Planning and Management Process

-- How Strategic Planning Can Contribute to the Current Exercise of Policy Evaluation Activities in Japan -- Yoshiaki Ryu, Ryo Sasaki

Articles

A Study on Participatory Secondary Evaluation for the Public: The Case of JICA's Evaluation on "Poverty Reduction and Rural Development Project in Sulawesi, Indonesia"

Satoshi Morita, Kana Takamatsu, Koichi Miyoshi

Incidences of Public Education Spending in Developing Countries Takako Yuki

Research Notes

The Tao of Evaluation Craig Russon

A Study on Outcome Measurement with Social Indicators

-- A Method for the Comparison of Time-series Indicators -- Tatsuya Ono

A New Evaluator Model for Local Government Reform

-- An Empowerment-oriented Evaluator for Local Government Reform -- Junro Nishide What Measures Should Be Taken for the Water Cycle Based on the Consciousness of Residents?

Masaaki Mutsuro

Development

Evaluation on Cooperation in the Area of Gender Statistics in Indonesia Toshiko Hamano

The Japan Evaluation Society