

# 日本評価研究

## Japanese Journal of Evaluation Studies

Vol. 10, No. 1, March 2010

### 特集：エビデンスに基づく実践の世界的動向と 日本における取り組み

特集「エビデンスに基づく実践の世界的動向と日本における取り組み」に寄せて  
佐々木 亮 大島 巖

エビデンスに基づく医療（EBM）の展開から学ぶもの  
－EBM普及に伴い経験した課題と解決策－ 正木 朋也 津谷 喜一郎

教育におけるエビデンスに基づく政策－新たな展開と課題 岩崎 久美子

精神保健福祉領域における科学的根拠にもとづく実践（EBP）  
の発展からみたプログラム評価方法論への貢献  
～プログラムモデル構築とフィデリティ評価を中心に～ 大島 巖

「エビデンス」の利用に関する検討～技術移転と追試過程を中心に～  
津富 宏

開発援助分野におけるRCT導入に見られる旧くて新しい課題  
－理論なきRCTから理論検証のRCTへ－ 青柳 恵太郎

エビデンスに基づく開発援助評価  
－援助評価の歴史、ランダム化実験の起源、スクリヴェンとバナージェの考え方の比較－  
佐々木 亮

#### 実践・調査報告

フィリピンでのNGO教育事業参加型評価  
～ファシリテーターの役割とステークホルダーのエンパワーメント  
田中 博

#### 研究ノート

地方自治体へのバランスト・スコアカード適用に関する研究  
－「財務の視点」に着目したフレームワークの検討－ 佐藤 幹

日本評価学会

Japan Evaluation Society

『日本評価研究』編集委員会

Editorial Board

編集委員長  
Editor-in-chief

三好 皓一(立命館アジア太平洋大学)  
Koichi MIYOSHI

副委員長  
Vice-Editor-in-chief

西野 桂子(ジーエルエム・インスティテュート)  
Keiko NISHINO

常任編集委員  
Standing Editors

牟田 博光(東京工業大学)  
Hiromitsu MUTA

山谷 清志(同志社大学)  
Kiyoshi YAMAYA

編集委員  
Editors

青山 温子(名古屋大学)      大沢 真理(東京大学)  
Atsuko AOYAMA      Mari OSAWA

大島 巖(日本社会事業大学)  
Iwao OSHIMA

岡本 義朗(三菱UFJリサーチ&コンサルティング(株))  
Yoshiaki OKAMOTO

小野 達也(鳥取大学)      窪田 好男(京都府立大学)  
Tatsuya ONO      Yoshio KUBOTA

佐々木 亮(国際開発センター)      佐藤真理子(筑波大学)  
Ryo SASAKI      Mariko SATO

渋谷 和久(国土交通省)      鈴木 絲子(神戸女子大学)  
Kazuhisa SHIBUYA      Itoko SUZUKI

田中 弥生(大学評価・学位授与機構)      松岡 俊二(早稲田大学)  
Yayoi TANAKA      Shunji MATSUOKA

村松 安子(東京女子大学)  
Yasuko MURAMATSU

事務局  
Office

〒140-0002 東京都品川区東品川4-12-6 日立ソフトタワーB 22F  
財団法人国際開発センター内  
特定非営利活動法人日本評価学会  
編集事務担当 鈴木 直美  
TEL: 03-6718-5931, FAX: 03-6718-1651  
E-mail: jes.info@idcj.or.jp

『日本評価研究』は、和文年2回、英文年1回または2回刊行します。

# 日本評価研究

第10巻 第1号 2010年3月

## 目次

特集：エビデンスに基づく実践の世界的動向と日本における取り組み

佐々木 亮 大島 巖

特集「エビデンスに基づく実践の世界的動向と日本における取り組み」に寄せて……………1

正木 朋也 津谷 喜一郎

エビデンスに基づく医療（EBM）の展開から学ぶもの

—EBM普及に伴い経験した課題と解決策— ……………3

岩崎 久美子

教育におけるエビデンスに基づく政策—新たな展開と課題 ……………17

大島 巖

精神保健福祉領域における科学的根拠にもとづく実践（EBP）の発展からみた

プログラム評価方法論への貢献

～プログラムモデル構築とフィデリティ評価を中心に～ ……………31

津富 宏

「エビデンス」の利用に関する検討～技術移転と追試過程を中心に～……………43

青柳 恵太郎

開発援助分野におけるRCT導入に見られる旧くて新しい課題

—理論なきRCTから理論検証のRCTへ— ……………53

佐々木 亮

エビデンスに基づく開発援助評価

—援助評価の歴史、ランダム化実験の起源、スクリヴェンとバナージェの考え方の比較— ……………63

実践・調査報告

田中 博

フィリピンでのNGO教育事業参加型評価

～ファシリテーターの役割とステークホルダーのエンパワーメント ……………75

## 研究ノート

佐藤 幹

地方自治体へのバランスト・スコアカード適用に関する研究

—「財務の視点」に着目したフレームワークの検討— .....95

## 第10回全国大会開催報告

第10回全国大会プログラム（実績） .....107

共通論題セッション報告 .....110

シンポジウム報告 .....113

自由論題セッション報告 .....114

## 委員会活動

企画委員会 .....118

国際交流委員会 .....119

広報委員会 .....120

春季第7回全国大会及び10周年記念シンポジウムのご案内 .....121

日本評価研究刊行規定 .....122

日本評価研究投稿規定 .....124

日本評価研究執筆要領 .....126

日本評価研究査読要領 .....129

Publication Policy of the Japanese Journal of Evaluation Studies .....131

Information for Contributors (For English Papers) .....133

Writing Manual of the Japanese Journal of Evaluation Studies (For English Papers) .....135

Referee-Reading Guideline .....137

## 【巻頭言】

## 特集「エビデンスに基づく実践の世界的動向と日本における取り組み」に寄せて

佐々木 亮

国際開発センター

大島 巖

日本社会事業大学

日本評価研究第6巻第1号（2006年3月）の「特集 エビデンスに基づく評価の試み」において、私たちは各分野における最新の動向をレビューし、日本における普及の可能性について論じた。それから3年が経過し、世界的に見れば、エビデンスに基づく実践が各分野で徐々に普及するとともに、実施の困難さや倫理的問題、現場への適用上の問題などからその是非に関する議論も深まってきた。また「エビデンス」は日本語で「科学的根拠」とも訳されるが、何をもちょう科学的とするかなどその意味も多様に解釈される現状がある。さらに、多くのエビデンスを産出しシステムティック・レビューによりその質を確保し、かつ、それを蓄積して提供するシステムが構築されたうえでもなお、蓄積された結果を実際に現場で適用する際に生じる問題や、その他の要因を加味して総合的判断を行う上での課題など、エビデンスを「つかう」段階における新たな知見が得られてきた分野もある。

こうした世界的な状況を踏まえて、今回、再び特集号を組むことになった。各分野の動向に加えて、分野横断的な動きに関する論文が掲載されたのも今回の大きな特徴である。

また前回の特集号の編纂当時は、まだ日本では保健医療分野を除きほとんど実践例が見られなかったが、3年の間に日本でもついに本格的な取り組みがいくつか始まった。本特集号ではその取り組みの進捗についても報告し、日本の状況に即した議論も展開している。

以下、簡単にそれぞれの論文について解説したい。

正木／津谷の『エビデンスに基づく医療（EBM）の展開から学ぶもの－EBM普及に伴い経験した課題と解決策－』は、ランダム化比較試験（RCT）の報告の質向上のためのCONSORT声明など、エビデンスに基づく実践を一貫してリードしてきた保健医療分野における最新動向を論じており、改めて、評価研究を牽引する力強さを感じることができる。

岩崎の『教育におけるエビデンスに基づく政策－新たな展開と課題』は、「エビデンスに基づく政策」という概念に基づいて、経済協力開発機構（OECD）が積極的な政策提言を行っている状況を論じており、日本の教育政策にとって大きな示唆を与えている。

大島の『精神保健福祉領域における科学的根拠にもとづく実践（EBP）の発展からみたプログラム評価方法論への貢献～プログラムモデル構築とフィデリティ評価を中心に～』は、EBPツールキットプロジェクトという新しい技術移転の方法という観点から、プログラム評価方法論の発展の可能性を指摘している。

津富の『「エビデンス」の利用に関する検討～技術移転と追試過程を中心に～』は、従来から論じられてきた情報提供（「つたえる」）という機能に加え、組織づくり、人づくり、リーダーづくりや技術移転支援専門組織の創設など具体的な「つかう」ための考察がなされている。

青柳の『開発援助分野におけるRCT導入に見られる旧くて新しい課題－理論なきRCTから理論検証のRCTへ－』は、同分野におけるRCT導入に関する諸論点を整理して、それらを乗り越えるための取り組みを論じ、最後に国際的潮流の中で日本が取り組むべき課題について考察している。

佐々木の『エビデンスに基づく開発援助評価－援助評価の歴史、ランダム化実験の起源、スクリヴェンとバナージェの考え方の比較－』は、「エビデンスに基づく開発援助評価」には3つの起源があると指摘し

たうえで、開発援助評価において独占的というわけにはいかないが、ランダム化実験デザインが利用できるし利用すべき余地が確かに存在すると結論している。

本特集号に掲載されたこれら6本の論文が、日本におけるエビデンスに基づく実践の本格的普及に貢献することを、再び願ってやまない。

## 【総説】

## エビデンスに基づく医療（EBM）の展開から学ぶもの —EBM普及に伴い経験した課題と解決策—

正木 朋也

北里大学大学院医療系研究科

masakit-ky@umin.ac.jp

津谷 喜一郎

東京大学大学院薬学系研究科

tsutani-ky@umin.ac.jp

### 要 約

保健医療領域において1950年代から始まったランダム化比較試験（RCT）の総数は60万件を越えた。1990年代中頃以降の世界的なエビデンスに基づく医療（EBM）の動きの中で、改めてRCTに対する関心が増大している。一方、このEBMの動きの中で、いくつかの課題が同定されそれに対する解決策も開発された。これらの課題は、既報で紹介したエビデンスを「つくる」局面というよりむしろ、「つたえる」と「つかう」局面で浮上したものであった。本稿では「つたえる」局面での、RCTの報告の質向上のためのCONSORT声明とその多方面への展開、および「つかう」現場における叙述的情報やpost-RCTのひとつの流れとしてのベストケースについて紹介し、さらに、投資に対する継続的評価の必要性および患者や市民の役割についての最新動向についても触れた。

### キーワード

EBM、質改善、CONSORT声明、ベストケース、市民参加

#### 1. はじめに

既報において、「エビデンスに基づく医療」（evidence-based medicine: EBM）と政策評価の系譜についての考察を試みた（正木・津谷 2006）。そこではまず、EBMという概念が誕生した歴史背景を概観し、その社会的要請がいかん達成されたかをみた。また、情報インフラの役割を担っているコクラン共同計画（The Cochrane Collaboration）について概説し、システマティック・レビュー（systematic review）とその結果をまとめてデータベースとしたコクラン・ライブラリを紹介した。さらに、情報の質の確保に必要なとなる要点をレビ

ューし、研究デザインと批判的吟味、研究結果の要約を効率よく伝えるための構造化抄録の重要性、情報の共有化と社会認識の問題についても触れた。

EBMの実践にはおおまかに二つのアプローチがある。ひとつは、臨床現場での目の前の患者およびその家族らを対象としたアプローチである。もうひとつは、保健医療およびその周辺全体を対象とした公衆衛生を念頭におくアプローチである。いずれの場合においても、エビデンスを「つかう」には対象となる個々の患者あるいは集団の好み（patient preference）、価値システム（value system）、および現場特有の状況なども踏まえた総合的な判

断が必要となる。そのためには、研究により質の高いエビデンスをまず「つくり」、つぎにエビデンスにはグレードがあることも含めてうまく「つたえ」、さらに好みや場の状況も含めて適切に「つかう」ことが必要である。

本号の特集テーマであるEBP (evidence-based practice: EBP) は保健医療のみならず広くエビデンスに基づく実践的な活動全体を扱う。発表論文数などから類推すれば、保健医療を除く他の多くの領域では、未だRCTの実施とエビデンスを「つくる」ことに傾注した普及・啓蒙が行われているのが現状であろう<sup>1)</sup>。

一方、保健医療領域において1950年代から始まったランダム化比較試験 (randomized controlled trial: RCT) の総数はコクラン共同計画のデータベースの登録数から伺い知ることができる。ここ数年の増加が著しく、既報作成時から4年弱の間に約14万件が追加され、本稿執筆時点で60万件を越えていた<sup>2)</sup>。日本においても医中誌データベースの「ランダム化比較試験」のタグ付は年々増加し、2009年10月時点で1万件を越えている<sup>3)</sup>。昨今の登録件数の伸びからも、1990年代中頃以降の世界的なEBMの動きの中で、改めてRCTに対する関心も増大し、実際にさまざまな領域で多数のRCTが実施されている現状が伺える。

しかし、このEBMの普及・展開の過程は必ずしも順風満帆ではなく、いくつかの課題が同定され、またそれに対する解決策も開発されてきた。経験した課題の多くは、エビデンスを「つくる」局面というよりむしろ、「つたえる」と「つかう」局面におけるものであった。

本稿では、今般保健医療およびその周辺領域で経験したいくつかの課題とその解決策について紹介する。このような経験は、他の分野・領域において、その現状に則した評価活動に活かすことが可能であろう。

はじめに、まず「つたえる」局面における、RCT報告の質向上のためのCONSORT声明とその多方面への展開について報告する。また、「つかう」現場からのRCTへのフィードバックについて述べ、さらに、調査研究への投資に対する継続的評価の必要性および患者や市民の役割についての最新動向と将来展望についても触れた。

## 2. エビデンスの「つたえ」方の標準化 —CONSORT声明など種々のイニシア ティブとその展開—

ここでは、いろいろな領域で取り組んだ研究報告の質改善の取り組みについて、(1) RCT、(2) RCT以外の研究デザイン、(3) メタアナリシスやシステマティック・レビューなどの統合型研究、および、(4) 相補代替医療関連の4領域に分け、その契機となったCONSORT声明 (Begg 1996, 改訂版2001) が、それぞれの領域でどのように展開したかをレビューする。その上で、EBMにおける「つたえ」方の標準化と報告書の質の確保の流れを整理し、その全体像と系譜について考察も加える。

### (1) RCTの質改善：CONSORT声明

1990年代半ば、RCTの報告の質を改善しようとする2つの独立した先導的グループが共同して、CONSORT (Consolidated Standards of Reporting Trials; 臨床試験報告に対する統合基準) 声明を発表した。初版CONSORT声明 (1996) は、RCT報告書の著者がチェックリストとフローチャートを利用して報告書の質を改善することをめざした。作成に携わったメンバーは、国際的な臨床試験実施者、統計学者、疫学者、生物医学雑誌編集者らの国際的なグループであった。その後、初版に対する批判も取り入れて2001年4月に改訂勧告として公表した。この改訂版CONSORT声明は22項目のチェックリストとフローチャートからなる (Moher他 2001a; 津谷他 (訳) 2002)。

チェックリストは、「タイトル・抄録」、「はじめに」、「方法」、「結果」、「考察」のカテゴリ毎に、報告書を書く際、具体的にどのような内容を記述すべきかの要点について、全体の通し番号を付し列挙してある。例えば「方法」のカテゴリでは、参加者の適格基準、意図された介入の詳細、目的と仮説、アウトカム評価項目の定義、症例数、ランダム化の順番、割振りの具体的な実施方法、マスキングの方法、統計学的手法について、3~12番までの10項目に整理されたチェック内容がリストされている。

フローチャートでは、「組入れ」、「割付け」、「追跡」、「解析」の各段階の参加者数について、

その例数を示し、追跡不能や除外された場合にはその理由を明示することを提案し図示することを推奨している。

併せて公表された解説と詳細 (Altman他 2001; 津谷他 (訳) 2002) には、具体的な例示およびコメントが加えられ、さらに、専門用語の解説を兼ねた用語集も含まれている。また、末尾にはこれらCONSORTグループの会議開催にあたり、米国国会図書館や製薬企業名を記し資金提供を受けたことを明示している。このことは、公開の場で公明正大に行う活動主旨にも整合している。

声明本文にコメントとして、報告書は可能な限り平明に記述すべきであるとし、CONSORT声明が将来試験をデザインする際の助けとなるのみならず、論文評価に際し査読者および編集者の指針にもなりうることを述べている。このことは、バイアスが入り込まない配慮を試験実施者および執筆側側に求めつつ、同時に、広く将来の評価者やエビデンスの利用者らにもCONSORT声明が寄与することを示している。読者側が、優れたデザインの試験と結果の疑わしい試験とを容易に見分けられる具体的な枠組みを提案し、実際に使える形として広く世界に提供したことは意義深い。

CONSORT声明の初版 (1996) は、その限界にも拘わらず、実際にこれを採用した雑誌におけるRCT報告の質の改善に寄与したことが報告されている (Egger, Jüni and Bartlett 2001; Moher, Schulz and Altman 2001a, 2001b, 2001c)。CONSORT声明は常に発展中の文章であり、新たなエビデンスや批判的なコメントの集積に応じて、継続的な評価、改良・変更が必要であることを付け加えている。

現在、英語の他、中国語、オランダ語、フランス語、ドイツ語、イタリア語、日本語、韓国語、ポルトガル語、ロシア語、スペイン語に翻訳され、CONSORTグループに関する他の情報も含めインターネット上で公開されている<sup>4</sup>。

なお、上述のとおりCONSORT声明は、報告書に含めるべき項目が漏れないという点で極めて有効であることは実証された。しかし、報告書にCONSORT声明の枠組みを用いたからといって研究の質が保証されるわけでないことは自明である。CONSORT声明は報告書作成時点で単独で利用するものではなく、研究計画の立案段階から活

用してはじめてその効用が期待できる。すなわち、報告書を書く以前の問題として、選択する研究デザインや、対象の選択基準、試験実施とその後のデータの取り扱いなど、実施試験全体の質に応じて研究報告内容の質も大きく変わることについて十分留意したうえでCONSORT声明を利用することにその意義がある。

その有用性が認識され各方面への拡張がなされ、RCT領域では、2004年11月にクラスター・ランダム化試験 (cluster randomized trial) について (Campbell, Elbourne, Altman 2004; 津富・津谷 (訳) 2008)、またRCTにおける害について (Ioannidis, Evans, Gøtzsche 他 2004; 八重・大橋 (訳) 2008)、が作成された。さらに、非劣性および同等性への拡張 (2006年3月)、ハープ介入RCTへの拡張 (同年3月)、中薬<sup>5</sup>のRCTへの拡張 (2006年5月)、雑誌および会議録の抄録に含まれるべき項目についての拡張 (2008年1月) がなされた。

## (2) ランダム化比較試験以外の研究デザイン

CONSORT声明の成功に刺激され、エビデンスを正しく「つたえる」ためにRCT以外の研究デザインにおいても同様のガイドラインの作成が始まった。

行動介入および公衆衛生学的介入を評価した非ランダム化研究報告の質改善 (TREND声明 2004)、診断確度に関する研究報告のためのルール (STARD声明 2003)、また、疫学における観察研究の報告の強化のためのガイドライン (STROBE声明 2007) および観察研究のメタアナリシスに関する提案 (MOOSE声明 2004) がそれにあたる。以下、簡単にそれらの概要を示し、研究報告の質向上にこれらも大きく貢献してきたことをレビューする。

TREND (Transparent Reporting of Evaluations with Nonrandomized Designs) 声明は、その名のとおり、非ランダム化デザインにて実施された研究内容の透明性確保を目的としたものである (Des Jarlais 他 2004)。

EBMの観点からは一般に、RCTに劣るデザインから得られるエビデンスにはほとんど価値がない、あるいは無価値であるというみなされることが多い。しかし、「エビデンスに基づく公衆衛生

(evidence-based public health: EBPH)」の取り組みや公共政策評価などの場においては、純粋なRCTを導入することがむしろ困難な場合の方が多く、したがって、EBPHにおいては、RCT以外の手法を用いた研究が必要であるとの強い主張もある(Victra 2004)。彼らは、公衆衛生的介入を評価するには、RCTの実施が非現実的または非倫理的であることが多いことを指摘し、「妥当的デザイン」(plausibility design) および「適切性デザイン」(adequacy design) と称した非ランダム化デザインから因果関係を推論するための手法について検討している。

また別の研究者らからも、グループランダム化試験(group-randomized trial) についての問題点を指摘しつつも、RCTの代わりとして適切な場合があることなどの議論がある(Donner, Klar 2004, Murray 2004, Varnell 2004)。この他にも、準実験デザイン(quasi-experimental design)、非ランダム化試験(nonrandomized trial)、自然の実験(natural experiment) などがあるが、これら非ランダム化デザインを用いた研究を体系的に活用するのであれば、可能な限りその質を改善する努力をすべきである。

このような背景のもとで、非ランダム化研究においても、研究の理論や介入条件および比較条件などについて明示するなど、報告書の質を高めるための「非ランダム化研究の透明性の高い報告基準」(Transparent Reporting of Evaluations with Nonrandomized Designs)、いわゆるTREND声明が公表された。

この声明では特に、研究の理論や介入条件および比較条件に言及し、存在するバイアスの調整にどのような手法を用いたかを明確にすることに重点がおかれている。想定されるデザインが多岐にわたるためか、この声明にフローチャートは含まれないが、「タイトル・抄録」、「はじめに」、「方法」、「結果」、「結果」、「考察」の大項目のもとで具体例とともに22項目のチェックリストとしてまとめられている。

また、STARD (Standard for Reporting of Diagnostic Accuracy) 声明は診断確度を報告する際の基準であり、本文中に25項目からなるチェックリストとフローチャートの例が示されている

(The STARD Group 2003, Bossuyt他 2004)。新規の診断技術と既存の診断技術を評価する際の報告書作成についてチェックリストの項目毎に具体的事例を示した解説も同時に公表した(Bossuyt他 2003)。他の声明同様、いかにバイアスを防ぐかという視点に加え、結果の一般化可能性(generalisability) について読者側で評価可能な情報を含む報告書を作成できるようガイドすることがその主目的である。

さらに、非ランダム化デザイン、診断確度報告の質改善の他、疫学研究においてよく用いられる観察研究の報告書の質向上のためのガイドラインとしてSTROBE声明が公表された(Elm他 2007)。

観察研究とは具体的に、既存のデータベースの活用、横断研究、症例集積、ケース・コントロール研究、歴史提唱研究、コホート研究などのデザインにより実施された「病因または有効性に関する研究」などを示す。

観察研究の報告はデザインの性質上往々にして介入と結果との因果関係の立証が難しく、また研究の強さと弱さおよび一般化可能性の評価が困難である。STROBEイニシアティブは、コホート研究(cohort study)、ケース・コントロール研究(case-control study)、および横断研究(cross-sectional study) の三つを主要な試験デザインと定めて検討を行い、報告書の「タイトル」、「抄録」、「はじめに」、「方法」、「結果」および「考察」に関連する22項目のチェックリストを公表した。

### (3) 統合型研究：QUOROM声明など

RCTの報告書の質改善を目的としたCONSORT声明に対して、1999年にはメタアナリシス報告書の質改善のためのQUOROM (Quality of Reporting of Meta-analyses) 声明が発表された。

QUOROM運営委員会は統合型研究の質向上のための指針として、システムティック・レビューやメタアナリシス報告における質の向上をめざし、声明本文とともにチェックリストとフローチャートを公表した。チェックリストは、メタアナリシス報告の「抄録」、「はじめに」、「方法」、「結果」、「考察」の各項目に整理してある。それらは、文献検索、選択、妥当性評価、データ抽出、研究の特性、定量データの統合に関する副見出しを含

む21項目で構成されている。またフローチャートには、特定されたRCTの件数、採用件数、除外件数とその除外理由についての情報を整理できるよう工夫してある。

その後約8年を経てQUOROMはPRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) 声明として名称変更し進化した。PRISMA声明は27項目のチェックリストと4段階のフローチャートからなり、その解説補足資料も同時に公表している。QUOROMからPRISMAへの名称変更の理由のひとつとして、提唱者らはシステマティック・レビューとメタアナリシスの両方を包含したことをあげている。

メタアナリシスは、システマティック・レビューにおいてレビュー対象とする研究結果を統合するために用いる統計学的手法である。この統計手法はシステマティック・レビューに用いられることもあれば、用いられないこともある。すなわち、名称変更には、QUOROMが対象としていたメタアナリシス報告書の質改善のための声明が単純に改訂されたものではないことをアピールする意味と目的が込められている。

システマティック・レビューはいまや、費用対効果 (Ladabaum他 2001)、診断 (Deeks他 2001)、予後 (Altman 2001)、遺伝的関連性 (Ioannidis他 2001)、および政策決定 (Lavis他 2005) などで広く行われており、実施中にレビュー・プロトコルを変更しなければならない状況も生じ得る。PRISMA声明は、そのような変更が不適切であるとして指摘するのではなく、むしろその報告と説明記述を促すよう配慮したチェックリストを提供している。

因みに、観察研究のメタアナリシスに関する提案としてはMOOSE (Meta-analysis Of Observational Studies in Epidemiology) がある (Stroup他 2000)。

医療や公衆衛生の諸課題を検討する際、RCTによる介入研究を実施することが困難な場合は多い。そのような現場、観察研究 (observational study) の情報を最大限活用することが肝要である。しかし、観察研究のメタアナリシスには研究固有のバイアスや研究デザインの差異から生じる特有の問題が生じやすい。この問題について検討グループは、観察研究のメタアナリシス論文のシステマテ

ック・レビューを実施し2004年にMOOSE提案としてチェックリストを提案・公表した。

この他、診療ガイドライン (clinical practice guidelines: CPG) に関するルール (COGS提案 2003) およびエビデンスの質評価に関するルール (GRADE 2004) がある。

ガイドライン標準化協議会 (Conference on Guideline Standardization: COGS) は診療ガイドラインの標準化に関わる提案を行った (Shiffman他 2003)。いわば、ガイドライン作成のガイドラインである。

COGS提案作成にあたり診療ガイドラインの構成要素について、既存ガイドラインなどを元に要素項目を抽出し、抽出された項目名が診療ガイドラインに必須な要素であるかについて各種手法により評価し、かつ、ガイドライン作成に携わる組織の代表者らによる評価を経て、18のトピックにまとめられた。

ガイドラインにばらつきが生じる原因のひとつとして、ガイドライン執筆経験の少ない人々が作成に携わっていることがあげられる。この点を鑑み、COGSは共通の枠組みを示すことにより、各診療領域のガイドラインの質を高め、より包括的なものにする事目的とした。CONSORT声明同様にCOGS提案も必要な要素を特定する目的で非常に有効であるが、ガイドラインそのものの質を保証するものではないことには留意する必要がある。

#### (4) 相補代替医療関連

介入としてユニークな特徴を持つ相補代替医療 (complementary and alternative medicine: CAM) 領域での拡張は、今後の医学領域以外におけるRCTの参考ともなるものであり、やや詳しく説明しよう。

CAM領域では以前よりエビデンス産出のニーズが高く、1996年には米国国立衛生研究所 (National Institute of Health: NIH) の代替医療室 (Office of Alternative Medicine: OAM) の資金援助により、コクラン共同計画にComplementary Medicine Fieldが設置された。なお、OAMは1993年に設立されたものだが、5年後の1998年に国立相補代替医療センター (National Center for

Complementary and Alternative Medicine: NCCAM)に昇格している。

CAMは大きく「もの系」と「ひと系」に分けられる。前者の代表は薬物系治療で、日本や中国では生薬・中薬と漢方処方・漢方方剤がこれにあたり、後者の代表は鍼である。これら領域では早くからその評価にRCTが取り入れられ、またCONSORT声明の動向にもいち早く対応している。

1997年11月にはNIHで鍼のコンセンサス形成会議が開かれ、それを受けた声明が発表された<sup>7</sup>。その後2007年にはSAR (Society for Acupuncture Research)<sup>8</sup>、統合医療センター (Center for Integrative Medicine)、メリーランド大学の共催にてコンセンサス形成会議10周年を記念するSAR学術集會が開催された。会議では、鍼研究に対するNIHの助成額が増加するなかで研究対象が様々な変化してきた過程、研究の現状、エビデンスのために解決すべき問題が明らかにされるとともに、今後の鍼研究への提言がなされた (高澤 2008)。

一方、2001年には、既に鍼の臨床試験におけるデザインと報告に関する統一規格が提案された。STRICTA (Standards for Reporting Interventions in Controlled Trials of Acupuncture) グループとIARF (International Acupuncture Research Forum) の推奨がそれである。

鍼灸治療を西洋医学コミュニティが需要する過程において、その基礎と臨床の両面からの研究が進められ、特に鍼の臨床試験とその評価に大きな進展がみられた。

鍼の臨床試験は医薬品のそれをモデルとしてきたが、その適用は順調には進まなかった。鍼では医薬品開発のRCTで用いるプラセボに該当する空治療を患者、施術者間でマスクすることはが難しい。また、鍼灸療法の配穴 (刺鍼部位、複数選択が可能)、刺鍼深度、刺激方法、刺激時間、使用鍼、さらには、流派や術者の熟練度、時代や地域による習慣の違いなどが多岐にわたるなど、介入そのものの標準化が困難であるという問題があった。加えて、結果の解釈についても尺度の共通認識は確立されておらず、必要情報を適切に記述する様式についての流派を越えた取り決めも存在していなかった。

2001年、経験を積んだ鍼灸師と研究者らの国際

的なグループがイギリスに集まり、STRICTA推奨を起草し、雑誌編集者らも巻き込み、これを各国の当該領域の雑誌に採用することを促した。CONSORT声明同様、STRICTA (鍼の臨床試験における介入の報告基準) のためのチェックリストを整備した。特に、外部妥当性に直接影響する、鍼治療の方法とコントロール群の設定にあたり、当時妥当と考えられた鍼の臨床試験デザインをとりまとめて推奨した。

CONSORT声明は、その後、ハーブ介入RCTへの拡張 (2006年3月) と、中薬のRCTへの拡張 (2006年5月) がなされている。

以上、(1)~(4)の各項目で示した領域における展開から、各イニシアティブは研究デザインそのものの発展を促すとともに、関係領域相互に影響を及ぼしつつ、「つたえ」方の標準化を進化・発展させてきたことが伺える。

このように、現在もなお多方面から研究報告の質を高めるための努力が続いている。これらの声明や推奨勧告の殆どは邦訳されており、解説などとともに別書籍にまとめられている (中山・津谷 2008)。それぞれのチェックリストやフローチャートなどの詳細についてはそちらも参照頂きたい。

### 3. ひとを動かすベストケース

#### (1) 叙述 (ナラティブ) に基づく医療 :

##### Narrative-based Medicine: NBM

エビデンスの「つたえ」方の標準化の進展に伴い、エビデンスのエッセンスを凝縮して「つたえる」二次情報誌の盛衰が生じた。その背景についての議論のひとつに、EBMという言葉が広く一般に浸透するとともにそれへの慣れや飽きと倦みに関わる話がある。また、定式化されたEBMのプロセスなど、教条主義に対する倦みもありEBM離れも生じつつあった。何にせよ、EBMが普及すればするほど、本来のEBMの概念が適切に伝えられることなく、いわば定式が形骸化したEBMが流布する状況が生じていたことが伺える。

1990年代中頃からは、EBMは全盛時代を迎え、EBMあるいはエビデンスベースドを冠する書籍

や雑誌、セミナーなど枚挙にいとまがない状況であった。このこと自体は、EBM推進のための追い風となる好ましい状況であり、多くの推進者の努力に支えられ、確実にEBMが日常診療の基盤として浸透していった。医療現場でエビデンスの有無は総合判断を行うための判断材料の極めて重要な一角を占め、その「科学的な証拠に基づいている」という安心感とともに広くEBMが重視されるようになった。

しかし、その反面、時に、権威者の前で若手がエビデンスの有無についての論争をはじめたために、人間関係が悪化するなど、問題となるケースも生じたようである。

このような現場の経験やエビデンスが産出されていない領域における判断材料として、EBMとは一見対極にあるようにもみえる症例報告の有用性が再認識された。そこでは、EBMが普及・浸透するさなかにおいて、流行に流されることなく、EBMをも踏まえた医学的決疑論の重要性が尊重された。

また、必ずしも大きな流れとはいえないが、EBMに対峙する呼称を携え、EBMを補完する実践的な活動として、ナラティブ (narrative-based medicine: NBM) も現れた (Greenhalgh and Hurwitz 1998, 斎藤・岸本 2003)。

ここでは、EBM盛衰との兼ね合いも踏まえて、ベストケースとNBMについて簡単に補足しておきたい。

精緻に統計学的手法を用いたメタアナリシスやforest plotができない場合でも、叙述情報 (ナラティブ) が現場での判断材料として役立つケースがあることが次第に見直されてきた。

患者の語る物語 (ナラティブ) には解釈の独自性、およびその解釈が成り立ちうる個人的社会的背景が含まれている。こうした独自の背景を理解することこそが、医療者にとって必要不可欠であるという元来からの考えがある。

医療は科学としての医学に立脚しているのはもちろんであるが、治療にあたり患者のおかれている状態をナラティブから解釈することと、実証科学的な方法論から得られたエビデンスとをバランスよく組み合わせることによって、真に高いレベルの医療が可能となる。

NBMはEBMに対峙したりそれを軽んじたりするものではなく、むしろ実証主義的な方法論の限界を補完するものである。すなわち、患者のおかれた状態や疾病に対する認識を質的な観点から解明し、治療に取り入れEBMを補おうとする立場である。医療者自ら、行う診療がエビデンスに基づくものであることは暗黙の了解であるとしても、ナラティブにも十分な配慮をしているかを改めて自問することは、その結果として統合された診断・治療の改善に必ず役立つはずである。

## (2) ベストケースはpost-RCTのひとつの流れ

ベストケース (best case) は、dramatic、extraordinary、unusual caseなどとも呼ばれ、広義には必ずしも著効例のみを示さず、劇的な、並外れた、特異な、驚くほどの、尋常ではない症例を一般にこのように呼ぶ。こうした個別の症例から得られる知見は臨床上也重要であり、ベストケースの観察から、臨床試験では知ることができないような稀でかつ重要なエビデンスを得られる可能性が秘められている。

並外れた良い (あるいは逆に悪い) 結果が得られた症例の背景を詳細に記録し、叙述的な内容も含めて、効果に寄与した可能性のある全ての背景情報を探索的に検討することからエビデンスを見いだそうとする試みである。コントロールとなるケースと対比させてその原因と考えられる差異を見いだす手法は、疫学研究におけるケース・コントロール研究の方法論に通じるところがある。

Best/worst caseを上手に表現することは集団を対象とした臨床試験からは、容易に得ることのできない種類の貴重な所見を効率よく抽出し、それらを「つたえる」ことに対応する。ただし、こうして得られたエビデンスのレベルは一般にEBM上は低いものと考えざるを得ない。したがって、ベストケースから得た所見は、あくまでも仮説と考え、可能な限り、一般化可能性を証明するための努力を行うべきであろう。効果量 (effect size) の推定が可能となる臨床試験を実施することにより、その効果が得られた背景についての新たな有益な知見の発見につながることもある。

ベストケースについての叙述的内容の方が、時に、現場ではより役立つ可能性があることが見直

され、EBMを推進するうえでどのような症例報告書くべきかという議論に発展した (Jenicek 2001; 西・川村 (訳) 2002)。

症例報告は、医療の現場で観察された唯一無二の経験として記録されるとともに、より広く医学の知識・経験として、一般化された知恵の一部として扱われるべきである。ベストケースは、post-RCTを踏まえたEBMのひとつの流れとして定着しつつある。

国内では、日本東洋医学会 (The Japan Society for Oriental Medicine: JSOM) において、「葛根湯プロジェクト」を皮切りに、ベストケース・タスクフォースの活動が始まっている。

#### 4. EBMのさらなる進展：将来展望

##### (1) エビデンス利用とアウトカム評価：継続的評価の重要性

昨今、大規模な長期臨床研究により、治療や検診を受けなくても5年10年後の生存率などの結果に大きな差がないことなどが明らかにされる場合がある (Thomas et al. 2002; Lu-Yao et al. 2002)。その際、かかったコストは適切であったか、また、侵襲的な介入がなされた場合、患者らのQOLや精神的苦痛などを減ずる方策が、具体的にどのようにその後の治療に反映されたかについての評価も必要となる。個々のRCTの結果は比較的短期間での介入効果に興味が置かれることが多く、長期的インパクトやそれに引き続いて得られるアウトカムについての情報が提供されないため、それらを明らかにするための枠組みも必要である。

アメリカでは1993年の「政府業績成果法」(Government Performance and Results Act: GPRA)の成立を受け、既に巨視的視点からの評価が行われている。例えば、1999年に米国国立衛生研究所 (National Institute of Health: NIH) が行った調査報告によれば、NIHが投資した29の疾病領域別の研究費と障がい調整生命年数 (disability adjusted life years: DALY<sup>10</sup>) との間には正の相関が認められたが ( $r=0.62$ ,  $p<0.001$ )、疾病の発生率、有病率、入院日数については、相関は認められなかった (Gross et al. 1999)。

このように個々の介入効果のエビデンスを「つくる」ことに加えて、今後は、そのエビデンスが実際に社会全体にどのように寄与しているかという公益性の評価をするための制度や仕組みづくりも必要であろう。

対費用効果も含めて公益性が明確でない場合やなんらかの「害」や無駄が見いだされた場合には、その原因を究明し即座に対応しなければならない。2009年11月に行われた民主党政権下での行政刷新会議による「事業仕分け」においても、本来、こうした中長期的かつ継続的な評価結果が整っている状態で、それぞれの「場」における個々の利害を超えた議論が行われることが望まれる。

##### (2) 「つたえ」あうこと：フィードバックを受けるとの重要性

エビデンスの質についての理解が伴わない場合の議論は難しく、事業仕分けの議論の難しさとも通じる部分がある。すなわち、主旨や背景を共有できれば、単に事業の「廃止」を求めている訳ではない場合でも、メディア報道による「〇〇事業廃止」、「〇〇に関する予算削減」といった単純化された情報として一般に「つたわる」傾向がある。エビデンスも事業仕分けも結論のみが単純化かつデフォルメされて「つたわる」ことが総合判断をゆがめる原因となり得る。

報道機関と多様な教育レベルや生活環境におかれている一般市民との間において、「つたえる」側と「つたえられる」側のメディアリテラシー<sup>11</sup>の格差が生じることは容易に推察できる。公共のメディアは「つたえる」べき相手の平均的な像を想定し、それに応じた最も適切な「つたえ」方を選択しているはずであるが、受け手のもつ情報量のギャップや「場」の違いによる考え方の多様性がある状況における主旨伝達にはより慎重かつ十分な配慮が必要である。

本来、最新情報を正確に提供する役割を担うマスメディアによる報道も、限られた時間の中では内容を要約し、また時にはデフォルメした情報として伝えざるを得ないこともある。そのような場合には、意図せず、元の問題を助長したり、新たな問題を誘発したりするといった想定外のリスクを抱え込むことも生じ得る<sup>12</sup>。一般市民のリスク

を煽り、結果的に社会不安を誘導する事態は絶対に避けなければならない。

この点に配慮すれば、「つたえ」なければならない相手に、相互に「つたえ」あうことのできるしくみが必要となる。

公共政策の評価・判断を行う場合も、利害関係者が相互に顔を合わせてその内容理解を共有する場が重要となる。公聴会やタウン・ミーティングなどの市民集会の場で当事者らが直接意見を交換する場を設けることは基本であろう。

さらに、これまでの新聞・雑誌、ラジオやテレビといった一方向性のメディアに加えて、インターネット普及にともなう双方向性のコミュニケーション環境の利用も有効と考えられる。ホームページやブログに加えて、最近急速に利用が拡大しているツイッター（twitter）<sup>13</sup>も民意を相互に直接「つたえ」あうためのインフラとして役立つことが政治家らの間でも徐々に理解され浸透しつつある。

今後のEBP普及にあたり、その評価・判断根拠となる質の良いエビデンスを「つくる」と同時に、一般市民にもそれを上手に「つたえる」必要がある。その役割を担うマスメディアには、常に「報道の質の改善」を念頭においた、より積極的な対応が望まれる。一方で、市民側もメディアリテラシーを高め、目的とする判断母体に対する的確なフィードバックを行うことが重要である。

次の項では診療ガイドライン作成の経緯と背景を振り返りつつ、受益者である患者や市民参加の役割と重要性について述べる。

### (3) 患者・市民の役割

2009年9月、厚生労働科学研究班「診療ガイドラインの新たな可能性と課題：患者・一般国民との情報共有と医療者の生涯学習」主催のシンポジウム「診療ガイドライン・医療政策の場における患者・市民の役割 ～英国NICEと日本の現状～」が開催された。先述のとおり、診療ガイドラインはEBMが広く展開していく中で、各診療領域のガイドライン作成のためのガイドラインである。ガイドライン作成の要となるガイドラインにおいてこうした患者・市民参加に関わるテーマが取り上げられた背景について触れておきたい。

診療ガイドライン研究班は2001年から3年度毎にテーマを定めて検討を進めてきた。第1期はEBMを指向した「診療ガイドライン」作成のための方法論の検討が行われた。第2期ではガイドライン作成の基盤整備の一環として患者・医療消費者の参加に向けての検討が行われた。第3期には診療ガイドラインの新たな可能性と課題を検討テーマとし、特に、患者・一般国民との情報共有と医療者の生涯学習をテーマとした検討が進められてきた。

2004年の第2期から既に、患者や医療消費者の参加にかかわる視点が盛り込まれていたが、2007年からの第3期において、より明確に患者・一般国民との情報共有の重要性を取り上げ、診療ガイドラインの新たな役割として、医療者と患者（と家族、介護者）の情報共有・コミュニケーションの基点としての位置づけについての検討が進められている。具体的には、患者らも交えて一緒に決めるshared decision making（協働的意志決定<sup>14</sup>）を行うことの重要性を明示した診療ガイドラインの改訂作業が進んでいる。

このような流れは日本独自のものではない。1999年に設立されたイギリスのNational Institute for Health and Clinical Excellence（NICE）においては、設立当初よりガイドライン作成時に患者・一般市民（すなわち受益者）の参加の必要性が提唱され、「専門家としての患者（受益者）」の意見をガイドライン作成段階から組み込む仕組みとして運用している。

各診療ガイドライン作成時には、そのメンバーとして患者・一般市民の数名を加えるルールとし、NICE内のPatient Involvement Unit（PIU）が患者団体や関連学会を通じてメンバーの公募を行っている。個人応募も可能であり、応募者は経歴と小論文を提出し、ガイドラインのトピックと自分の関係や考えなどを述べる。書類選考を経て、NICEのガイドライン作成メンバーが応募者に電話インタビューを行い、当該診療域とのかかわり、バランスや人物の総合評価の後に最終決定する。

さらに、診療ガイドラインが作成される毎にプレスリリースが発行され、新聞やテレビなどの各メディアにて大きく取り上げられている。このように、イギリスでは医療従事者のみならず、患

者・一般市民の間でもガイドラインの認知度を高く維持するための仕組みがある。

#### (4) EBM展開の方向性

ベストケースや患者・市民参加の話がでてくると、改めてEBM展開の方向性に疑問を呈する方が出てくるかもしれない。この疑問への回答には、1980年代よりEBMの概念を構築・推進してきたMcMaster大学のSackettらの概念説明と用語の使い方の微妙な変化が参考となる。

1990年代中頃EBM実践についてSackettらは以下のように述べていた。「EBMの実践は、個人の臨床的専門技能と系統的研究から得られる最良の入手可能な外部の臨床的エビデンスとを統合することを意味する。」“The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research (Sackett et al. 1996).”

その4年後の2000年に、EBMハンドブックを改訂した折、後半部分の記述を以下のように変更した。「...臨床的専門技能と患者の価値観をもって、最良の研究から得られたエビデンスを統合」“...the integration of best research evidence with clinical expertise and patient values (Sackett et al. 2000).”

すなわち、系統的研究から得られる最良の入手可能な「外部の臨床的エビデンス」の部分が「患者の価値観」に置き換わっている。このことは、エビデンスを「つかう」機会が増えるにしたがって、「場」の状況に応じた対応が求められることを示している。言い換えれば、RCTやシステムティック・レビューによる系統のあるいは精緻な研究から得られるいわば「上澄みの科学的エビデンス」のみを追求する姿勢から、「場」の状況に応じて患者の好みや価値観を尊重し、これを取り込んだ（合成した）総合判断の重要性が増すことを示唆している。

エビデンスを「つかう」段階においては、システムティック・レビューからは決して得ることができない、また、容易に説明することもできない個々の患者やその家族のニーズをも含めた「場」の状況に応じた柔軟な判断も重要であることが改めて見直されつつある。

こうした変化は、生命倫理や人間の尊厳といったEBM普及以前からの議論への逆行と捉えるのではなく、むしろ、実際にエビデンスを蓄積したうえで実践をも踏まえた、より現実的な問題を検討可能なレベルに達したものと考えるべきであろう。高次の意志決定や評価判断を行う際の具体的かつ実践的に解決すべき課題としてとらえ、今後ともさらに議論を深める必要があるだろう。

ここで重要なことは、「堅い」とされる科学的にも検証された真に正しいと考えられるエビデンスだけで人が動くものではないということである。この点を、改めて認識したうえで、いかに対応すべきかを実践的視点から検討する段階に至ったものと考えられる。エビデンスが蓄積された現時点において、EBMの実践の場から得られた経験を真摯に受け止め、これを踏まえたより現実的な対応を模索・検討しなければならない。今一度、EBMが本来目指してきたSackettらの基本概念に立ち返り、人を対象とした実践的かつ全人的な価値判断を含めたエビデンスの活用に基づいたEBMのさらなる展開が望まれる。

このように、EBMの普及に伴い、また時代のニーズにも応じて、EBMそれ自体に対する考え方が現実の社会にも整合し充実しつつ発展してきたことが理解できる。これまでに保健医療領域で得られた課題と対策検討の経験は、追従する他の多くの領域のEBP実践の場において、その将来像を見据えた検討に役立つであろう。

## 5. おわりに

EBMの普及とその実践上の検討課題から得られた教訓は、EBPを積極的に推進している立場の方からみると、やや意外ともいえる展開となっているかもしれない。確かに、EBP推進の初期段階では、いわば「RCT/システムティック・レビュー原理主義」あるいは「RCT/システムティック・レビュー至上主義」といった立場をとることが必要な場合もあった。

近年、疾病・薬剤疫学研究から得られる情報や過去の経験を踏まえたモデルとシミュレーションを駆使した（仮想的な）効果予測に基づく医薬品

開発 (model-based drug development: MBDD) が始まっている。これにより、適用範囲や対象をより限局したRCT実施も可能となり、いわばティラーメイドの医療を可能とする精度が高く堅いエビデンスが得られる可能性もある。実際、これまでの集団 (population) を対象としたRCTによる全体集団的エビデンスとは異なり、対象を絞った明快かつ限局的利用も可能な、キレの良いエビデンスを「つくる」ことは好ましいことである。

しかし、仮に精緻なデザインや理論的モデルを用いたキレのよい明確なエビデンスが得られ、かつそれが正確に「つたえ」られたとしてもなお、エビデンスを「つかう」段階に至れば、その明確なエビデンスをしても「場」の状況に応じた意志決定において、どの程度重視されるかは個々のケースで変わり得る。これがEBMにおける本来のエビデンスの使われ方であることについてEBPを推進する立場にある者は皆、認識を新たにしておくべきであろう。

すなわち、これまでEBPの推進者が追求してきたエビデンスを教条主義的に使わせるのではなく、EBMが本来のコンセプトに従って、必然的に患者の好みや価値観をうまく統合すべきであるとの方向に向かってきたことについて再認識する必要がある。EBPを推進しようとする者が現段階でこのことを認識しつつ将来の発展を思料することが重要である。

患者や市民の好みや価値観を上手に統合するためには、EBPの実践の場に関わる全てのひとの積極的関与が必須である。しかし、その合意形成は非常に複雑なものとなることは容易に想像できる。その際、質の高いエビデンスが存在していることが前提となることに変わりはない。重要なことは、それをいかに上手に「場」の関係者に「つたえ」有効かつ有益に「つかう」かである。

「場」の関係者が多い場合には、最終的に全員が納得できる合意形成に至らないケースの方がむしろ多いかもしれない。その場合にも、EBPを推進する者は最後まで「つかう」場の状況を見守りつつ、常に協調する姿勢をもって臨むべきである。建設的、発展的、肯定的、前向きな合意形成のための総合判断に必要なエビデンスの強度をその質とともに正確に「つたえる」ことが肝要である。

適切にエビデンスを「つたえ」、「つかう」場に即したセカンド・オピニオンを示し続けてゆくことになるのかもしれない。

一方で、こうした“shared decision making”に対する課題も認識されはじめている。例えば、年齢や教育レベルによってはむしろ父親的態度 (paternalism) が望まれることがある。しかし、どのような場合にこれを適用すべきであろうか？あるいは、限られた時間的・経済的制約のもとで全ての関係者の意向を汲み取って治療方針を決定することの是非についての課題もある。精神疾患患者や自殺企図をもつ患者の自主性 (autonomy) は認めるべきであろうか？ 実際、限りある医療資源 (人員、財源、時間) で、全ての患者やその家族の全ての希望には対応しきれない。個々の判断の多様性が過大となれば、医療制度の方針決定が困難になる。どのような場合にいかに短時間で合意形成を導くかは今後もエビデンスを適用する現場における大きな検討課題である。

仮にエビデンスと個々の価値観を合成するパラメータが文化や時代とともに変化し続けるとしても、わが国におけるEBPの普及を推進し、その適切な利用を可能とする人材を育成・涵養することが、次世代に向けての我々の責務であることに変わりはない。

## 謝辞

本稿作成にあたり、診療ガイドラインや医療政策における患者・市民の役割にかかわる資料提供を頂いた京都大学大学院医学研究科教授 中山健夫氏に感謝致します。

## 注記

- 1 シンポジウム「学際領域における評価のデザイン-RCTとシステマティック・レビューの現状 (2006/2/18)」にて栄養、運動、温泉、教育、司法、国際援助の各領域における現状報告があり、そこからも同様の状況が推察された。
- 2 コクランライブラリCENTRALに登録されたRCTの数は600,472件 (2009/10/23) であった (既報 2006/1/3の調査 463,763件から3年10ヶ月弱で約14万

- 件の増加)。
- 3 2009年10月30日開催の医中誌編集委員会配布資料(非公開)を参照した。
  - 4 CONSORT声明の考え方や、これを用いている雑誌のリスト、活動に関わるニュースレターなどを閲覧できる(CONSORT Website: <http://www.consort-statement.org/>, accessed Feb, 2010)。なお、本稿執筆時点(2010/2/22)で2004年の拡張版については、中国語と日本語に翻訳中との記載があった。
  - 5 「中薬」とは日本の漢方薬に対応する中国語である。原論文ではChinese herbal medicine (CHM: 中草薬)と記されている。処方内容や製剤技術が異なるため、日本では一般に「中薬」として漢方薬と区別する。
  - 6 Webのみの提供でありダウンロード可能なwordテンプレートの掲載がある(<http://www.prisma-statement.org/statement.htm>, accessed Feb, 2010)。
  - 7 NIHのウェブサイトにて参照可能(<http://consensus.nih.gov/1997/1997Acupuncture107html.htm>, accessed Feb, 2010)。
  - 8 鍼と東洋医学の世界に科学的な手法で刺激を与えることを目指し1993年に設立されたNPO組織。鍼やハーブ療法などを研究する際の方法論について毎年シンポジウムを開催している(<http://www.acupunctureresearch.org>, accessed Feb, 2010)。
  - 9 日本東洋医学会EBM特別委員会ベストケース班(Best Case Task Force)により企画運営される研究プロジェクト([http://www.jsom.or.jp/html/ebm\\_bc/project.html](http://www.jsom.or.jp/html/ebm_bc/project.html))。
  - 10 Chris Murrayらが提唱した傷病や障がいの程度や期間によって重み付けをした生存年数指標。DALY=YLL+YLDで表される。ここでYLL: Years of Life Lostは死亡数と(年齢階層別の)平均余命の積、YLD: Years Lost due to Disabilityは(障がい発生数)×(障がいの重度)×(余命の損失年数)。障がいの重み付けの恣意性や障がい後の余命に与える環境因子が先進国と途上国では異なるが、これらを考慮しない指標であるとの指摘と批判がある。
  - 11 リテラシー(literacy)には識字などの、読み手の立場の状態のみならず、書き手や説明する側の「つたえ」る力も含まれる。
  - 12 既報にて述べた「説明(過不足)のジレンマ」問題がここでも生じ得る。

- 13 2006年7月にアメリカのベンチャー企業が始めたサービス。フォームに140文字までのテキストを入力し「つぶやき」(tweet)として公開するのが特徴。素早く端的に公開の場における双方向コミュニケーションを可能とする。
- 14 現時点で定訳はないが、中山はその意味を踏まえて「共働的意志決定」という訳語を提案している。

## 参考文献

- 斎藤清二・岸本寛史(2003)『ナラティブ・ベイスト・メディスンの実践』、金剛出版
- 中山健夫・津谷喜一郎(編著)(2008)『臨床研究と疫学研究のための国際ルール集』、ライフサイエンス出版
- 正木朋也・津谷喜一郎(2006)「エビデンスに基づく医療(EBM)の系譜と方向性—保健医療評価に果たすコクラン共同計画の役割と未来—」、『日本評価研究』、6(1):3-20
- Altman, D. G., Schulz, K. F., Moher, D., et al. (2001). CONSORT GROUP (Consolidated Statements of Reporting Trials). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*, 134, 663-694. 津谷喜一郎, 栗原千絵子訳(2002)『臨床評価』30:163-208
- Altman, D. G. (2001). Systematic reviews in health care: systematic reviews of evaluations of prognostic 39 variables. *BMJ*; 323,224-228.
- Begg, C., Cho, M., Eastwood, S. et al. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT Statement. *JAMA*, 276(8), 637-639.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., et al. for the STARD Group. (2003). The STARD Statement for reporting studies of diagnostic accuracy: Explanation and elaboration. <http://www.consort-statement.org/Initiatives/newstard.htm> (accessed Feb, 2010). 中山健夫訳(2008). In 中山健夫, 津谷喜一郎編著『臨床研究と疫学研究のための国際ルール集』、ライフサイエンス出版、179-193
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., et al. for the STARD Group. (2004). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Family Practice*, 2004 21(1), 4-10.

- Campbell, M. K., Elbourne, D. R. and Altman, D. G. for the CONSORT Group (2004). Education and debate, CONSORT statement: extension to cluster randomized trials. *BMJ*, 328, 702-708. 津富宏訳、津谷喜一郎監訳 (2008). In 中山健夫、津谷喜一郎編著『臨床研究と疫学研究のための国際ルール集』、ライフサイエンス出版、105-117
- Deeks, J. J. (2001). Systematic reviews in health care: systematic reviews of 38 evaluations of diagnostic and screening tests. *BMJ*; 323, 157-162.
- Des Jarlais, D. C., Lyles, C., Crepaz, N. and the TREND Group. (2004). Improving the reporting quality of Nonrandomized evaluations of behavioral and public health interventions: The TREND Statement. *Am J Public Health*, 94, 361-366. 中山健夫訳 (2008). In 中山健夫、津谷喜一郎編著『臨床研究と疫学研究のための国際ルール集』、ライフサイエンス出版、118-134
- Donner, A., Klar, N. (2004). Pitfalls of and controversies in cluster randomization trials. *Am J Public Health*, 94, 416-422.
- Egger, M., Jüni, P., Bartlett, C. for the Consort Group. (2001). Value of flow diagrams in reports of randomized controlled trials. *JAMA*, 285, 1996-1999.
- Greenhalgh, T., Hurwitz, B. (1998). *Narrative-Based Medicine: Dialogue and Discourse in Clinical Practice*. London: BMJ Books. (斎藤清二訳) (2001) : 『ナラティブ・ベイスト・メディスン:医療における物語と対話』、金剛出版
- Gross, C. P., Anderson, G. F., and Powe, N. R. (1999). The relation between funding by the national institutes of health and the burden of disease. *N Engl J Med*, 340 (24), 1881-1887.
- Ioannidis, J. P., Ntzani, E. E., Trikalinos T.A., et al. (2001). Replication validity of genetic association studies. *Nat Genet*, 29, 306-309.
- Ioannidis, J.P., Evans, J.W., Gøtzsche, P.C., et al. for the CONSORT Group (2004). Better report of harms in randomized trials: an extension of the CONSORT Statement. *Ann Intern Med*, 141, 781-788. 八重ゆかり訳、大橋靖雄監訳 (2008). In 中山健夫、津谷喜一郎編著『臨床研究と疫学研究のための国際ルール集』、ライフサイエンス出版、118-134
- Jenicek, M. (2001). Clinical case reporting in evidence-based medicine: 2nd edition. London: Arnold. (西信雄、川村孝訳) (2002) : 『EBM時代の症例報告』医学書院
- Ladabaum, U., Chopra, C. L., Huang, G., et al. (2001) . Aspirin as an adjunct to screening for prevention of sporadic colorectal cancer: a cost-effectiveness analysis. *Ann Intern Med*, 135, 769-781.
- Lavis, J., Davies, H., Oxman, A., et.al. (2005) . Towards systematic reviews that inform health care management and policy-making. *J Health Serv Res Policy*, 10, 35-48.
- Lu-Yao, G., Albertsen, P. C., Stanford, J. L., et al. (2002). Natural experiment examining impact of aggressive screening and treatment on prostate cancer mortality in two fixed cohorts from Seattle area and Connecticut. *BMJ*, 325, 740.
- Moher, D., Schulz, K. F., and Altman, D. for the CONSORT group. (2001a). The. CONSORT. Statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *Ann Intern Med*, 134, 657-662. <http://www.consort-statement.org/> (accessed Feb, 2010) 津谷喜一郎、小島千枝、中山健夫 (訳) (2002) 「CONSORT声明: ランダム化並行群間比較試験報告の質向上のための改訂版勧告」、『JAMA<日本語版>』、118-124 [http://homepage3.nifty.com/cont/CONSORT\\_Statement/consort2001J.pdf](http://homepage3.nifty.com/cont/CONSORT_Statement/consort2001J.pdf) (accessed Feb, 2010)
- Moher, D., Schulz, K. F., and Altman, D. for the CONSORT group. (2001b). The. CONSORT. Statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *JAMA*, 285, 1987-1991.
- Moher, D., Schulz, K. F., and Altman, D. for the CONSORT group. (2001c). The. CONSORT. Statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *JAMA*, 357, 1191-1194.
- Murray, D. M., Varnell, S. P., and Blitstein, J. L. (2004). Design and analysis of group randomized trials: a review of recent methodological developments. *Am J Public Health*, 94, 423-432.
- Shiffman, R. N., Shekelle, P., Overhage, J. M. et al. (2003) Standardized reporting of clinical practice guidelines: A proposal from the conference on guideline standardization. *Ann Intern Med*, 139, 493-498.
- The STARD Group (2003). The STARD initiative: Towards complete and accurate reporting of studies of diagnostic

- accuracy. <http://www.consort-statement.org/Initiatives/newstard.htm> (accessed Feb, 2010). 中山健夫訳 (2008). In 中山健夫、津谷喜一郎編著『臨床研究と疫学研究のための国際ルール集』、ライフサイエンス出版、172-178
- Stroup, D. F., Berlin, J. A., Morton, S. C., et al. (2000). Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*, 283, 2008-2012.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., et al. (1996). Evidence-based medicine: what it is and what it isn't. *BMJ*, 312, 71-372.
- Sackett, D. L., Straus, S., Richardson, S., et al. (2000). *Evidence-based medicine: how to practice and teach EBM. 2nd edition*, London: Churchill Livingstone.
- Thomas, D. B., Gao, D. L., Ray, R. M. et al. (2002). Randomized trial of breast self-examination in Shanghai: Final results, *JNCI J Natl Cancer Inst*, 94(19), 1445-1457.
- Varnell, S. P., Murray, D. M., Janega, J. B. et al. (2004). Design and analysis of group-randomized trials: A review of recent practices. *Am J Public Health*, 94, 393-399.
- (2010.2.27受理)

## Lessons Learned from the Worldwide Spread of Evidence-Based Medicine (EBM)

Tomoya Masaki

Graduate School of Medical Science  
Kitasato University  
masakit-ky@umin.ac.jp

Kiichiro Tsutani

Graduate School of Pharmaceutical Sciences,  
The University of Tokyo  
tsutani-ky@umin.ac.jp

### Abstract

The total cases of randomized controlled trials (RCT) from the 1950s in the area of medicine and healthcare have exceeded 600,000. Interest in RCT has been increasing since the mid-1990s with the global movement of evidence-based medicine (EBM). Several problems were identified in the EBM movement, leading to the development of new solutions. It was found that these problems were discovered at “use” and “cascading evidence” stages rather than the “generate” stage of the evidence usage whose categories were introduced in the previous reports. This report introduces CONSORT statements for the improvement in quality of RCT reports and their utilization in many aspects at the “stream down” stage. Feedback from narrative based information and the informative best case as one of the current movements of post-RCT at the “use” stage are also discussed. Furthermore, the latest trends toward the need for continuous evaluation of public investment, and the patient's and the citizen's roles in EBM development are discussed.

### Keywords

evidence-based medicine (EBM), quality improvement, CONSORT statements, best case, citizen participation

## 【研究論文】

## 教育におけるエビデンスに基づく政策—新たな展開と課題

岩崎 久美子

国立教育政策研究所

iwasaki@nier.go.jp

## 要 約

エビデンスの概念は、わが国の教育分野では、一般的に知られているとは言い難いが、昨今、「エビデンスに基づく政策」(evidence-based policy)という言葉が政策を論議する文献等で多く散見されるようになってきている。本論では、このような背景にあつて、第一にコクラン共同計画やキャンベル共同計画が基本とする、ランダム化比較試験(RCT)の系統的レビュー(メタ・アナリシス)で産出される厳密な定義でのエビデンスを中心に、産出、普及、活用の上で生じる課題を明らかにする。第二に広義の科学的根拠という意味で、経済協力開発機構(OECD)が、近年「エビデンスに基づく政策」という言葉を用いて政策提言を積極的に行つてきている背景を人的資本論に基づきながら論じる。そして、最後に教育におけるエビデンスに基づく政策についての方向性を示唆する。

## キーワード

エビデンスに基づく政策、人的資本論、教育投資、経済協力開発機構(OECD)、キャンベル共同計画

## 1. はじめに

エビデンスとは、実践や政策決定の際に用いられる科学的根拠を表す言葉である。

この言葉は、わが国の教育分野では、一般的に知られているとは言い難いが、昨今、「エビデンスに基づく教育」という言葉が政策を論議する文献等で多く散見されるようになってきている(Mosteller et.al.,2002, Pawson,2006, OECD 2007など)。定義が難しいこのエビデンスという言葉は、教育ではどのように扱われているのか。

エビデンスという言葉が最初に普及したのは医学分野である。医学分野では、ランダム化比較試験(RCT)に基づいた一次研究を集め、質評価の

ための系統的レビューを行い、統計学的手法により統合(メタ・アナリシス)することで産出されるものを、最良のエビデンスと評価する。医学でのこのようなエビデンスをめぐる動きを顧みれば、最新、かつ最良のエビデンスを継続的に提供しようとする先駆的試み、コクラン共同計画プロジェクトが始まったのは1992年である。このコクラン共同計画は、「エビデンスに基づく医療」(evidence-based medicine: EBM)の世界的な動きと連動し、医学、薬学、公衆衛生領域などの臨床の場面のみならず、医療行政の領域にも影響を与えた(津谷 2005、pp.23-38)。医学では、現在、研究成果から導かれたエビデンスが、科学的な判断材料として、医師の治療方針や政策判断の根拠

として活用されている。このコクラン共同計画に範をとり、社会科学分野でも、キャンベル共同計画（津富 2005、pp.1-22）という類似の組織が2000年に発足し、その活動が推進されてきた。キャンベル共同計画には、刑事司法、社会福祉、教育の3つの部会が設けられ、コクラン共同計画と同様の手続きにより、エビデンスの産出・普及が行われている。しかし、医学と異なり、社会科学分野では、コクラン共同計画が基本とするランダム化比較試験（RCT）に基づく実験計画を立てることは難しく、厳密なエビデンスの量に限界があることが指摘されてきた<sup>1)</sup>。とりわけ、教育分野では、子どもを対象に実験を行なうことは、倫理的問題を提起することが多く、エビデンスとされる良質の研究成果を産出する条件設定が難しい。実際、キャンベル共同計画の3つの部会の中でも、手続き（プロトコル）を経てレビュー後にエビデンスとしてウェブ上に公開されるのは、教育の部会が一番少ないのである<sup>2)</sup>。

そのため、コクラン共同計画やキャンベル共同計画が基本とする系統的レビュー（メタ・アナリシス）を経た厳密なエビデンスというよりも、目的に応じエビデンスの厳密度を柔軟に考えて活用するという現実的な考え方が教育分野では取り入れやすく、エビデンスという言葉は、広義の科学的根拠という意味で用いられることが多い。最近では、経済協力開発機構（OECD）が、政策を裏付ける根拠といった観点から教育の効果を測定

し、エビデンスという言葉を用いて様々な政策提言を行うほか、2004年から2006年にかけて、教育におけるエビデンスに基づく政策研究を扱った4回のセミナーを開催し、その成果を*Evidence in Education-Linking Research and Policy*という報告書にとりまとめている（OECD 2007）。

本論では、以上の観点から、第一にコクラン共同計画やキャンベル共同計画が基本とする系統的レビュー（メタ・アナリシス）を経たエビデンスを中心に、産出、普及、活用の上で生じる課題を考察する。第二に広義の科学的根拠という意味で、経済協力開発機構（OECD）が、エビデンスに基づく政策という言葉を用いて政策提言を積極的に行っている背景を人的資本論に基づきながら論じる。そして、最後に教育におけるエビデンスに基づく政策の方向性を示唆することにした。

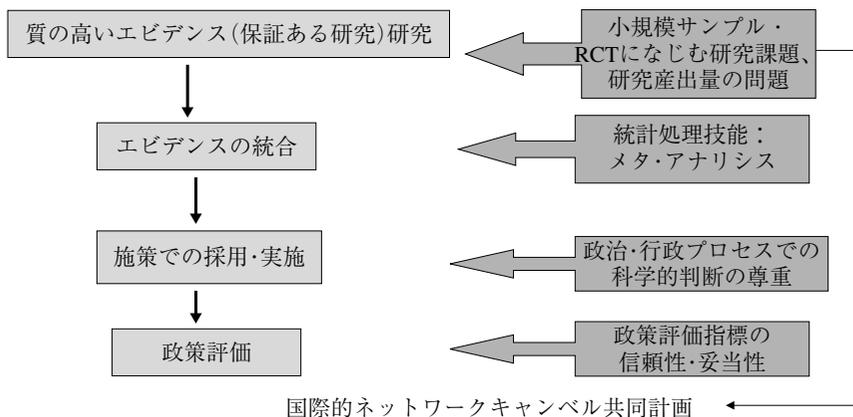
## 2. エビデンスの産出・普及・活用の課題

### (1) エビデンスの産出

エビデンスには、産出、普及、活用するという3つの段階がある。このうち、産出については、キャンベル共同計画が最も厳密な手続きを経て、エビデンスを産出している。

キャンベル共同計画では、玉石混交の多様な研究成果の中から、ランダム化比較試験に限定し、質のよい研究を求めてレビューという振り分け作

図1 キャンベル共同計画に見るエビデンスの産出と課題



（出所）筆者（岩崎）作成

業を行い、メタ・アナリシスを経てエッセンスとして抽出されたエビデンスのみを、キャンベル共同計画のウェブに公開している。このような過程を経てレビューされている教育関連のテーマの例としては、次のようなものが挙がる。

- ・放課後のプログラムの生徒のアウトカムに対するインパクト
- ・親の関与と小学生の学業成績
- ・子どもの攻撃行動の減少を目的とした学校現場の社会情報スキル向上のための介入
- ・就学前および幼稚園におけるバイリンガル教育が、第二言語学習者の英語習得に与える効果
- ・アメリカおよびカナダにおける高校中途退学率を下げるための諸方法の効果

コクラン共同計画やキャンベル共同計画が基本とする、介入に基づくランダム化比較試験を集め、システマティック・レビューを経たエビデンスは、もっとも質の高いエビデンスである。

しかし、研究によっては、このような厳密な手法でなくても、エビデンスとしての意義があるものも多い。そこで、エビデンスの質に関し、階層構造を持った格付けが各種なされている。たとえば、米国医療政策研究局（Agency for Health Care Policy and Research: AHCPR）によれば、最もレベルの高いレベルⅠaのエビデンスは、ランダム化比較試験をシステマティック・レビュー（メタ・アナリシス）したもの、レベルⅠbは、ランダム化比較試験（メタ・アナリシスなし）によるもの、レベルⅡaは、よくデザインされた非ランダム化比較試験によるもの、レベルⅡbは、他のタイプがよくデザインされた準実験的研究によるもの、レベルⅢは、よくデザインされた非実験的記述的研究による比較試験、相関研究、ケースコントロール研究など、レベルⅣは、専門家委員会のレポートや意見や権威者の臨床試験、と規定されている。

教育分野でのエビデンス産出の最も大きな課題は、レベルⅠに相当するランダム化比較試験による質の高い一次研究が乏しく、産出し難いということにある（Fitz-Gibbon 2000、p.86）。

ランダム化比較試験は、対象者をランダムに、実験群と対照群の2群に分け、外部要因による影

響を除去し、実験群に特定のプログラムを実施した後、その差について統計学的検定を行う。このようなランダム化比較試験を行う場合について、次のような8つの制約が挙げられている（龍・佐々木 2004、pp.52-58）。

- ①社会要因による影響の除去が難しい。
- ②実施コストが高い。時間がかかる。
- ③倫理的制約を敏感に考慮する必要がある。
- ④大きなサンプル数を必要とする。
- ⑤実験の参加の同意をとるが、同意しないで離脱する人もいる。
- ⑥プログラムの最初から評価実施者が参加する必要がある。
- ⑦実験と本格適用の違いがありえる。
- ⑧アサインメントによるセレクション・バイアス発生の可能性。

これらの制約について、実際に筆者が実施した小学校での実験研究に基づき、教育でのランダム化比較試験によるエビデンスの産出の難しさを例示してみたい（生涯にわたる生活リズム形成研究会 2007、pp.5-6）。

この研究は、「早寝・早起き・朝ごはん」の効果検証を行なうことを意図し、小学校6年生2クラスを対象に、ランダム化比較試験による実験プログラムを計画したものである。プログラムは、早朝登校、検温、ラジオ体操、一律メニューの朝食提供、歯磨きからなり、歯の健康状態改善、体温上昇、集中力増加などの効果を想定した（実験計画の詳細についてはp.21【資料】を参照のこと）。

しかし、一般の日本の学校で、ランダムに2群を設定することが現実的には難しかったため、進級時のクラス分けがほぼ均等になされているという前提を準用し、1クラスを実験群、1クラスを対照群とした。そのため、本実験は、米国医療政策研究局の格付けで言えばレベルⅡaのランダム化されていない比較試験に留まることに留意してほしい。

- ①社会要因による影響の除去が難しい。

教育に関わる研究では、特にこの社会要因による影響の除去が難しい。生活習慣については、学校場面でのプログラム介入以前に、家庭要因が非常に大きな影響を持つ。早朝登校により、早起きの時間を統制しても、通塾などの事情もあり早寝

の強制は難しい。また、走る、歩くなどの登校の様態が異なるため、学校までの距離や通学方法が体温に影響を与えるなどの指摘がなされた。

②実施コストが高い。時間がかかる。

3週間にわたる毎日の朝ごはんの材料費は、文部科学省の研究委託費（「早寝・早起き・朝ごはん」プロジェクト）によった。朝食の調理・配膳は、PTAのボランティアが行った。ただし、検温、データ取得などは養護教諭や栄養士の追加業務となった。通常の授業以外に実験を行うことで、学校からは最大3週間という期間の提示がなされたが、効果が生じるには十分な期間とはいえなかった。

③倫理的制約を敏感に考慮する必要がある。

子どもたちのプログラムへの参加の許諾をとるために、学校側でPTAへの説明会を実施した。保護者からの質問等には、学校側が対応したが、当初すべての保護者から賛同が得られたわけではない。PTAを通じ、説明を繰り返し行い、併せて、途中離脱の自由を明記した同意書を保護者から取得することで、最終的に実施に至った。

④大きなサンプル数を必要とする。

子どもを対象にした統制下での実験を設定する場合、大きなサンプル数の確実な確保は難しい。厳密なランダム化比較試験であれば、なおさら条件が難しくなる。そのため、限られたランダム化比較試験による精緻な研究成果を広く集めるためには、キャンベル共同計画のような国際的な取組みが必要となる。

⑤実験の参加の同意をとるが、同意しないで離脱する人もいる。

2クラスのうち、保護者全員の同意書が得られたクラスを実験群にせざるを得ず、実施を優先する上で、ランダムな実験群設定にはならなかった。しかし、1クラス全体に対する介入プログラムだったため、通常の学校生活に準じており、実験群の途中離脱者はいなかった。

⑥プログラムの最初から評価実施者が参加する必要がある。

3週間にわたる学校での実施であったため、教育現場を混乱させないよう、学校にプログラムやデータ取得の一部を委託した。このことが、データ取得の精密性に影響した。

⑦実験と本格適用の違いがありえる。

実験を本格適用しようとした場合、学校文化によって、実施状況が著しく異なる。実験した学校は、校長が学校改善のために、新しい試みを行なう意欲が強く、校長主導で実験が導入された。しかし、本格的適用には、条件が整ったところばかりではないため、実験と同様の効果が検出されるとは限らない。校長のやる気といった態度変数が、科学的介入以上に、子どもたちに影響を与えることも考えられる。

⑧アサインメントによるセレクション・バイアス発生の可能性。

小学校の2クラスは、学級編成時にできるだけ等質になるよう編成される。そのため、便宜上、2クラスをそれぞれ実験群と対照群とした。しかし、小学校では、学級編成後、担任の指導力や特性などで異なる学級文化が醸成される。教員たちからは、実験群よりも対照群のクラスの方が落ち着いた雰囲気があるとの印象を持たれており、介入以前から実験群の方が成績の良いことが推定されていた。また、学力指標の集中力テスト（百マス計算）を実施するにあたっては、対照群の担任と子どもたちが、実験群にならなかったことに対し逆の意欲を持ったとの解釈が教員からなされた。実際、体温や歯肉については、実験群の変化が認められたが、学力指標については、介入しなかった対照群が実験群よりも成績が良いという結果になった<sup>3</sup>。

以上のように、ランダム化比較試験を教育の現場で実施する場合、統制された実験環境の設定は労多く、また、得られた結果に介入の影響以外の要因が介在する可能性が高い。学校などの教育現場で実験を実施するためには、研究上の野心と実施可能性との間のバランスをとることも時に必要であり（Gueron 2002, pp.25-26）、場合によっては、定性的な研究を含む多角的なアプローチが志向される。このことが、教育の分野で、エビデンスという言葉や、ランダム化比較試験に限定せずに、政策を裏付けるデータ一般といった意味合いで広く解釈することが多い所以とも言える。

## (2) エビデンスの普及

エビデンスの普及に関しては、キャンベル共同

## 【資料】 研究計画

## 1. 目的：

生活リズムの安定が集中力を増し、学力向上や健康増進に寄与することを比較により実証する

## 2. 対象：東京都内小学校6年生2クラス

## 3. 群の構成：

①実験群（1クラス28名）

②対照群（1クラス29名）

## 4. 取組期間：3週間

## 【実験群のプログラム】

7:20 登校、検温

7:30 ラジオ体操

7:45 朝食

8:10 歯みがき

## 5. デザイン：比較研究

6. 実施期間：2006（平成18）年6月26日から  
2006（平成18）年7月14日

## 7. 評価項目：

①体温Ⅰ（赤外線サーモカメラ）

②体温Ⅱ（腋下体温計）

③体重

④歯科検診による歯肉に変化（写真撮影）

⑤集中力（百マス計算／「内田クレペリン検査」）

⑥忘れ物チェック（上履き、教科書）

⑦心理尺度（文部科学省「児童生徒の心の健康と生活主観に関する調査協力者会議」平成15年3月作成・一部CBCL（Child Behavior Check List）参照）

計画のような国際的な取り組みと同時に、エビデンスをとりまとめ、その情報を提供する機関がある。

たとえば、代表的な機関としては、キャンベル共同計画の枠組みにそって厳密なエビデンスを政策に活用しようとする米国のWWC（What Works Clearinghouse）と現実的に政策に活用可能なデータをエビデンスと考える英国のEPPIセンター（Evidence for Policy and Practice Information and Co-ordinating Centre）がある。

WWCは、2002年「教育科学改革法」（the Education Sciences Reform Act of 2002）のもとで新たに設立された教育科学研究所（Institute of Education Sciences）のイニシアチブにより、教育

における効果についての科学的エビデンスの信頼おける情報源を教育者、政策立案者、研究者、及び一般市民に提供する目的で設立された。具体的には、因果関係が認められたエビデンスをレビュー、要約し、ウェブサイトで情報提供している。WWCは、教育科学研究所が重視する、ランダム化比較試験、及び良質の準実験による効果的な介入による質の高いエビデンスを重視することで、質を保証し、知見の蓄積を行う。また、科学的基準を維持するため、透明性の高い意思決定ルールとプロトコル（科学的実験等を実施する際の手順や手続き）を整備している。WWCの成立の目的と運営方針は、教育科学研究所がWWCを設立する際に、キャンベル共同計画と共同契約を結んだことから明らかとなり、キャンベル共同計画のエビデンスの厳密な基準の影響を受けている。2004年から2006年の間に行われたレビューの例としては、早期教育、初期の読解力、小学校の算数カリキュラム、中退防止策、国語（英語）教材、道徳教育、ミドルスクールの数学のレビューなどが挙げられている。教育で「何が有効か」についての科学的根拠として、WWCはウェブ上で質の高いエビデンスを公開している（Boruch/Herman 2007, pp.53-61、田辺 2006, pp.31-41）。

一方、英国のEPPIセンターは、ロンドン大学の社会科学調査研究所（the Social Science Research Unit）の一部に位置づけられ、政府を含む多くの資金援助団体からの資金供与により活動している。EPPIセンターでの政策や実践に役立つエビデンスを系統的に集める作業は、1993年に始まっている。EPPIセンターの目的は、公益性の高い政策、実践、研究上の重要な課題について系統的レビューを行なうことである。その目的において、①利用者の課題に基づくレビューの幅広い概念枠組みや、あらゆるタイプの研究エビデンスを含む系統的レビューのための方法と手段の開発、②異なる学問分野におけるエビデンス基盤の強化、そして、③エビデンスに基づく政策や実践能力を開発するための支援と研修の提供を行っている。特記すべきことは、EPPIセンターでは、ランダム化比較試験のような定量的なデータのみならず、定性的なデータの統合も試み、様々な研究の利点をできるだけ高い比率で利用しようとする

ることである。厳密なエビデンスの質保証を目的にするWWCに比べ、EPPIセンターでは、より現実的に幅広いエビデンスを集積している。EPPIセンターのレビューの手続きは、図2のとおりである（Gough 2007、pp.63-70）。

エビデンスは産出されたのちに、実践や政策に活用されることが期待される。そのためには、WWCやEPPIセンターのような、エビデンスを普及するための仲介機関というものが重要となる。我が国においては、教育政策に資する研究エビデンスを集約し提供する専門的仲介機関が存在していない。そのため、必要な研究成果を政策に提供することが難しく、エビデンスに基づく政策が醸成しない一因となっている。エビデンスの普及・活用を目指すとするれば、今後、我が国でも制度的な整備が必要となるであろう。

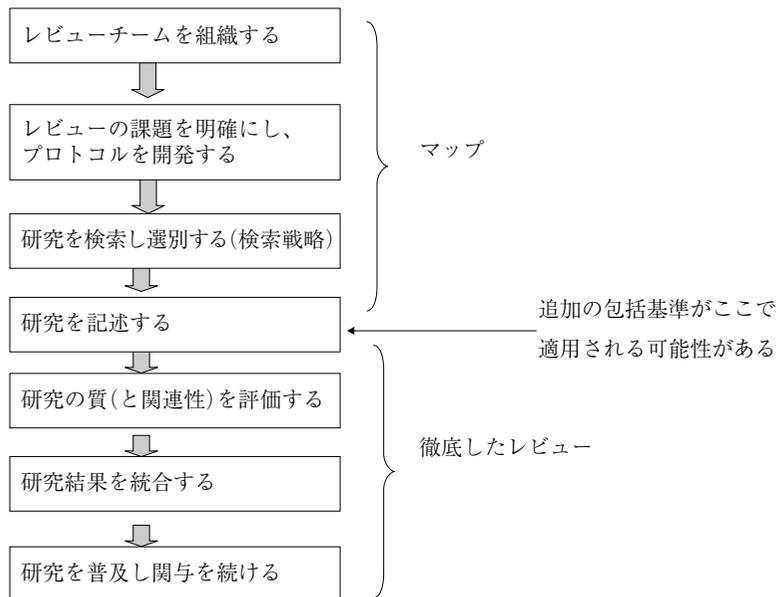
### (3) エビデンスの活用

エビデンスを活用するのは、研究者、実践家や各国の政策立案者である。このうち、エビデンスに基づく政策が喧伝されている中で、政策立案者

が政治、行政プロセスでの科学的判断を尊重するかどうかは、文化的風土が左右する。エビデンスに基づく政策は、アングロサクソン系の国々、例えば、米国、英国、カナダといった国々で唱道されている。

文化的要因と同時に、エビデンスに基づく政策が教育の分野で確立しづらい理由としては、教育学という学問分野の科学的基盤の脆弱さも否めない。キャンベル共同計画の2009年の会議での「どのようにエビデンスに基づく教育学を開発するか」という講演では、医学での大改革のカギとして、①「科学に基づく専門的訓練」、②「質の高いより多くの研究」、③「研究と実践との間の密接な連携」、④「系統的な記録管理」、⑤「研究にアクセスできるシステム」、⑥「技術進歩」の6つの観点が挙げられている。これを対照にした教育学の現状は、①「科学に基づく専門的訓練」は実施されており、②「より多くの質の高い研究」と④「系統的な記録管理」は今後改善が見込め、⑥「技術進歩」は進行中であるが、③「研究と実践との間の密接な連携」⑤「研究にアクセスできる

図2 EPPIセンターのマッピングと統合



(出所)「エビデンスによる政策と実践のための情報連携センター（英国）」（豊浩子訳）OECD『エビデンスと教育』明石書店 p.110.

システム」は、実現されていないとの評価がされている (Noonan 2009)。

### 3. OECDによるエビデンスへのアプローチ

#### (1) 教育投資への裏付けとしてのエビデンス

このようなランダム化比較試験の系統的レビューを経たレベル I aのエビデンスを基本とするキャンベル共同計画とは別の文脈で、近年、OECDの教育政策分析や学習の社会的成果のプロジェクトの報告書や政策文書で、エビデンスという言葉を目にすることが多くなった。

OECDがこのようなエビデンスの産出に関心を持つようになってきたのは、教育や訓練への教育投資を裏付ける根拠を必要とする各国の共通課題があるからである。

それは、第一に、グローバル化の進展の中で、先進諸国が、人件費が安い国々と比べて、競争的で優位な経済的地位を維持するためには、高い生産性、付加価値、技術革新をもたらす知識や熟練したスキルを持つ労働力を育成し、高熟練、高賃金の雇用を創出することが喫緊の政策課題となっていること、第二に、社会の高度化や技術進歩により、知識基盤社会が目指される中で、長期にわたる公教育・訓練を必要とする専門的な知的労働者への需要が拡大し、半熟練や専門的訓練を要しない仕事の割合が減少してきていること、第三に、知識やスキルが性別、年齢層、階層などで異なる分布が認められるのであれば、その現状により、それぞれに応じた教育や研修への投資を行うことが効率的であるとの費用対効果の視点が新たに生じてきていること、などが挙げられる。つまり、景気後退、グローバル経済競争の激化、失業率の高まりなどの状況において、他の国に対して労働者ひとり一人の生産性や質をあげることで競争優位性を確保しなければ、生活水準の低下や失業の増加は免れない、との認識が先進諸国の多くに生じてきている。知識、学習、情報、専門能力が重要な資源と考えられるとすれば、政策上、教育や訓練への投資がこれまで以上に望まれ、教育の重要性が増しているとの共通認識が持たれるようになってきたのである。

これらの共通課題に対して、教育投資を公財政に求める場合、先進諸国の財政金融関係省庁（日本であれば財務省）は、公財政からの教育費支出に対し、妥当で説得力あるエビデンスの提出を求める。その際のエビデンスとは、教育・訓練の成果と有効性が貨幣価値を基準に算定されるということにある (OECD 2009, pp.16-18)。このことが、エビデンスという言葉がOECDの文書に現れる所以となっている。つまり、キャンベル共同計画が研究者側からの政策へのデータ提供とすれば、OECDのエビデンスに基づく政策への関心は、政策立案者の求めに応じてエビデンスの産出を推進しようとする国際的取組みに向かっているといえよう。

#### (2) 人的資本論の新たな展開

このように、OECDの文書によれば、知識基盤社会に移行していく中で、今後知識の持つ価値が大きくなっていくこと、また、技術革新を支え、国際的な競争力を維持し高めるためには、教育・訓練に対する重要性は高まっていくことが予見されている。

同時に、1960年代にシュルツ (Schultz, T.W.) やベッカー (Becker, G.S.) などの経済学者により展開された、教育や訓練への投資が熟練労働力の社会的ストックを増やし、生産性を向上させるという「人的資本」論が、OECDの国際プロジェクトの基礎理論として、半世紀を経て再び重要な理論として注目されるようになってきている。

人的資本論は、シュルツやベッカーが論じた1960年代当初、一般的に受け入れられたものではなかった。シュルツの「人的資本」の理論を下敷きに、教育と訓練における収益率の分析を行ったベッカーは、1964年に刊行された『人的資本』の第1版で、人的資本論という考え方が、多くの人々に嫌悪感抱かせ、人的資本への教育その他の投資の経済的効果の否定、収入と人的資本投資との相関が実は能力によるのではないかとの疑問、分析対象にバイアスがかかっているなどの批判があったことを記している (Becker 1993, pp.11-13)。しかし、1992年の第3版の序文には、第1版から30年以上たつて、このような論議を醸し出してきた「人的資本論」が経済学のみならず他の学問領域

や一般にも受け入れられてきていること、そして、特に、ブッシュ (Bush, G.H.W.) や、それに続くクリントン (Clinton, B) といったアメリカの大統領がその選挙演説で、労働者の教育とスキルの向上の重要性をとりあげ、「人的資本への投資」という言葉で、大学教育やOJTへのより多くの投資を公約したことは信じられないと記している (Becker 1993, xix-xx)。

このことに表されるように、先進諸国の多くで、経済発展の原動力として「人的資本」という言葉への関心が増大し、政策立案者自ら教育や訓練の現状把握と施策への予算化の重要性を認識するようになってきている<sup>4</sup>。その背景には、若年者の失業率の高まりなどを背景に、雇用の安定が喫緊の課題となってきたこと、また、「シュンペーター的ルネッサンス」(Schumpeterian renaissance) (OECD 2004, pp.7-11) という言葉で表現されるように、知識経済の中で、イノベーションが、経済発展のための国際競争力の原動力や社会変容の鍵であることが喧伝されるような新たな状況が生じてきていることが挙げられる。

「人的資本論」の定義をめぐることは、教育や訓練への投資が、労働者の生産性上昇により経済成長の基盤 (社会的便益) をもたらし、労働者個人の生涯所得の増大や安定雇用の保証 (私的便益) を促す (ウッドホール 2005, pp.260-261) といった古典的な定義を超え、個々の能力の適切配置・開発のための教育や訓練、健康管理などに使われた費用の社会的・私的収益を併せて考慮する考え方も出てきている (OECD 2006, pp.192-195)。たとえば、OECDの報告書*The Well-Being of Nations*では、人的資本を「個人的、社会的および経済的な福利の創出を促す、個人が体現する知識、スキル、コンピテンシー、および属性」と広く定義している (OECD 2001, p.18)。

若年失業者問題に悩むヨーロッパ各国などで、職業教育、訓練、労働経験を与えるプログラムを多く策定しているのは、人的資本への投資という考え方が根底にあるからである。より多くの者に雇用を確保するためには、失業手当よりも、教育・訓練の方が失業状態から雇用へと転換させるために効果的ということであろう。このような教育投資を政策的に裏付けるエビデンスを、政策立

案者は求めているのである。

#### 4. 教育の成果を表すエビデンスとは

##### (1) エビデンスとしての教育インディケータ

人的資本論の適用は、時間、金銭的投資と、その結果としての経済的またはその他の収益が生じる直接的な線形モデルを示唆する。そのため、分析担当者は投資利益の推計に既存のツールを使うことができ、政策立案者は人的資本形成に関する支出を正当化できる (OECD教育研究革新センター 2008, pp.56-62)。教育投資の正当性を、国際的な枠組みでの比較調査により裏付け、加盟国が政策に用いることができるようにするため、教育分野でのエビデンスとして、教育投資とその成果の因果関係を明らかにする試みが、OECDによりなされている。

OECDでは、加盟国の教育に関する国際比較可能な教育指標を開発するために、INES事業 (INES: Indicators of Education Systems) を1988年に開始し、各国の教育のデータを収集し分析してきた<sup>5</sup>。このINES事業の一環として、15歳児を対象に「生徒の学習到達度調査」(PISA: Programme for International Student Assessment) による国際的な調査が、2000年以降3年ごとのサイクルで、読解力、数学的リテラシー、科学的リテラシーの3分野を中心に行なわれている。OECDでは、現在、これに加えて、16歳から64歳を対象にした「国際成人力調査」(PIAAC: Programme for the International Assessment of Adult Competencies) を企画している。

この「国際成人力調査」(PIAAC) は、直接評価 (DA: Direct Assessment) で測定される読解力 (literacy)、数的思考力 (numeracy)、IT環境下での問題解決能力 (problem solving) などの能力にかかわる検査結果と、それを解釈する背景質問票 (BQ: Background Questionnaire) から構成される。背景質問票は、学歴、学習歴、職業、職業で求められるスキルなどを確認し、読解力、数的思考力、問題解決能力などの能力との因果関係と、その形成要因を探る。「国際成人力調査」(PIAAC) の目的とは、「現状把握と施策への予算化」のための

表1 教育・訓練の金銭的・非金銭的アウトカムの可能性

	(A) 個人のアウトカム	(B) 公共のアウトカム
(1) 金銭的アウトカム	収益、所得、富、生産性	税金、社会移転コスト、ヘルスケア・コスト
(2) 非金銭的アウトカム	健康状態、生活への満足度	社会的統合、信頼、正常に機能する民主主義、政治的安定

(出所) OECD/CERI 2008 p.67

政策への関心に応えることであり、その内容としては、第一に、国内でどのように職業で求められるスキルが分布しているか、第二に、どのようにスキルの水準の向上が可能か、第三に、スキルは経済的、社会的になぜ重要であるか、第四に、マクロレベルでスキルはどのような影響を与えるか、また、長期にわたってどのようにスキルは開発されるのか、の4つが挙げられている（PIAAC 2009）。

## (2) 教育投資のアウトカム

政策課題への対応、つまり、政策に伴う教育投資のためのエビデンスの提示のためには、教育投資に伴う教育・訓練教育の成果（アウトカム）を想定することが求められる。

教育や訓練といった教育投資に対して、アウトプットされる社会的成果として想定されるものをまとめたのが表1である。この表によれば、アウトカムは、金銭的アウトカムと非金銭的アウトカムの2つに大別される（OECD教育研究革新センター 2008、p.67）。

人的資本論では、表1に表される金銭的アウトカム、特に個人の所得や生産性を重視するが、表1では、そのような経済的アウトカムではない健康状態や生活への満足度などの社会的アウトカムも併せて想定されている。しかし、社会的アウトカムの多くは、複雑な要因による帰結の場合が多く、単純な因果関係で測定することが難しい。そのため、所得や生産性といった関係から、教育と訓練の効果を測定する古典的な人的資本論の考え方に基づくモデルが、教育投資に対するエビデンスとして有効性が高いものとして評価される傾向があるのである。

いずれにせよ、OECDの教育インディケータ事

業は、政策課題への回答の提示とともに、予算化や行政評価のためのデータの裏付けとしてエビデンスを提供する事業といった捉え方もできるのである。

## 5. おわりに：教育におけるエビデンス活用の課題

「エビデンスに基づく政策」は、当初、教育分野では、キャンベル共同計画の枠組みで、ランダム化比較試験の系統的レビューによる研究成果の産出という形をとった。それは、研究者からの政策へのアプローチであった。それに対し、OECDでのエビデンスに基づく政策関心は、政策立案者の需要によるアプローチである。このことは、研究成果をレビューする目的が「知識支援」か「決定支援」かの違いにも対応する。政策立案者からの需要は、費用対効果と財源配分の根拠といった「決定支援」の側面であり、このような「決定支援」には、厳密なエビデンスの解釈を超えて、質的研究を基盤にしたランダム化比較試験によらない研究、ステークホルダーや専門家パネルなどの見解といった、政策立案者の問いに応じてエビデンスを修正する必要が生じる場合もある（ポーパら2009、pp.14-17）。

現状は、「エビデンスに基づく政策」(evidence-based policy) の名の下に、何がエビデンスとして重要かの定義やどのように活用されるに関するOECD内部の合意はとれていない。キャンベル共同計画のような厳密なエビデンスの解釈と、政策ツールとして公共政策を決定づけるエビデンスの最適形態や効果的活用を目指す解釈と、OECDのプロジェクト内でも見解が割れている（OECD

2007, pp.22-24)。

また、エビデンスとして測定されるものが、本当に教育の効果として重要なものを測定できるのかという議論も、いまだ根強い。エビデンスとして最も提示しにくいのは、直接観察されず、貨幣価値尺度でも測定できない社会的利益である (Psacharopoulos 2006, pp.120-121) といった観点も、人的資本論に基づき、教育の効果が収入や生産性などで提示されようとする流れの中で、あらためて主張されるようになってきている。

しかし、様々な課題はあるものの、教育分野でのエビデンスに基づく政策への関心は高まってきている。それは、第一に、透明性ある政府、情報公開といった世の中の動きに対し、政策決定者が政策の妥当性を根拠あるデータで国民に説明する道義的義務とそれに対する社会認識が高くなってきていること、第二に、昨今の財政状況から、実証性のない社会・経済的施策が結果として経費の無駄に帰結するというアカウンタビリティに応じた政策効率性の問題が生じてきていること、そして、第三に、省庁や地方公共団体の組織文化の中での長年の経験や蓄積に基づき行われてきた従来の政策立案に対し、知識伝達やコミュニケーションスタイルの変化、政策決定システムの透明性志向により、政策立案の根拠としてデータに基づく意思決定・判断の明晰さが望まれてきているからである。この点については、ODA分野などの、介入、効果検証の評価に基づく政策評価が早くから実施されてきた公共分野でも、昨今では、マサチューセッツ工科大学におかれた「貧困アクションラボ」などを中心に、エビデンスに基づく評価を活用することで、これまで主流であった簡便な評価から、科学的手法による評価へとバランス化が図られてきている (佐々木 2006, pp.43-54)。

キャンベル共同計画のような科学的手法に基づく研究成果を産出するためには、教育分野での研究能力、特に定量的研究に対するレベルの低さが指摘され、また、その活用のためには、教育研究・政策とイノベーション間の連携の弱さが指摘されている (OECD 2007, pp.16-18)。しかし、厳格な科学的手法によらないながらも、OECDでは、各国政府を基盤とした国際的プロジェクトにより、より現実的な政策課題に即座に応えるため

の、政策への「決定支援」のためのデータを提示しようとする。このようなOECDの動きの基底には、グローバル社会での国際競争力の強化と知識基盤社会でのイノベーションの必要性を背景に、人的資本論に基づく教育・訓練への教育投資を求める裏付けとしてのエビデンスの産出をOECDのINES事業が行うという戦略的図式が存在する。

OECDに代表される国際的な動きの影響を受けて、我が国でも、今後、エビデンスという言葉は、政策用語として定着し、教育政策研究の新たな方向を提示する概念となっていくことも予想される。その際、研究者の立場から言えば、政策がデータに裏付けられていく文化が徐々に醸成していくことにより、エビデンスの産出・普及・活用を目指した政策科学としての教育研究の充実が図られることが期待されるであろう。

## 注記

- 1 コクラン共同計画のライブラリーを用いたランダム化比較試験 (RCT) の引用は、医学分野では50万件数を超えるが、キャンベル共同計画の下位部会で最も多い刑事司法分野でさえ85件である。また、キャンベル共同計画のレビューのうち54%がランダム化比較試験 (RCT) 以外のものである。(Noonan, E. (2009). How can we develop evidence-based pedagogy? (The Campbell Collaboration, Decade of the Mind Conference, 2009/9/11 会議資料)
- 2 2005年3月の時点では、レビューされたものは刑事司法が42件、社会福祉が34件、教育領域は9件である (岩崎久美子 (2005) 「キャンベル共同計画教育領域の動向」『キャンベル共同計画の日本への適用－実証研究の系統的レビュー・システムの開発－』(研究代表者 山田兼尚 平成15-16年度) 科学研究費補助金報告書:150-151)。
- 3 身体的変化、心理変化、行動変容のうち、介入後の変化が確実にあったのは、身体的変化 (体温は統計的有意差あり。歯肉の変化は歯科医の判定) であった。学力指標である百マス計算では、実験群も対照群ともに終了する時間の平均値が早くなり、学習効果と推定された。
- 4 実際、OECDで現在企画している「国際成人力調査」(PIAAC) のワーキングペーパーの冒頭には、「こ

こ数十年、経済発展の原動力として人的資本への関心が増大し、政策立案者自ら教育や訓練の現状把握と施策への予算化の重要性を認識しつつある」と書かれている (PIAAC BQ JRA V5.0-Conceptual Framework, PIAAC-NPM (2009\_02\_01) BQ V.5.0 Conceptual Framework, doc.(working paper))。

- 5 ほぼ毎年OECD, *Education at a glance*の名称で出版されている。日本版は、『図表でみる教育—OECDインディケータ』(明石書店)として刊行されている。

## 参考文献

- 市川昭午編 (1987)『教育の効果』、東信堂
- 岩崎久美子 (2005)「実証研究に基づく政策と実践を求める社会的背景」『キャンベル共同計画教育領域の動向』、前掲報告書、207-215
- 岩崎久美子 (2010)「教育分野でのエビデンスの産出」、『薬理と治療』、38 (1) : 11-25
- 国立教育政策研究所編 (2002)『生きるための知識と技能—OECD生徒の学習到達度調査』、ぎょうせい
- 佐々木亮 (2006)「ODA分野における『エビデンスに基づく評価』の試み：「貧困アクションラボ」の動向」、『日本評価研究』6 (1) : 43-54
- シャーマン,L.W.ほか編 (津富宏・小林寿一監訳) (2008)『エビデンスに基づく犯罪予防』(財)社会安全研究財団
- シュルツ,T.W.(清水義弘訳) (1964)『教育の経済価値』、日本経済新聞社
- 生涯にわたる生活リズム形成研究会 (2007)「子どもの生活リズム向上のための調査研究」、文部科学省委託事業報告書
- 田辺智子 (2006)「エビデンスに基づく教育—アメリカの教育改革とWhat Works Clearinghouseの動向」、『日本評価研究』、6 (1) : 31-41
- 丹後俊郎 (2002)『メタ・アナリシス入門』、朝倉書店
- 津谷喜一郎 (1996)「コクラン共同計画とシステムマテイク・レビュー」、『薬理と治療』、25 (1) : 11-24
- 津谷喜一郎 (2000)「コクラン共同計画とシステムマテイク・レビュー—EBMにおける位置付け—」、『公衆衛生研究』、49 (4) : 313-319
- 津谷喜一郎・津富宏・正木朋也 (2003)『人文社会科学でも実験は可能だ—キャンベル共同計画と政策評価の現状と将来—』、東京大学大学院薬学系研究科 医薬経済学講座
- 津谷喜一郎 (2005)「日本におけるコクラン共同計画のあゆみ—日本のキャンベル共同計画の発展へのアドバイス」、『キャンベル共同計画の日本への適用—実証研究の系統的レビュー・システムの開発—』、平成15-16年度科学研究費補助金報告書 : 23-38
- 津富宏 (2005)「キャンベル共同計画の概要」、前掲報告書 : 1-22
- ハルゼー,A.H.他編 (清水義弘監訳) (1963)『経済発展と教育』、東京大学出版会
- ハルゼー,A.H./ローダー,H./ブラウン,P./ウェルズ,A.S.(住田正樹・秋永雄一・吉本圭一編訳) (2005)『教育社会学—第三のソリューション—』、九州大学出版会
- ブードン,R.(杉本一郎・山本剛郎・草壁八郎訳) (1983)『機会の不平等』、新曜社
- ホープ,C.,メイズ,N.ポベイ,J.(伊藤景一・北素子監訳) (2009)『質的研究と量的研究のエビデンスの統合—ヘルスケアにおける研究・実践・政策への応用—』、医学書院
- ライチェン,D.S.,サルガニク,R.H.編 (立田慶裕監訳、今西幸蔵、岩崎久美子、猿田祐嗣、名取一好、野村和、平沢安政訳) (2006)『キー・コンピテンシー—国際標準の学力をめざして—』、明石書店
- 龍慶昭・佐々木亮 (2004)『「政策評価」の理論と技法』(増補改訂版)、多賀出版
- 労働政策研究・研修機構 (2008)『OECD国際成人技能調査 (PIAAC) に関する報告』、JILPT資料シリーズ No.37
- OECD教育研究革新センター (NPO法人教育テスト研究センター監訳、坂巻博之、佐藤郡衛、川崎誠司訳) (2008)『学習の社会的成果—健康、市民・社会的関与と社会関係資本』、明石書店
- OECD (稲川英嗣、御園生純監訳、一木玲子、大塚裕介、高橋聡訳) (2009)『世界の教育改革3 OECD教育政策分析』、明石書店
- OECD (岩崎久美子、菊澤佐江子、藤江陽子、豊浩子訳) (2009)『教育とエビデンス—研究と政策の協同に向けて—』、明石書店
- Becker, G. S.(1993). *Human Capital* (Third edition). Chicago: The University of Chicago Press.
- Boruch, R. and Herman, R.( 2007) .What Works Clearinghouse, United States. In OECD. *Evidence in*

- Education—Linking Research and Policy*. Paris: OECD, 53-61.
- CEDEFOP (European Centre for the Development of Vocational Training). (1982). *Alternance Training for Young People: Guidelines for Action*, Thessaloniki: CEDEFOP.
- Davis, P.(1999). What is Evidence-based Education? *British Journal of Educational Studies*, 47(2), 108-121.
- Fitz-Gibbon, G.(2000). Education: Realizing the Potential. In Davis, H.T.O., Nutley, S.M. and Smith, P.C.(Eds). *What Works-Evidence-Based Policy and Practice in Public Services*, Bristol: Policy Press.
- Feinstein, L., Sabates, R., Anderson, T.M. Sorhaindo,A and Hammond, C.(2006). What are the Effects of Education on Health? In Desjardins, R. and Schuller, T. (Eds.), *Measuring the Effects of Education on Health and Civic/social Engagement*.Paris: OECD.
- Glass,G.V., McGaw,B. and Smith,M.L.(1981). *Meta-Analysis in Social Research*.Calif.: SAGE Publications.
- Gough, D.(2007). The Evidence for Policy and Practice Information and Co-ordinating (EPPI) Center, United Kingdom. In OECD. *Evidence in Education—Linking Research and Policy*. Paris: OECD. 63-70.
- Gueron, J.M.(2002). The Politics of Random Assignment: Implementing studies and Affecting Policy. In Mosteller, F. and Boruch, R.(Eds.). *Evidence Matters: Randomized Trials in Education Research*, Brookings Institution Press.
- Noonan, E. (2009). How can we develop evidence-based pedagogy? (The Campbell Collaboration, Decade of the Mind Conference, 2009/9/11 会議資料)
- Nutley, S.M., Walter, I. and Davies, H.T.O.(2007). *Using Evidence: How Research Can Inform Public Services*. Bristol: Policy Press.
- OECD (2000).*Knowledge Management in the Learning Society*. Paris: OECD.
- OECD (2001). *The Well-being of Nations—The Role of Human and Social Capital*. Paris: OECD.
- OECD (2004).*Innovation in the Knowledge Economy-implications for education and learning*. Paris: OECD.
- OECD, Statistics Canada.(2005). *Learning a Living: First Results of the Adult Literacy and Life Skill Survey*. Paris: OECD.
- OECD ( 2009 ) . PIAAC BQ JRA V5.0-Conceptual Framework. *PIAAC-NPM (2009\_02\_01) BQ V.5.0 Conceptual Framework, doc.*(working paper).
- Pawson, R.(2006). *Evidence-based Policy: A Realist Perspective*. London: SAGE Publications.
- Psacharopoulos, G.(2006). The Value of Investment in Education: Theory, Evidence and Policy. *Journal of Education Finance*, 32(2),120-122.

(2010.2.13受理)

## **Evidence-based Policies in Education: New Directions and Issues**

Kumiko Iwasaki

National Institute for Educational Policy Research

iwasaki@nier.go.jp

### **Abstract**

The world of evidence-based policies and practices has not always been widely understood in the field of education. However, this has appeared to change in recent times with a considerable surge in the number of studies that discuss the educational policies of various countries across the world. In this paper, firstly, issues related to production, mediation and the use of evidence will be clarified, mainly based on the rigorous definitions provided by the Cochrane and Campbell collaborations, which are known to gather data through randomized control trials (RCTs), undertake systematic reviews, and conduct exhaustive meta-analyses. Secondly, the fact that OECD has had to frequently make policy recommendations using the world of evidence-based policies,-that is, making political decisions with the help of scientific data,-will be discussed in juxtaposition with the human capital theory. Finally directions for evidence-based policies will be suggested.

### **Keywords**

Evidence-based policies, Human capital theory, Educational investment,  
Organization for Economic Co-operation and Development (OECD), Campbell Collaboration



## 【研究論文】

# 精神保健福祉領域における科学的根拠にもとづく実践(EBP)の 発展からみたプログラム評価方法論への貢献 ～プログラムモデル構築とフィデリティ評価を中心に～

大島 巖

日本社会事業大学社会福祉学部

iwao\_oshima@yahoo.co.jp

## 要 約

近年、精神保健福祉領域では、科学的根拠にもとづく実践プログラム（EBP; Evidence-Based Practices）に大きな注目が集められ、その中から従来のプログラム評価の方法論に、新たな視点や評価のアプローチが提起されている。本稿では、EBP研究・実践の新しい動向と成果を、優れたプログラム効果に関連するEBPプログラムの効果的援助要素の研究、効果的プログラムモデルのフィデリティ評価の発展、対人サービス領域の新しい技術移転の方法であるEBPツールキットアプローチに焦点を当てて検討した。その上で、これらEBPサービス研究の経験が、プログラム評価の方法論に与える寄与の可能性と今後の課題を考察した。

## キーワード

科学的根拠にもとづく実践プログラム（EBP）、フィデリティ尺度、サービス普及研究、EBPツールキットアプローチ、プログラム理論

## 1. はじめに

近年、精神保健福祉領域では、科学的根拠にもとづく実践（Evidence-Based Practice; EBP）プログラム（以下EBPプログラム）への関心が高まっている。EBPプログラムとは、利用者の援助効果（アウトカム）を向上させる一貫した科学的証拠（エビデンス）のある実践プログラムをいう（Drakeら 2003; 2005, Corriganら;2008）。薬物療法などの身体療法にも匹敵する効果をもたらすことがランダム化比較試験（Randomized Clinical

Trial; RCT）やその他の介入研究で明らかにされ、成果（アウトカム）に関するエビデンスが蓄積されている。このような中、世界的な治療ガイドラインにはEBPプログラムが位置づけられるようになり、ニーズのある利用者すべてがこれらプログラムを利用できるようにする取り組みが試みられている（大島 2007a; 2007b）。

こんにち保健・医療領域を中心とした対人サービス領域では、パラダイムシフトとも言える大きな潮流として、科学的根拠にもとづく医療（Evidence-Based Medicine; EBM）や、関連分野の

科学的根拠にもとづくアプローチに注目が集められている（縣 2000、矢野 1999）。同時に、EBMに関わる臨床研究やそのエビデンスを社会的に共有化するための取り組みが活発に行われ、プログラム評価の発展に示唆を与えている。

まずEBMでは、RCT研究を中心に蓄積した知見を体系的にメタ分析で検討するシステムティックレビュー（Systematic Review; SR）の方法が確立した。また、評価知見を幅広く社会が共有するための治療ガイドライン等のアプローチ法も一般化した（正木ら 2006）。

一方で、EBMは薬効判定や身体療法の評価など臨床評価が主流であり、社会プログラムに対する評価方法論である体系的なプログラム評価の方法論は一般的には適用されない（大島 2007a）。これに対して、精神保健福祉領域を中心に発展したEBPプログラムは、基本的には社会問題の解決を目指した社会プログラムである。EBPプログラムの評価には、プログラム評価の方法論が大きな役割を發揮する（大島 2007a）。すなわち、EBPプログラムへの評価には、EBMで主に注目されるアウトカム評価以外にも、サービス提供体制や実施方法の検討（プロセス評価）や、ニーズを持つ対象者を特定し（ニーズ評価）、それらの人たちにプログラムを行き渡らせるプログラム理論評価を含むサービス普及研究が体系的に適用され、さまざまなプログラム評価の枠組みや方法論が体系的に活用される（Rossiら 2004）。その中で、EBMと同時に発展して来たEBPプログラムに対する評価研究の取り組みから、従来のプログラム評価の方法論にはない、新たな評価アプローチ法が生み出されているように思われる。

本稿では、精神保健福祉領域において発展して来たEBPのプログラム評価研究や、サービス普及研究の新しい動向を文献的に検討するとともに、プログラムアウトカムに関連するプロセス評価法の発展と、サービス普及研究とその具体的アプローチ法としての発展したツールキットプロジェクト（Mueserら 2003）の意義と役割を示し、プログラム評価の方法論への寄与と今後の課題を検討する。

## 2. 科学的根拠にもとづく実践（EBP）への注目とその背景

### (1) 国際的研究動向

#### ①科学的根拠にもとづく医療（EBM）の発展と科学的根拠にもとづく実践（EBP）

まず保健医療福祉領域における、現代の新しい潮流となったEBMと、EBPプログラムの関連と差異について整理しておく。

まずEBMとは、直感やあやふやな経験にもとづくのではなく、科学的に明確なエビデンスに基づいて最適な医療や治療法を選択し実践するための方法論、あるいは行動指針である（縣, 2000; 矢野, 1999）。カナダのマクマスター大学のGuyattらのグループが初めてEBMの用語を用い、この概念を発展させて以来（Evidence-Based Medicine Working Group 1992）、EBMは短期間のうちに世界的な新しい大きな潮流になった（縣 2000、矢野 1999）。

一方、EBPプログラムとは、利用者の援助効果を向上させる一貫した科学的証拠のある心理社会的援助プログラムや社会プログラムをいう（Drakeら 2003、Corriganら 2008）。このプログラムは、EBM同様に有効性に関する十分なエビデンスと実施についての社会的合意が得られている。システムティックレビュー（SR）が行われてエビデンスが蓄積され、EBPプログラムが治療ガイドラインに組み入れられている（Drakeら 2005、Corriganら 2008）。

しかしながらその一方で、EBPの成果が実証されて20年や30年が経過するにも関わらず、普及が不十分でサービスの恩恵に与られない人々が数多く存在したり、プログラム基準を満たさない不適切な実践しか行えない厳しい現状が常に議論の対象になった（Drakeら 2009a; 2009b）。EBPプログラムは、薬物療法や他の身体療法と同様に、患者・利用者が適切に利用・活用する権利を有しており、一方では、医療機関・実施機関や社会は、限られた援助資源の中でEBPプログラムを優先的に提供する責務がある（Drakeら 2001）。

このように、EBPプログラムは、EBMに比較して社会システムの中での実施や普及と、限られた社会資源の中での優先的実施という、プログラム

評価が対象とする社会的・政治的文脈が常に課題になる。

②科学的根拠にもとづく実践プログラム（EBP）  
 のエビデンスと治療ガイドライン

精神保健福祉領域における代表的なEBPプログラムとしては、包括型ケアマネジメント（ACT; Assertive Community Treatment）、家族心理教育、IPS（Individual Placement and Support）援助付き雇用などがある（Corriganら 2008、Mueserら 2003）。たとえば、家族心理教育について、Mariら（1994）が行ったメタ分析の結果では、ランダム化比較試験で条件に合致した6研究のデータをプールしたメタ分析の結果、対象者総数350人で家族心理教育群が対照群に比べて再発率が減少する割合（オッズ比（括弧内は95%信頼区間））は、介入後9ヶ月で0.22（0.10-0.37）、2年で0.17（0.10-0.35）となり、家族心理教育群の再発率が低いエビデンスが蓄積されている。

このような成果を受けて、精神障害に関する国際的治療ガイドラインでは、これらの心理社会的介入プログラムが明確に位置づけられている（大島ら 2007b; 2009）。すなわち、包括型地域生活支援プログラムACTと、家族心理教育、援助付き就労プログラムは、いずれのガイドラインでも、科学的根拠（エビデンス）のレベルも高く、推奨度が高く設定されている（表1）。一方、集中型ケー

スマネジメント、ピアサポートプログラム、援助付き住居プログラムなどのプログラムは、エビデンスがまだ十分に蓄積されておらず推奨度はガイドラインによって様々である。

③EBPプログラムのサービスギャップとその対応策

上記のとおりEBPプログラムは優れた科学的根拠があるにもかかわらず、プログラムの実施・普及の程度は、地域精神保健福祉が進んだ地域でも一般的には大変に低い（Lehmanら 1998、Drakeら 2003）。Lehmanら（1998）は、アメリカの2州における719名に対する体系的な調査から、PORT（Patients Outcomes Research Team）勧告という国家レベルの科学的根拠にもとづく治療ガイドラインに準拠して治療や支援の必要があると判断されたもののうち、実際の治療・支援が実施された数を明らかにした。その結果、薬物療法はいずれも90%以上がPORT勧告に基づく治療を受けているのに対して、薬物療法と同程度の介入効果のエビデンスがあり、PORT勧告にも位置づけられている家族心理教育や包括型ケアマネジメントACT、援助付き雇用については、ニーズをもつ対象者の10-30%程度にしかPORTに準拠する援助が提供されないことを明らかにした。

この結果は、EBPプログラムを臨床・実践現場で実施し、普及・定着させることの困難性を如実に示している。すなわち、EBPプログラムを適切

表1 統合失調症に対する各ガイドラインにおける推奨のグレード

	APA:統合失調症治療ガイドライン(2004)	統合失調症PORT改訂版(2004)	NICE:統合失調症臨床ガイドライン(2003)	エキスパートコンセンサスガイドライン(1999)
家族心理教育	◎	◎	◎	○
包括型地域生活支援プログラムACT	◎	◎	◎	○
集中型ケースマネジメント	○/△	-	△	○
ピアサポートプログラム	○	-	-	○
援助付き就労プログラム	◎	◎	○	○
援助付き住居プログラム	△	-	-	○

(注) ◎エビデンスに基づく推奨、○委員会・エキスパートの推奨、△委員会・エキスパートの限定的な推奨  
 (出所) 大島ら 2007、2009c

に実施・普及していくためには、プログラムの実施システム、さらにはそれを人的・財政的・技術的にバックアップする社会システムの変更を要するが、それが容易でないことを示唆している。EBPプログラムは、実施システムやそれを取り巻く支援システムに関わる社会プログラムである(大島 2007a)。このプログラムを社会に適切に位置づけていくためには、プログラム理論などを用いて、社会の実施システムやバックアップシステムを含む周回なプログラム設計を行うことが求められる。

EBPプログラムがニーズをもつ人たちに行き届いていない不適切な状態は、サービスギャップ(service gap)(Ravivら 2009、Stefflら 1985)、あるいはscience-to-service gapと呼ばれる(Drakeら 2009)。これに対して、EBPプログラムのサービスギャップを改善するために、近年欧米諸国では取り組みが始められた。それは、まず効果的なプログラムモデルを定式化・標準化し、その上で、EBPプログラムの有効な実施体制・サービス提供体制のあり方や工夫を示し、実施の困難・障壁を明らかにし、組織的な実施・普及体制を整えるための取り組みである(Drakeら 2003、Mueserら 2008)。

アメリカ連邦政府では、サービスギャップを埋め、EBPの実施・普及を進めるために、EBPツ-

ルキットプロジェクトを1999年より開始した(アメリカ連邦保健省薬物依存精神保健サービス部 SAMHSA 2009、Mueserら 2003)。このプロジェクトは、現在、全米に広まりつつある(Corriganら 2008)。EBPプログラムの実施・普及を進めていくためには、障害をもつ人々への直接的なプログラムの提供だけでなく、実施システムやそれを取り巻く支援システムに関わるレベルまでを含めたプログラム設計が重要である。このようなプログラム設計を明確に示し、プログラムの各要素が適切に機能するための「工夫」などを盛り込むとともに、EBPプログラムに関わる各種の利害関係者への情報提供と、プログラムへの関与の仕方のガイドラインを示したEBPツールキットの取り組みは注目に値する(大島 2007)。

## (2) 日本におけるEBPプログラムのエビデンス蓄積、治療ガイドライン

日本のEBPプログラムに対するRCT研究・比較研究について、医中誌Web(ver.4)とMedline(Ovid)を使用し、2002年以降に出版された論文を検索した(大島ら 2008b)。その結果、認知症に関するプログラム研究、家族会支援(重い精神障害対象)に関する研究がわずかに公表されただけであった(表2)。日本において、ACTやIPS援助付き雇用に関する取り組みがまだ緒につ

表2 「国際的に推奨される地域自立支援プログラム」に関する日本の臨床研究・総説論文

職種別テキストの種類	国際的に推奨される地域自立支援プログラム				
	包括型地域生活支援プログラム ACT	ケースマネジメント	ピアサポートプログラム	就労支援	住居(統制語)
医中誌Web(精神障害/会議録除く)★A	88	419	4	140	102
★Aのうち(総説)	0	2	0	1	1
★Aのうち(RCT/準RCT/比較研究)	0	3	0	3	4
★Aのうち(RCT/準RCT)	0	0	0	0	1
Medline(mental disorders)★B	126	412	1479	140	297
★Bのうち(Japan)が含まれるもの★C	0	4	16	3	4
★Cのうち(review articles)	0	2	2	1	0
★Cのうち(Clinical trial, all)	0	0	1	0	0

(注) 2002年～2007年の文献(最終アクセス:2007.12.8)

(出所) 大島ら 2008b

表3 日本のEBP-RCT研究の取り組み状況 (筆者らの取り組み)

番号	プログラム・研究名	研究代表者等	無作為化の方法	追跡期間	主なアウトカム指標	主な知見
1	NCNP国府台病院・家族心理教育プログラム評価	Itoら (2002)	利用者単位の入院日による系統的無作為化	退院後 9M	再入院率、家族のEE、家族負担尺度、家族機能尺度、他	再入院率の減少、EE改善、家族機能改善
2	12施設共同研究・家族心理教育プログラム評価	大島ら (2004b)	利用者単位の入院日による系統的無作為化	退院後 9M, 2Y	再入院率、精神症状、本人参加準備性・ソーシャルサポート、家族負担尺度、家族機能尺度、他	精神症状の改善、本人参加準備性・ソーシャルサポートの改善、家族の本人への拒否的感情を低め、ケアへの自己効力感を高める
3	NCNP国府台病院・ACT包括型地域生活支援Prgプログラム評価	伊藤ら (2008)	利用者単位の無作為化(乱数表)	退院後 9M, 2Y	地域滞在期間、QOL、サービス満足度、自己効力感、他	地域滞在日数の増加、サービス満足度の向上
4	公立地域生活支援センター・IPS援助付き雇用Prgプログラム評価	大島ら (2008a)	利用者単位の無作為化(乱数表)、ウェイトイングリスト法	開始後 6M	就労率、就労週数、自己効力感、QOL、地域滞在日数、他	就労率、就労週数の増加
5	17施設共同研究・心理教育ツールキット導入プログラム評価	福井ら (2009)	施設単位の層化無作為化(乱数表)、ウェイトイングリスト法	導入後 2Y	スタッフの知識、ケア意識・態度・行動、施設環境、患者動態、他	スタッフの家族心理教育の認識向上、家族ストレスへの注目、施設環境の改善

いたところであり、日本の精神保健福祉システムや文化に根ざした知見 (比較研究) の蓄積は、これからの課題であることがわかる。

表3は、EBPプログラムについて、著者らが取り組むEBPプログラムのRCT研究である (福井ら 2009、Itoら 2002、伊藤ら 2008、大島ら 2004b、大島ら 2008a)。研究成果の公表がこれからの段階のものが少なくないが、日本でもEBPプログラムのエビデンスを積み重ねる段階に入ったことを示す一例として提示した。

2001年以降に発行された統合失調症の治療ガイドラインと精神科関連職種「精神医学」教科書を検索した (大島ら 2008b)。日本で出版された治療ガイドラインとしては、精神医学講座担当者会議 (2004) が監修した「統合失調症治療ガイドライン」がある。また精神医学の教科書としては、医学教育のためのテキスト4冊、看護教育のためのテキスト4冊、精神保健福祉士・作業療法士教育用のテキスト3冊を把握した。

「統合失調症治療ガイドライン」は、ただ一つ体系的にEBPプログラムの治療指針を提示している。しかしこのガイドラインでは、SRに基づくエビデンスや病期別の対応に関する記述が不十分

である。一方、精神科関連教科書については、一部の教科書を除いて、EBPプログラムの記述が不十分である。また、EBPプログラムのエビデンスについて基本的には触れられていない。

### 3. サービス普及研究とEBPツールキットプロジェクト

#### (1) 国際的研究動向

##### ① サービス普及研究とは

まずEBPプログラムとの対比から、EBMで重視されるのは、限られた特定の属性をもつ対象者に対して、厳密なランダム化比較試験 (RCT) を行い、アウトカムに関する知見をSRで評価し、エビデンスのレベルを定めてそれを蓄積する (矢野 1999)。これに対して、EBPプログラムについては、このアプローチだけでは必ずしも十分ではない。通常の臨床場面・実践場面を想定して、効果的なプログラムをニーズをもつさまざまな対象者に対して、多様な治療・支援環境の中で実施するためのエビデンスも用意する必要がある (Rosenheck 2001)。すなわち、多様な背景や属性

をもつ対象者に対して、必ずしもよく訓練された高い技術レベルのスタッフでなくとも、一定のガイドラインに基づいて実施すれば、一定の成果が納められることを明らかにすることが求められる。

これに関して、Rosenheck (2001) は、4レベルの効果評価研究、サービス研究を定義する。すなわち、第1レベルは効能研究 (efficacy research) で、高いレベルの統制された環境で介入効果を明らかにする研究である。第2レベルは効果研究 (effectiveness research) で、現実社会により近接する環境で統制された介入効果を明らかにする研究である。第3レベルの有効性研究 (effectiveness studies) は、現実社会に研究の知見を移転するための方法を明らかにする研究である。最後の第4レベルの普及プロセス研究 (dissemination process research) は、研究から実践に移し替えるための組織体のプロセスに焦点を当てた研究である。サービス普及研究は、上記の第3レベル、第4レベルを中心に行われる研究である。効果的プログラムモデル構築に関する実証的研究、行政組織に関する研究、医療経済研究・財政分析、サービス組織のスタッフ・指導者の知識・技能・教育に関する研究、組織過程に関する研究、合意形成に関する関係者の意識に関する研究などが含まれる。

サービス普及研究は、プログラム提供の実施プロセスに関わる研究であり、プログラム評価のプロセス理論に深く関わるアプローチと行うことができよう。

## ②プログラムモデルの確立とフィデリティ評価

EBPプログラムでは、効能研究や効果研究によって有効性が明らかにされた心理社会的介入プログラムのプログラムモデルを同定し、その定式化・標準化を行うことを重視する。プログラムの実施・普及のためには、その標準的な効果的普及モデルの確立が不可欠だからである。

効果的普及モデルの定式化・標準化のためには、個別プログラムに対してプロセス評価を実施して、効果的なプログラム援助要素 (critical components) を指定する (Bond 2000、McGrewら 1994)。効果のあがるプログラムモデルは、効果をもたらすプログラムの援助要素から構成される

と考えられるからである。たとえば、包括型ケアマネジメント (ACT) では、チームアプローチ、チーム会議を毎日実施、24時間対応サービス、ケースマネジャーが直接サービス提供、一定量以上のサービスの量、関わりの頻度、チームに看護師がいる、チームに精神科医がいるなどが、アウトカムに結びつくプログラム要素であることが実証されている (Bond 2000、McGrewら 1994)。これらの要素は、プログラムのゴールとの関係で、ゴールの実現 (ACTの場合は再入院日数の減少など) に結びつく援助要素を相関分析などによって抽出される。効果的普及モデルの実施マニュアルは、このように検討された効果的なプログラム援助要素を中心に、よく精査して具体的に記述する (Solomonら 2009)。

フィデリティ尺度 (fidelity scales) は、あるプログラムが効果的なプログラムモデルの基準に準拠している程度を測定する尺度である (Bond 2000)。良いアウトカムを予測する効果的なプログラム援助要素を組み合わせる尺度構成するフィデリティ評価のための尺度である。フィデリティ尺度は、実践が意図されたとおりに導入されているかどうかを評価する系統的な方法であり、EBPプログラム、およびサービス普及研究では不可欠の評価法として近年注目されている (Bondら 2000、Corriganら 2008、Drakeら 2005)。

サービス普及研究では、フィデリティ尺度を含むプロセス評価を体系的に実施しながら、EBPプログラムの実施・普及に影響する諸要因、困難・障壁の分析などが行われる。このような実証研究の結果に基づいて、EBPプログラムの改善のための方策が実証的に検討される。

フィデリティ尺度を使うことによって、①活動の実施状況が明らかになり、②新たなプログラム実施を追求する現場へ具体的なフィードバックが提供でき、③行政機関に対して、プロジェクト目標の達成状況の情報を提供することが可能となり、④情報を公共のために公開することもできる (Bondら 2000、Corriganら 2008)。すなわち、サービスの質のモニタリングに活用できるとともに、プログラムモデルを発展・改善させるために用いられることになる。

### ③EBPツールキットプロジェクトの発展

前述したアメリカのEBPツールキットプロジェクトは、EBPプログラムの深刻なサービスギャップを埋めて、EBPの実施・普及を促進するために、アメリカ連邦厚生省薬物依存精神保健サービス部（SAMHSA）が、全国のサービス研究者と連携しながら取り組んでいる（アメリカ連邦保健省薬物依存精神保健サービス部SAMHSA 2009、Corriganら 2008、Drakeら 2003、Mueserら 2003）。このプロジェクトは、EBPのプログラムモデルを確立しその効果的普及モデルを適切に実施・普及するための用具類（ツールキット）を開発し、関係者の合意を得ながら全国に普及させる取り組みである。

EBPツールキットには、EBPのプログラムモデルをさまざまな角度から記述・紹介するパンフレットや研修テキスト、研修ビデオ、紹介ビデオが含まれている。このほか、利用者や家族・他の支援者、プログラムの実践家や指導者、精神保健行政担当者というEBPプログラムに関わる利害関係者（stakeholders）が、それぞれの立場から当該のプログラムに関わり、プログラムを実施していくための工夫や実施ガイドラインが盛り込まれている。さらに、前述したフィデリティ尺度や、アウトカムモニタリングや効果評価に用いる簡便な評価尺度も用意されている。

ツールキット開発者たちは、単にEBPプログラムが普及すれば良いという立場を取っていない（Corriganら 2008、Mueserら 2003）。真に効果をもたらすプログラムの援助要素を同定し、それをフィデリティ尺度で測定しながら、プログラムの質を確保し、より効果の上がるプログラムモデルを構築しようとする（Corriganら 2008、Mueserら 2003）。ツールキットの作成に当たっては、いくつかの州の実施機関が試行版ツールキットを使用してその有用性を検証する研究が行われている（Mueserら 2003）。それは、当該プログラムモデルやツールキットの実用可能性、あるいは実施可能性を検討するサービス普及研究のモデル的取り組みと考えられる。このように、実施・普及システムの構築のレベルにおいても実証的な取り組みが行われている。

### (2) 日本におけるサービス普及研究、フィデリティ評価研究の発展

現在、精神保健福祉領域で主に使用されているフィデリティ尺度は、プログラム単位でプログラム全体を評価するものが主流である。一方で、そのプログラムを受けている利用者に対して、必要なプログラム援助要素が届いているのかどうか注目した個別の利用者単位に測定するフィデリティ評価も重要であろう。それは、実践的に利用者単位にサービス提供状況をモニターする必要あるとともに、利用者別にサービス改善の課題を明らかにし、その成果との関係を検討することが、利用者支援の立場から重要だからである。さらには、プログラム全体で、サービス機能を評価するのは容易ではないが、利用者単位ではサービスの内容や質を適切に把握することができる。

このように、現在、発展途上にあるフィデリティ評価法の一つの重要な検討課題は、個別利用者ごとのフィデリティ評価法の開発と発展にあると考える。

筆者らの研究グループは、精神保健福祉領域のいくつかの心理社会的介入プログラムにおいて、個別利用者単位のフィデリティ評価尺度を開発して、その有用性を検討している（福井ら 2004、Oshimaら 2004、大島ら 2009a, 2009b、園ら 2009）。

個別利用者フィデリティ尺度の開発に当たって、特に留意すべき点は以下の3点であろう。まず、①そのプログラムで標準的に提供すべきサービス内容を同定すること、②利用者別にそのサービスの必要の有無を判定することが不可欠であり、ニーズ評価法の確立が必要になること、③日常的な関わりを評価するため、サービス提供に関わる関係者（主にケースマネジャーと利用者・家族）が、信頼性のある評価をできるように配慮することである。なお、③に関連するが、個別利用者フィデリティ尺度を評価する評価者は、①個別利用者の援助に関わるスタッフ、②サービスを受ける利用者・家族、③援助場面を観察できる評価専門職である。③については、日常的な関わりを詳細に把握するのが難しい場合が少なからずあるであろう。このため、個別利用者フィデリティ尺度の主な評価者は、援助に関わるスタッフと、利用者・家族ということになる。それぞれの評価の

一致・不一致の程度の評価も重要な意味を持つことになるであろう。

#### 4. EBPプログラムの発展からみたプログラム評価方法論への示唆

EBPプログラムへの評価アプローチは、実践領域から発展した評価研究でありながら、実践で検証されながら発展するプログラム評価の新しい形態を提起し、そこからプログラム評価の新たな可能性が示唆されているものと考えられる(大島 2007)。

EBPプログラムからみたプログラム評価への示唆について、まず、フィデリティ評価法の活用が重要であろう。EBPアプローチで重視されるサービス普及研究では、フィデリティ評価を中心にプロセス評価が大きな役割を果たす(アメリカ連邦保健省薬物依存精神保健サービス部(SAMHSA 2009, Corriganら 2008)。フィデリティ尺度は、アウトカムとの関係からプログラムプロセスである効果的なプログラム援助要素を抽出して、尺度構成する(Bondら 2000)。日常的な実践活動の中から、アウトカム評価とプロセス評価を融合して分析し、より効果のあがる援助要素を検討して、さらには実証的により良いプログラムモデルに発展させるアプローチは新しい取り組みであろう。実証的・形成評価法の有力な1アプローチに位置づけることができよう。

さらに、実証的なプログラム形成評価を進めていく上で日常実践場面において利用者・対象者に対する支援の結果を、個別にモニタリングし、工夫や配慮を対応の仕方の「改善(カイゼン)」に役立てて行くアプローチとして注目する必要がある。その際、前項で取り上げた個別利用者フィデリティ評価は有用であろう。

EBPプログラムからのもう一つの示唆に、EBPツールキットプロジェクトの経験(Drakeら 2003, Mueserら 2003)がある。EBPプログラムの実施・普及を進める中で経験的に生み出されたEBPツールキットの取り組みは、まず第一に、当該のEBPプログラムのプログラム理論の実証的再検討に役立つであろう(Rossiら 2004)。EBPプログラムの実施・普及を進めていくためには、エビデン

スが蓄積された、効果的な直接的な援助プログラムのサービス提供(中核的な「EBPプログラム」)だけでなく、その実施・普及を支える実施システムやそれを取り巻く支援システム(実施・普及モデルとしての「EBPプログラム」)を取り入れた検討が必要になる。経験的に生み出されたEBPツールキットに基づいて、プログラムの実施・普及のためのシステム要素と機能を同定し、より機能する、効果的な実施・普及モデルとしてEBPプログラムのプログラム理論再構築が可能になる可能性がある。

また、EBPツールキットの取り組みは一般的なプログラム理論、特にプロセス理論の構成要素の再考に検討素材を提供する可能性がある。Rossiら(2004)は、プログラム理論のうちプログラム構造に関する規定であるプロセス理論には、ニーズをもつ利用者にプログラムを提供するプログラム機能に注目したサービス利用計画(Service utilization plan)と、そのプログラム機能を提供する組織の枠組みを示した組織計画(Organizational plan)の二つの理論が含まれるとした。これらの二つの理論が、効果性に着目して構成された中核的な「EBPプログラム」と、EBPツールキットから派生した実施・普及モデルとしての「EBPプログラム」において、どのように構成されるのか、検討を進める必要があるだろう。

EBPツールキットのもう一つの示唆は、効果的な社会プログラムを実施・普及していく上で必要な技術移転の方法に、新しいアプローチ法を提起していることである。特に、EBPプログラムに関わる利害関係者(stakeholders)ごとに必要な情報提供を行い、それぞれの立場で当該プログラムに関わり、実施していくための工夫や実施ガイドラインが盛り込まれている点は参考になると考える。

#### 5. まとめ

精神保健福祉領域を中心に発展したEBPプログラムに対して、アウトカムに関するエビデンスの蓄積を背景に、活発に研究が進められている。EBPプログラムのアウトカム評価知見が世界的に共有化され、治療ガイドラインなどに反映される

中、日本でもようやくRCTによる知見の蓄積と、治療ガイドラインなどでの結果の共有化が始まった。

EBPプログラムに対する評価アプローチには、基本的にプログラム評価の方法論が適用されていることを明らかにし、EBPプログラムに対する評価の取り組みから、プログラム評価の方法論に新たな示唆を与えていることを示した。具体的には、EBPプログラムでは、その効果性を背景に、より効果を生み出すプログラム援助要素を抽出し、フィデリティ尺度を構成していること、そして効果的なプログラムがニーズをもつ人たちに普及しないことに対して、実施・普及のためのEBPツールキットを作成することがプログラム評価に関わる新機軸と考えられた。

フィデリティ尺度構築の方法については、より効果的なプログラムモデルを構築するために用いられる実証的な形成評価法として有用であること、EBPツールキットについては、より効果性を高めるための中核的プログラム理論とともに、実施・普及に関わるプログラム理論の統合的再構築の必要性が示唆されること、そして両者の統合的理論を検討することが課題であること、これらが効果的な社会プログラムの新しい技術移転のアプローチ法として注目されることが示唆された。

## 参考文献

アメリカ連邦保健省薬物依存精神保健サービス部 (SAMHSA) 編 (2009)、日本精神障害者リハビリテーション学会監訳 (責任訳者：大島巖) 『アメリカ連邦政府EBP実施・普及ツールキットシリーズ』、日本精神障害者リハビリテーション学会、東京

伊藤順一郎・大島巖・西尾雅明・園環樹・他 (2008) 「日本におけるACTプログラムのRCT研究」、『平成19年度厚生労働科学研究補助金 (こころの健康科学研究事業) 重症精神障害者に対する新たな訪問型の包括型地域生活支援サービス・システムの開発に関する研究報告書』(主任研究者：伊藤順一郎)、1-20

大島巖 (2007a) 「保健福祉評価～分野別評価の現状と課題」、三好皓一編『評価論を学ぶ人のために』、世界思想社、208-223

大島巖 (2007b) 「統合失調症治療のガイドラインの作

成とその検証に関する研究：統合失調症をもつ人たちへの地域での自立支援プログラムに関する国際的治療ガイドラインのレビュー」、『平成18年度厚生労働省精神・神経疾患研究委託費総括報告書・統合失調症治療のガイドラインの作成とその検証に関する研究』(主任研究者：渡邊義文)

大島巖・伊藤順一郎・園環樹・堀内健太郎・鈴木友理子・西尾雅明 (2009a) 「電子カルテ様データベース・援助サービスコードシステムの開発～コードシステムの構築とその活用」、『平成16年度～19年度科学研究費補助金・基盤研究B研究成果報告書、心理社会的介入プログラムのプロセス評価方開発と効果的な援助要素の分析』(主任研究者：大島巖)、95-112

大島巖・高山莉理子・園環樹・賛川信幸・小川雅代 (2009b) 「個別利用者フィデリティ尺度 (ICFS) ～ケアマネジャー評価ACTフィデリティ評価』、『平成16年度～19年度科学研究費補助金・基盤研究B研究成果報告書、心理社会的介入プログラムのプロセス評価方開発と効果的な援助要素の分析』(主任研究者：大島巖)、67-76

大島巖・瀬戸屋雄太郎・福井里江・吉田光爾・賛川信幸・園環樹 (2009c) 「地域自立支援」、『平成20年度厚生労働省精神・神経疾患研究委託費総括報告書、統合失調症治療のガイドラインの作成とその検証に関する研究』(主任研究者：渡邊義文)、101-112

大島巖・梅原芳江・久米知代・星ゆかり・近藤昭子・伊藤順一郎・西尾雅明 (2008a) 「公設地域活動支援センターにおけるIPS援助付き雇用 (個別職業紹介とサポートプログラム) 導入とその評価 (2)」、『平成19年度厚生労働科学研究補助金分担研究報告書、精神障害者の一般就労と職場適応を支援するためのモデルプログラム開発に関する研究』(分担研究者：大島巖、主任研究者：西尾雅明)、17-28

大島巖・瀬戸屋雄太郎・福井里江・吉田光爾・賛川信幸・園環樹 (2008b) 「統合失調症治療のガイドラインの作成とその検証に関する研究・自立支援プログラム～日本における地域自立支援プログラムに関する治療ガイドライン・スタンダード：治療ガイドライン、教科書などのレビューから』、『平成19年度厚生労働省精神・神経疾患研究委託費総括報告書、統合失調症治療のガイドラインの作成とその検証に関する研究』(主任研究者：渡邊義文)

- 大島巖・伊藤順一郎・福井里江・瀬戸屋希・他 (2004b) 「心理教育プログラムのニーズアセスメントと効果評価に関する全国試行調査～その1 効果評価の総括：統合失調症を持つ人本人への効果」、『平成15年度厚生労働省精神・神経疾患研究委託費報告書、統合失調症の治療およびリハビリテーションのガイドライン作成とその実証的研究』（主任研究者：浦田重治郎）：127-134
- 縣俊彦編 (2000) 『EBM～医学研究・診療の方法論』、中外医学社
- 全国精神科講座担当者会議監修、佐藤光源・井上新平編 (2004) 『統合失調症治療ガイドライン』、医学書院
- 園環樹・大島巖・賛川信幸・伊藤順一郎 (2009) 「ACTにおける家族支援のあり方：サービス記録と利用者アウトカムの分析」、『平成16年度～19年度科学研究費補助金・基盤研究B研究成果報告書、心理社会的介入プログラムのプロセス評価方開発と効果的な援助要素の分析』（主任研究者：大島巖）、129-146
- 福井里江・賛川信幸・香月富士日・大島巖・他 (印刷中) 「統合失調症を持つ人々を対象にした科学的根拠に基づく心理社会的介入プログラム普及促進のためのツールキット開発とその有効性の評価（その29）～本試行研究におけるRCTスタッフ調査2年後アウトカム」、『平成20年度厚生労働省精神・神経疾患研究委託費総括報告書、精神政策医療ネットワークによる統合失調症の治療及び社会復帰支援に関する研究』（主任研究者：塚田和美）
- 福井里江・大島巖・瀬戸屋（大川）希・他 (2004) 「統合失調症に関する家族心理教育プログラムの家族の視点からみたプロセス評価（第二報）ープログラム実施要素の家族による認知度と介入効果の関連」、『精神医学』、46: 487-492
- 正木朋也・津谷喜一郎 (2006) 「エビデンスにもとづく医療（EBM）の系譜と方向性：保健医療評価に果たすコクラン共同計画の役割と未来」、『日本評価研究』、6(1): 3-20
- 矢野栄二編 (1999) 『医療と保健における評価』、南江堂
- Bond, G.R., et al. (2000). Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research* 2, 75-87.
- Corrigan, P.W., Mueser, K.T., Bond, G.R., Drake, R.E. and Solomon, P. (2008). *Principles and practice of psychiatric rehabilitation: An empirical approach*. New York: Guilford Press.
- Drake, R.E., Essock, S.M. (2009). The science-to-service gap in real-world schizophrenia treatment: the 95% problem. *Schizophrenia Bulletin*. 35, 677-678.
- Drake, R.E., Goldman, H.H. (2003). *Evidence-Based Practices in mental health care*. Washington DC.: American Psychiatric Association.
- Drake, R.E., Merrens, M.R. and Lynde, D.W. (2005). *Evidence-Based Mental Health Practice*. New York: WW Norton & Company.
- Evidence-Based Medicine Working Group (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 268, 2420-2425.
- Ito, J., Oshima, I., Tsukada, K. and Koishikawa, H. (2002). Family psychoeducation with schizophrenic patients and their families from the viewpoint of empowerment. Kashima H, Falloon IRH, Mizuno, M., Asai, M. (eds), *Comprehensive treatment of Schizophrenia: Linking neurobehavioral findings to psychosocial approaches*. Springer, 100-106.
- Lehman, A.F., et al (1998). Patterns of usual care for schizophrenia: initial results from the Schizophrenia Patient Outcomes Research Team (PORT) Client Survey. *Schizophrenia Bulletin*. 24(1), 11-20, discussion 20-32. 1998.
- Lehman, A.F., Kreyenbuhl, J., Buchanan, R.W., Dickerson, F.B., Dixon, L.B., Goldberg, R., Green-Paden, L.D., Tenhula, W.N., Boerescu, D., Tek, C., Sandson, N. and Steinwachs, D.M. (2003). The Schizophrenia Patient Outcomes Research Team (PORT): Updated Treatment Recommendations 2003. *Schizophrenia Bull* 30, 193-217.
- Mari, J.J., Streiner, D.L. (1994). *An overview of family interventions and relapse on schizophrenia: meta-analysis of research findings*. *Psychological Medicine* 24, 567-578.
- McGrew, J.H., Bond, G.R., Dietzen, L., et al (1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Consulting Clinical Psychology* 62, 670-678.
- Mueser, K.T., Torrey, W.C., Lynde, D., Singer, P. and

- Drake, R.E.(2003). Implementing evidence-based practices for people with severe mental illness. *Behavior Modification* 27, 387-411.
- Oshima, I., Cho, N. and Takahashi, K.(2004a). Effective components of a nationwide case management program in Japan for individuals with severe mental illness. *Community Mental Health Journal* 40(6), 525-537.
- Raviv, A., Vago-Gefen, I., Fink, A.S. (2009). The personal service gap: factors affecting adolescents' willingness to seek help. *J Adolescence*. 32, 483-499.
- Rosenheck, R.A.(2001). Organizational process: A missing link between research and practice. *Psychiatric Services* 52, 1607-1612.
- Rossi, P.H., Lipsey, M.W. and Freeman, H.E.(2004). *Evaluation: A systematic approach (7th edition)*, Thousand Oaks, Sage. (大島巖他訳 (2005)『プログラム評価の理論と方法～システマティックな対人サービス・政策評価の実践ガイド』、日本評論社)
- Solomon, P., Cavanaugh, M.M. and Draine, J.(2009). *Randomized controlled trials: Design and implementation for community-based psychosocial interventions*. New York: Oxford University Press.
- Stefl, M.E. and Proserpi, D.C.(1985). Barriers to mental health service utilization. *Community Mental Health Journal* 21, 167-178.

(2010.2.8受理)

## **Contribution to methodology of program evaluation from the perspectives of development of Evidence-Based Practice (EBP) research: Program model constructions, fidelity assessments, and methods for their technology transfers**

Iwao Oshima

Japan College of Social Work  
iwao\_oshima@yahoo.co.jp

### **Abstract**

Innovative approaches for research and practices of Evidence-Based Practices (EBP) that have gained the worldwide attention in the field of mental health have been greatly developed, and the approaches seem to throw a new light on methodology of program evaluation. In this article, we reviewed current research trends and results of EBP service research, focusing on critical components of effective EBP program models associated with favorable program outcomes, development of fidelity scales for the effective models, and EBP toolkit approaches that are new measures for technology transfers in the field of human services. We finally discussed possible contribution of EBP research experiences to the methodology of program evaluation and issues to develop the perspectives.

### **Keywords**

Evidence-Based Practices (EBP) , fidelity scales, service dissemination research,  
EBP toolkit approaches, program theory



## 【総説】

## 「エビデンス」の利用に関する検討 ～技術移転と追試過程を中心に～

津富 宏

静岡県立大学

tsutomi@u-shizuoka-ken.ac.jp

### 要 約

あるプログラムが有効であるというエビデンスがつくられたとしても、エビデンスを利用すること、つまり、そのプログラムの有効性を保持ないし確認・更新しつつ実施することは容易ではない。そこで、本稿は、二つの視点で、エビデンス利用の過程を捉える。一つの視点は、技術移転、すなわち、オリジナルのプログラムの有効性を保持する過程として捉えるものであり、もう一つの視点は、追試、すなわち、オリジナルのプログラムの有効性を確認・更新する過程として捉えるものである。技術移転として捉えると、プログラムの実施に当たって、オリジナルのプログラムの有効性を保持するには、単なる知識伝達ではなく、エビデンス利用に向けて協働的なコミュニケーションが図れるよう、組織自体や組織と外部との関係を変えることが大切であることが明らかとなった。追試として捉えると、コア要素を、アウトカムの予測変数として捉え、異なるコンテキストにおける外的妥当性を高める要素（ローカルな追加や普及促進要素）を探求することが大切であることが明らかとなった。さまざまな特定分野において、技術移転の視点と追試の視点に立って、エビデンス利用を促進するため、情報提供、組織づくり・人づくり・リーダーづくりや技術支援、外部評価を担うpurveyorを創設することが次のステップである。

### キーワード

エビデンス、技術移転、追試

#### 1. はじめに エビデンス利用の難しさ

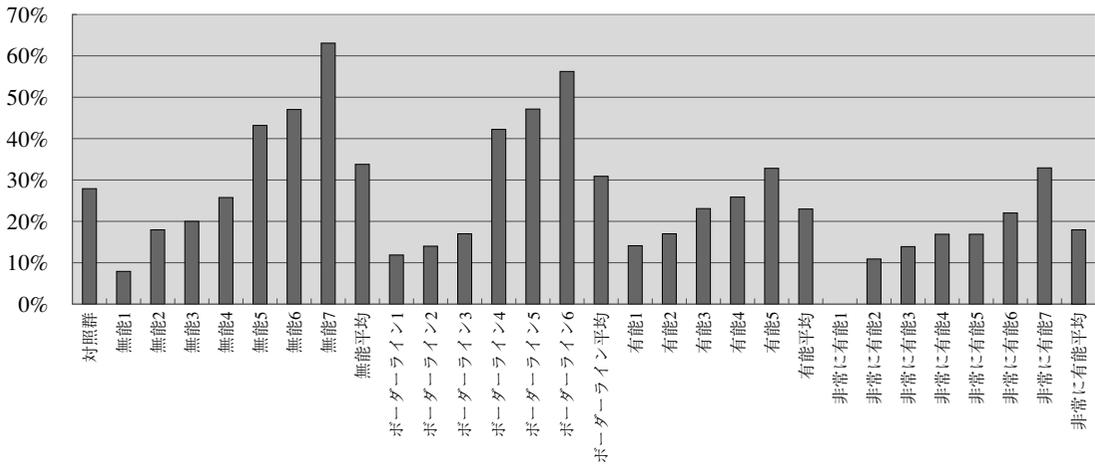
エビデンスとは、通常、「ある介入があるアウトカムに対して所期の効果を及ぼす」という因果命題に関する、実証的検討の結果、つまり、その因果命題が真である、あるいは、偽であるという言明である。

しかし、本稿においては、「ある介入があるアウトカムに対して初期の効果を及ぼす」という命

題が真であるという言明、平たく言えば、「ある介入は有効である」という言明を指すこととする。ただし、「ある介入は有効である」という言明は、既存の研究によって支えられた言明であって、今後の研究によって、この言明が塗り替えられうるのは当然である。

そこで、本稿では、エビデンスの利用を、「ある介入は有効である」という言明を与えられた介入を、その有効性を保持しつつ、あるいは、その

図1 Functional Family Therapyのセラピストが担当する青年の18ヶ月重大再犯率



(出所) Barnoski (2004) Exhibit 6から

有効性を確認・更新しつつ、実施する過程であると定義する。

キャンベル共同計画が公表した最初の系統的レビューである、Scared Straight (以下、SS) に関する系統的レビュー (Petrosino, Turpin-Petrosino, and Buehler, 2003) は、固定効果モデル (オッズ比1.68 (信頼区間1.20-2.36)) においても、ランダム効果モデル (オッズ比1.72 (信頼区間1.13-2.62)) においても、ボランティア受刑者による非行少年の教育プログラムであるSSが有害である、すなわち、非行行動を増加させるという結論に達した。しかし、このレビューを発表後、このSSに関する否定的な結論を出した、当の本人であるPetrosinoに対し、SSを導入したいので、プログラムの実施手法を教えてくださいという問合せが幾度となく来たという (私信、2004)。おそらく、この問合せをした人々は、肝心の系統的レビューを読んでいない。このエピソードは、系統的レビューという厳密な手法を用いて、エビデンスを正しく産出しても、そのエビデンスが正しく伝達され、利用されるわけではないことを象徴している。本稿の目的は、なぜ利用されないかを、利用側の課題としてではなく、普及側の課題として捉えることにある。

Functional Family Therapy は、犯罪・非行行動の減少に有効であるというエビデンスが示された

家族介入の技法であるが、Barnoski (2004) は、セラピストの職務能力をあらかじめ測定して、セラピストを4群 (無能群、ボーダーライン群、有能群、非常に有能群) に分け、「有能群」や「非常に有能群」のセラピストが実施すれば、無介入の対照群に比して、重大再犯率は減少するものの、「無能群」や「ボーダーライン群」のセラピストが実施すれば、重大再犯率がむしろ増大することを示している (図1参照: 対照群の再犯率と、各群のセラピスト各自が担当したクライアントの再犯率と、各群のセラピスト全員の平均の再犯率を示したもの)。つまり、有効であるというエビデンスが示された手法であっても、実施の仕方 (この例においては、セラピストの職務能力) によって、オリジナルのプログラムの有効性が保たれたまま利用されるとは限らないことを示している。なお、プログラムが実施されるにあたって、エビデンスどおりの有効性が保たれるかどうかは、実証の問題である。これは、たとえば、洗剤が開発されたとき、実際の使用条件によって、その洗浄力がどの程度左右されるかが、実証の問題であると同様である。

さらに、Petrosino and Soydan (2005) の系統的レビュー (表1) は、評価チームが、自ら介入のデザインと実施に関わっている程度が高い場合のみ、その介入の効果値が大きい (0.40) ことを

示している。このことは、介入が、オリジナルの開発チーム（ここでは、評価者と開発者が一体）の手を離れると、介入の有効性が下がることを意味している。

表1 介入のデザインと実施に対する、評価チームの影響と、「介入後の最初の事後効果」の関係

評価チームの影響	効果値	研究数
高	0.40	59
中	0.03	69
低	0.02	152
全体	0.10	271 <sup>a</sup>

<sup>a</sup> 19ケースが欠損

(出所) Petrosino and Soydan (2005) Table 2

この三つの事例は、エビデンスの利用がいかに困難な過程であるかを例証している。本研究では、この困難さを乗り越えていくために、二つの視点で、エビデンス利用を捉える。一つの視点は、技術移転、すなわち、エビデンスに支えられた（＝有効性が示された）プログラムを開発済みであると考え、そのプログラムの、異なるコンテキストにおける実施の目的を、その有効性を保持する（＝エビデンスどおりに実施する）過程であると捉える。もう一つの視点は、追試、すなわち、エビデンスに支えられた（＝有効性が示された）プログラムを開発中のものであると考え、そのプログラムの、異なるコンテキストにおける実施の目的を、プログラムを進化させ、有効性を深化向上する（＝エビデンスを確認・更新する）過程であると捉える。

## 2. 技術移転としてのエビデンス利用

有効であるというエビデンスが得られた介入が開発されたとき、それは、どのような条件において、その有効性を保ったまま実施しうるのであるうか。

Fixsen and Blasé (2006) は、「情報頒布、それ自体は、有効な実施につながらない」として、情報の頒布の例として、研究文献、資料の配布、実

施ガイドラインの配布などを挙げている。さらに、「いくらよく行われたとしても、訓練のみでは、有効な実施につながらない」として、知識伝達や訓練は、有効な実施を担保しないと主張している。

また、Walter et al. (2005) は、技術移転に関するいくつかのアプローチをレビューして、有効なアプローチとして、協働（研究と、政策・実務コミュニティの結びつきを強化するアプローチ）と、促進（技術的、財政的、組織的、情緒的なサポートを提供するアプローチ）を挙げている。この知見は、組織のあり方自体を、研究成果が活用される方向へと変えて行く必要があることを示している。

さらに、Kerner and Hall (2009) は、「実務家の行動の変容は、知識をもつことではなく、特定の患者に対する具体的なエビデンスの応用が有益であるという確信を持つことと関係している」と述べ、実務家本人に特定の知識が有用であることを納得させる必要があると主張している。

同様に、Shojania et al. (2006) は、（実施の質ではなく、健康をアウトカムとして測定し）「チームの変更とケースマネジメントが確かな改善をもたらしたのは、とりわけ、医師の承認を得ずに、ケースマネジャーが投薬を変更できた場合である」ことを見出し、コミュニケーションのあり方が、現場の主体性を生かす方向へ変わることの重要性を強調している。

劇的なのは、教員研修に関するJoyce and Showers (2002) の研究である。彼は、4種の研修技法（理論研修・討議、理論研修・討議・他者による実演、理論研修・討議・自分自身による実習と振り返り、理論研修・討議・実際の教室におけるコーチング）の、3つのアウトカム（知識、技術の実演、教室での使用）に対する影響を吟味した。言うまでもなく、教室での使用に対して影響を持たなくては、研修は意味がない。なお、4種の研修技法のうち、最初の3種はいずれも、教室を離れた場で行われる研修である。その結果、彼は、現場（教室）での教育技術の利用というアウトカムに相応のインパクトを持つのは、実際の教室におけるコーチングを伴う研修のみであることを明瞭に示した（表2）。

Fixsen et al. (2009) は、こうした一連の研究を

表2 研修の手法と、研修の効果

研修手法	アウトカム		
	知識	技術の実演	教室での使用
理論と討議	10%	5%	0%
上に加え、実演	30%	20%	0%
上に加え、実習と振り返り	95%	60%	5%
上に加え、教室でのコーチング	95%	95%	95%

(出所) Joyce and Showers (2002)

まとめ、実施を可能とするコア実施要素 (Core Implementation Components) として、職員の選抜、業務前・業務中訓練、継続的コーチングと相談助言、業績評価、決定支援データシステム、組織的支援、機関の一体化を挙げ、一貫して提供する役割を果たす、purveyor (技術移転支援専門組織) が必要であると論じている。purveyorとは、たとえば、multisystemic therapyを移転するためのMST Services, INC<sup>1</sup>やエビデンスに基づいた教育手法を移転するためのSuccess for All Foundation<sup>2</sup>などがそれに当たる。

さらに、Mihalic et al. (2004) は、エビデンスに基づいた非行・犯罪防止プログラムを移転するために行われているコロラド大学のプロジェクト Blueprints for Violence Prevention を振り返って、実施が成功するために欠いてはならない要素として、有効な組織、職能を有するスタッフ、プログラム移転を支持する強力なリーダー (program champion)、積極的な技術支援、外部評価を挙げている。Homel et al. (2004) もまた、結果として失敗に終わったUK Crime Reduction Programmeを振り返って、同様の結論に達している。

以上、技術移転としてのエビデンス利用についてまとめると、現場の「行動」を変えるには、単なる、知識伝達は無効であり、エビデンス利用に向けて協働的なコミュニケーションが図れるよう、組織自体や組織と外部との関係を変えることが大切であることが明らかとなった。つまり、個人レベルでの変革は困難であり、組織レベルの変革が必要であり (Nutley, et al., 2007)、そのためにも、技術移転には資源投入が必要となる (Investing to Deliver) (Homel et al., 2004)。

模範的な事例としては、Ogden et al. (2005; 2009) が紹介している、ノルウェーへのParent Management Training, Oregon Model の移転がある。ノルウェーは、purveyorとしての、Norwegian Center for Child Behavioral Development を創設した。このセンターでは、セラピストの訓練を行ってプログラムの実施の質を担保し、媒介変数としてのフィデリティ研究など実施に関する研究を行ってプロセス評価を行い、ランダム化比較試験による有効性研究を行ってインパクト評価を行うという、システムティックな取組みを1997年以降一貫して進めている。

### 3. 「追試」としてのエビデンス利用

エビデンスを利用することは、他のコンテキストにおいてもそのエビデンスが有効であるかどうかを確認・更新する追試であり、さらには、追試を通じたより良い介入の開発でもありと考えられる。このような考えは、Weisz (2000) の臨床現場に根ざした治療開発モデル (Clinic-Based Treatment Development Model)、Glasgow (2009) の実践に基づくエビデンス (practice-based evidence<sup>3</sup>) といった考え方とも軌を一にする。端的には、エビデンス利用とは、Winter and Szulanski (2001) のいう、真のコアを発見するための探求 (exploration to discover the arrow core) であり、Lenard-Barton (1988) のいう、技術革新としての実施 (implementation as innovation) である。以下、この関心に沿って実証研究を見ていきたい。

表3 Chorpita (2003) の「蒸留」のためのクロス表の一部分の例 (一部改変)

恐怖症の治療プログラム	プログラム要素							
	認知	暴露	報酬	緩和	賞賛	弛緩	コーピング言明	心理教育
A	あり	あり	あり	あり	あり	あり	なし	なし
B	あり	あり	あり	なし	あり	なし	あり	なし
C	なし	あり	なし	なし	なし	なし	なし	なし
(以下に続く)	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

(出所) Chorpita (2003)

Blakely et al. (1987) は、①フィデリティの高い実施者は低い実施者よりも効果的であること、②現場におけるプログラムの「改変」は有効性と無関係であること、③現場におけるプログラム要素の「追加」は有効性を高めることを見出し、ローカルな追加の重要性を指摘している。

また、Dearing (2009) は、フィデリティを、(プログラムの属性としてではなく) アウトカムの属性 (fidelity as a property of outcome) として定義した。つまり、フィデリティ評価とは、アウトカムの予測変数を確定するためのプロセス評価であり、予測性の高い予測変数を残していくことによる介入改善である。この考えは、Dearing and Meyer (2006) が、program fidelity (オリジナルのプログラムへの忠実度) に対するものとして提示した、effect fidelity (効果を高める要素としてのフィデリティ) という概念の延長上にある。つまり、追試を通じて、プログラムが、(新しい局面においても) なぜ有効でありうるのかを理解することこそ、コア要素の確定に有用であるという考え方である。こうなると、追試において検討されているのは、内的妥当性を高める要素ではなく、外的妥当性を高める要素となる。Blakely et al. (1987) に似るが、Dearing and Meyer (2006) は、追加的な要素によって外的妥当性が高まるという主張も行っている。

また、同様に、Rovniak et al. (2005) は、バンデューラの社会的認知理論に従って、モデリング、目標設定、自己モニタリング、フィードバックという四つの観点を取り出して、「理論的フィデリティ」を定義し、二種の運動 (ウォーキング)

プログラムについて理論的フィデリティを査定した上で、理論的フィデリティの高いプログラムのほうが低いプログラムよりも有効であることを見出した。この研究においても、フィデリティは、オリジナルプログラムの再現として定義されているわけではなく、プログラムの連続量的な「品質」(アウトカムの予測変数) として位置づけられている。

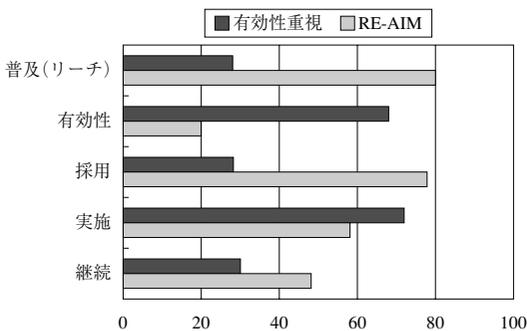
さらに、Chorpita (2003) は、複数の研究 (プログラム) に含まれている要素をクロス表にして一覧にすることでコア要素を確定できると主張し、この手法を「蒸留 (distillation)」と名づけている。クロス表の一例は、表3であるが、この例では、A、B、Cをはじめとする、恐怖症の治療に有効なプログラムについて、認知、暴露、報酬、緩和、賞賛、弛緩、コーピング言明、心理教育といった、治療プログラムの各要素を有しているかどうかを一覧表にしたものである。Chorpitaの行った実際の分析では、恐怖症の治療に有効なプログラムの90%以上に共通して含まれている要素 (コア要素) として暴露 (exposure)、40%以上に含まれている要素 (共通要素) としてセラピストによる賞賛、40%未満しか含まれていない要素 (オプションな要素) として弛緩、コーピング言明、心理教育が抽出された。複数のプログラムの共通要素としてコア要素を定義する、このChorpitaのアプローチもまた、コア要素を特定のプログラムにとらわれずに、アウトカムの予測変数として定義しようとする試みの一つであると考えられる。

Kumpfer et al. (2002) は、文化的マイノリティ

集団にプログラムを応用するにあたって、プログラムを改変することの効果を検討し、文化背景を考慮していない一般バージョンと、文化背景を考慮した修正バージョンを比較すると、一般バージョンのほうが効果がわずかに大きかったものの、修正バージョンのほうが対象家族のリクルートと保持においては41%も優れていたことを示している。この知見は、プログラムの実施において、純粹な効果を取るのか、普及（ターゲットへのリーチ）を取るのかという選択を迫る。コア要素は、そもそも有効要素なのか普及促進要素なのかという問いである。

そこで、Kumpfer et al. (2002) は、RE-AIMモデル (cf. Glasgow, 2009; Ory et al., 2007) を提案する。プログラムのトータルとしての効果を、Reach（普及）－Effectiveness（プログラム自体の有効性）－Adoption（組織レベルでの採用）－Implementation（現場レベルでの実施）－Maintenance（継続）の5つの要素を掛け合わせたものとして定義するモデルである。図2は、プログラムの有効性を重視した場合と、RE-AIMモデルに立ちReachやAdoptionを重視した場合を対比したものである。この架空例では、有効性重視のモデルの方がプログラム単体のエフェクトサイズは大きいものの、5つの要素を掛け合わせれば、普及や実施にかかわる諸要素を重視するRE-AIMモデルの方が、より大きなインパクトをもたらしていることが分かる。

図2 RE-AIMモデルの架空例



(出所) Kumpfer et al. (2002)

以上、エビデンス利用を追試として捉えると、コア要素とは、アウトカムの予測変数であって、異なるコンテキストにおける外的妥当性を高めるもの（ローカルな追加や普及促進要素）であるとも考えられることが明らかとなった。このように捉えると、フィデリティとは、オリジナルのプログラムの単なる再現度合ではなくて、むしろ、オリジナル・プログラムが修正可能であることを前提にした、(より広範な) ターゲットに対するプログラム・インパクトの度合であると定義される。率直に言って、後者の定義の方が、社会に資するのは当然である。

#### 4. 最後に

以上、有効であるというエビデンスが示されたプログラムの実施を二つの視点から論じてきた。一つは、プログラムを、有効性を保持しつつ実施する過程である技術移転と捉える視点であり、もう一つは、プログラムを、有効性を確認・更新しつつ実施する過程である追試として捉える視点である。

第一の視点からは、プログラムの実施に当たり、オリジナルのプログラムの有効性を保持するには、現場の「行動」を変えることが必要であり、そのためには、単なる知識伝達ではなく、エビデンス利用に向けて協働的なコミュニケーションが図れるよう、組織自体や組織と外部との関係を変えることが大切であることが明らかとなった。

第二の視点からは、プログラムの実施に当たって、オリジナルのプログラムの有効性を確認・更新していくには、コア要素を、オリジナルプログラムの単なる再現要素ではなく、むしろ、アウトカムの予測変数として捉えて、異なるコンテキストにおける外的妥当性を高める要素（ローカルな追加や普及促進要素）を探求することが大切であることが明らかとなった。

第一の視点に立ち、技術移転を可能とするのは、職員の選抜、業務前・業務中訓練、継続的コーチングと相談助言、業績評価、決定支援データシステム、組織的支援、機関の一体化といったコア実施要素を、一貫して提供する役割を果たす、

purveyor（技術移転支援専門組織）であることが指摘されてきた。これまで論じられてきてはいないものの、第二の視点にも立ち、エビデンス利用の過程を、エビデンス確認・更新の過程、すなわち、（より広範な）ターゲットに対するプログラム・インパクトの度合を高める追試の過程として定義し、purveyorにこの度合を高めることを追求させることは可能であろう。

今後は、わが国の、さまざまな特定分野において、技術移転の視点と、追試の視点を持ったpurveyorを創設し、「エビデンスをつかう」過程を促進することが次のステップであろう。<sup>4</sup>たとえば、筆者は、元少年院の教官として、犯罪者処遇分野におけるエビデンスの紹介を行ってきた（e.g. 津富, 1999a,b,c; 2000）が、今後は、本稿の知見をもとに、「エビデンスに基づく矯正のためのコンソーシアム」（Consortium for Evidence-Based Corrections）の設立を提唱したい。

#### 注記

- 1 <http://www.msts-services.com/index.php>
- 2 <http://www.successforall.net/>
- 3 エビデンスに基づく実践（evidence-based practice）ではない。
- 4 精神保健領域ではそうした取組みはすでに始まっている。大島巖のHP（<http://ioshima.com/>）を参照されたい。

#### 参考文献

- 津富宏（1999a）「犯罪者処遇の評価研究（一）」、『*刑政*』1283:58-67
- 津富宏（1999b）「犯罪者処遇の評価研究（二）」、『*刑政*』1284:50-60
- 津富宏（1999c）「犯罪者処遇の評価研究（三）」、『*刑政*』1285:46-55
- 津富宏（2000）「EBP（エビデンス・ベイスト・プラクティス）への道」、『*犯罪と非行*』124:67-99
- Barnoski, R. (2004). *Outcome Evaluation of Washington State's Research-Based Programs for Juvenile Offenders*. Washington Institute of Public Policy.
- Blakely, C. G., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B., and Emshoff, J. G. (1987). The Fidelity-Adaptation Debate: Implications for the Implementation of Public Sector Social Programs. *American Journal of Community Psychology*, 15, 253-268.
- Chorpita, B. F. (2003). Identifying core practice components across evidence-based psychosocial treatments for youth. Texas Department of Mental Health and Mental Retardation Children's Mental Health Evidence-Based Treatment Consensus Conference, Austin, TX. <http://www.dshs.state.tx.us/mhservices/pdf/IdentifyingCorePracticeComponentsAcrossEvidenceBasedPsychosocialTreatmentsforYouth.pdf> [2009年11月23日アクセス]
- Dearing, J. W. (2009). Applying Diffusion of Innovation Theory to Intervention Development. *Research on Social Work Practice*, 19(5), 503-518.
- Dearing, J. W., and Meyer, G. (2006). Revisiting Diffusion Theory. In Singhal, A., and Dearing, J. W. (Eds.), *Communication of innovations: A journey with Ev Rogers*, Thousand Oaks, CA: Sage, 29-60.
- Fixen, D. L., Blasé, K. A. (2006). "What works" for implementing "what works" to achieve consumer benefits. PowerPoint Presentation at the Treatment for Homeless Program Technical Assistance Workshop. [http://conferences.jbs.biz/tx\\_for\\_homeless\\_06/presentations/Wednesday,%20Jan%2025/Managing%20Change,%20Fixen.pdf](http://conferences.jbs.biz/tx_for_homeless_06/presentations/Wednesday,%20Jan%2025/Managing%20Change,%20Fixen.pdf) [2009年11月23日アクセス]
- Fixsen D. L., Blasé, K. A., Naoom, S. F., and Wallace. F. (2009). Core Implementation Components. *Research on Social Work Practice*, 19(5), 531-540.
- Glasgow, R. E. (2009). Critical Measurement Issues in Translational Research. *Research on Social Work Practice*, 19(5), 560-568.
- HomeI, P., Nutley, S., Webb, B., and Tilley, N. (2004). *Investing to Deliver: reviewing the implementation of the UK Crime Reduction Programme (Home Office Research Study 281)*. Home Office Research, Development and Statistics Directorate.
- Joyce, B., and Showers, B. (2002). *Student Achievement through Staff Development*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Kerner, J. F. and Hall, K. L., (2009). Research

- dissemination and diffusion: translation within science and society. *Research on Social Work Practice*, 19(5), 519-530.
- Kumpfer, K. L., Alvarado, R., Smith, P., and Bellamy, N. (2002). Cultural Sensitivity and Adaptation in Family-Based Prevention Interventions. *Prevention Science*, 3, 241-246.
- Leonard-Barton, D. (1988). Implementation as mutual adaptation of technology and organization. *Research Policy*, 17, 251-267.
- Mihalic, S., Fagan, A., Irwin, K., Ballard, D., and Elliott, D. (2004). *Blueprints for Violence Prevention*. Office of Juvenile Justice and Delinquency Prevention.
- Ogden, T., Forgatch, M. S., Askeland, E., Patterson, G. R., and Bullock, B. M. (2005). Implementation of Parent Management Training at the National Level: The Case of Norway. *Journal of Social Work Practice*, 19, 317-329.
- Ogden, T., Hagen, K. A., Askeland, E., and Christensen, B. (2009). Implementing and evaluating evidence based treatments of conduct disorder problems in children and youth in Norway. *Research on Social Work Practice*, 19 (5), 582-91.
- Ory, M. G., Mier, N., Sharkey, J. R., and Anderson, L. A. (2007). Translating science into public health practice: Lessons from physical activity interventions. *Alzheimer's & Dementia*, 3, S52-S57.
- Petrosino, A. (2004). 私信.
- Petrosino, A. and Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasiexperimental research. *Journal of Experimental Criminology*, 1, 435-450.
- Petrosino, A., Turpin-Petrosino, C., and Buehler, J. (2003). 'Scared Straight' and other juvenile awareness programs for preventing juvenile delinquency (Updated C2 Review). In: *The Campbell Collaboration Reviews of Intervention and Policy Evaluations (C2-RIPE)*, November, 2003. Philadelphia, Pennsylvania: Campbell Collaboration.
- Rovniak, L. S., Hovell, M. F., Wojcik, J. R., Winett, R. A., and Martinez-Donate, A. P. (2005). Enhancing Theoretical Fidelity: An E-mail-based Walking Program Demonstration. *American Journal of Health Promotion*, 20, 85-95.
- Shojania, K. G., Rangi, S. R., McDonald, Kathryn M., Grimshaw, J. M., Sundaram, V., Rushakoff, R., J., and Owens K., D. (2006). Effects of Quality Improvement Strategies for Type 2 Diabetes on Glycemic Control: A Meta-Regression Analysis. *JAMA*, 296, 427-440.
- Walter, I., Nutley, S. M., and Davies, H. T. O. (2005). What works to promote evidence-based practice? A cross-sector review. *Evidence & Policy*, 1(3), 335-364.
- Weisz, J. R. (2000). Lab-clinic differences and what we can do about them: The clinic-based treatment development model. *Clinical Child Psychology Newsletter*, 15(1), 1-3, 10.
- Winter, S. G., and Szulanski, G. (2001). Replication as Strategy. *Organizational Science*, 12(6), 730-743.

(2010.2.19受理)

## **Dissecting the process of using evidence as a process of technological transfer and a process of replication**

Hiroshi Tsutomi

University of Shizuoka

tsutomi@u-shizuoka-ken.ac.jp

### **Abstract**

Even when a program is shown to have evidence that it is effective, using the evidence, or implementing the program while retaining or confirming/renewing its effectiveness is not easy. This article proposes two views of the process of using evidence. One view is of the technological transfer, the process of retaining the effectiveness of the original program. The view of the technological transfer suggests that the implementation of the original program while retaining its effectiveness cannot be achieved by knowledge provision, but must include a change in intra- or inter-organizational environment to be more collaborative. The other view is of the replication, the process of confirming/renewing the effectiveness of the original program. The view of replication suggests that core components of a program can be defined as predictors of the outcome, or components which can enhance external validity of the program in different contexts, such as local additions and factors promoting applicability of the program. The next future step for the promotion of using evidence is to establish a purveyor (an entity responsible for the development of organizations, staff, and leaders, technical assistance and the third-party evaluation as well as the provision of information) in various fields, equipped with the view of the process of using evidence for both technological transfer and replication.

### **Keywords**

evidence, technological transfer, replication



## 【研究論文】

# 開発援助分野におけるRCT導入に見られる旧くて新しい課題 —理論なきRCTから理論検証のRCTへ—

青柳 恵太郎

国際協力機構

Aoyagi.Keitaro@jica.go.jp

## 要 約

開発援助分野では2000年代に入った頃からプロジェクトのインパクトを精緻に推計するという取り組みが強化され始めた。その一端としてRCTを用いたインパクト評価の事例も多く蓄積されてきている。こうした動きは主として開発経済学者によって主導されてきているが、RCTが広く用いられるようになるにつれて、長年に亘り評価研究において議論されてきたRCT導入の是非、及び初期の経済学に見られた「理論なき計測」に関わる議論が開発援助分野を舞台として再度問われていることが確認できる。他方、RCTが先行している疫学等の分野に比べて制度化が遅れていることが指摘できる。翻って、こうしたRCTの国際的展開の中で日本の取り組みを見ると、積極的とは言いがたいものがあった。しかしながら、近年はRCTを適用することも視野に入れつつ、インパクト評価全般への取り組みが強化され始めている。

## キーワード

RCT、インパクト評価、開発経済学、評価研究、開発援助

## 1. はじめに

開発援助分野では、厳格なインパクト評価 (rigorous impact evaluation) の文脈で2000年前後からランダムマイゼーション (randomization、randomized controlled trial: RCT) の波が押し寄せたが、およそ10年を経た今、ひとつの転換点を迎えつつある。Christopher Blattmanが、Impact Evaluation 2.0と題したレポートを著していることから、その変化の一端が伺える (Blattman 2008)。本稿では、開発援助分野において過去10年間に行われてきたインパクト評価、特にRCTを巡る国際的議論を要約すると共に、これらの議論を乗り越

えて見えてきた一つの方向性を紹介する。また、国際的潮流を見据え、現在の日本の開発援助分野に見られる取り組みについて、最新動向に触れる。これらを通じて、開発援助分野におけるRCTの過去を振り返り、未来を展望することを本稿の目的とする。

このわずか10年のRCTの展開の中で、古くから議論されてきた大きな二つの論点が再現されている。長年に亘り評価研究において議論されてきたRCT導入の是非、及び初期の経済学に見られた「理論なき計測」に関わる議論が開発援助分野を舞台として再度問われていることが確認できる。他方で、RCTにおけるランダム割付の実施過程で

生じる問題、及びその報告様式については、他分野において過去に議論がされてきておりながら、開発援助分野においては未だ十分に注意が払われていないことが指摘できる。

以下、本論は次の様に構成される。第2節において、開発援助分野へのRCTの普及過程を概観し、現在の到達点を確認する。次に第3節で、これまでに議論されてきたRCT導入にまつわる諸論点を整理する。続く第4節、第5節において、この10年間で見てきた課題とそれらを乗り越えるための取り組みを紹介する。最後に、第6節において、国際的潮流の中で日本がいかなる取り組みをしているかについて触れる。

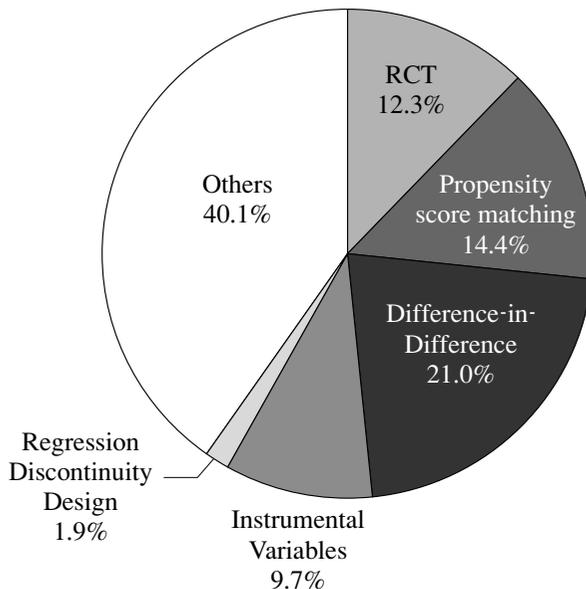
## 2. 開発援助コミュニティへのRCT普及過程

インパクト評価の話題となると、かつては操作変数やマッチングを用いた計量経済分析が真っ先に想起されたが、近年ではRCTが、その是非はと

もかく議論・関心の中心を占めるに到っている。図1は世界銀行が整備しているインパクト評価のデータベースに登録されている評価報告書の中で用いられた分析手法をまとめたものである。このデータベースへの掲載基準は明示的にcounterfactualを想定した評価をしているかという点のみであり、世界銀行のプロジェクト以外も登録されている。2010年1月現在、45カ国、204件の事例が蓄積されているが<sup>3</sup>、このうち26.0%（53件、全分析の12.3%）と、インパクト評価の4件に1件の割合でRCTが用いられている。RCTはアメリカの政策評価などでは60年代末から既に行われていたが、開発援助分野では、その有用性は認識されていたものの90年代後半まで、ほとんど実施されることはなかった。それが今では主要な分析手法の一角を占めるまでに到ったことは、特筆すべき変化である。

このような変化をもたらす契機となったのは、メキシコで実施された条件付現金給付プロジェクトの代名詞ともいえるPROGRESA<sup>4</sup>のRCTによる

図1 厳格なインパクト評価に用いられた分析手法



(注) OthersにはPipeline comparison、Other matching methods、Simulated counterfactual、Single difference、Difference in means、Duration model、otherが含まれる。図は、全分析 (n=431) に占める各手法の割合を示している (1事例につき複数の分析手法が用いられていることもある)。

(出所) 世界銀行Impact Evaluation Databaseより筆者作成。

評価、及びAbdul Latif Jameel Poverty Action Lab (J-PAL)<sup>5</sup>といった開発経済学者からなる組織的な啓蒙的活動・実践の開始、そしてそれらを下支えるミレニアム開発目標 (Millennium Development Goals; MDGs) やパリ宣言に代表される成果を求める開発援助潮流の変化である。

PROGRESAは、開発分野のプロジェクト評価におけるRCT適用の先駆的事例である。この評価では、プロジェクトの対象に選ばれた地域において介入のタイミングをランダムに変えることで介入群と統制群を生み出し、その効果を正確に検証する事を可能にしている<sup>6</sup>。評価作業にはワシントンDC所在の国際研究機関であるIFPRI (International Food Policy Research Institute)をはじめ、多くの研究者が関与し、主に人的資本蓄積への貢献を検証した実証分析が多数報告されている。これらの評価分析はPROGRESAの多面的効果を認める極めて強いエビデンスを提示しており、その結果、類似の条件付現金給付プロジェクトは南アメリカ諸国の間に広まる事となった。またメキシコ政府自身によっても同プロジェクトはスケールアップされ、重要な社会政策の一部をなすこととなった。PROGRESAがこれほどまでに展開した背景として、プロジェクトの効果について単に詳細な評価が行われたというだけではなく、それがRCTという厳格な手法によるものであったという事実がある事は明らかである。

これ以降、或いは同時期から世界銀行をはじめとする開発援助機関、北米を中心とした大学・研究機関、シンクタンク等が次々と厳格なRCTによるインパクト評価を実施し、多くの知見を蓄積し始めている。とりわけ、前述のMITとハーバード大学の開発経済学者を中心に設置されたJ-PALは、RCTの強力な推進機関としての地位を確固たるものとしており、多くのエビデンスを産出している。

このようなエビデンス重視の取り組みが開発援助コミュニティにおいて真剣に議論されるようになった背景として、開発援助を取り巻く環境変化が指摘できる。周知の通り、2000年9月に開催された国連ミレニアムサミットにおいて開発援助コミュニティの共通目標として、ミレニアム開発目標がまとめられた。そこでは具体的な数値目標が

明示され、その達成のためには援助効果の向上が必須であるとの認識が共有されるに到った。援助効果向上の議論は、さらに2005年のパリ宣言によって加速化されている。同時期に、アメリカのシンクタンクCenter for Global Development (CGD)のように、従来の開発援助評価を痛烈に批判する組織も現れた。CGDは、「これまでの開発援助評価はプロセス評価に偏重しており、インパクト評価が軽視されている。またインパクト評価が実施されたとしても、用いられる手法 (評価デザイン) に問題があるために結果の信頼性は著しく低い。その結果、評価によって、どのようなプロジェクトが如何程の効果を生んだかが明らかにされていない」といった趣旨のEvaluation Gapと呼ばれる議論を展開し、開発援助評価実務へも無視できない影響を与えた (Savedoff et al. 2006)。このような成果、そしてその成果に繋がるエビデンスを求める援助潮流に対して、評価の側から提示されたツールのひとつが厳格なインパクト評価である。

こうした取組強化に呼応する形で、インパクト評価の制度化も検討されてきた。DACのEvalunetから派生したインパクト評価に関わる組織の国際ネットワークであるNetwork of Networks on Impact Evaluation (NONIE) では、2009年10月にインパクト評価ガイドラインを整備している (Leeuw and Vaessen 2009)。また、上述のCGDによる批判の中で、インパクト評価を司る組織の設立が提言されているが、International Initiative for Impact Evaluation (3ie) として具現化している。3ieは、インパクト評価実施のための資金提供や学術雑誌の刊行等、エビデンスの蓄積に関わる各種取り組みを行っている。特記すべきは、キャンベル共同計画の方法論に則り、系統的レビューを開始している点である。まだ始まったばかりということもあり、2010年1月現在、*Water and Sanitation Interventions to Combat Childhood Diarrhoea in developing countries*、及び *Behavior change interventions to prevent HIV among low-income girls and women living in low and middle income countries* と題したレビューが二件公開されているのみであるが、これに加えて全7件のプロトコルが公開されている (Waddington et. al. 2009、Padian 2009)。

### 3. RCT導入を巡る論争

前節で概観したように、開発援助分野におけるRCTの適用事例は確実に増加傾向にあり、またCGDやJ-PALによる啓蒙活動等によってRCTの適用可能性も強く認知され、制度化も進むに到っている。しかし、新しい概念が導入され、普及・発展していく過程において、批判的反応が見られるのは歴史の常である。開発援助分野も例外ではなく、研究者、実務者、受益者を問わず様々な視点から批判的見解が提示され、またそれに対する擁護も行われている。中には的を外した愚論も見受けられるものの、これらの議論を通じてRCTの正確な理解が促され、同時にその限界も認知されてきている。

インパクト評価、特にRCTに関わる議論についてはこれまでに多くの識者が整理を試みている(青柳 2007、和田・青柳 2008、Blattman 2008、Deaton 2009、Easterly 2009、Leeuw and Vaessen 2009、White 2009等)。ここでは従来の開発援助の批判者として名高いWilliam Easterlyの手際のよい整理を基に、追加的な論点を加え、これまでの議論を振り返ることとする。

#### (1) 倫理面への配慮は十分か？

第一は、統制群を得るために、偶然に左右された人をプロジェクトから排除するということが、非倫理的ではないのかという指摘である。この点に関しては、すでにRCT擁護派の見解が広く受け入れられているように思われる。実際のプロジェクト実施時には、資金をはじめとする様々なリソースが限られており、潜在的な受益者全員をカバーできないことが多い。そのような状況下においては、政治的理由やアクセスの容易さといったような恣意的な要因によって受益者を決めるのではなく、むしろランダム割付をする方が公平かつ透明性を持つ手続きであると認められるのである。Easterlyは、この見解に同意しつつも、統制群・実施群間の軋轢が生じる可能性を指摘し、必ず事前の同意を得ることの重要性を説いている。これについてはJ-PALなどは、たとえ統制群に割り当てられる可能性があったとしても透明性が確保されるため、受益者は好意的に受け入れることが多

いという見解を持っており、議論は収束していない。

著者から次の点につき補足しておく。上記の見解はリソースが限られているという前提で議論を行っている。小規模なNGOの活動などでは当てはまることも多いかもしれないが、大規模な政府開発援助となると必ずしもリソースが不足しているという状況が生じるとは限らない。特に医療のような基本的人権に関わる支援については、リソースが潤沢である場合には統制群を構成することを慎重に検討しなくてはならないのは言うまでもない。

ただし、やや自己矛盾をするようであるが、効果が明らかになっていない段階でリソースが潤沢にあるという状況は本来的には誤っている。プロジェクト効果が十分に明らかになっていない場合や、効果的と考えられる介入が複数あるときなど、パイロットプロジェクトで効果を検証するべきであって、エビデンスなく実施が行われるのであれば、その構造の方をまずは問題にするべきである。効果の検証がないまま大規模に介入を行う方が、小規模に統制群を構成するよりも、はるかに社会倫理に反するであろう。

#### (2) 得られた評価結果は一般化できるのか？

第二は、ある特定の状況下で行われたプロジェクトが効果を生むということが分かったとしても、果たして同一の介入が別の状況でも同様に機能するのであろうかという、外的妥当性に関わる論点である。実際には、プロジェクトのデザインのみならず、実施環境、文脈が重要であり、一般化の可能性が小さいという批判がされる。また、似たような環境下へのスケールアップを視野に入れたパイロットプロジェクトであったとしても、小規模であればきめ細やかな対応をすることができるが、大規模になったときに同じような対応ができるのかという疑問が呈されている。後者はプロジェクトの実施過程による失敗 (implementation failure) までは検証できないという批判と言い換えてもよいであろう<sup>7)</sup>。

擁護派は、様々な条件下で多くの事例を積み重ねることで、実際に機能するプロジェクトが証明されてくると述べている。しかしながら、エビデ

ンスとして確立するためにはどれだけの検証を積み重ねていかななくていけないのかは知る術がない。この点に関しては、今後議論が深められていくものと思われる。詳細には立ち入らないが、Imbens (2009) は構造推計との関係でRCTの有用性を説いており、擁護派からの外的妥当性に対する反論として傾聴に値する。

### (3) 理論の検証なきRCT

第三は、プロジェクト効果の有無を検証するだけではなく、なぜプロジェクトが機能したのか(しなかったのか)が明らかにならない限り、有用性は低いという「理論の欠落」を指摘する批判である。RCTはプロジェクトの効果を検証するためではなく、人間行動に関する理論を検証するために用いられる時、その真価を発揮すると主張される。

Blattman (2008) は、理論の必要性には明示的に言及をしていないものの、プロジェクト効果の有無のみを検証し、費用便益 (Return on Investment: ROI) をはじき出すインパクト評価を Impact Evaluation 1.0 (IE1.0) と呼び、一世代前の古い概念という形で整理をしている。そして、これからはヴァージョンアップを果たした Impact Evaluation 2.0 (IE2.0) の世代となるべきことを指摘している。IE2.0においては、例えばインセンティブ構造に関するプロジェクトのプロセスや実施方法をランダムに変えて、RCTによる検証を行うことが提唱されており、Blattmanの立場も従来のRCTはプロジェクトが機能した理由が明らかにならないという点において、理論の欠如を批判する論者と方向性を一にするものである。

### (4) 恣意性を排除できるのか？

第四は、RCTによる結果の恣意性に関わるものである。RCTはバイアスのないインパクトの推計値を得ることができるが、実際にはアウトカム指標を変えたり、サブグループ分析をすることによって、“好ましい”結果を提示することが可能である。これは分析者の倫理観によるものであるが、様々な思惑がある場合には、RCTであっても結果を恣意的に操作するインセンティブから自由ではない。

### (5) RCTは開発課題に対する解決策を提示できるのか？

最後の論点はRCTの狭い適用範囲に起因する論点である。RCTが対象とできるプロジェクトは限られており、マクロ経済政策といった課題を扱う事はできない。しかしながら、現実の開発課題はマイクロなものに限ったものではない。

以上がEasterlyによってまとめられたRCTに関わる主要論点である。これらの論点の多くは、評価研究の文脈で繰り返し問われ続けてきたものばかりであり、ここに古くからある議論の再現を見ることができる。

## 4. RCTにおける理論の役割

前節において、開発援助分野のRCTに関わる主要論点を紹介したが、評価研究の系譜との大きな相違は第三の指摘、理論の検証に関わるものであろう。評価研究においては、RCTを擁護する立場の中には、理論の構築・検証は無意味なものとして退け、インプットとアウトカム (インパクト) の部分のみを押さえ、RCTによってバイアスなき推計値を得るということに価値を見出すという流れが存在する。一方で、ロジックモデルによって理論を描き、RCTによって効果を検証するということも広く行われている。しかしながら、乱暴な表現をすれば、ロジックモデルは常識に沿った因果連鎖 (causal chain) を描いたものにすぎず、背後に人間行動の原理に関する仮定をおいていない。誤解を恐れずに言えば、開発援助分野を支える学術的基盤の中心が経済学であったことは疑いない。こうした背景もあってか、人間行動の原理が組み込まれていないロジックモデルによる因果連鎖の描写は理論とはみなさない傾向がある。この観点から見ると、評価研究自体は理論から一定程度距離を置いてきたとみることができる。

開発経済学者によって主導されてきたRCTを用いた開発援助評価であるが、明示的には述べられていないものの、その初期段階では評価研究に近い立場を取っていたとみることができる。この立場に立つ限り、プロジェクト効果の有無は検証で

きるが、効果に繋がるメカニズムはブラックボックスのままとなり続けてしまうのである。上記の批判は、まさにこのブラックボックスをいつまで見ずにいるのかというものであった。

しかしながら、現在の開発援助分野におけるRCTは新たな展開を見せつつあり、こうした理論なきRCTに真っ向から対立する様相を呈している<sup>8</sup>。経済理論から導出される行動モデルを組み(理論予測を行い)<sup>9</sup>、その検証をRCTによって行うことで、どのようなプロジェクトが機能するのかという問いを超え、なぜ機能するのかという問いに答えようとしているのである。

マイクロファイナンスから具体的事例の一つを紹介しよう。そこでは、マイクロファイナンスへのアクセスが世帯所得を向上させるのか、どのような貸付制度を採用すれば返済率を上げることができるのかといったことが評価の主な関心となるが、後者に関しては、グループ貸付制度が一つの解決策と考えられていることが多い。しかし、グループ貸付を行い、そのパフォーマンスを評価することでは、マイクロファイナンスの中心的課題である逆選択とモラルハザードのどちらが深刻な悪影響を有しているのかというメカニズムにまでアプローチすることはできない。そこでKarlán and Zinmen (2009)は、経済理論から予想される個人の行動を明確に峻別できる形でプロジェクト設計をし、RCTを用いて、両者の相対的な重要性を識別している<sup>10</sup>。その結果、逆選択の影響は相対的に小さいことを示している。

こうした理論と実証の関係は、かつて経済学に見られた抽象的な経済理論への不信、それにもとづく「理論なき計測」の支持に端を発する論争と同一視することができよう。理論なき計測からは、経済政策の指針を引き出す事ができないとして、コウルズ委員会から痛烈な批判がなされたことがあった。今まさにRCTを用いた実証分析という場に舞台を移して、理論の意味が問われているのである。

## 5. 現行のRCTの課題

何が機能するのかという検証から、なぜ機能す

るのかという理論検証へと分析の関心がシフトする一方で、RCT実施上の課題も指摘されている。すなわち、RCTの実施手順が適切に行われていない、或いは報告書等において十分な記載がなされていない事例が散見されるのである。

このような批判は、プリンストン大学の著名な実証経済学者のAngus Deatonによって先鞭が付けられた(Deaton 2009)。Deatonは、具体的事例として開発援助分野におけるRCT適用の草分けの事例として言及されることの多いEdward Miguelらによる虫下し薬の効果に関する論文について、ランダム割付とはみなせない形で介入群と統制群が決められていることを指摘している。この研究は、学校を単位にプロジェクトの実施が割り当てられているが、学校名のアルファベット順に割付が行われていたのである。ここで指摘されるべき点は、割付に際し学校名を用いた事ではなく、その事実を論文の中で正確に記述していないこと、そして適切な分析がなされていないことである。実際、Deaton自身も、完璧なRCTを行う事は通常困難であるとし、学校名の利用自体には批判的ではない。ただし、一旦学校名を割付に用いたならば、操作変数を用いるなどして、適切な統計的処理を施す必要がある事を説いている。

Bruhn and McKenzie (2008)は、小規模サンプルを対象にRCTを用いた開発実証経済学の論文を18篇取り上げ、より包括的にランダム割付の実施手順、及び実施手順の記載のされ方をサーベイしている。さらにRCTを実施した研究者に対して質問票調査を行っており、実際の実施過程について詳細を聞き出している。

表1は彼らの報告書サーベイの結果をまとめたものである。①ランダム割付がどのような環境下(公開か否か)で行われたか、②実施群・統制群間の不釣合を回避するためにどのような手法がとられたか、③両群間の釣合のテストをしているかといった点を確認しているが、非常に雑な記述しかなされていないことが明らかとなっている。②については、どのような手法を用いたかは記述があるものの、例えばブロッキングの場合には、その詳細な手続きが記述されていないといった問題が指摘できる。さらに、Balancing testの結果、再割付を行っている事例も見られるが、どのような

基準で再割付の必要性を判断したのかが明らかになっておらず、恣意性への疑問が残る。また、第3節(1)において、RCTは実施対象者の選別が偶然に左右されるために潜在的受益者にとって受け入れられやすいとの見解を紹介したが、実際にはどのような形で割付がされたか明示的に記載がされていないことが大半を占めており、透明性の確保も確認をする事が困難である。ここでは詳細には触れないが、研究者への質問票調査からも、類似の対応が報告されている。

疫学などのRCTが先行している分野では、報告書の記載事項等が事細かに定められているのが通例である。しかしながら、この点については開発援助分野においては制度化が遅れている。今後こうした批判を受け、スタンダードが規定されていくものと推察される。

## 6. RCTに関わる日本の取り組み

最後に日本の取り組みについて言及をする。残念ながら、これまでに述べてきたRCTの促進・普及過程、論争、実務的課題等につき、日本からの目立った貢献はない。具体的実践例としても、厳

格なインパクト評価と呼べるような評価事例は数例しかなく、RCTを用いたものに到っては現在までも皆無という状況である<sup>11</sup>。これらはRCTに対して否定的な見解を持っているということの意味しているわけではない。そもそも他の手法によるものも含め厳格なインパクト評価を実務の中でどのように位置づけ、活用していくのかという議論自体がなされてこなかった状況の反映である<sup>12</sup>。その意味ではインパクト評価へは、より消極的な姿勢を示してきたと言わざるを得ない。

旧国際協力銀行(旧JBIC)の取り組みが唯一の例外である。旧JBIC評価室、及び開発金融研究所では、国際的動向を積極的に注視し、円借款事業を対象としたインパクト評価の検討・実施を行ってきた。しかしながら、これまでに実施してきたインパクト評価の多くはプロジェクト終了後に事後的にデータ収集を行い、マッチングをはじめとする統計的手法を用いたものであり、RCTについての実施事例はない。円借款事業の中心は、大規模なインフラ整備であるため、そもそも現在のインパクト評価の手法が案件になじまないという面もあり、適用可能な範囲内で統計的手法による評価を試行的に行うという段階に留まらざるを得なかったのである。

表1 RCTの事例要約

Paper	Randomization Unit	Sample Size	Number Treated	Public or Private	Stratification Used	Matched Pairs?	Number of Strata	Strata or pair dummies used?	Table for assessing balance?	# variables used to check balance	Test of significance for balance?
<i>Published/forthcoming Papers</i>											
Ashraf et al. (2006a)	Microfinance clients	1777	710	n.a.	No	No			Yes	12	Yes
Ashraf et al. (2006b)	Barangay (area)	10	5	n.a.	No	Yes		Yes	Yes	12	Yes
Banerjee et al. (2007)	School	98	49/49	n.a.	Yes	No	n.a.	No	Yes	4	Yes
	School	111	55/56	n.a.	Yes	No	n.a.	No	Yes	4	Yes
	School	67	32/35	n.a.	Yes	No	n.a.	No	Yes	4	Yes
Bertrand et al. (2007)	Men wanting a Driver's license	822	268/264	Public	Yes (C)	No	23	Yes	Yes	22	Yes
Bobonis et al. (2006)	Preschool cluster	155	59/51/45	n.a.	No	No			Yes	24	Yes
Field and Pande (2008)	Microfinance group	100	38/30	Public	No	No			No (A)		
Glewwe et al. (2004)	School	178	89	n.a.	Yes	No	n.a.	No	Yes	8	Yes
Miguel and Kremer (2004)	School	75	25*3	n.a.	Yes	No	n.a.	No	Yes	21	Yes
Olken (2007a)	Village	608	202/199	n.a.	Yes	No	156	Yes	Yes	10	Yes
	Subdistrict	156	n.a.	n.a.	Yes	No	50	Yes	Yes	10	Yes
<i>Working Papers</i>											
Ashraf et al. (2007)	Household	1260	6 groups	n.a.	Yes	No	5	Yes	Yes	14	Yes
Björkman and Svensson (2007)	Community	50	25	n.a.	Yes	No	n.a.		Yes	39	Yes
Duflo et al. (2007)	School	113	57	n.a.	Yes	No	n.a.	No (E)	Yes	15	Yes
Dupas (2006)	School	328	71	n.a.	Yes	No	n.a.	No	Yes	17	Yes
Glewwe et al. (2006)	Township	25	12	n.a.	No	Yes			Yes	4	Yes
He et al. (2007)	School division	194	97	n.a.	Yes	No	n.a.	No	Yes	22	Yes
Karlan and Valdivia (2006)	Microfinance group	239	104/84	n.a.	Yes	No	n.a.	No (D)	Yes	14	Yes
Kremer et al. (2006)	Spring	200	50/50/100	n.a.	Yes	No	n.a.	No	Yes	28	Yes
Olken (2007b)	Village	48	17	n.a.	Yes	No	2	Yes	Yes	8	Yes

Note:

n.a. denotes information not available in the paper.

A: Paper says check was done on a number of variables and is available upon request.

C: It appears randomization was done within recruitment session, but the paper was not clear

D: Dummies for location are included, but not for credit officer which was the other stratifying variable.

E: Dummies for district are included, but not for the number of households in the area which were also used for stratifying within district.

(注) 表中に引用されている論文については、出所論文のリファレンスを参照のこと。

(出所) Bruhn and Mckenzie (2008)

しかしながら、2007年に旧JBIC評価室の下、「インパクト評価研究会」が組織され、円借款事業評価を主たる対象に、案件形成時からベースラインデータ収集体制整備、インパクト評価適用可能案件の類型化、職員の能力強化といった課題が議論された。2008年のJICA統合後も評価部において議論は継続され、技術協力プロジェクトも視野に入れ、RCTも含むインパクト評価ガイドラインの作成に向けた取り組みを開始することが合意されている。

評価部においては、他機関がほとんど実施をしていないインフラ分野でのインパクト評価であると共に、NONIEが農業生産性に関わるインパクト評価を年次テーマに掲げたことから、JICAの比較優位を検討し、灌漑案件のインパクト評価に戦略的に取り組む方針を打ち出している。現在この方針にそって、アジア4ヶ国において統計的手法を用いたインパクト評価を実施中である。また灌漑セクターに限らず、RCTの導入も視野に入れた準備調査の取り組みも一部で始まっている。

こうした取り組み分野、評価設問を検討していくに当たっては、インパクト評価の技術的な側面のみならず、オペレーションの視点から評価結果をどう活用するかという点を十分に踏まえることが肝要である。すでに多くの場面で言及されているように、インパクト評価がなじむ分野はそもそも限られているし、また投入できるリソースも無限ではない。戦略的にインパクト評価の適用範囲を検討していくことが求められる。RCTもそこで検討される手法上の選択肢のひとつにすぎない。なぜこの分野・案件でインパクト評価をするのか、どの手法を用いるのか（これは換言すれば、エビデンスにどの程度の質を求めるのかということになる）、いかなる評価設問を立て、何を明らかにしたいのかといったようなインパクト評価結果の活用を十分に検討しなくては、プロジェクトの効果の有無を示し、アカウントビリティを果たすというためだけの評価を産出するだけになってしまう。

開発経済学者が主導するインパクト評価の議論では、内的妥当性、外的妥当性、理論の検証といった面に議論が集中し、開発援助組織としてどう評価結果を活用するのかという視点が抜け落ちる

ことが多い。誤った実施手続きや手法の選択は、誤った評価結果（バイアスのかかった評価結果）に繋がる。したがって、これらの論点が重要なことは疑いないが、その評価結果をどう戦略的に用いていくのかという実務側のニーズをより一層掘り起こしていくことが求められるであろう。実務者と研究者双方のニーズ・関心の接点に立ってRCTの導入が検討される事を期待したい。

## 注記

- 1 本論において示す考えは個人的見解であり、筆者の属する組織を代表するものではない。
- 2 厳格なインパクト評価の文脈とあえて断るのは、開発援助分野においてはDAC5項目を用いた評価基準や、ニュー・パブリック・マネージメント (New Public Management: NPM) の一環として行われるモニタリング・評価等が行われてきており、プログラム評価の中で論じられるようなアウトカム・インパクトの各種評価手法が必ずしも主流を占めてきたわけではないためである。RCTは厳格なインパクト評価のための一つの手法と理解されている。なお、開発援助コミュニティの中では、インパクト評価という用語は複数の定義づけがなされており、共通理解の上に立った議論を困難とする一因となっている。ここでは、counterfactual (反実仮想) を明示的に想定し、ネットインパクトをバイアスを極力除去した形で推計する評価を指して厳格なインパクト評価という用語を用いる。すなわち、プロジェクトとその効果との因果関係、帰属問題を論じているものを対象としている。ただし、表記上の煩雑さを避けるために、文脈から誤解がないときは「厳格な」という表現は省略することとする。
- 3 なお、2007年4月時には登録された評価結果は27カ国、77件であり、この3年弱の間に大幅な増加を見せている。
- 4 PROGRESAは2002年にOportunidadesと名称が変更された。
- 5 J-PALの和文による包括的な解説については佐々木(2006)を参照のこと。また近年の活動展開については本特集号の佐々木論文を参照されたい。
- 6 このような段階的介入のランダム化は、予算制約により潜在的受益者を全てカバーしきれない時に用い

られる事の多い形態である。不透明な理由により介入のタイミングが決められるよりも、ランダムに実施される事で透明性を確保し、かつ倫理的問題も回避する事に成功している。

- 7 EasterlyはRCTの批判としてimplementation failureの検証性を指摘しているが、これはRCTに限った話ではなく、パイロットプロジェクトの評価一般に関わる問題として整理するべきである。
- 8 もちろん経済学者の中でも理論の扱いについて議論が行われている。例えばDeaton (2009)、Heckman and Urzua (2009)、Imbens (2009) らの論争を参照されたい。
- 9 合理的な経済人を仮定する新古典派的モデルは経済学が扱うモデルの一つに過ぎない。限定合理性を仮定した行動経済学理論に基づく行動モデルなど、かつてに比べ経済学が描く人間像は、はるかに豊かなものとなっていることを付記しておく。
- 10 プロジェクトの詳細及びリサーチデザインに関しては原論文を参照されたい。また不破 (2008)、高野 (2008) 等で原論文の手際のよい要約がなされている。
- 11 但し日本の研究者が独自に行った事例はある。以下では対象を政府開発援助 (ODA) 分野に限定する。
- 12 筆者の推測の域を出ないが、日本に開発援助コミュニティにおいてインパクト評価の浸透が進んでいない理由として、二点指摘する事ができるように思われる。第一は、そもそもインパクト評価の概念が十分に、或いは正確に理解されておらず、何に使えるのかという議論にまで到っていないこと、第二は、統制群、比較群を必要とする事の倫理的拒絶感が強いことである。

## 参考文献

- 青柳恵太郎 (2007) 「インパクト評価を巡る国際的動向」、『国際開発における評価の課題と展望』、国際開発高等教育機構、87-153
- 高野久紀 (2008) 「マイクロファイナンスとフィールド実験：逆選択・モラルハザードとグループ貸付の検証」、経済産業研究所「開発援助の経済学」研究会、[http://www.rieti.go.jp/jp/projects/development\\_aid/column\\_08.html](http://www.rieti.go.jp/jp/projects/development_aid/column_08.html) (2009年11月18日アクセス)
- 佐々木亮 (2006) 「ODA分野における『エビデンスに基づく評価』の試み：「貧困アクションラボ」の動向」、『日本評価研究』、6 (1) : 43-45
- 不破信彦 (2008) 「実証開発経済学の分析手法の最近の動向について —計量経済分析における「内生性」問題を中心に—」、『農業経済研究』、79 (4) : 233-247
- 和田義郎・青柳恵太郎 (2008) 「日本のODAへのインパクト評価導入に関する一考察」、湊直信・藤田信子編著『開発援助の評価とその課題』、国際開発高等教育機構、111-136
- Blattman, C. (2008). *Impact Evaluation 2.0*, mimeo.
- Bruhn, M. and McKenzie D. (2008). In Pursuit of Balance: Randomization in Practice in Development Field Experiments, *BREAD Working Paper*, 189.
- Deaton, A. (2009). Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development. *NBER Working Papers*, 14690.
- Easterly, B. (2009). *Development Experiments: Ethical? Feasible? Useful?* <http://aidwatchers.com/2009/07/development-experiments-ethical-feasible-useful/> (2009年11月18日アクセス)
- Heckman, J. and Urzua, S. (2009). Comparing IV With Structural Models: What Simple IV Can and Cannot Identify, *NBER Working Papers*, 14706.
- Imbens, G. W. (2009). Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009). *NBER Working Papers*, 14896.
- Karlan, D. and Zinman, J. (2009). Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment. *Econometrica*, 77 (6), 1993-2008.
- Leeuw, F. and Vaessen, J., (2009). *Impact Evaluation and Development: NONIE Guidance on Impact Evaluation*, Washington D.C., Network of Networks for Impact Evaluation.
- Padian, N. (2009). Behavior change interventions to prevent HIV among low-income girls and women living in low and middle income countries, *3ie Synthetic Reviews*, 008.
- Savedoff, W. D., Levine R., and Birdsall N. (2006). *When will We Ever Learn? Improving Lives through Impact Evaluation*. Washington D.C., Center for Global Development.

Waddington, H., Snilstveit, B., White, H., and Fewtrell, L. (2009). Water, Sanitation and Hygiene Interventions to Combat Childhood Diarrhea in Developing Countries. *3ie Synthetic Reviews*, 001.

White, H. (2009). Some Reflections on Current Debates in Impact Evaluation. *3ie Working Papers*.

(2010.2.2受理)

## **Challenges with the Implementation of RCT in International Development Aid Evaluation – from the RCT without theory to RCT with theory –**

Keitaro Aoyagi

Japan International Cooperation Agency  
Aoyagi.Keitaro@jica.go.jp

### **Abstract**

Rigorous impact evaluation in the field of international development aid has been increasingly implemented since around 2000. In line with the growing importance of rigorous impact evaluation, a number of RCT is conducted. Leading development economists are discussing the pros and cons of implementing RCT, which has been a controversial issue in evaluation studies for a long time. In addition, criticism on “measurement without theory” in the field of economics in the past seems to have been revived. However, compared with other areas such as epidemiology, RCT in development economics needs to be standardized and institutionalized. From Japan’s standpoint rigorous impact evaluation including the implementation of RCT is considered.

### **Keywords**

RCT, Impact Evaluation, Development Economics, Evaluation Studies, International Development Aid

## 【総説】

# エビデンスに基づく開発援助評価 —援助評価の歴史、ランダム化実験の起源、スクリヴェンと バナージェの考え方の比較—

佐々木 亮

財団法人 国際開発センター

sasaki.ryo@idcj.or.jp

## 要 約

貧困アクションラボがリードする「エビデンスに基づく開発援助評価」には、少なくとも3つの起源がある。それらは、ランダム化実験デザインの是非を議論してきた評価研究の系譜、独自の発展を遂げてきた開発援助評価の系譜、そして新しく当該分野をリードし始めた経済学の系譜である。それぞれの歴史的背景および現状を論じたうえで、ランダム化実験デザインの優位性と制約に関するスクリヴェンとバナージェの考え方の比較を行う。結論は、開発援助評価において、独占的というわけにはいかないが、ランダム化実験デザインが利用できるし利用すべき余地が確かに存在するということである。そしてランダム化実験デザインを適用することにより、すべてというわけにはいかないが「機能する援助」が確かに存在することが証明されてきており、今後は、効果が証明された援助活動に対してより多くの資源を投入していくことが求められていくであろう。

## キーワード

エビデンス、国際開発援助、貧困アクションラボ、スクリヴェン、バナージェ

### 1. 問題の所在

開発援助の効果に関して長い間、根本的な疑問が呈されてきた。「援助は機能しているのか？」(Does Aid Work?) という問いがそれであり、これに対してカッセン (Cassen, R. & associates 1986, 1994.) は、多くの場合は機能している一方で別の多くの場合は機能していないと述べている。特筆すべきは貧しい国ほど機能していないということだと結論した。

続いて世銀が作成したいわゆる Wapenhans report (World Bank 1992) では、世銀自身による事後評価をレビューして、世銀ローンを利用した3分の1のプロジェクトは成功していないと結論した。根底には、プロジェクトの事前評価が十分には行われない一方で、組織文化としてローンを積極的に是認しようという風土 (“Loan approval culture”) があると指摘した。

さらにその後、アジア経済危機を受けて行われた「援助を査定する：何が機能し何が機能しない

のか、そしてなぜか」(World Bank 1998)では、途上国政府自身が作る政策体系が出来ていなければ援助の効果は非常に疑わしくなると結論した。これを受けて世銀は、ローン供与条件の名の下に、非援助国の政策体系の構築支援に乗り出した(Policy-based loanの導入がそれにあたる)。

それで結局、開発援助は機能しているのか？これは今でも引き続き聞かれる質問であるが<sup>1</sup>、なぜ我々は今でも同じ質問に答えられないのであろうか。有力な理由のひとつは、そもそも対象国も分野も方法も多様な援助全体に関して、効果があったかどうかを論じることが無理なのではないかという指摘である。それよりも、さまざまな援助事業の中から本当に効果があるものと効果がないものを峻別することにより、限られた援助資金を効果があると証明されたものだけに集中しようという発想の方が健全だと言える。いわゆる「選択と集中」である。その峻別作業に貢献するために、2003年に設立されたのがマサチューセッツ工科大学(MIT)に「貧困アクションラボ」(Abdul Lateef Jameel Poverty Action Lab.通称J-PAL)である。数人のエスタブリッシュされたエコノミストが設立した貧困アクションラボは、ランダム化実験デザイン(Randomized experimental design)<sup>2</sup>を用いた厳格な援助評価を通じて、援助資金負担者(政府、国際機関、NGOなど)がその限られた資源を本当に効果があると証明された援助事業に振り向けることに貢献してきた。今まさにJ-PALが、「援助は機能しているか」という長年の問いに答えを出そうとしていると言える。その答えとは、「機能する援助は確かに存在する」ということであり、厳格な手法を用いて確かに援助が機能しているというエビデンスを特定することにより、機能する援助に資源を振り向けていくことを目指している(Duflo, E. 2009)。

## 2. 「エビデンスに基づく開発援助評価」の3つの起源

貧困削減ラボがリードする「エビデンスに基づく開発援助評価」には、少なくとも3つの起源があるように思われる。それらは、ランダム化実験

デザインの是非を長年にわたって議論してきた評価研究の系譜、開発援助という分野のなかで独自の発展を遂げていた開発援助評価の系譜、そして新しく当該分野をリードし始めた経済学の系譜である。それぞれは別々に発展しながら、ときおり深く交流して影響し合ってきた形跡が観察できる。本論文ではそれぞれの発展および交流の影響についてレビューして議論していく。

## 3. 評価研究の系譜：ランダム化実験デザインの是非に関する議論

そもそもランダム化実験デザインは、1920年代に、イギリスのロザムステッド農業試験場(世界でもっとも古い農業試験場と看做されている)に採用された農業研究者のフィッシャーによって確立された<sup>3</sup>。肥料の効果を、土地の肥沃度合いに影響されずに純粋に測定するためにこのデザインを考案したわけで、実験デザインの始まりは農業分野だったわけである。フィッシャーは、『リサーチワーカーのための統計的手法』(Fisher, R.A., (1925). *Statistical Methods for Research Workers*)と「実験のデザイン」(Fisher, R.A., (1935). *The Design of Experiments*, 1935)を執筆し、農業分野ではその有用性が広く認められて普及し、標準的なりサーチ・デザインとして認知されていった。

その後、ウィリアム・G・コクランとガートルド・M・コックスによって『実験デザイン』(Cochran, W.G. & Cox, M.G. (1950). *Experimental Designs*)が執筆され、統計学者のリファレンスとして長く参照された。

また、保健・医療分野では1950年代前後から、ランダム化実験デザインが紹介されて適用されており、1972年にアーチボルト・L・コクランが執筆した『効果と効率：ヘルスサービスにおける無作為反応』(Cochrane A.L. (1972). *Effectiveness & Efficiency: Random Reflections on Health Services*)によって、ランダム化実験デザインを用いた臨床試験結果の体系的な収集と分析結果の普及が提案された。なおこの時期にイギリス国内では犯罪学の分野で広くランダム化実験デザインが用いられていたという記載があり(Farrington 2003, p.164)、

ランダム化実験デザインの普及はアメリカ国内よりも早かった可能性がある。

続いて教育分野では、1960年代にキャンベルとスタンレーが『リサーチのための実験と準実験』(Campbell, D.T. & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*) を執筆し、これをきっかけにランダム化実験デザインが教育分野に広く普及していった。同書は教育分野のみならず社会科学分野一般に広く普及し、現在に至るまで、同書が社会科学分野におけるリサーチ・デザインの標準的テキストとして今でも用いられている。

その後、1980年にロッシ、フリーマン、ライトが『評価：体系的アプローチ』(Rossi, P.H., Freeman, H.E. & Wright, S.R. (1979). *Evaluation: A Systematic Approach*) を出版し、この中でランダム化実験デザインと準実験デザインが詳細に解説されており、同書がプログラム評価の基本概念や手法に関する標準的なテキストとなったのはよく知られているとおりである。

1980年代になると、インパクト評価の事実上のスタンダードとみなされたランダム化実験デザインに対する批判が始まった。パットン (Patton, M) が1980年代に行った一連の研究で、ランダム化実験デザインを用いて行われた連邦政府レベルの施策の評価結果がほとんど意思決定に利用されなかったと結論したことが引き金になったと言われている。その反省からパットンは『実用重視の事業評価入門』(Patton, M. (1978). *Utilization-focused evaluation*) という著作を書いていくことになった(長尾が2001年に翻訳している)。

そして1980年代中期に、アメリカの社会科学全体に非常に大きな影響を残すことになったパラダイム論争が起こった。評価研究以外でも名前を広範に知られることとなったグーバとリンカーンが執筆した『自然的研究』(Guba, E. & Lincoln, Y. (1985). *Naturalistic Inquiry*, 1985) と『第四世代評価』(Guba, E. & Lincoln, Y. (1989). *Fourth Generation Evaluation*) の中で、われわれはランダム化実験デザインを使用することを明確に拒否すると述べて、代わりに各種の定性的手法の有用性を指摘した。なお、インタビューや集団討論などを用いることにより、会話を通じて真実を明らか

にしていこうという彼らの立場は「構築主義」(Constructivism) として広く知られている。また、構築主義は、住民参加を重視する参加型開発の考え方と近似性を持つとも言える。

そして、ランダム化実験デザインに代表される定量的アプローチと、インタビューやケーススタディに代表される定性的アプローチの論争は、1990年代を通じて激しさを増していった。しかし、この論争は1990年代の終わりにはある程度の決着に至ったことが観察されている。それは、定性的手法も定量的手法も長所と短所を有しているのだから、ひとつの評価の中で両方の手法を使うべきという合意であり、これを「混合手法」(“Mixed methods”) と呼ぶ (Patton, M. 1990; Baker, J.L. 2000)。評価実務者の間で「混合手法」が幅広く受け入れられたことによって、定性・定量を巡る議論は収束したかに見えた。

#### 4. 開発援助評価の系譜：開発援助という特殊領域において独自の発展を遂げた評価手法

次に援助評価の歴史を概観してみる。開発援助は国内の各行政分野と違った特徴を有しており(たとえば、資金負担者(納税者)と受益者が物理的に遠く離れていて情報交流がないなど)、その特徴を反映して評価手法も独自の発展を見せた部分がかかなりあると言える。

1959年にUNESCOの依頼に基づいてハイエスが「開発プロジェクトを評価する」(Hayes, S.P. (1959). *Evaluating Development Project*) というモノグラフを執筆して広く普及したが、これにより初めて援助分野において「評価」(Evaluation) が、経済分析 (Economic analysis) から独立した活動として認識されることになったと言える。その中でハイエスは、事前事後比較デザインやウィズ・ウィズアウト (With-without) デザインなどの基本的なインパクト評価のデザインを紹介しているが、ランダム化実験デザインについては言及がない。

1960年代に入ると、前述のとおり、キャンベルとスタンレーが社会科学分野にランダム化実験デ

ザインを本格的に紹介したが、それが援助分野ではどのように受け取られたかは明確な資料はない。おそらく、キャンベルとスタンレーの著書自体はほとんど省みられなかったはずである。

しかし、1970年代に入ってUNESCOの要請に基づいて、ロッシ、ライト、フリーマンが「途上国における社会プログラムを評価する」(Rossi, P.H., Wight, S.R., & Freeman, H.E. (1979). *Evaluating Social Programs in Developing Countries*) を出版し、同書は広く普及した<sup>5</sup>。そして、その本ではキャンベルとスタンレーの伝統を受け継いで、ランダム化実験デザインと準実験デザインを詳細に解説しており、同書によって開発援助分野にランダム化実験デザインが本格的に紹介されたと言える。ただし、当時はその考え方や手法が広く知られたわりには、ランダム化実験デザインが開発援助の現場で実際に適用されることはきわめて稀な手法であったようである<sup>6</sup>。なお、先にUNESCOから依頼された本を執筆したあとに、それをアメリカ国内でも出版すべきではないかと考えて、3. で解説した「評価：体系的アプローチ」というタイトルで出版したと同書のはしがきで述べている(ただし2冊は双子ではないとも断っている)。

1980年代には、前述したように定性的な評価手法や簡便なデータ収集手法が提案されて普及していった。これを受けて開発援助の分野でも簡便な手法が普及したし(USAID 1997)、さらに構築主義の発想に近い参加型評価手法が普及していった(Chamber, R. & Associates 1989)。

1990年代になると開発援助評価は二つのグループに分かれていった(Sasaki, R. 2006)。ひとつはOECD-DAC(1991)が提案したDAC評価基準(DAC evaluation criteria)を受け入れたグループであり、北欧諸国、ドイツ、日本、世銀などもこれに属する。もうひとつはいわゆるパフォーマンス・メジャーメント(業績測定)(Osborne, P. & Gaebler, T. 1993; Hatry, P.H., 1999)のグループで、定量的指標と数値目標を用いた民間セクターに由来する目標管理を基本的な発想としており、アメリカ国際開発庁、カナダ国際開発庁、UNDPなど一部の国際機関が属する。前者は、ランダム化実験デザインや準実験デザインの利用によって特定されるインパクトだけではない複数の視点を導入

した点でより洗練された評価枠組を採用したと言えるが、一方でインパクトがなくても全体評価としては「成功」とされる可能性を高めたと言える。一方で、後者はランダム化実験デザインの適用の困難性を回避して、非常に単純な手法に回帰したわけであるが、そのために援助行為以外の影響をもインパクトに誤って含むリスクを内包することになった(佐々木 2003)。ただし両者は排他的というわけではなく両方を併用する援助機関もある(Sasaki 2008)。

つまり、1990年代末から2000年代冒頭に現れた状況とは、かつてないほど洗練された開発援助評価の手法と、究極まで単純化して簡便さを優先させた手法に分裂した状況だったと言える。そして、いずれにせよ、冒頭で解説したような「援助は機能しているのか」という単純な質問に対して直接的に答えることが難しくなった状況であったと言える。佐々木(2008)は、2004年度に全世界で発行されて入手可能だった1034冊の援助評価報告書からランダムサンプリングによって選択した102冊をメタ評価したが、その結果は、世界の援助評価報告書の質は「優秀」からは程遠いと言わざるを得ず、さまざまな点で改善の余地があるという結論であった。

## 5. 経済学の系譜：ランダム化実験デザインの経済学への導入

1990年代末に「混合手法」の普及によって安定したかに見えた評価研究の状況に転機が訪れた。また独自の発展を享受してきた開発援助評価にとっても同じ転機が訪れた。その転機とは経済学分野からの発信であり、評価研究コミュニティや開発援助コミュニティにとってはまったく予想外のことであった。

経済学が用いる主な分析道具は重回帰分析などの統計分析手法(別名は計量経済学)であるが、1990年代を通じてこの主要な分析道具に関する信頼を自問自答するようになっていた。たとえば経済学博士のラリー・オールは次のように述べている。「標準的な計量経済のテクニックを使って、自分が出したいと思うあらゆる結果を出せた。い

わば『フィッシング』(釣り)であり、(計量経済モデルで得られた)どの式がどの式よりも優れているとは誰も言うことができない。』(Orr, L.L. (1999). *Social Experiments*. p.xi)。ただし、これはコンピュータの飛躍的な能力向上と統計分析ソフトのインターフェースの改善によって引き起こされた当然の結果であったとすることができる。

こうした危機的状況の中で経済学者は解決策を模索した。解決策のひとつとして提案されたのは、いわゆる実験経済学 (Experimental Economics) である。被験者を集め、新たに導入しようとしている経済制度に沿った環境を実験室の中で再現し、経済的動機付けを導入する経済実験がなされるようになったのである (西條 1997)。そして、経済学者のカーネマン (Kahneman, S.) とスミス (Smith, V.L.) が、「実験経済学という新研究分野の開拓」という理由により、2002年にノーベル経済学賞を受賞した。経済学分野で、実験的な試行の重要性が認められたわけである。ただし実験経済学は、ランダム化実験デザインが必須というわけではなく、理想的な環境を人工的に作り出して、その中で想定された経済的インセンティブが有効かどうかを試行することを定義としている。

同時期に、もうひとつの解決策として経済学者から提案されたのはランダム化実験デザインの導入であった。マサチューセッツ工科大学とハーバード大学の経済学者が、国際援助の分野でランダム化実験デザインを適用する専門機関である「貧困アクションラボ」(Poverty Action Lab.) を設立して、ランダム化実験デザインを用いた調査報告書を大量に産出するようになったのである (佐々木 2006)。同ラボに所属する経済学者によると、事後的な回帰分析などの計量経済学的方法では、本来は効果がないにも関わらず各種のバイアスによって、効果があったという評価結果が出されることがしばしばあることを指摘し (Kremer 2005, p.10)、ランダム化実験デザインこそが本当に効果があったのかを明らかにする単純明快な方法だと結論している。そして、ランダム化実験デザインを用いた評価には、21世紀の社会政策に革命を起こす可能性があるとしている (Duflo, E., & Kremer, M. 2003)。青柳 (2010) が指摘するように、こうした貧困アクションラボの主

張は、経済学分野において急速に支持を集めている主張であると言える。

経済学分野において注目を浴びたランダム化実験デザインに関する議論は、国際援助分野へ波及し、そして今度は評価研究の分野へ輸入され始めた。すでに一通りの議論を済ませたと考えていた評価研究のコミュニティにおいて、ランダム化実験デザインの利用に信頼を寄せる研究者グループが息を吹き返し、再びランダム化実験デザインの有用性を主張し始めた。この状況を受けて、ランダム化実験デザインを巡る論争 (今回は名前を変えて「エビデンスに基づく実践」(Evidence-based practices) と呼称されている) が再び活発化している。

## 6. 貧困アクションラボの発展

すでに述べたように、貧困アクションラボは、ランダム化実験デザインを用いた評価を通じて途上国の貧困削減に貢献することを目的として、2003年にアメリカのマサチューセッツ工科大学に設立された。サウジアラビアにおけるトヨタ車販売で財を成したAbdul Latif Jameelの寄付を元に設立されたため、正式名称をAbdul Latif Jameel Poverty Action Lab (略してJ-PAL) と言う。設立の経緯や活動内容については、佐々木 (2006) で詳細に解説されているので参照されたい。またなぜランダム化実験デザインという単一のデザインのみを採用しなければならないかの理由については同ラボのデュフロ、クレマーらの論文で解説されている (Kremer, M. 2005; Duflo, E. & Kremer, M. 2003)。

同ラボは、2003年の設立以来、世銀などからの委託を受けて、世界各地でランダム化実験デザインを用いた108件の評価を実施したり現在実施中である。対象国は、ケニア、ウガンダ、インド、パキスタン、フィリピン、インドネシア、コロンビア、エルサルバドルな世界の30カ国以上に広がっている。分野としては、保健、教育、マイクロファイナンス、地方分権化、女性の政治参加、腐敗対策、地方開発など多岐に渡る。そして保健や教育など以前からランダム化実験が適用されてき

た以外の分野で適用結果が見られることが特筆に値する。また、通常は適用が困難と考えられるマイクロファイナンスや女性の政治参加にランダム化（無作為割当）を用いることができることを実証してきたことになる。

また、同ラボは、設立以来、着実に組織を拡大している。2007年6月にはインドの研究機関内に、南アジア支部（J-PAL South Asia）を設立した。続いて2007年12月には、パリ経済大学に、ヨーロッパ支部（J-PAL Europe）を設立した。さらに2009年10月には、チリのカトリカ大学にラテンアメリカ支部（J-PAL Latin America）を設立した。いずれも、それぞれの地域内における貧困削減プログラムおよび社会プログラムの効果の有無を、ランダム化実験によって明らかにすることにより、政府やNGOの政策改善に資することを目的としている。なお、日本を含む東アジアおよびアフリカではまだ地域支部が設立されていない。

2009年10月の国連総会において、いわゆるミレニウム開発目標の達成に関して議論するセッションが持たれ、貧困アクションラボのデュフロ教授が、蓄積された知見に基づいて「7つのベストバイ（お買い得）」（Duflo, E. (2009). *Best Buys for MDGs*）と題する発表を行った。世界各地のランダム化実験によって確かに効果があると証明された7つの介入行為に関する発表である<sup>7</sup>。また、アメリカ国際開発庁に対しても、開発事業の効果の改善のためのアドバイスを行っている。以上に見たように同ラボは順調に規模を拡大し、2009年で設立6周年を迎えた。

## 7. ランダム化実験デザインの是非を巡る考察：スクリヴェンとバナージェの考え方の比較

次に、評価研究者と経済学者を代表する論者の意見を考察する。

評価研究の第一人者（Shadish, W. et al, 1991）であり評価研究における唯一の哲学者（Stake, R.E. 1982）と呼ばれることもあるスクリヴェンは、因果関係を証明するとはどういうことかに関して哲学的な研究を重ねてきた（たとえば

Scriven, M. 1975など）。そしてスクリヴェンは、ランダム化実験デザインが因果関係を証明するための最良のデザインであるという主張を一貫して鋭く批判してきた。たとえば、アメリカ教育省の「一人の子供も落ちこぼれにしない法」（No Child Left Behind Act of 2001）においてランダム化実験デザインがもっとも望まれる調査デザインであると明文化されたことに端を発して、それに対する賛否によりアメリカ評価学会が二分された際には、スクリヴェンはランダム化実験デザインを批判する勢力の急先鋒として論陣を張った（詳細は、Davidson, S.I. & Christie, C.A. (2004). *The 2004 Claremont Debate*を参照）。スクリヴェンが展開してきた批判を体系的に整理した論文として「因果探索を巡る論理」（Scriven, M. (2007). *The logic of Causal Investigation*）がある。その中でスクリヴェンは、教育や保健を含む社会施策の評価において、無条件にランダム化実験デザインが最良のデザインであると主張するのは論理的な誤りがあると指摘している。

一方で、経済学者であり貧困アクションラボの所長であるバナージェ（Banerjee）は、自身の著作（2007）において、ランダム化実験デザインでなければ本当に効果があるかどうかを判断することはできないと明言している。そして、開発援助分野においてランダム化実験デザインを普及させるために同ラボの活動を開始したと述べている。

スクリヴェンとバナージェは専門分野が違うこともあり、二人の間で直接的な論争が交わされた形跡はない。しかし、著者（佐々木）が2006年に同ラボを訪問してバナージェと議論する機会があったので<sup>8</sup>、著者のメンターであったスクリヴェンの主張を直接質した（Sasaki, R. (2006). *Discussion with MIT's Poverty Action Lab.*）。以下に、スクリヴェンの批判の概要と、それに対するバナージェの反論を整理した。さらに、開発援助評価を長年にわたって研究対象としてきた著者（佐々木）の考察をそれぞれに記載した。つまり、エビデンスに基づく開発援助評価の3つのルーツ（評価研究の系譜、開発援助の系譜、経済学の系譜）に対応する3人の意見を整理することにより、議論の結論を得ることを目指した。この整理作業を通じて分かることは、それぞれの主張にはじつはそれほ

どの違いはないということである。

### (1) 貧困アクションラボのランダム化実験デザインは盲検法が欠落している

スクリヴェンは、保健医療分野で用いられているランダム化実験デザインは二重盲検法 (Double blind design) が適用されているが同ラボのデザインにはそれが欠落していると指摘する。つまり「無盲検」(“Zero” blind) であり、いわゆるホーンソン効果が入り込む余地があるので、純粹に介入の効果进行明らかにできるとは言えないと指摘する (Scriven M. 2007)。

これに対してバナージェは次のように答えた。その指摘はそのとおりだが、保健医療の実験 (Clinical trial) ではないのだから、真薬と偽薬 (Placebo) を使うことはできない。かわりに実験グループと統制グループが同一の情報を共有することにより、無盲検である影響を最小化しようとしていると反論している。

結論としてバナージェは自身が用いているデザインの限界を認めたことになると言えるが、社会科学分野にランダム化実験デザインを導入したキャンベルでさえも盲検法の適用を論じていないわけであり、バナージェの言うように、社会施策を対象とする限り「仕方がない」(No other way) と言わざるを得ない。

### (2) 統計的有意と社会的有意は違う

スクリヴェンは、二群の差が統計的に有意なだけでは不十分であり、社会的あるいは実践的に有意 (Socially or practically significant) でなければ介入は効果があったとは言えないが (Scriven M. 2007)、貧困アクションラボは、統計的有意を持って介入は効果があると判断していると指摘する。

これに対してバナージェは、そのアイデアは拒否しないが、統計的に有意でなければ社会的に有意であることもありえないと反論している。つまり統計的有意は、社会的に有意かどうかを検討するために最低限満たされるべき条件である。また同ラボは、地元の人たちの意見を聞いて社会的に意味があるかを判断していると反論しているが、それでは地元の人たちが成功・失敗を決定す

るのかという問いに対しては、彼らは「相談される」(Consulted) と答えている。

つまり効果があるかどうかの判断は、まずは専門的な統計分析によってなされるべきでそれをしてないで地元の人たちの判断に委ねることは、ランダム化実験デザインの利点である厳格さを損ねると反論していると理解できる。

### (3) 人々を偶然によって二分することは倫理的問題があるし、その処置に関して事前承認を得ることは困難である

スクリヴェンは、途上国に住む親は、自分の子供が統制グループに入るかも知れない処置を承認しないだろうとして、反倫理の問題および事前承認 (Informed consent) の取得の困難さをランダム化実験デザインの避けがたい制約として指摘している (Scriven M. 2007)。

これに対してバナージェは、現場の経験から言えることは、じつはランダム化こそがフェアなのだとは反論する。第一に、もともと援助資源は全員をカバーできるほど用意されていることは稀であり、恣意的に適用者を決めることを避けてランダム化 (同じ確率に基づく宝くじのアイデア) を適用することはたいへんフェアなのだとは指摘する。第二に、今まではランダム化の代わりに「われわれドナーの基準によると」と説明して、幹線道路沿いの村や一日で視察できる村が選定されることが多く、住民としては非常に不平等でアンフェアだと思っていたと聞かされることがある。そうした状況の中でランダム化のアイデアを打診すると歓迎されることが多いと指摘している。

結論として、援助機関側が想定するほどランダム化への抵抗は少なく、ランダム化実験デザインの適用に対する制約にはならないという指摘は説得力がある。実際のところ、援助に携わる者は著者を含めて皆経験しているように、援助側の都合で対象地域や対象者を選んできたわけであり新鮮な指摘である。

### (4) 「エビデンス」という単語が、定量手法を用いる研究者に独占されている

スクリヴェンは、「エビデンスに基づく実践」という考え方は完全に受け入れられる考え方だ

が、エビデンスの定義がランダム化実験デザインの適用によって得られた結果のみに限定されていることが問題だと指摘する (Scriven M. 2007)。そして、定性的手法も厳格に適用することによってエビデンスを産出することができるとして、エビデンスという言葉の定義の再検討を要求している。

これに対してバナージェは、第一に、どのような分析手法も否定する意図はないとする。単純に、ランダム化実験デザインがその有用性にも関わらず他の手法に比べて非常にわずかにしか用いられて来なかったことから、現状よりも頻繁に用いられるべきだと主張しているだけだとしている。第二に、定性的手法、特に詳細な観察記述 (Rich description (Stake, 1982) は「なぜそれが起こったか」を説明するので補完的な役割を演じることができる」と指摘している。さらに言えば、ランダム化実験デザインは限られた情報しか提供しないから、詳細な観察記述は、説明の中心的な方法 (Center piece of explanation) となっていると指摘している。

スクリヴェンも、クック (Cook, T. 2000) の論文を引用して、定性的手法と定量的手法は相互補完的に利用できるし利用すべきだと結論しており、スクリヴェンとバナージェの間にじつは大きな認識の差はないと言える。

#### (5) ランダム化実験デザインの適用が適切ではない (あるいは意味がない) 介入行為のタイプが存在する

スクリヴェンは、ランダム化実験デザインが時間と資源の制約によりランダム化実験デザインが適用できないタイプの介入があることを指摘することにより、常にランダム化実験デザインが最良であるとは言えないと指摘する (Scriven M. 2007)。

これに対してバナージェは、ランダム化実験デザインが常に最良と言うつもりはなく、今までの経験からランダム化実験デザインの適用が不可能だったり不適切だったりするタイプの介入があったことに同意するとしている。それらは、(i) すでに終わった事業 (あるいはすでに開始されている事業)、(ii) 全国を対象とするような大規模事業 (ランダム化実験デザインの名が示すとおり実験

的な小規模の事業に向く) である。さらに、(iii) 事前に内容が確定しておらず実施しながら決定していくフレキシブルな事業は、ランダム化実験デザインに向かないとしている。そして開発援助にはそうした事業が予想外に多いと指摘する。最後に、(iv) 介入効果が強力に出ることが分かっている事業があえてパイロット事業に選定されていることがあり、これもランダム化実験デザインの趣旨に沿わないと指摘している。

貧困削減ラボの公開資料での宣言と比べて、バナージェは実際には意外と冷めていると言わざるを得ない。ランダム化実験デザインの適用経験を重ねる中で、ランダム化実験デザインの適用が困難な特定の介入タイプが理解されてきたのだと言える。

ここまでの議論から分かることは、両者の主張にそれほど違いがあるようには見えないということである。そして、両者が述べる結論はますます違いがなくなる。スクリヴェンが論文の結論部分において、もともとランダム化実験デザインは定性的手法の助けを借りて運用される混合手法であると言えるからランダム化実験デザインが用いられるべき場合は確かにあると述べている一方で (Scriven M. 2007)、バナージェはランダム化実験デザインによって評価活動が独占されるべきだと考えているのではなく、他の手法に比べてあまりに用いられてこなかったのもう少し頻繁に用いられるべきだと主張しているだけだと結論している。つまり、開発援助評価において、独占的というわけにはいかないが、ランダム化実験デザインが利用できるし利用すべき余地が確かに存在すると言うことである。

## 8. 結論および日本への示唆

本論文では、エビデンスに基づく開発援助評価の出現に至る3つの系譜を論じたうえで、その是非について検討した。結論は、開発援助評価において、独占的というわけにはいかないが、ランダム化実験デザインが利用できるし利用すべき余地が確かに存在すると言うことであった。そしてランダム化実験デザインを適用することにより、す

べてというわけにはいかないが「機能する援助は確かに存在する」ことが証明されてきており、今後は、エビデンスによって効果が証明された援助活動に対してより多くの資源を投入していくことが求められていくであろう。

なお、2006年の日本評価研究の特集号（エビデンスに基づく評価の動向）の論文（佐々木 2006）のなかで、貧困アクションラボの日本支部を設立して本家を追い越すほどの活動をすべきという提案をしたが、日本のいずれの大学や研究機関においてもそうした具体的な動きはまだ見られない。その後の3年間の間に、ラテンアメリカ支部、ヨーロッパ支部、南アジア支部が設立され、さらに複数の途上国で同ラボに継続的な協力を約束する組織が現れている。日本はこうした世界的な潮流に呼応することに成功していないと言わざるを得ず、今後のアクションが望まれるところである。ただし、2008年度から2009年度にかけて、国際協力銀行（JBIC、現JICA）において「インパクト評価研究会」が組織され、貧困アクションラボの活動と同様の活動が日本でも可能かどうか真剣に検討されたことは言及されねばならない。研究会の結論として、ランダム化実験デザインを用いた評価を実施するためのマニュアルや工程表の作成が続けられることになっており、日本におけるエビデンスに基づく開発援助評価の展開に関して今後の進展が期待される場所である。

## 注記

- 1 たとえば、日本では2009年の政権交代後のいわゆる「事業仕分」の対象として、政府開発援助に関連する各費目が組上に上げられている。
- 2 Randomized Controlled Trial (RCT)とも言う。
- 3 フィッシャーが確立する以前に無作為割当を適用した実験がすでに行われていた可能性については、フォーセルンド (Forselund, L., Chalmers, I., & Bjorndal, A. (2007)) を参照のこと。
- 4 今でもインターネットから入手可能である。
- 5 ただし、同時にアメリカ国内向けに出版された Evaluation: A Systematic Approach(1st ed.)(1979)の方がよく読まれたようである。
- 6 ただし実際の適用事例は、1970年代初期にUSAID

支援の南米の案件ですでに確認されている (Rawlings, 2003、佐々木、2006)。

- 7 「7つのベストバイ」とされた介入は次の7つ。(i) 学校における回虫駆除薬配布 (健康指標だけでなく教育指標も改善するし費用が格段に安い)、(ii) 基礎教育への支援 (どのレベルの教育よりも費用対効果が高い)、(iii) 蚊帳床への投資 (保健効果が高い)、(iv) HIVに関する教育 (低費用で社会便益が大きい)、(v) 代議士選挙における女性枠の設定 (ほぼ無費用でより女性のニーズを反映させられるようになる)、(vi) 予防注射 (低費用で幼児死亡率の低減を実現)、(vii) かしこい補助金支給 (農業技術の向上への適切な補助金が貧困を削減する)。
- 8 (財) 国際開発高等教育機構の調査により派遣された際のインタビューによる。インタビュー実施者は、佐々木亮 ((財) 国際開発センター)、青柳恵太郎 ((財) 国際開発高等教育機構) (当時) である。なお、(財) 国際開発高等教育機構の湊・国際開発研究センター所長が調査を指揮した。

## 参考文献

- 青柳恵太郎 (2010) 『開発援助分野におけるRCT導入に見られる古くて新しい課題—理論なきRCTから理論検証のRCTへ—』、日本評価研究、10 (1)
- 西條辰義 (1997) 『新たな経済学の構築に向けて：実験経済学アプローチ』大阪大学 <http://www.iser.osaka-u.ac.jp/~saijo/lec/micro/97/expecon.pdf>
- 佐々木亮 (2002) 『政策評価トレーニングブック』、多賀出版
- 佐々木亮 (2006) 『ODA分野における「エビデンスに基づく評価」の試み：貧困アクションラボの動向』、日本評価研究、6 (1) : 43-54
- 佐々木亮 (2009) 「評価に関する3つの根本的な論争—内的妥当性、外的妥当性、評価論理に関する論争と日本の評価研究への示唆—」、『評価クォーターリー』10、行政管理研究センター
- 龍慶昭・佐々木亮 (2000) 『政策評価の理論と技法 (増補改訂版)』、多賀出版
- Baker, J. L. (2000). *Evaluating the impact of development projects on poverty: A handbook for practitioners*. Washington, DC: The World Bank.
- Banerjee, A. and Duflo, E. (2008). *The Experimental*

- Approach to Development Economics*. <http://econ-www.mit.edu/files/3159>
- Banerjee, A. (2007). *Making Aid Work* (*Boston Review Books*). Cambridge, MA: The MIT Press.
- Campbell, D.T. and Stanely, J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Cassen, R., and Associates (1986). *Does aid work?* (1st ed.). Oxford: Oxford University Press.
- Cassen, R. and Associates (1994). *Does aid work?* (2nd ed.). Oxford: Oxford University Press.
- Chamber, R., Pacey, A., and Thrupp, L. A. (Edit) (1989). *Farmer first: Farmer innovation and agricultural research*. London, UK: Intermediate Technology Development Group Publishing.
- Cochran, W.G. and Cox, G.M. (1950). *Experimental Designs*. NY: John Wiley & Sons, Inc.
- Cochrane, A.L. (1972). *Effectiveness and Efficiency: Random Reflections on Health Services*. London, UK: Royal Society of Medicine Press Limited.
- Donaldson, S.I. and Christie, C.A. (2004). The 2004 Claremont Debate: Lipsey vs. Scriven : Determining Causality in Program Evaluation & Applied Research: Should Experimental Evidence Be the Gold Standard?, *the Journal of MultiDisciplinary Evaluation* (No. 3 Oct. 05) . The Evaluation Center, Western Michigan University.
- Duflo, E. (2009). *Best buys for the MGDs*. Presentation paper for the U.N. General Assembly. <http://www.povertyactionlab.org/MDG/Duflo%20-%20Best%20buys%20to%20reach%20the%20MDGs.pdf>
- Duflo, E. and Kremer, M. (2003). *Use of Randomization in the Evaluation of Development Effectiveness*. Paper prepared for the World Bank Operations Evaluation Department (OED) Conference on Evaluation and Development Effectiveness in Washington, D.C. 15-16. <http://econ-www.mit.edu/files/765>
- Farrington, D.P. (2003). British randomized experiments on crime and justice. *The Annals of the American Academy of Political and Social Science*, 589, 150-167.
- Fisher, R.A. (1925). *Statistical Methods for Research Workers*. NY: Hafner Publishing Company Inc.
- Fisher, R.A. (1935). *The Design of Experiments*. NY: Hafner Publishing Company Inc.
- Forselund, L., Chalmers, I., and Bjorndal, A. (2007). When was random allocation first used to generate comparison groups in experiments to assess the effects of social interventions? In *Economics of Innovation and New Technology*, 2007, 16 (5), 371-384.
- Guba, E.G. and Lincoln, S.L. (1989). *Fourth Generation Evaluation*, Thousand Oaks, CA: Sage Publication
- Guba, E. and Lincoln, Y. (1985). *Naturalistic Inquiry*, Thousand Oaks, CA: Sage Publication.
- Hayes, S. P. (1959). *Evaluating development projects*. Belgium: UNESCO.
- Hatry, P. H. (1999). *Performance measurement: Getting results*. Washington, DC: Urban Institute.
- Hjertholm, P., and White, H. (2000). Foreign aid in historical perspective: Background and trends. In F. Tarp (Ed.). *Foreign aid and development*. 80-102. New York: Routledge.
- Kremer, K. (2005). *Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons*. Massachusetts, MA: Poverty Action Lab. [http://www.economics.harvard.edu/faculty/kremer/files/Randomized\\_Evaluations.pdf](http://www.economics.harvard.edu/faculty/kremer/files/Randomized_Evaluations.pdf)
- OECD-DAC. (1991) . *DAC criteria for evaluating development assistance*. Paris: OECD.
- Orr, L.L. (1999). *Social Experiments: Evaluating Public Programs With Experimental Methods*, Thousand Oaks, CA: Sage Publications.
- Osborne, D., & Gaebler, T. (1993) . *Reinventing government: How the entrepreneurial spirit is transforming the public sector*. New York, NY: Plume.
- Patton, M.Q. (1978). *Qualitative Evaluation and Research Methods, 2nd edition*. Thousand Oaks, CA: Sage Publications.
- Patton, M.Q. (1990). *Qualitative Evaluation and Research Methods, 2nd edition*. Thousand Oaks, CA: Sage Publications.
- Rossi, P.H., Lipsey, M.W. and Freeman, H.E. (1979). *Evaluation: A Systematic Approach* 1st ed. Thousand Oaks, CA: Sage Publications.
- Sasaki, R. (2006). *Discussion with MIT's Poverty Action Lab*. Tokyo: FASID [http://www.fasid.or.jp/chosa/oda/pdf/apply\\_6\\_2.pdf](http://www.fasid.or.jp/chosa/oda/pdf/apply_6_2.pdf)

- Sasaki, R. (2008). *Metaevaluation by Formal Evaluation Theory of Aid Evaluation Work*. Western Michigan University. [http://evaluation.web.fc2.com/pdfdata/sasaki\\_1\\_maintext\\_aug2009.pdf](http://evaluation.web.fc2.com/pdfdata/sasaki_1_maintext_aug2009.pdf)
- Stake, R. E. (1982). The two cultures and the evaluation evolution. *Evaluation News*, 3, 10-14.
- Scriven, M. (1975). Causation as explanation. *Nous*, 9, (1).
- Sriven, M. (2007). *The Logic of Causal Investigations*. RMIT University.
- USAID. (1987). *Rapid, Low-Cost Data Collection Methods for A.I.D.* Washington D.C.: USAID.
- World Bank. (1992). *Effectiveness implementation: Key to development impact (the Wapenhans report)*. Washington, DC: The World Bank.
- World Bank. (1998). *Assessing aid: What works, what doesn't, and why*. Washington, DC: The World Bank. (2010.2.8受理)

## **Evidence-based development aid evaluation: The history of aid evaluation, the origin of RCT, and the comparison of the thoughts of Scriven and Banerjee**

Ryo Sasaki

International Development Center of Japan  
sasaki.ryo@idcj.or.jp

### **Abstract**

“Evidence-based Development Aid Evaluation” is a hot topic in the aid evaluation field and this movement has been led by Poverty Action Lab (J-PAL) since its establishment. By reviewing the history, three origins can be identified about this movement. The first is the evaluation study which root is deeply embedded the Campbell and Stanley’s proposal; the second one is the aid evaluation which has been uniquely developed due to the unique characteristics of the field; and the third one is relatively newcomer which is development economics field. After discussing each origin, the thoughts of Scriven, program director of the Evaluation Center, Western Michigan University (-2008), and Banerjee, director of J-PAL, are examined about advantages and constrains of randomized experimental design. One conclusion is: there are some rooms for employing RCT even though it cannot be dominant. It should be considered to put some more resources on the aid activities which effectiveness is verified by this approach.

### **Keywords**

Evidence, Development aid, J-PAL, Scriven, Banerjee



## 【実践・調査報告】

## フィリピンでのNGO教育事業参加型評価 ～ファシリテーターの役割とステークホルダーのエンパワーメント

田中 博

参加型評価ファシリテーター

(特活) ヒマラヤ保全協会

nepalippine@gmail.com

### 要 約

利害関係者が評価プロセスに参加する参加型評価は、評価の質を高めプログラム改善に効果的である反面、客観性確保や参加する関係者の代表制等の課題がある。また参加型評価において外部評価者は、従来型評価における査定者ではなく、ファシリテーターの役割を果たす。日本のNGOソルト・パヤタスがマニラ近郊で実施したプロセス型の教育支援事業の参加型評価に、筆者はファシリテーターとして関わった。評価プロセスを実施する中でスタッフを中心にエンパワーメントや改善効果が観察された一方、評価技術の不足や主観的に判断してしまう等の課題も見受けられ、ファシリテーターが様々な対応を行った。この事例からプロセス型事業における参加型評価の有効性及び、参加型評価の長所を活かし短所を補うための、ファシリテーターに求められる三つの役割の重要性が確認された。

### キーワード

参加型評価、ファシリテーター、NGO、評価手法、エンパワーメント、プロセス型事業

#### 1. はじめに

私はNGOワーカーとして活動する中で、プロジェクト評価の重要性を、それも外部者による従来型評価ではなく、利害関係者による参加型評価こそ有効であると痛感してきた。

所属するヒマラヤ保全協会では、ネパールにおいて森林保全や農村開発を行ってきた。プロジェクトサイトでは、緊急性の高い問題が山積みであり、だからこそ外部のNGOが支援をしている訳であるが、小さな団体の限られた予算と人材では、プロジェクト形成時に十分な調査や、明確な目標

設定を行うことは難しい。また、プロジェクトを進める中で、現場の状況は、治安や経済動向の変化などでめまぐるしく変わり、新たな問題が発生していく。そもそも、現場の人々が持っている生活の複雑なリアリティは、自然環境や伝統的な農業、信仰や文化などに深く根付いており、容易に外部者（特に外国人）には理解しがたい。支援を成功させ、住民の生活を向上させていくには、常に村人や現地スタッフと話し合いをしながら、柔軟に目標や活動内容を修正していく必要があった。それどころか、村人など当事者からよく話を聞いて情報を共有し、一緒に考えていけばいくほ

ど、解決への方向性は外部者ではなく、当事者の中から生み出されてくることに気がついた。後に大学院で勉強し、そのプロセスこそが、参加型評価と呼ばれていると知ったのである。

日本のNGOプロジェクトの多くは、ヒマラヤ保全協会がそうであるように、小規模の社会開発であり、ブループリント型でなくプロセス型である。また少ない予算と人材でやりくりしている現状がある。プロセス型事業は、受益者のニーズにあわせて柔軟に計画を修正し、きめ細かい活動が出来る長所があるといわれている。その通りであるが、そのためには受益者など多様な利害関係者と密接に連携を進めながら、現状把握と議論、合意形成を繰り返していく、高度なコミュニケーションと調整能力が必要とされる。残念ながら、多くの日本のNGOには、内部にそのための人材や予算が不足していると感じざるを得ない。

解決には、プロジェクトへの参加型評価の実施が効果的だと考える。目標や計画があいまいだったり、常に変更されたりするプロセス型事業への評価には、客観的な目標達成度を問う従来型の評価はなじまない。内部事情に不案内な外部評価者が下した評価結果に、内部の利害関係者は共感できないことも考えられる。当事者が納得しなければ、プロジェクトの改善は望めない。

しかし利害関係者自らが、自分たちがやってきた活動を真摯に振り返り、成果と課題を見極め、その原因を明らかにしていくならばどうなるだろうか。良い点を伸ばし、不十分な点を補っていきこうと、やる気が出るのではないだろうか。一方、内輪だけで議論をしていると、思い込みで判断したり、専門性に欠けたりする懸念もある。その問題に対しては、評価の進行を手助けする、適切な外部協力者（ファシリテーター）の支援によって解決されるのではないだろうか。それならば、ファシリテーターには具体的にどのような役割が期待されるのだろうか。日本のNGO事業の参加型評価実施を事例に、プロセス型事業と参加型評価の有効性、ファシリテーターの役割について考えてみたい。

## 2. 日本のNGOと評価

### (1) NGOのプロジェクト評価の全体像

そもそも日本のNGOはどの程度、プロジェクト評価を実施しているのだろうか。また、評価に対してどのような考えを持っているのだろうか。少し古い数字だが、(特活) アーユスが、2000年に開発NGOのプロジェクト評価の動向調査アンケートを実施している (アーユス 2003)。海外における開発協力を実施している190団体に調査票を郵送し、98の有効回答を得ている (回収率51.6%)。回答団体の内訳を予算規模から見ると、「100万円以上500万円未満」から「5億円以上」の範囲で、40%以上が「1,000万円以上、5,000万円未満」に収まっている。スタッフ数では、国内の専従スタッフがいなかったり、3人未満という団体が過半数を占めており、欧米の国際NGOと比較すると小規模な団体が多数派であるといえる。

表1と2をご覧ください。まず評価の実施状況

表1 日本のNGOによるプロジェクト評価の実施状況

	回答数	構成比
行った	82	83.7%
特に行っていない	16	16.3%
総数	98	100.0%

(出所) アーユス (2003)

表2 評価のメリット

評価のメリット	回答団体に対する割合
プロジェクトの問題点が明確になり、プロジェクトが修正された	61.0%
プロジェクトの目的や内容が明確になり、プロジェクト運営がやりやすくなった	54.9%
対象プロジェクトの終了時を含めた、その後の活動計画の策定に役立てることができた	53.7%
評価による教訓を新しいプロジェクト計画に反映することができた	52.4%
会員を含めてプロジェクトに対する理解が深まり、組織の活性化につながった	43.9%
組織全体の方針の見直しが行われた	32.9%
社会的信用が高まった	18.3%
資金調達が可能になった	6.1%
その他	1.2%

(出所) アーユス (2003)

況だが、82団体（83.7%）が何らかの形で、プロジェクト評価あるいはそれに類する活動を行っていた。評価を「特に行っていない」としたのは16団体（16.3%）である。評価を行ったことのメリットとして多くの団体が指摘するのは、「プロジェクトの問題点が明確になり、プロジェクトが修正された」（61.0%）であり、次いで「プロジェクトの目的や内容が明確になり、プロジェクト運営がやりやすくなった」（54.9%）、「対象プロジェクトの終了時を含めた、その後の活動計画の策定に役立てることができた」（53.7%）と、半数を超える団体が回答している。

プロジェクト評価の経験あるなしに関わらず、評価活動を行う上でどのような問題点があるかという質問では、一番多かったのは「プロジェクトの成果は数量化して測ることができないものが多い」である（58.2%）。プロジェクトを実施した結果生じる人々の意識の変化、意欲、人と人とのつながりなどが数量化しにくい、というのである。他の問題点としては、「スタッフの数が足りない」（35.7%）、「業務が忙しくて時間がない」（22.4%）、「評価の方法や手法がわからない」（20.4%）、「費用がかかりすぎる」（16.3%）、「プロジェクト開始時のデータがないので比べる対象がない」（14.3%）と続いている（表3）。

表3 評価実施上の問題点

評価実施上の問題点	回答数	回答団体に対する割合
プロジェクトの成果は数量化して測ることができないものが多い	57	58.2%
スタッフの数が足りない	35	35.7%
業務が忙しくて時間がない	22	22.4%
評価の手法や方法がわからない	20	20.4%
費用がかかりすぎる	16	16.3%
プロジェクト開始時のデータがないので比べる対象がない	14	14.3%
評価できる人が内部にいない	11	11.2%
評価を頼める人が外部にいない	7	7.1%
評価をしてもその結果が生かされない	5	5.1%
その他	5	5.1%
団体数	98	

（出所）アユース（2003）

## （2）日本のNGOプロジェクトの特徴と評価

一般にNGOのプロジェクトは社会開発事業が主流といわれているが、この調査結果からもその特徴が観察できる。社会開発事業では、住民に直接はたらきかける活動が主流であり、また最初から活動計画が細かく決定されたブループリント型ではなく、外部状況や関係者の意識にあわせて、柔軟に計画を修正するプロセス型プロジェクトが多い。このような事業の場合、住民の意識や行動変容などは実施前・実施後の比較や、計画の達成度を数量化して測ることが難しい。佐藤（2007）は、因果律に基づいてプロジェクトの貢献度が測りにくいこと、目標が曖昧なので、数量化した評価になじまないことを理由に、社会開発的な介入の「評価」は困難であると指摘している。もう一つの特徴としてプロジェクト実施に手一杯で、評価に費やす資金や専門的な人材が不足している現実もみえる。上記アンケートでは今後プロジェクト評価に取り組む予定のない団体も2割ほどあり、「予算やスタッフ面で制約がある」ことが理由の一つとしてあげられている。このような状況を鑑みると、これから日本のNGOでより効果的に評価を実施していくには、NGOプロジェクト特有の二つの性格を十分考慮した上での評価の設計、実施が求められているといえよう。

## 3. 参加型評価とは何か

### （1）参加型評価の定義と従来型評価との違い

参加型評価の定義は、JICA（2001）によると「参加型評価とは、最終受益者を含めた幅広い関係者（stakeholders）が、評価計画の作成、情報の提供・収集・分析、プロジェクトの当初計画の修正などに可能な限り参加して行う評価である（後略）」となっている。またUNDP（1997）では「参加型評価は住民を中心とする評価である。プロジェクト関係者と受益者は単に評価の対象となるだけでなく、評価を行うプロセスの中で中心的な役割を担う」とある。

次に参加型評価は、一般的な従来型評価と、どう違うのだろうか。従来型評価は、客観性、中立性、正確性を重視しており、利害関係者、特に資

金提供者や政策決定者に対するアカウンタビリティの確保（中略）に優れた特性を出し、またプロジェクト自体の良し悪しや価値を判断することに焦点が当てられ、その役割を果たしてきた。これに対して、参加型評価では、プロジェクト・スタッフや利害関係者を評価に参加させることによって、彼等のプロジェクトに対する主体的な対応の促進、評価能力の向上により直接的にプロジェクトの改善をはかる（三好・田中 2001）、という特徴がある。誰が（Who）、何を（What）、どうやって（How）、いつ（When）、何のために（Why）評価するのか、両者の違いを簡潔にまとめた表4を参照いただきたい。

表4 従来型評価と参加型評価の比較

	従来型評価	参加型評価
Who	外部の専門家	コミュニティ住民、プロジェクトスタッフ等利害関係者 外部者はファシリテーターを務める
What	事前に設定した指標に基づき、投入に対する成果を計る（定量的データ中心）	利害関係者が自ら、成果を含めた指標を決める（定量的&定性的）
How	「科学的客観性」に焦点 結果への共有に時間がかかる	現場に適した単純な手法、現地住民の参加による素早い共有
When	通常プロジェクト終了時、適宜中間評価	中間、終了時にとらわれず柔軟に実施可能
Why	説明責任、支援継続の判断	主体的な活動・改善に向けて住民のエンパワー

（出所）Narayan-Parker（1993）に筆者加筆

## （2）参加型評価の長所と短所

参加型評価の特徴から、いくつかの長所（メリット）と短所（デメリット）を掲げることができる。長所としては、評価のプロセスが学習過程として作用し、利害関係者の評価対象への理解が深まり、当事者意識や責任感が醸成され、態度、行動変容につながる。また評価手法など関係者へのキャパシティ・ビルディング（アーユス 2003、源 2008）、対象プログラムへの主体的関与、関係者間の結束、相互理解の促進などの利点があり、ひいてはマネジメント能力の向上につながる（JICA2001）、などがあげられる。また参加型評価では評価結果は評価に参加した関係者と共有

し、次の行動につなげていくことが目指され、受益者を含む被援助国側に対するアカウンタビリティも重視される（JICA 2001）。

エンパワーメントの定義は多数あるが、人々が潜在的に持つ「力」を顕在化させるプロセス（太田 2007）と考えると、参加型評価は評価プロセスにおいて、利害関係者の能力向上、主体性の強化、効果的フィードバック（情報の共有）などを通じて、関係者のエンパワーメントを促進し、結果としてプロジェクトの改善が期待できるわけである（表5）。

表5 参加型評価のメリット、デメリット

メリット	デメリット
マネジメント能力強化/キャパシティ・ビルディング	利害関係者の評価知識・技術にばらつき
オーナーシップ（主体性）の強化	自己評価なので客観性・中立性に欠ける
効果的フィードバック/双方向アカウンタビリティ	誰が参加するか、という代表制の問題

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

このようにエンパワーメントや改善が期待されること、内部事情を知る当事者による判断ができること、また内部のみで行えば外部の専門家雇用など追加的な出費を必要とせず、通常の仕事の延長で評価を実施することも可能である。これがNGOプロジェクト評価に、参加型評価が適していると考える大きな理由である。

その一方で短所としては、たとえなんらかのトレーニングをしたとしても、利害関係者の評価知識、評価技術にはばらつきがあり、評価の質が低くなるという批判や、従来型評価と比較すると客観性や中立性に欠けるという指摘がある。また評価に参加すべき人々がきちんと参加しているか、排除されていないか、といった参加者の妥当性や代表制の問題がある（源 2008）といわれている。

参加型評価に限らず、従来型評価も含めてどのようなアプローチにもそれぞれの長所、短所が存在する。参加型評価を行う際には、評価対象や目的に応じて、長所を最大限に活かしながら、短所を出来るだけ押さえていく現実的な調整作業が必要だと思われる。そこで、評価者の位置づけとし

て、参加型評価ファシリテーターがクローズアップされてくる。

### (3) 参加型評価とファシリテーターの役割

ファシリテーターとは、「単なる司会進行や講師（教師）とは違い、学習者・参加者の知識や経験を引き出しながら、円滑かつ効果的に会議やワークショップ、あるいは人々のさまざまな社会活動などを促進・援助する人」（国際協力用語集2004）と定義されているが、参加型評価におけるファシリテーターの役割とはどのようなものなのだろうか。

三好・田中（2001）は、「参加型評価では、プロジェクトに直接関わっているプロジェクト・スタッフやプロジェクト関係者が、評価者として評価グループを編成し、評価を実施することが基本となる」として、「参加型評価では、外部評価者の役割は従来型評価とは大きく異なることになる。一般に評価参加者の評価能力が十分でない参加型評価では、外部評価者は評価の実施方法の知識や手続きを促進するファシリテーターとしての役割が求められることになる。それゆえに、外部評価者が参加型評価を実施する場合には、従来の査定者としての役割ではなく、ファシリテーターとしての役割を担うという意識変革が重要になる」としている。またEstrella and Gaventa（1998）は、参加型評価の良いファシリテーターの役割として、「様々な参加者の異なる経験や理解をまとめる、主に触媒や刺激者」であると述べている。

評価とは価値の判断（査定）を伴うが、このように参加型評価では、判断を下すのはあくまでプロジェクトの利害関係者である。ファシリテーターはそのプロセスを側面から支援する者といえる。しかし直接の査定者ではないものの、評価プロセスを通じて利害関係者に深く関わる中で、結果に大きな影響を与えると考えられる。私はこのファシリテーターの働きが、参加型評価の長所を活かし、短所を補うために大きな役割を果たすと考えている。日本のNGOがフィリピンで実施した教育支援事業の評価事例を材料に考察したい。

## 4. 特定非営利活動法人ソルト・パヤタスの活動

### (1) 活動の沿革と概要

特定非営利活動法人ソルト・パヤタスは1995年、日本人女性がスタディツアーでフィリピンを訪問した際、現地住民組織からの要請と、ツアー主催NGOからバックアップを受け、ボランティア2名が、19名の子どもの奨学金支援を開始したことを機会に設立された。マニラ近郊ごみ投棄場周辺地域、パヤタス地区及びカシグラハン地区周辺のスカベンジャーの家族を対象に、学校に行きたいけれども貧しい子どもに、日本でスポンサーを募る奨学金支援が中心の活動である。これを柱に補習授業とダイケアセンターの運営からなる教育支援及び、女性への収入向上、緊急医療支援、スタディツアー等を実施してきた。2009年2月現在、2つの地域で支援する子どもは180名、収入向上に参加する母親は22名となっている。年間予算はおよそ900万円（2008年度）で、国内には有給専従スタッフはなく、マニラ事務所にはフィリピン人数名を雇用している小規模な日本のNGOである。

### (2) これまでの活動の経緯

14年間の活動の結果、教育支援事業においては延べ550人の子どもに就学・復学の機会を提供してきた。徐々に活動が広がり現地で事務所を立ち上げ、日本人及び現地スタッフが働いている。近年は奨学金で進学した大学生が補修や保護者会などに参画し、活動の中核を担ってきている。その一方で、プロジェクト開始時に十分な事前調査が行えず、ベースライン数値把握や明確な目標設定がなかった。そのため、これまでの事業の効果を客観的に把握・共有する事がないまま活動を継続してきた。将来計画を立案するための教訓を得られないまま、今後の見通しが立てられない状況となってきたのである。

### (3) 評価実施に至る背景

この状況を打開するために、ソルト内部で教育支援事業の評価の必要性が認識された。そこで、2009年2月の現地調査（2週間）を含む、外部ファ

シリテーター（筆者）による教育支援事業の評価を実施することになった。

現地調査の前に、ソルト・パヤタスの日本人代表者とファシリテーターの打ち合わせが行われ、ソルト側からは「プロジェクト形成時までさかのぼり、実績把握・計画の妥当性・活動の有効性について調査したい」という希望とともに、「外部評価者による従来型評価では、プロセス型であるソルト事業にふさわしくないのではないか」といった不安が寄せられた。協議の結果、オーナーシップ向上や、エンパワーメントが期待できる参加型評価を採用すること、得られた教訓・提言を共有し、2010年に向けて新3ヵ年計画を策定する、いわゆる改善目的の評価を行っていくことが合意された。

派遣にあたっては、国際協力機構（JICA）の「NGO海外プロジェクト強化のためのアドバイザー派遣」制度を活用した。これはNGOがプロジェクトの改善のため、外部専門家を現地に派遣する際に、渡航費や滞在費、謝金などをJICAが支援する制度で、専門家を派遣したくても人材がない、外部者を雇用する資金がないといった日本のNGOにとって、たいへん有意義な制度である。

## 5. 評価実施のプロセス

### (1) 評価に関するブリーフィング

通常、評価の実施ステップには、①評価設計、②データ収集・分析、③評価結果の報告と共有の3つの段階があり、参加型評価の進め方もこれに準じる。源（2008）は、参加型評価の実践（想定される共通事項）という名目で、参加型評価の基本的な進め方をまとめている（表6）が、これに従って評価プロセスを実施した。

最初に、代表を含む日本人メンバー数名で、日本側の評価チームを結成した。以降、これを本稿ではソルト日本と呼ぶことにする。またソルト・パヤタスのマニラ事務所の専従スタッフをソルトマニラと呼ぶ。この両者が、今回の評価の重要なアクターとなった。評価プロセス開始前に、初めての評価を参加型で行うために、ファシリテーターによる評価一般や、参加型評価に関する短い講

表6 参加型評価の実践（想定される共通事項）

I 評価設計
1. 誰が参加するのかを決める
2. 利害関係者で評価の目的を共有する
3. 評価対象の内容を共有する
4. 評価設問を検討する
5. 評価のデザインやデータ収集法を検討する
II データ収集&分析
III 評価結果のまとめと報告

（出所）源（2008）

義（ブリーフィング）をソルト日本、ソルトマニラ対象にそれぞれ行った。

### (2) 評価設計

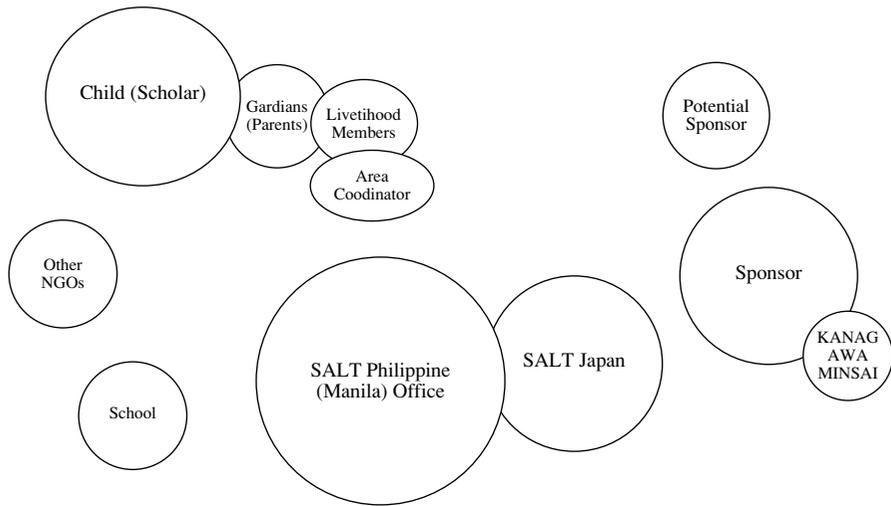
#### ①誰が参加するのかを決める

一般に開発協力プロジェクトには受益者を含め多数の利害関係者が存在するが、参加型評価には「誰が評価に参加するか」「大切な関係者が排除されていないか」という代表制の問題が存在する。この課題をクリアするために、評価設計の最初の段階でファシリテーターの指導による入念な関係者分析を行った。

ソルト日本は現地調査前に日本国内で会合を持ち、ソルトマニラは現地調査時に、スタッフ参加により、それぞれベン相関図<sup>1</sup>を作成しながら関係者分析を行った。できあがったソルト日本による図1、ソルトマニラによる図2を、ソルト日本とマニラのメンバーで互いに比較しながら、多数の利害関係者のどの範囲までが、評価に参加するかを検討した。

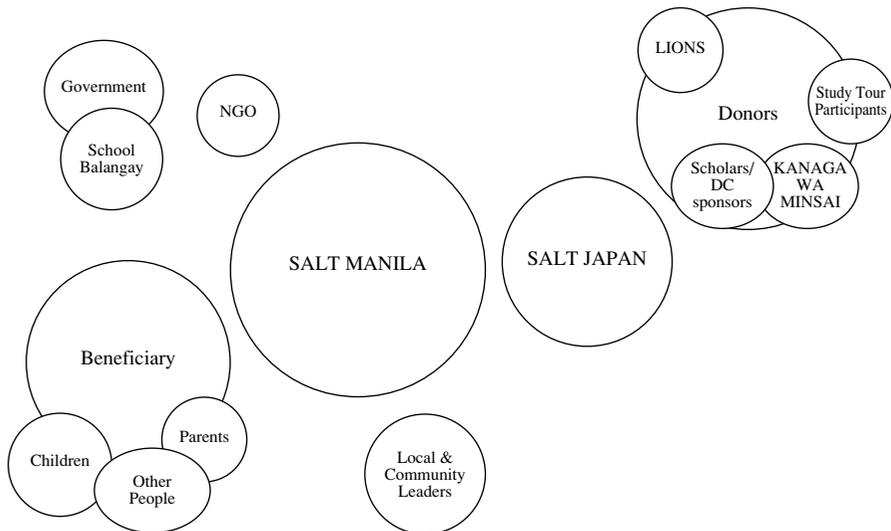
ソルト日本の図解（図1）には、たくさんの関係者が登場するが、自身であるソルト日本、パートナーのソルトマニラが、大きく中心に描かれ、お互いに近く配置されている。右側には奨学金の里親などのスポンサーや、助成金提供者である「かながわ民際協力基金」の二つが合体して、大きな島を作っている。左上には、奨学生の子どもたち、保護者、収入向上事業の参加者など事業の受益者層が合わさって大きな固まりになっている。これに対して他のNGOや学校、潜在的資金

図1 ソルト日本による関係者分析



(出所) 評価プロセスに従い筆者作成

図2 ソルトマニラによる関係者分析



(出所) 評価プロセスに従い筆者作成

提供者は小さく、距離をおいて表現されている。

次にソルトマニラ作成の図解(図2)を見ると、同様に中心に自身であるソルトマニラとソルト日本が大きく、かつ近く描かれている。右に資金提供者(ドナー)が、里親やデイケアセンターの支援者、ライオンズクラブなどと一緒に、大き

な群を形成している。左下は子どもや保護者など受益者群だが、これも大きなカテゴリーとして描かれている。地域の行政、他のNGO、公立の学校は小さくて比較的遠くに描かれていることがわかる。

二つの図解を比較すると、まず重要な利害関係

者としてソルト日本とソルトマニラの存在が共通し、これらは近い関係にあると相互に認識されている。他に重要な関係者としては、プロジェクトの受益者である子どもや保護者も、大事なアクターとして両方の図解に現れている。さらにプロジェクトに資金を提供する資金提供者が大切な存在として両者に認識されている。その一方、行政や学校、他のNGOは利害関係者として双方に認識はされているが、重要性は低いと思われる。

そこで、ソルト日本・ソルトマニラ・受益者・資金提供者の四つが事業にとって重要なアクターと考え、評価に参加する関係者はこの4つとした。次いで、どの関係者がどのように評価に参加するか議論を行った。その結果、①ソルト日本は評価設計の段階から、データ収集・分析及び結果の共有まですべてに参加、②ソルトマニラは日本側から提案される評価設計案の共有、データの収集と分析及び結果の共有、③受益者はデータ収集の情報提供者として、また結果の共有で参加する（設計・分析には参加しない）、④資金提供者は結果の共有（報告）を受ける形で参加すると合意した。表7をご覧ください。

表7 利害関係者の誰が、どの段階で評価に参加するか

	評価設計		データ収集		結果の共有
	設計	共有	収集	分析	
ソルト日本	○	○	○	○	○
ソルトマニラ	-	○	○	○	○
受益者	-	-	○	-	○
資金提供者	-	-	-	-	○

(出所) 評価プロセスに従い筆者作成

なお、開発事業への利害関係者の参加の度合いについては、援助機関によっていくつか類型されている。代表的なものの一つがアジア開発銀行(ADB)による「参加の深さ」である(表8)。低いレベルでは住民は情報共有の対象や相談の相手であるが、中間レベルでは共同の調査や意志決定を行うことが可能になり、最高レベルでは住民自身がマネジメント能力を身につけ、エンパワーメ

表8 参加の深さ (Depth of Participation)

レベル1	情報共有 (Information Sharing)
レベル2	相談 (Consultation)
レベル3	共同アセスメント (Joint Assessment)
レベル4	共同決定 (Shared Decision Making)
レベル5	協働 (Collaboration)
レベル6	エンパワーメント (Empowerment)

(出所) ADB (2001)

ントされることを目指している。今回の評価プロセスをあてはめると、評価設計への参加はレベル3(共同アセスメント)や4(共同決定)の段階、データ収集における収集や情報提供のみはレベル1(情報共有)か2(相談)、分析の段階はレベル4、また結果の共有はレベル1程度の参加に該当すると考えられる。

参加型評価の定義によると、ターゲットである受益者が評価の設計や分析に参加することが重要だといえる。しかしながら初めての評価であり、限られた時間で多数の受益者を巻き込むことの困難が予想され、まずスタッフ・レベルの参加を優先することとなった。

## ②利害関係者間で評価の目的を共有する

次に関係者間での、評価結果をどのように活用するか、という評価目的の共有である。当初はソルトマニラから、評価に対する漠然とした不安も感じられた。そこで、ソルト日本からいわゆる改善志向の評価にしたい、という提案がなされた。両ソルトの協議の結果、「よりよい新プロジェクト立案のために、現在と過去のプロジェクトを振り返り、その教訓から学ぶ」という内容で合意を得た。受益者にも、データ収集時に必ずこの評価目的を伝えるようにして、調査への協力を訴えるようにした。

## ③評価対象の内容を共有する

具体的な評価の計画を策定する前に、対象プロジェクトの詳細を十分に理解しておくことが重要である。ところが、この事業は開始時において明確な目標設定がなされておらず、目的と手段の関係性も十分に議論がされないままに、奨学金支給

表9 ソルト教育支援事業のプロジェクト目標とロジックモデル

プロジェクト目標	子ども自らが問題解決能力を身につけ、貧困から脱却する		
期待される成果1	子どもが小学校を卒業し、可能な限り上の学校へ進学できる	活動	奨学金、補習授業、保護者会等
期待される成果2	未就学児童が小学校へ行けるようになる	活動	奨学金、補習授業、保護者会等
期待される成果3	子どもの自己肯定感が高まり、他者への関心、分ち合い、相互の尊重の価値観が形成される	活動	デイケアセンターの運営等

(出所) 評価プロセスに従い筆者作成

表10 評価設問、調査対象・データ収集方法、調査結果・分析

	評価設問	調査対象/データ収集方法	調査結果/分析
1	何名が小学校を卒業した／しなかったか（1995～現在）？その理由は？	事務所の記録確認、エリア・コーディネーターに確認	239名の奨学生のうち178名が小学校を卒業（約75%）した。
2	教育を続けられない共通要因は何？	OB/OG、教師に面接	設問10と一緒に考察する
3	子供は事業で貧困から脱却したか？	プロジェクト目標のため、調査終了時に総合的に判断する	
4	教育を継続させる共通要因とは？	OB/OG、教師に面接	決意、夢、家族／ソルトの支援
5	DCセンター支援を継続するべきか	保護者会で質問	継続望ましいが、収益に難あり
6	保護者会の活動は有効か？	保護者会で質問	有効だが、改善の余地あり
7	DCセンターは小学校入学に貢献したか？	保護者、教師	貢献している
8	補習授業は有効か？	補習教師に面接	有効だが、改善の余地あり
9	支援者と手紙交換は役に立っているか？	保護者会で質問	役に立っている
10	教育を阻害する外的要因は何か？	OB/OG/教師面接、スタッフKJ法	家庭、地域社会、経済社会問題
α	日本から支援が切れても継続性はあるか	保護者会で質問、スタッフ面接	現状では難しい
β	活動全体の効率性はどうか？	スタッフに質問	効率性は高いと思われる

(出所) 評価プロセスに従い筆者作成

などさまざまな活動が行われていた。また14年の間にスタッフの交代や事務所移転などで、事業全体のプロセスを俯瞰できる資料が存在しなかった。そのままでは評価ができないので、既存の報告書や会員向けニュースレターを見直したり、日比のスタッフの記憶をたどったりして、開始当時に関係者間の暗黙の了解として存在したプロジェクト目標を再設定したところ、「子ども自らが問題解決能力を身につけ、貧困から脱却する」ことであることが確認された。それをもとに期待される成果や、そのための活動など事業のロジックモデルを再構築した（表9）。

#### ④評価設問を検討する

評価を通じて何を知りたいか（評価設問）を、ロジックモデルと関連で議論する中で、日比のスタッフの意見に基づいて、10問を決定した（表10の左欄）。設問の数が多いのは、初めての評価であるため、活動を続けながら関係者が疑問に感じていた事項が多数寄せられたといえる。このように、設問は現場のスタッフの問題意識を反映したものであるが、「何名が小学校を卒業したか／しなかったか（設問1）」「保護者会の活動は有効か（設問6）」などのように、活動の妥当性や目標達成度を問う設問が多い。これに対して事業の効率

性や、継続性（自立発展性）を測る設問がスタッフの口からは出てこなかった。そこで、評価がより多面的な視点から行われるように、ファシリテーターから提案をして、効率性や持続性に関する設問の追加を行った。「日本から支援が切れても継続性はあるか」、「活動全体の効率性はどうか」である（表10左下の $\alpha$ と $\beta$ 二つ）。

#### ⑤評価のデザインやデータ収集方法を検討する

その後、設問に沿う形で調査対象及びデータ収集方法を、引き続きソルトマニラとソルト日本の共同作業で検討した。例えば、評価設問1である「何名が小学校を卒業したか／しなかったか」に答えるには、「事務所の記録を確認する」「記録で不明な点は、(現場に詳しい) エリア・コーディネーターに確認する」という要領である。このように評価設計はファシリテーターの進行のもと、ソルトマニラと日本のメンバーが主体的に行った。二週間の調査日程の内、半分の一週間が評価設計に費やされた。当初は、ソルトマニラは日本側からの評価設計の提案を共有する計画だったが、実際には提案に対して積極的な議論や修正が行われ、予想以上にマニラスタッフの主体的な参加が実現した。また本来ならば、データ収集方法が決定後、達成度を測る指標も参加型で設定されることが望ましい。しかし時間の制約上、各評価設問に該当する適切な指標を、すべて設定するには至らなかった。

### (3) データ収集・分析

#### ①参加型によるデータ収集

データの収集も、評価設計で決めた通りにソルト日本とソルトマニラの共同で実施した。

収集を始める前に、マニラスタッフは社会調査の経験が少ないため、ファシリテーターによる基本的調査手法の説明と、ロールプレイによる面接調査方法研修をスタッフに行った。

実際の収集は評価設問が多数あるため、保護者や学校教師など、収集する対象ごとに、設問を整理して行った。例えば、「評価設問6：保護者会の活動は有効か」、評価設問8「補習授業は有効か」、「評価設問9：支援者との手紙交換（クリスマスカードなど）は役に立っているか」や、追加された

設問 $\alpha$ 「日本からの支援が切れても継続性はあるか」といった質問は、子どもの保護者が全てデータ収集の対象であった。そこでソルトが定期的実施している保護者会に、通常通り保護者に集ってもらい、ソルトマニラより評価目的を説明の上、複数のグループに分かれて討論を行った。その際、日本人のファシリテーターだけでなく、ソルトマニラのスタッフも分担してグループに入り、質問をした。グループ討論が終わると、忘れないうちに日比スタッフで集まり結果を表にまとめ、それぞれのグループ討論の結果を共有する中で、成果や問題点について一緒に考えていった。

#### ②データ分析

収集したデータの分析も現地でソルト日本とマニラが共同で行い、学びや教訓が今後の活動や計画策定に活かされることをめざした。設問1の「(1995年からの奨学生で) 何名が小学校を卒業したか／しなかったか」を例にあげて解説する。これは、有効性を問う設問であるが、調査対象とデータ収集方法をどうするか議論した結果、「事務所の記録を確認する、それでも不明な部分はエリア・コーディネーターに照会する」とこととなった。その結果14年間で242人の奨学生のうち、小学校を卒業、あるいは現在も就学が継続できている人数は181名で比率は約75%であることが判明した。

この結果を皆で考察したところ「成功である」「いや、そうではない」とマニラスタッフの意見が分かれた。ソルトによるとフィリピン一般では小学校卒業率は70%といわれており、それと比較して上出来という見解と、資金や手間の多大な投入を行ったにもかかわらず、この程度では不十分という声もあった。また卒業できなかった25%の理由は、家庭の問題や本人の意向など多岐にわたっているが、別団体の奨学金を獲得したのでソルトの支援を離れた例もあり、「必ずしも失敗にはならない」という意見もでた。

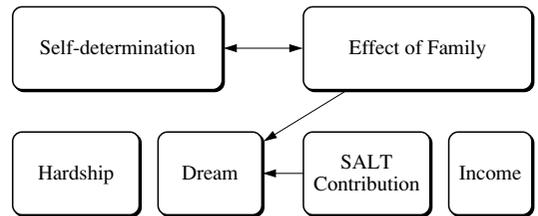
このような議論を通じてプロジェクト開始時にサイトに於けるベースライン調査が無く、また達成目標も何%と明確に設定をしていなかったため、比較の対象が無く判断が難しいことが明らかになった。このため次回のプロジェクト立案の際には、ベースライン調査や目標設定を行う必要性

が共有された。さらに事務所の記録を探したり、集計したりする中で、これまでのデータ保管手段の問題点や改善の必要性も認識されるという利点もあった。

もう1つのデータ分析の例をあげると、設問4「教育を継続させる共通要因とは何か」についての調査と議論である。厳しい環境で教育からドロップアウトする子どもが多い中、支援を受けた後も進学を続けている奨学生OB/OGも存在する。行き詰まってしまう事例と、何が違うのだろうか。成功例である数名に、ライフストーリーの聞き取りを行い、それぞれの事例からキーワードを抽出し、KJ法<sup>2</sup>でまとめた。すると「本人の決意」「将来の夢」「家族のサポート」「ソルトなど外部の支援」などが、共通要因として浮かび上がってきた。これらは、現場で子どもたちの動向に触れているマニラスタッフも納得する内容であり、今後のプロジェクト立案に活用できる教訓であると感じられた(図3参照)。

このように評価プロセスへの参加を通じて、関係者間でプロジェクトの実態(実績、成果と課題)やその背景に対する理解が深まり、改善に向けての意思統一が図れるところが参加型評価の大きなメリットだと思われる。

図3 教育を続ける外部要因



(出所) 評価プロセスに従い筆者作成

### ③分析データに基づく価値判断

続いて、ソルトマニラ全員とソルト日本で収集データ分析結果に基づき、教育支援事業の最終的な評価結果の抽出を行った。手始めに、ソルトマニラのみで、DAC評価五項目に基づく評価結果の決定をスタッフ皆で行った。五点満点(5:最高 4:良い 3:普通 2:悪い 1:非常に悪い)で表11の結果となった。妥当性/有効性においては、活動は「受益者のニーズに基づいている」という高めの評価となった一方で、プロジェクト目標の「貧困から脱却する」という設定自体が不明確、もしくは高すぎるので「厳しい現場の実態とかけ離れており、フラストレーションが溜まる」という厳しい数字になった。効率性は材料が少なく判断が難しいが、「活動に優先順位をつけ、無

表11 ソルト教育支援事業の参加型評価結果

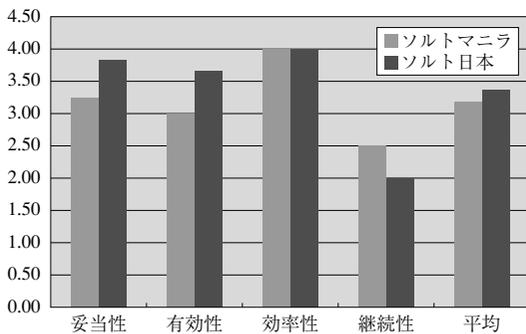
評価項目		ソルトマニラ		理由 (マニラ側)	ソルト日本	
妥当性	受益者のニーズに沿っていたか?	4	3.25	貧しい人を対象に行われている	4	3.83
	目標は適切に設定されたか?	2.5		方向は良いが、目標が高すぎる	3.5	
	プロジェクトアプローチは適切か?				4	
有効性	期待される成果は達成されたか?	4	3	はい。でも充分でない。	4	3.66
	プロジェクト目標は達成されたか?	2		教育支援だけで達成は難しい	3	
	期待通り貧しい人々に裨益したか?				4	
効率性	成果に見合う投入、無駄遣いしないか?	4		優先順位をつけ無駄遣いしない	4	
継続性	日本の援助がなくても大丈夫か?	2.5		お金不十分だが人材育っている	2	
インパクト	プロジェクト正負のインパクトは?	正: 家族へ波及効果、日本と交流。負: 住民にソルトへ依存心?				
総合評価 (五段階)		3.19			3.37	

(出所) 評価プロセスに従い筆者作成

駄遣いをしていない」と、評価は4になった。継続性は、ソルト日本からの送金や収入向上事業からの収入など決算の内訳を分析して、「資金的な自立度は不十分だが、人材は育っている」と2.5の評価。正のインパクトとしては「教育を受けた子どもが家族に教えている」「日本の人々との交流が進んでいる」という波及効果があげられ、負のインパクトとして「成績が上がらないのはソルトが原因(補習の質が低い)と思う受益者がいる」といった依存心が感じられる点が指摘された。

単純平均で総合評価を出したところ、3.19がソルトマニラによる結論となった。現地調査帰国後、東京都内において再度ソルト日本で会合を行い、収集データを共有・分析して日本側の評価結果を抽出した(平均3.37、表11右)。結果判断は互いの結果を見ないで別々に行ったにも関わらず、両者の評価結果が類似している点が興味深い(図4のグラフ)。

図4 マニラと日本の評価結果の比較



(出所) 評価プロセスに従い筆者作成

#### (4) 評価結果のまとめと報告

評価の最後の段階が、評価結果のまとめと報告である。参加型評価の場合、データの分析を関係者の参加で行うため、分析が終わった段階である程度の結果が共有されている場合が多い。今回も現地調査最終日のミーティングにおいて、評価結果のおおまかな合意はソルトマニラと日本の間で出来ていた。しかしながら、結果の細かな分析や、物理的に距離の離れた利害関係者への報告は、評価調査の終了後に、引き続き実施せざるを得ない。

2009年8月現在、ファシリテーターの協力によって、ソルト日本が国内支援者や資金提供者や一般市民対象の報告書を作成中である。ソルトマニラや受益者住民への最終的な評価結果の共有は、2009年4月にソルト代表が再訪比して行ったと聞いている。

## 6. 評価実施中に観察された課題とファシリテーターの対応

5章では、ソルト・パヤタス教育支援事業の参加型評価の実施プロセスを紹介した。それでは、プロセスを実施する中で外部ファシリテーターはどのような役割を演じたのだろうか。筆者自身がファシリテーションを進める中で、評価設計、データ収集・分析、結果の報告/共有の各ステップにおいて、観察された課題がいくつかあった。それを解決するために、どんな対応をとったのか、またその結果どのようになったかを表12にまとめた。

表12 評価プロセスの各段階で観察された課題とファシリテーターの対応

段階	評価実施関係者が直面した課題	ファシリテーターの対応	対応した結果どうなったか
評価設計	関係者が評価の進め方がわからない	評価設計からOJTで一緒に行う	関係者の当事者意識が高まった
	関心低い事項は評価設問でてこない	ファシリテーターが適宜設問を追加	調査計画のバランスが改善
	データ収集対象・方法を先入観で限定	他の選択肢や可能性を紹介	多様な情報源が明らかになった
データ収集	データ収集方法がわからない	簡単な社会調査の研修を実施	効率的なデータ収集が実現
	収集データの信頼性が低い	データ分析時に信頼性を確認	信頼性を意識するようになった
分析/判断	データでなく、主観で判断する傾向	常にデータを共有し判断する指導	根拠をもって発言する姿勢見られる
結果共有	現場データは他関係者に共有困難	図解や表でまとめ・共有する	結果の他関係者へ説得力が増す

(出所) 評価プロセスに従い筆者作成

これはファシリテーターである筆者が、自分の観察と行動、その反応を振り返りまとめたものである。主観的な側面が含まれることは、予めお断りしておきたい。

### ①評価設計の段階

最初の評価設計の段階において、ソルト日本やマニラのメンバーが「参加型で評価をしたくても、評価の進め方がわからない」という事態があった。彼等は評価の専門家ではなく、また初めての評価であったので当然といえる。これに対しては、設計時にとどまらず、データ収集や分析、結果共有までファシリテーターが評価に関するブリーフィングを行っただけでなく、関係者分析のワークショップを進行するなど、評価の進め方を常時OJTで指導した。このように設計時から関係者が主体的に評価に参加する中で、当事者意識を高めていくことを目指した。マニラスタッフは最初の段階では座って受動的に話を聞くことが多かったが、プロセスの進行に従って、自らアイデアを出したり、積極的に調査結果のプレゼンテーションを行ったなど、目に見えて態度の変化が表れた。

また評価設問を設定する段階で、関係者（ソルト日本とマニラ）の問題意識からは、事業の有効性や妥当性を問う設問のみが提案される傾向があった。これに対してはファシリテーターから、効率性や継続性に関する設問の追加を提案した。「そんな事考えたことがなかった」という反応もあったが、その結果、調査全体のバランスが改善されたといえる。

それからデータ収集方法について議論を行った際、収集対象や収集方法を広く吟味することがなく、自分たちの先入観のみで限定してしまう傾向が見られた。例えば「奨学生の動向を知るには、誰々に聞くのが一番」、「受益者の意見を知るにはアンケート調査を行うべき」といった主張である。時には当事者だけが知りうる、貴重な情報源を見つけることができる場合があり、そこが参加型評価の長所であるのだが、広い視野から見てより良い方法を見逃す恐れもある。その場合はファシリテーターから「動向を知るには、個人の記憶だけに頼るよりも、事務所に保管された文書記録を見る方が正確ではないか」、「限られた時間で人々の

意見を捉えるには、アンケートよりもグループ討論の方が効率的と思うがどうか」といった、他の選択肢や新しい可能性を提案するように心がけた。その結果、多様な情報源から効果的にデータ収集を行うことが可能となったと思われる。

### ②データ収集と分析の段階

次にデータ収集の段階に移る。ソルトマニラのスタッフは業務を通じて、日常的に受益者に接しているが、社会調査の経験や技能が豊富にあるわけではない。そこでファシリテーターを講師に、簡単な調査手法の講義をスタッフ対象に行い、ロールプレイによる面接調査の研修も実施した。これによってスタッフの多くがデータ収集をできるようになり、ファシリテーターが単独で収集するよりも、多量のデータ収集が可能となった上、受益者から生の声を聴くことによって、スタッフの問題意識が高まることも期待できる。

しかしながら短期間の研修のみでは、面接やグループ討論を通じて、客観的な事実を受益者から引き出す調査スキルの体得は難しい。ファシリテーターの観察でも、マニラスタッフによる収集データの信頼性は、当初高くはなかった。そのデータからバイアスのかかった評価結果を抽出してしまう恐れもある。そこで毎回データを収集する度に、スタッフでミーティングを持ち、データを共有・分析しながら、その信頼性についても皆で確認したり、他のデータとのトライアングレーション（同じ事象を別の方法で調査する）の提案をしたり心がけた。「この意見は、全員の総意なのか、特定の個人の発言なのか」、「発言者は、質問の意図を十分に理解した上で、意見を述べたのか」「他の調査でも、同じ結果が出ているのか、よく見てみよう」といった具合である。この共有と確認を繰り返す中で、スタッフ各自がデータの信頼性を意識して収集を行う姿勢が見られるようになった。

収集したデータの分析に関しては、評価設問ごとに収集したデータを大きな図表に記入して壁にはり、それをスタッフ全員で眺めながら討論を行った。データが示す事実に基づき分析・判断することが大切であるが、現場経験豊かなスタッフほど、特定の事象に強い思い入れがある場合が見ら

れ、「私の経験ではこうだった」と、主観に基づき判断してしまう傾向があった。当事者の主観も1つのデータではあるのだが、その点ばかりを拡大解釈しては、評価結果が偏ったものになってしまう。そこで、データを分析する際、発言者には必ず理由を問いかけ、どのデータに基づいているのかどうかを確認しながら、議論を進めるように心がけた。その結果、収集データに基づいて発言する習慣が、スタッフに徐々に身に付いたと思われる。

### ③結果の報告と共有段階

評価の最後の段階が、結果の報告と共有である。前述のように参加型評価では評価設計の段階から利害関係者と合意をしながらプロセスを進めているので、ことさらに最後の報告・共有のみに重点は置かれていない。逆にどの段階でも常に関係者に必要な情報が行きわたっているかどうか、チェックが必要である。もし不十分であれば早めに共有を進めることが重要となり、その役割はやはりファシリテーターが担うこととなる。

また、社会調査や分析に不慣れな関係者が、積極的に評価に参加し続けるには、いくつかの工夫が必要である。例えば、やさしい言葉遣い、親しみやすい会議の雰囲気作り、スライドや図解を多用したわかりやすい説明等であり、その準備に相応の時間を費やした。

さらに設計から報告に至る一連の評価プロセスは、評価初心者には複雑で難しい場合がある。かつそれを忙しく消化する中で、調査のための調査になってしまい、そもそもの評価目的を忘れがちである。そこで全体の流れを表にして会議室の壁に貼り、連日プロセスの進行状況を示しながら、同時に「プロジェクトを改善するために行う」という評価目的を確認する作業が、皆で主体的に評価を進める上で効果的であった。

評価調査に直接参加した関係者以外、すなわち日本国内の資金提供者などへの結果報告と共有も、評価の大切な仕事である。普通、日本の生活とかけ離れた開発途上国の現場の情報やリアリティは、なかなか日本の市民には伝わりにくい。今回、スタッフ間でのデータ共有や議論のためにたくさんの図解や図表を作成したところ、これが帰

国してから国内関係者に評価結果をわかりやすく伝えるのに有効であった。現場でデータ共有に有効なツールは、日本国内の結果共有でも役に立つことがわかったことは収穫である。

### ④評価プロセス全体を通じて

以上、述べたように、ファシリテーターは、評価プロセスの進行をする傍ら、プロジェクト関係者と深くコミットすることになった。全体を通して感じたことは、評価の進行に関する相談的な業務の多さである。初めての評価であり、また参加型で進めたため、各ステップにおいて、どのように仕事が進むのか不明であり、関係者に不安があったのだと思われる。その一方でプロセスに参加し意識が啓発される中で、関係者が「これまでの活動上の疑問が全て解決されるのではないか」といった評価に対する過剰な期待を持たせてしまう場合もあり、評価の限界についても説明を行うことがあった。

## 7. メタ評価（評価の評価）

メタ評価とは、評価の評価である。評価プロセス終了時に、ソルト日本及びマニラのスタッフにメタ評価（無記名）を実施した。質問項目は、参加型評価の特徴から五つ用意した。①評価の目的は達成されたか、②プロジェクトの将来の方向性は明確になったか、③評価によって、利害関係者のよりよい相互理解に貢献したか、④あなたや団体にとって、何か新しい学びがあったか、⑤評価ファシリテーターは適切に役割を果たしたか、である。（5：最高、4：良い、3：普通、2：あまり良くない、1：悪い）の五段階評価で採点してもらった。ソルトマニラ7名のすべての回答の平均は4.0で、ソルト日本5名の平均は4.1になっている（表13及び図5）。また自由記入欄に、ソルトマニラ、日本の面々が記入してもらったことは、表14にまとめたのでご参照いただきたい。

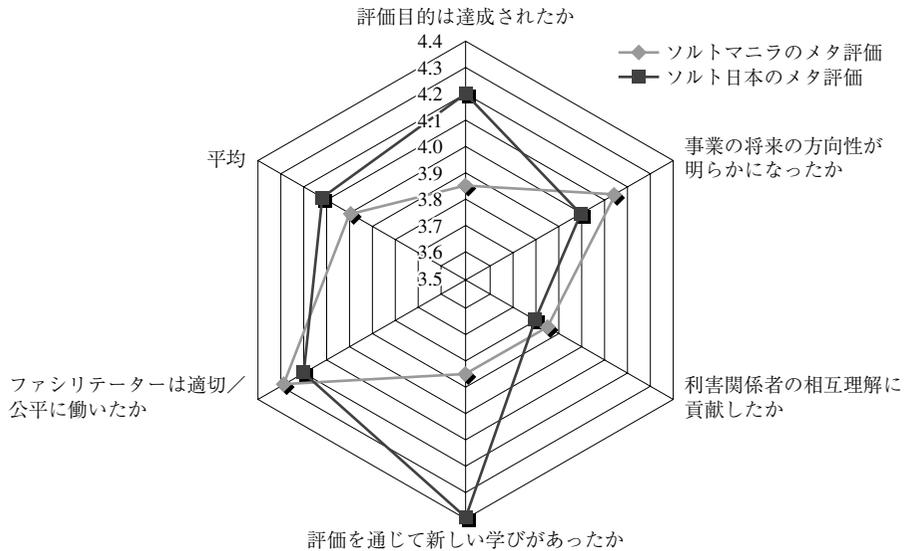
評価目的の達成度を聞く質問①と②では、①が日比の平均は4.05と「良い」で、②も4.05であった。関連する自由意見としては「評価を通じて課題が見えた（マニラ）」、「組織の弱み・強みが共

表13 ソルト日本及びマニラによるメタ評価結果

満点5	①評価目的達成	②方向性抽出	③関係者相互理解	④新しい学び	⑤ファシリテーター	平均
マニラ7名	3.9	4.1	3.9	3.9	4.3	4.0
日本5名	4.2	4.0	3.8	4.4	4.2	4.1
平均	4.05	4.05	3.85	4.15	4.25	4.05

(出所) 評価プロセスに従い筆者作成

図5 ソルトマニラと日本のメタ評価グラフ



(出所) 評価プロセスに従い筆者作成

表14 メタ評価結果自由記入欄 (抜粋)

ソルトマニラ	ソルト日本
異なる関係者に質問する事学んだ、ロジックモデルで成果を考える重要性わかった、評価設問に沿って質問する事を学んだ、関係者お互いの大切さを理解した、評価を通じて課題が見えた、ファシリテーターが良かった、等	目的や理念について日比の意見の違いわかった、外部者に評価を手伝ってもらって良かった、組織の強み・弱みが共有された、現地スタッフのモチベーション高まった、たくさんの気づきを得た、外部の協力で内部評価すれば結果を素直に受入れられる、等

(出所) 評価プロセスに従い筆者作成

有された」「現地スタッフのモチベーションが高まった(日本-3名)」と、現状認識の深化と関係

者の意識の高まりが実感されている。関係者の相互理解促進を問う③は、ポイントは3.9と3.8と普通の範囲だが、自由意見では「異なる関係者に質問すること学んだ」「関係者のお互いの大切さを理解した(以上マニラ)」、「日比の目的や理念の意見の違いがわかった(日本)」など重要な意見が寄せられている。

評価による学習効果を聞いた④はマニラ側3.9に比べ、日本側は4.4と0.5ポイントも高い。マニラの自由意見には「ロジックモデルで成果を考える重要性」や「評価設問に沿って質問すること」など評価の基本を学んだとある。次回以降、彼らのみで評価を実施できる可能性が高まったかもしれない。日本側でも同様の学びもあったと思うが、別の学びもあったようである。有給専従で構成さ

れているソルトマニラに比べ、ソルト日本は全員ボランティアである。二つの国の地理的距離もあり、マニラの現場からの情報は、忙しい日本人メンバーには十分に伝わらないこともあったのだろう。それが今回の評価を通じて一定程度、改善された実感が込められた数字というのが私見である。最後の質問は、筆者の働きぶりの評価だが、おおむね良い回答だった。しかし無記名といえども、目の前にいるファシリテーターへの批判的な意見は、書きにくいことも考えられる。

## 8. 結論

これまでの議論をふまえて以下の三つを本稿の結論としたい。

### (1) NGOプロジェクトにおける参加型評価とファシリテーターの役割

最初に、限られた事例の考察ではあるが、日本のNGOの多くが行っているようなプロセス型プロジェクトの改善の意味で参加型評価は効果的である、ということである。

ソルト・パヤタスの教育支援事業の評価を参加型で行うことによって、これまでの事業の一定の有効性や妥当性などの成果が明らかになったと同時に、目標設定の不明確さや、継続性の不足などの今後の課題が抽出された。これら成果と課題が参加型で議論され、関係者が自分たちの問題として認識したことにより、今後立案される新プロジェクトにおいて、今回の教訓が活かされ、より良い事業が実施される可能性は高まったと考えられる。

ソルト・パヤタスの日本人代表による報告書とメタ評価からは、「評価を下すまでのプロセスに、関係する全スタッフが参加したこと、評価設計、調査、結果分析など各段階において、細かく共有の時間が用意されたことにより、成果、教訓、課題について同時に考え、認識を共有することができた」「今回の評価終了時、現地スタッフから、『間をおかずに自分たちで新規事業のための調査と計画立案を行っていきたい』という希望があげられた」「参加型であるが故に、その体験を関係者が同時に体験でき、組織として今後活かす可

能性が高まりました」などが報告されている（ソルト2009）。

二番目は、参加型評価の長所を活かし欠点を補うために、外部ファシリテーターの存在が重要である、という点である。理屈では参加型評価は内部評価なので、外部者の支援がなくとも、内部の関係者だけで実施が可能である。ただし評価知識のばらつき、客観性確保、代表制の問題といった参加型評価の短所を、限られた内部の資源だけで克服することは困難である。外部者であり、評価知識を持ったファシリテーターが、関係者の評価プロセスを側面から支援することによって、これらの問題を最低限におさえ、参加型評価の利点である、能力開発や主体性の強化、効果的なフィードバックを確保して、プロジェクト全体を改善に導いていくことが可能となる。「第三者で、しかもNGOのことをよく知った専門家（ファシリテーター）であったからこそ、思いを引き出し、相手のモチベーションを高めるということが可能だったのではないかと思います。参加型評価は、自分たちだけでやれば、その結果に信憑性やその後の（改善への）影響が少なく、第三者の参加型評価の専門家を入れた時に効果が最大になると感じました」など、ソルト代表からのコメントである。

最後に、そのファシリテーターの役割の中身だが、ファシリテーターの他、カウンセラー及びトレーナーといった主に三つの仕事を総合的に担う必要がある。一般にファシリテーターは社会活動などを促進、援助するのが仕事であるが、参加型評価のファシリテーターの仕事内容を振り返ると、評価プロセス促進者の役割の他に、評価知識や技術を教えるトレーナーの役割や、評価全般に対する相談ニーズに応えるカウンセラー的な役割も必要であり（源2009）、単なる進行役にとどまらない、より広い能力が必要である。またこれらの役割を、必要に応じて臨機応変かつ柔軟に使い分けなければならず、評価技術を越えたコミュニケーション能力も求められる。

### (2) 残された課題

今後の課題としても、三つあげたい。

最初に、利害関係者の中での受益者の参加とエンパワーメントの問題である。今回はソルトにと

って初めての評価でもあり、評価設計から参加し、データ収集や分析に積極的に携わったのは、プロジェクトレベルのスタッフ、具体的にはソルト日本とソルトマニラのメンバーやスタッフにとどまった。結果としてスタッフレベルのエンパワーメントは一定程度、進展したと思われるが、プロジェクト現場には奨学生の子どもや保護者など、多数の受益者が存在する。今回受益者の参加は、データ収集の際の情報提供と、評価の報告を受けて結果を共有するレベルにとどまっている。ADBの参加の深さでいえば、レベル1の情報共有か、レベル2の相談にあたり、まだまだ積極的な参加とはいえない。参加型評価においては、受益者の参加を促進していくことは重要な視点であり、今後の課題であると考え。その一方で、多様な受益者の参加は現実的には困難な場合が多い。また評価目的によっては、受益者の参加の必要性が低い場合も考えられる。目的に沿って、どの利害関係者が、どのように参加することが望ましいのかを考える必要があるだろう。

次は、限られた時間と資金をどう使うか、というポイントである。内部評価の性格をもつ参加型評価は、深く参加するほど、関係者の作業量も増える。プロジェクト・スタッフは、日常業務としての事業運営も抱えており、評価の仕事はエクストラとなる。メタ評価でも「評価は大切だが、通常業務と平行して行うことは大変だ」というマニラスタッフの意見があった。今回二週間の現地調査の中で、一週間を評価設計に充てた。設計に時間をかけたことにより、データ収集は濃い密度で行われたともいえるが、ファシリテーターとしては、限られた滞在時間に収集や分析など全てのプロセスが終了できるか、大いに不安であった。参加型である以上、おおまかな予定はあっても、参加者の理解度や意見で進行する速度に大きな差がある。評価設計も含め余裕のある調査時間があることが望ましい。

また参加型評価といえども、外部ファシリテーターを活用する場合は派遣する費用負担が生じるので、多くNGOではその余裕がない。今回はJICAの「アドバイザー派遣制度」を活用して大変助かったが、評価に活用できるNGO支援の制度は少ないという現状もある。

この課題をじっくり考えると、単純にファシリテーターの工夫と努力の問題ではなく、NGOのプロジェクト・サイクルに関する制度化の問題であると気づく。事業の立案段階で、評価に必要な予算やスケジュールを計画的に組みこんでおけば、評価にかかる時間とお金の問題は、かなりの程度改善されるはずである。要は、計画・実施・評価・改善というPDCAサイクルを意識的に事業に組み込むか否かが、問われている訳である。

最後は、ファシリテーターの中立性の問題である。ファシリテーターの活躍で、参加型評価をよりよく実施することが出来る、というのが本稿の主張である。しかしながら、実際の評価現場では、さまざまな役割が必要であり、全ての役割を完璧にこなすことは不可能である。また進行役として、価値判断に中立であろうとしても、判断を手助けする中で、さまざまな助言を行うことになり、知らず知らずの内に、関係者の意志決定に影響を与えてしまう恐れがある。ファシリテーターは、常に自分の存在がどのように、評価プロセスに関わっているのか、慎重でなくてはならない。この件に関しては、太田（2007）は、社会開発における外部ファシリテーターの重要性を強調しながらも、「ファシリテーターの介入が善意によるものであればこそ、見落とされがちな落とし穴がそこにはある」と、利害関係者の自発性を尊重するファシリテーターでさえ、その恣意性から自由でないことに警鐘を鳴らしている。その弊害を押さえるためには、ファシリテーターがバイアスを持ち込んでいないかどうか、関係者によるチェックが必要である。今回のメタ評価では「ファシリテーターは適切に働いたか」と漠然と聞いたのみで、何がどう適切であった／なかったかはわからない。ファシリテーターへのメタ評価の方法論も、確立していく必要がある。

## 9. おわりに

参加型評価のファシリテーター専門家としての筆者が、今後の方向性として思っていることを述べて結びとしたい。

第一に、日本では参加型評価の実践事例がまだ

まだ少ないので、筆者の体験を含めもっとたくさん  
の事例を収集・研究していきたい。プロセス型  
事業に対する参加型評価の有効性も、複数事例に  
よる相互検討や、従来型評価との実践における比  
較の上で、さらに明らかにしていけるだろう。  
2009年9月に、日本評価学会の中で「参加型評価  
に関する研究分科会」を、学会員有志で立ち上げ  
たが、この会が内外の事例や理論を紹介し、活発  
な意見交換の場となることを願っている。

第二に、ファシリテーターの役割について、さ  
らなる追求である。ファシリテーターは、参加型  
評価の長所を伸ばし短所を補うだけでなく、全て  
の関係者のエンパワーメントに、もっと寄与して  
いけるのではないか。プロジェクト・スタッフだけ  
だけでなく、多数の受益者の人々が、参加の深さ  
におけるレベル5（協働）、レベル6（エンパワー  
メント）の段階に達していくために、外部ファシ  
リテーターはどのような貢献が可能なのか。おそ  
らくこの質問に答えるには、参加型評価だけでは  
なく、プロジェクトを形成して実施、その後評価  
していくという参加型のプロジェクト・サイクル  
全体の中での、ファシリテーターの役割を総合  
的に考察していく必要があるだろう。

最後に、「NGO海外プロジェクト強化のための  
アドバイザー派遣制度」を活用させていただいた  
国際協力機構地球ひろばの方々と、評価を決断・  
実施して私にファシリテーターの機会を与えて  
いただいた、ソルト日本の代表者やメンバーの方々、  
楽しい評価をともに行ったソルトマニラのスタッ  
フの面々を始め、忙しい日々の暮らしの中で時間  
を割いてくださった、フィリピンの受益者住民の  
皆さんに、多大の感謝の意を表して終わりとす  
たしい。

## 注記

- 1 ベン相関図は、重要性を円の大きさで、相互の関  
係性を円と円間の距離で表す。
- 2 文化人類学者川喜田二郎によって創設された、参

加型問題解決手法・発想法。マーケティングや品質  
管理、組織開発などに広く活用されている。

## 参考文献

- アーユス（2003）「日本のNGOによるプロジェクト評  
価の全体像と具体例」、『国際協力プロジェクト評  
価』、国際開発ジャーナル社、145-153
- 太田美帆（2007）「ファシリテーターの役割」、『テキ  
スト社会開発』、日本評論社、153-173
- 国際開発ジャーナル社国際協力用語集【第3版】（2004）  
国際協力事業団（2001）「参加型評価とは何か」、『国  
際協力と参加型評価』、国際協力事業団、1-21
- 佐藤寛（2007）「社会開発研究の課題～介入、計画、  
評価と社会の固有性」、『テキスト社会開発』、日本  
評論社、179-196
- ソルト・パヤタス（2009）『NGOプロジェクト強化の  
ためのアドバイザー派遣受入団体完了報告書』
- 源由理子（2008）「参加型評価の理論と実践」、『評価  
論を学ぶ人のために』、世界思想社、95-112
- 源由理子（2009）「第4講義 評価の設計」、『第三期評  
価士（初級）養成講座テキスト』、日本評価学会
- 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
- ADB（2001）. *Special Evaluation Study on Participatory  
Development Processes in Selected ADB projects in  
Agriculture, Natural Resources and Social Infrastructure  
Projects.*
- Estrella M., and Gaventa, J. (1998). *Who Counts Really?  
Participatory Monitoring and evaluation: literature  
review*, IDS.
- Narayan, D. (1993). *Participatory Evaluation: Tools for  
Managing Change in Water and Sanitation*, World Bank.
- UNDP (1997) *Who are the question makers? A Participatory  
Evaluation Handbook*, New York: Office of Evaluation  
and Strategic Planning.

(2010.2.8受理)

## **Participatory Evaluation of NGO Educational Project Facilitator Roles and Empowerment of Stakeholders**

Hiroshi Tanaka

Participatory Evaluation Facilitator  
The Institute for Himalayan Conservation  
nepalippine@gmail.com

### **Abstract**

Participatory evaluation, which is conducted by project stakeholders, holds advantages for effective assessment to improve programs. On the other hand, ensuring objectivity and the appropriate representation of stakeholders are challenges. In the participatory evaluation process, the evaluator plays the role of facilitator, not a position of assessor, as is the case of conventional evaluation.

The author facilitated the participatory evaluation of a process oriented educational project, which is implemented by a Japanese NGO, “Salt Payatas Foundation Philippines”. During the evaluation process, stakeholder empowerment and raising awareness for improvement were observed. However, insufficient evaluation skills and the tendency of subjective judgments were also encountered. Therefore, the facilitator made necessary measures corresponding to the situation.

This case study verifies; the effectiveness of participatory evaluation for process oriented projects, and also identifies three important roles of the facilitator in order to take advantage of participatory evaluation and make up for its weakness.

### **Keywords**

Participatory Evaluation, Facilitator, NGO, Evaluation Method, Empowerment, Process Oriented Project



## 【研究ノート】

# 地方自治体へのバランスト・スコアカード適用に関する研究 — 「財務の視点」に着目したフレームワークの検討—

佐藤 幹

広島大学大学院博士課程後期院生

d092567@hiroshima-u.ac.jp

## 要 約

バランスト・スコアカード（BSC、本文参照）は米国等の地方自治体では普及しているが、日本の地方自治体での適用例は少なく、かつ継続されていない。日本では「財務の視点」を民間企業におけると同様な取扱をしており、そこに無理があると思われた。そこで「財務の視点」に焦点を当てBSCのフレームワークを検討した。

民間企業の「財務の視点」の指標を整理した結果、その多くは株主満足を志向する企業の業績や成果を財務情報で表わすものであった。日本の地方自治体では、「財務の視点」の指標は財務情報による業績や成果を表すものではなく、財政の健全化を表すものがほとんどであった。また、米国等の地方自治体のフレームワークは民間企業のものとは異なっていた。

地方自治体の業績や成果を貨幣測定し財務情報で表すことは困難であり、一方、顧客である住民の福祉の向上のために、業績や成果を非財務指標で表すことが求められることから、「財務の視点」と「顧客の視点」を統合して、視点を3つとするフレームワークを用いるべきではないかとする仮説を導いた。

## キーワード

財務の視点、バランスト・スコアカード（BSC）、マネジメント、管理会計、地方自治体

## 1. はじめに

財政難を契機に1980年代から英国・米国等で、成果志向・顧客志向・市場機構の活用・分権化を運営原理とするニュー・パブリック・マネジメント（New Public Management: NPM）による行財政改革がすすめられた。日本では公共事業への過大な投資や多様な住民ニーズに応えた結果として、国の財政は逼迫し、地方自治体（地方公共団体）は厳しい財政難に陥った。このような状況を背景

として、経費の削減、説明責任の明確化等を目的とする行政評価<sup>1)</sup>に期待が寄せられてきた。

日本の行政評価は三重県で1996年から行われた事務事業評価が最初のものである。その後、他の自治体においても導入が進み、2007年の総務省の調査では、都道府県の98%（46団体/47団体）、政令指定都市では全ての団体で導入されている。

それにもかかわらず、国及び地方自治体の未曽有の財政難を象徴するかのよう北海道夕張市は破綻し、自治体の経営改善は喫緊の課題となり、

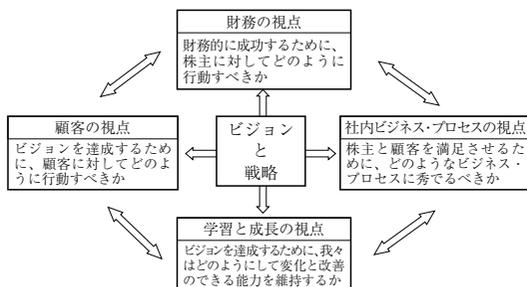
自治体の財政破綻を早い段階で防止することを目的として「地方公共団体の財政の健全化に関する法律」(以下、財政健全化法)が2007年6月15日に成立し、2009年度から全面施行された。さらに財政健全化の一層の強化等を目的に現在地方自治体の会計改革が進み、民間企業の会計方式の導入が行われつつある。

今まさに地方自治体への管理会計の適用を本格的に検討すべき時期が到来しているといえよう。また、近年、中央政府や地方自治体へのバランス・スコアカード(Balanced Scorecard: BSC)の導入も断続的に提唱されている状況にある。

BSCはKaplan and Norton(1992)で発表されたもので、当初は企業の業績評価ツールであったが、その後、戦略実行のツール、すなわち戦略的マネジメント・システムへと進化し、現在では戦略策定にも使われるほか、コミュニケーション・ツールや経営品質向上のためのツールともなっている。また、米国等においては、適用対象は民間企業だけでなく、NPO等の非営利組織や中央政府・地方自治体にも広がっている。さらに組織構成員の業績評価と報酬の結び付けにも活用されている。

そもそも、BSCは財務指標偏重の経営の反省のうえに立って、財務指標と非財務指標を複合的に活用するものである。加えて経営の視点を区分し、できるだけ少ない視点に絞込んで4つの視点から組織の業績を、よりよい方向に導くことを意図し、さらに4つの視点の関係性を可視化させて用いることが特長である(図表1)。

図表1 民間企業のBSCのフレームワーク



(出所) 吉川、2003、p.30を一部修正

地方自治体組織内の限られた資源で、そのミッション(使命)である住民の福祉を増進していくためには、従来型の行政運営を見直し、新たな自治体のマネジメント・システムを確立する必要がある。この10年間多くの地方自治体で行政評価が導入され、自治体の状況が変わるかと思われたが、大きな変化は見出せないままである。このままではさらに多くの地方自治体が破綻する可能性がある。財政難に陥っている地方自治体を再生へと向かわせることに加えて、歳入である税金を価値ある歳出にするべく実効性のあるマネジメント・システムの導入が必要である。そのためには管理会計理論の蓄積やツール、特にBSCの適用を促進するための研究を一層充実させるべきであると考えらる。

ただし、BSC適用の研究にあたっては官僚制の弊害などの一般的な阻害要因のほか、具体的な阻害要因についても検討する必要がある。例えば、行政評価の拡充のためにBSCを適用しようとする場合などである。戦略マップ<sup>2</sup>を含めたBSCは単なる業績測定ツールではなく、行政評価とは異なるものであり、BSC適用の障害となっている。BSC適用の目的は新たなマネジメント・システムの導入であることを明確に意識して適用する必要がある。

また、民間企業のフレームワークをそのまま適用しようとするのが障害となっている。地方自治体での「財務の視点」・「顧客の視点」は民間企業のそれらとは大きく異なり、そのあり方を十分に検討せず、そのまま適用しようとするのが阻害要因となる。この要因を解消することを課題として、本稿では「財務の視点」に焦点を当てフレームワークを検討するものである。

## 2. 先行研究

米国連邦政府の管理会計の研究としては小林(2002)、藤野(2003)があり、前者は米国地方自治体の管理会計も取上げている。また、双方ともにBSCについても若干の言及がある。

主に日本の中央政府を対象としたもので管理会計やBSCに論及したものとしては、山本(2000)、

山本（2001）がある。

しかし、日本の地方自治体への管理会計の適用に関する研究は少なく、中でもBSCの適用についての研究は極めて少ない。

櫻井（2002）は米国シャーロット市でのBSC導入の事例を紹介することにより、日本の中央政府・地方自治体における行政評価の展開にBSCが適用できる可能性を示唆したものである。

稲生（2004）は日本の地方自治体のBSC導入事例を詳細に紹介するものであるが、パブリック・ガバナンスを切り口とするものであること、また、民間企業においても経営可能な病院等を取上げるものである。しかし、BSCの視点に関する指標の意義や戦略マップにも言及しており参考になる。加えて福田（2004）は質問票調査を行い、全国の導入状況をまとめたものである。千代田区役所の具体的な事例を取上げるものとしては松尾（2006）がある。また、松尾（2009）は全国の地方自治体の行政評価を比較・分析したものであるが、千代田区役所や福岡市役所でのBSCの導入またはその検討についても取上げている。志村・石田（2009）は公共セクターへのBSCの適用について検討したものであり、「財務の視点」に着目してBSCのフレームワークのあり方を問うたものである。その他、石原（2004）、森岡（2006）、権（2006）、櫻井（2007）、櫻井（2008）、などをあげることができる。

どにかかる情報が例示されている。その中でもBSCの「財務の視点」としては、成果の貨幣的測定情報である収益性と将来の収益に結びつく可能性のある成長性の情報が中心となっている。

吉川（2004）は「財務の視点」における業績指標を大きく安定性、収益性、成長性、生産性、その他の5区分に整理している<sup>3</sup>が、BSCの適用のため、つまり組織のマネジメントを向上させて成果をあげるための指標に絞ったものではなく、経営分析に用いられる網羅的な指標を整理したものである。

## (2) 民間企業の「財務の視点」の指標の分類

現代企業は利益最大化を唯一の指標とするものではなく、多元的目標を具体的な業績測定指標によって指示し、創造的でしかも効率的なバランスのとれた経営を目指すものである。Kaplan and Norton（1996）は企業のライフサイクル（成長期、持続期、収穫期）とそれぞれの企業の戦略により、選定する指標は異なることを紹介している<sup>4</sup>。これは指標の分類の先例である。櫻井（2008）はこの分類を評価しつつも、収益の成長と新製品やサービスの組み合わせは究極的には収益性に、資産の有効利用と生産性は効率性に関連するものであるとし、収益性、コスト低減、効率性に区分して指標を分類し例示している（図表2）。

図表2 「財務の視点」の業績指標の分類例1

収益性	コスト低減	効率性
売上高	製造原価	総資産利益率
経常利益	総原価	使用資本利益率
売上高利益率	資本コスト	株主資本利益率
営業利益	販管費	棚卸資産回転率
貢献利益	物流費	1人当たり収益性
EVA	人件費	1人当たり付加価値
キャッシュ・フロー	販売促進費	1人当たり売上高
スループット	研究開発費	1人当たり機械設備

（出所）櫻井、2008、pp.220

## 3. 指標から見た「財務の視点」

### (1) 民間企業で用いられる財務指標

財務諸表における「利益」は企業の過去の経済的な業績や成果を貨幣的に測定した情報である。民間企業は経済的合理性を追求する団体・組織であり、最少の経済的犠牲をもって最大の経済的効果をあげることを目的にしている。つまり企業努力（価値犠牲）を費用の形で、企業の成果（価値獲得）を収益の形で把握して、両者を期間的に対応させることによって企業の成果たる利益を貨幣測定し計算上獲得するのである。

先行研究では財務情報として、収益性、安全性、効率性、成長性、生産性、キャッシュ・フローな

しかし、筆者は、櫻井（前掲）が効率性に分類している指標は「内部プロセスの視点」の指標であると考えるので、次に示すNivenの整理による成長性と収益性と価値創造の3つの分類<sup>6</sup>（図表3）の方が妥当であると考えられる。

図表3 「財務の視点」の業績指標の分類例2

成長性	収益性	価値創造
新製品の収益 ターゲットマーケット におけるシェア	グロスマージン 純利益 売上高利益率	投資利益率:ROI 経済付加価値:EVA

(出所) 松原訳、2004、pp.156-159より筆者作成

なぜなら民間企業では「財務の視点」は、主に株主満足の見点であるといわれ、そこで示される指標は財務諸表や財務分析中に現れる「売上高」・「利益」・「キャッシュ・フロー」や「売上利益率」・「貢献利益率」・「投資利益率」などが中心となるものであり、株主満足を志向する企業の業績や成果を財務情報で表わすことに重点が置かれている。したがって「コスト低減」といった組織の成果に対して消極的な指標群より、「成長性」・「価値創造」という積極的な指標群の方が役立ち、組織のパフォーマンスの向上をより適切にマネジメントできると考えるからである。

この章では民間企業の「財務の視点」の指標や指標群の分類を見てきたが、これによって、「財務の視点」は財務情報による指標・指標群を束ねたもので、それらの活用により、その業績や成果を向上させるというマネジメントの要となるものであることが分かった。つまり「財務の視点」は民間企業のBSCのフレームワークにおいて、欠くべからざるものなのである。

#### 4. 地方自治体での導入状況

日本では、BSCの体系構築の検討や構想を策定したものの、実際の運用にいたらなかった自治体がほとんどで、BSCを実際に適用するにいたった事例は数件にとどまっております、継続して適用している自治体はほとんどない。

稲生（前掲）によると、第1は導入済みで実際に適用している自治体、第2はBSCに基づく経営体系を構築したものの実施にいたらなかった自治体、第3は検討のみが行われた自治体があり、第1の例として横浜市、姫路市、練馬区、第2の例として千葉県、第3の例として札幌市と横須賀市を代表的な事例としてあげることができる<sup>7</sup>として

いる。福田（前掲）によると人口20万人以上の自治体のうち、BSCを「導入済み」が1件、「試行中」が1件となっており、また、「検討中」が10件、「検討したが導入しなかった」が5件であり<sup>8</sup>適用事例が少ないことがわかる。

なお、三重県病院事業庁のようにBSCを体系的に構築して実際に運用している事例はほかにもあるが、本稿では民間でも経営可能な病院・学校等は対象としていない。

##### (1) 地方自治体での事例分析

日本の地方自治体でBSCを本格的に適用した例は少なく、その中でも継続しているのは、筆者の調査では、地方自治体のうちでも基礎自治体である横浜市役所と姫路市役所の2つである。これに加えてかつてBSCを導入していた千代田区役所等の事例を取上げて分析する。

##### ①横浜市役所の事例

横浜市役所はホームページの統計ポータルサイトにおいてBSCの視点に模して市の統計情報を公開している。それには「市民の視点」、「財務の状況」、「業務プロセス」、「職員力と変革」の4つの視点が示されている。ここでは一般に使われる「財務の視点」に相当するものとして「財務の状況」という名称を用いている。

「財務の状況」はさらに収支、安定性、自立性、柔軟性、起債運営、生産性の6つの項目から構成され、図表4のように整理される。この表の下側

図表4 横浜市役所と民間企業の指標の比較

横浜市役所の指標					民間企業の指標	
財政の健全性					その他	
収支	安定性	自立性	柔軟性	起債運営	生産性	
プライマリー バランス	負債対標準 財政規模	自主財源比率	行政コスト合計 /収入合計	市債依存度	市民1人当り 行政コスト	
実質収支比率	流動比率	財政力指数	経常収支比率	市民1人当り の市債残高	市民10万人当り の職員数	
		地方税割合		市債の格付け		
		市税収納率				
成長性	収益性		価値創造			
新製品の収益	グロスマージン		投資利益率:ROI			
ターゲットマーケット におけるシェア	純利益		経済的付加価値:EVA			
	売上高利益率					

(出所) 筆者作成

は図表3を再掲して並列に示したものである。2つの表に掲げる指標等を比較すれば、かなり異質のものであると分かる。横浜市役所の「財務の状況」は、市民サービスを支える財務の健全性や自立性などを市民に把握してもらうためのものとされており、ほとんどの指標が財政の健全性を判断するためのものである。また、「市民1人当たり行政コスト」、「市民10万人当たりの職員数」などは生産性（効率性）についての指標であり、業績や成果を表す指標ではない。

### ②姫路市役所の事例<sup>10</sup>

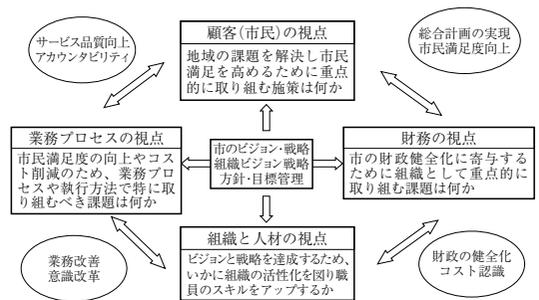
姫路市役所はBSC手法を適用した行政評価システムを運用することで、市民サービスや行政活動の費用対効果を高め、市民本位のサービス提供に努めている。この手法は、市民満足度は常に向上しているかという「顧客の視点」、コストは改善しているかという「財務の視点」、仕事の企画や効率化は行われているかという「プロセスの視点」、職員能力向上や組織活性化の「組織・人材の視点」の、4つの視点で経営改善を図っていく方法のことで、以下の5つの目的があるとされる。

- ・ 成果重視型の行政経営へのシフト
- ・ 市民に対しての説明責任
- ・ 職員の意識改革
- ・ マネジメントサイクルの定着化
- ・ 財政の健全化

また、2005年3月当時の総務局行政部行政改革課の資料によれば、「財務の視点」は市の財政健全化に寄与するために、組織として重点的に取り組む課題は何かという、自らへの問いかけであり、単当たりコストを確認して効率は改善したか、また、各種財政指標の状況から財政健全化が図られているかを確認して、市民満足の向上を目指すようマネジメントに活用するものである<sup>11</sup>としている（図表5）。

総務省公表の姫路市役所の資料<sup>12</sup>によると、2003年から事務事業評価と組織経営評価を同時に実施しているとされる。組織経営評価とは課・室の組織単位で、「顧客」、「コスト財務」、「プロセス」、「組織人材」というBSC手法の4つの視点から改革に取り組み、組織の経営力向上を図るマネジメント・システムであるとしている。行政シス

図表5 姫路市役所のBSCのフレームワーク



(出所) 内閣府経済社会総合研究所WEBサイト

テム改革室の組織経営評価シート（BSC総括表）に記載されている目標は「行政システム改革の推進により健全財政に寄与する」、「改革プランによる18年度経費削減額11億円以上をめざす」、「公共施設の建設から維持管理・運営までのトータルコストの削減」であり、要約すると財政健全化とコスト削減・効率化を目指すものである。

### ③千代田区役所の事例<sup>13, 14</sup>

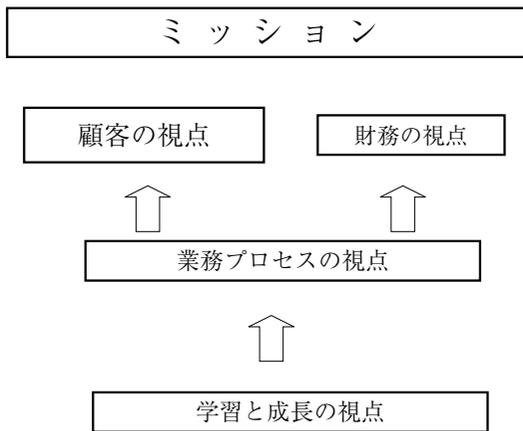
千代田区役所では民間企業に導入されていたカンパニー制にならない事業部制を導入していたが、事業部制を推進するため、組織の業績や成果の評価手法としてBSCを導入したが、現在は適用していない。

当時の戦略マップのアウトラインは図表6のとおりである。これをBSCのフレームワークとしてみると、ゴールとして「ミッション」は掲げられてはいるものの、やはり4つの視点で構成されている。最上位には「顧客の視点」と「財務の視点」が並列に配置されているが、「財務の視点」の占める割合が小さいところが特徴である。

図表6はKaplan and Norton（2000）<sup>15</sup>などを参考に、公共セクターにおける「顧客の視点」と「財務の視点」の関係やあり方について検討した結果たどりついたものであろうと思われる。

各々の視点について、「顧客の視点」は事業・施策の成果の視点からの評価、「財務の視点」は財政健全化の視点からの評価、「業務プロセス」の視点は業務改善の視点からの評価、「学習と成長の視点」は職員活力の視点からの評価として捉えられている。

図表6 千代田区役所のBSCの戦略マップ



(出所) 松尾、2006、p.160及び松尾、2009、p.195を参考に筆者作成

#### ④その他の事例（福岡市<sup>16</sup>、練馬区<sup>17</sup>）

福岡市役所では、BSCの導入が検討されたが、結果的には戦略マップやスコアカードの導入は行っていない。しかし、分権化による自律的な経営の観点から局単位に導入された「組織マネジメントシステム」の目標管理の項目として、最上位に顧客の視点に相当すると考えられる「施策推進に関する目標設定の視点」を設けたほか「資源に関する目標設定の視点」、「業務改革に関する目標設定の視点」、「人材育成・活性化に関する目標設定の視点」といったBSCの視点に対応したものを掲げてマネジメントに活かそうとしている。しかし、「資源に関する目標設定の視点」とは「財政健全化プランに基づく事業の見直し等に関すること」と「事務事業のコスト縮減等に関すること」を掲げており、横浜市役所と同様に業績や成果を表すものは掲げられていない。

練馬区役所では役所組織全体のBSCではなく、文書事務改善という個別業務について適用している。しかし、視点の設定に工夫が見られ参考になる。行政の目指すものは住民福祉の向上や地域社会の発展など非財務的な目標が多いため「財務の視点」の代わりに総合計画などの施策目標などで表される「業績の視点」を用い最上位の視点とし、次に「顧客の視点」を位置づけ、その次に業務処

理効率や人件費に関するものとして「業務プロセスの視点」を用い、最後に「学習と革新の視点」を配置している。このように個別業務に対してBSCを用いようとする民間企業のBSCのフレームワークの既成概念から逃れること、つまり「財務の視点」を適用対象の特性に応じて適切にカスタマイズできることが分かる。

#### (2) 事例における「財務の視点」の解釈

横浜市役所の「財務の状況」は、図表4に示したように、市民サービスを支える財務の健全性や自立性などを把握するためのもので、ほとんどの指標が財政の健全化を表すものであった。また、「市民1人当たり行政コスト」、「市民10万人当たりの職員数」などは生産性（横浜市は生産性と表記しているが効率性のことであると思われる。）についての指標であり、業績や成果を表す指標ではない。また、姫路市役所の「財務の視点」は横浜市役所同様、財政健全化とコスト削減を目指すものであるとしており、業績や成果を表すものとはなっていない。福岡市においても「財務の視点」は財政健全化とコスト縮減を目指すものと捉えていた。

財政危機が大きく意識されているためか、これらの事例では、「財務の視点」を財政健全化と効率性を目指す視点として生かそうとしており、民間企業のように組織の業績や成果を貨幣額により測定し、その結果をマネジメントに活用するという、本来のBSCの使い方とは異なるものである。また、このことは石原（前掲）で示される札幌市の事例でも同様である。

民間企業のBSCの「財務の視点」の指標は、組織の業績や成果を財務情報で表すものがほとんどである。民間企業の会計では組織の業績や成果を貨幣額により測定し表示することが基本である。しかし、地方自治体ではNPMの観点から顧客志向の立場をとったとしても、利益の概念（業績や成果を貨幣額で測定・表示するという意味において）のない公会計制度の中で、実施した事務事業等の成果を民間企業が採用しているような財務情報による指標を用いて表すのは困難である<sup>18</sup>。言い換えれば、地方自治体では利益の概念に基づいて、業績や成果を測定することができないのである。

現在、日本では中央政府の財政危機を受けて、少なからぬ自治体において、これまで経験したことのない財政難に陥り、北海道夕張市の破綻をきっかけに、2008年度から財政健全化法が一部施行されたところである。初年度となった2008年度では1,857の地方自治体のうち50程度の自治体が、財政健全化法等で示された基準値を上回って、財政状態が悪いという結果となった。

これら約50の自治体にとっては、財政の健全化こそが、重要であり、「顧客の視点」などその他の視点の前提的なものと捉えられているといえよう。地方自治体の役所等が破綻することは、住民にとって最も困ることの1つなのである。

また、地方交付税が交付されないほど豊かな税収のある地方自治体や初年度は財政健全化法等の基準値をクリアした自治体にあっても、財政健全化が自治体組織の存続のための基礎前提であることは間違いのないことである。

しかし、BSCを目的実現・目標達成のためのマネジメント・システムであると捉える観点からは、地方自治体組織の業績や成果の向上、つまり住民福祉の向上に注力するものでなければならない。そのためには業績や成果を表す非財務指標をしっかりと活用する必要がある。加えてこれらの指標群を束ねる視点の検討も必要であり、さらに視点の捉え方の観点からフレームワークの再構築が求められる。

## 5. 自治体BSCのフレームワーク

### (1) 日本の地方自治体への適用

従来から地方自治体の組織の業績や成果は、金銭的な評価つまり貨幣的測定が困難であるとされてきた。しかし、BSCは財務指標に加えて非財務指標を活用している点で、地方自治体にも有効ではないかと考える。また、山本（2001）によると企業の管理会計分野でもBSCなどのように非金銭的尺度を含む多次元的な業績測定システムに関心が集まっているが、もともとBSCは公的部門の業績測定にヒントをえたとしている<sup>19</sup>。しかし、日本の地方自治体での適用例が少ないということは、何らかの問題があるからであり、民間企業用

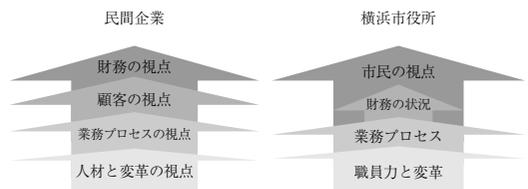
のフレームワークをそのまま用いようとするのも問題の1つではないかと考える。それゆえ本稿では、特に「財務の視点」の捉え方や扱いを検討すべきであることを示してきたところである。

志村・石田（前掲）によれば「顧客の視点」や「内部プロセスの視点」との因果連鎖の欠如を理由に、民間セクターではフレームワーク中、最上位に置かれている「財務の視点」（利益額等の指標に象徴されている）は、公共セクターのもの（財政の健全化）とは異質なものであるとし、公共セクター組織における「財務の視点」が、住民満足の上昇を判断するための組織の業績や成果を表していないので、公共セクター組織のフレームワークの視点として含めるべきではないとされている。

「財務の視点」の見直しについては筆者も同様の見解である。確かに財政の健全化は地方自治体組織や職員の努力の賜物ではあるが、顧客である住民の満足を上昇するという意味での本来の業績や成果ではなく、本来の業績や成果を上げるための基礎前提であり、組織のパフォーマンスの向上（すなわち住民福祉の向上）を目指しておこなうマネジメントのための成果指標群・視点として、「財務の視点」は適切ではない。

ここで再度、民間企業と横浜市役所の「財務の視点」等の位置づけを比較してみたい（図表7）。

図表7 「財務の視点」等の位置づけ



(出所) 左図 吉川、2006、p.109から  
右図 横浜市のWEBサイトより作成

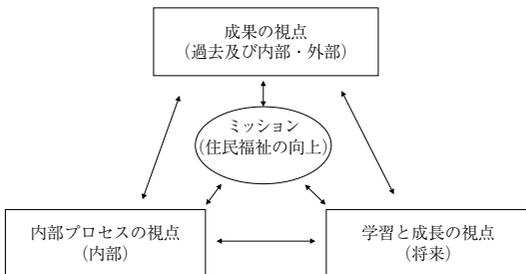
図表7は、これまでの議論を端的に可視化したものであるといえよう。地方自治体では「財務の視点」と「顧客の視点」は渾然一体であり、「財務の視点」は「顧客の視点」に内包されるべきものとも考えることもできる。組織の業績や成果を貨幣的に測定表示できない以上、少なくとも日本の

地方自治体では「財務の視点」は単独では用いないほうが良いのである。

(2) 仮説の提示

地方自治体の「財務の視点」の指標には財務情報による成果を表す指標がほとんどないことが分かった。民間企業では「財務の視点」は、主に株主満足の視点であり、「顧客の視点」はいうまでもなく顧客満足の視点である。しかし、民間企業においても「財務の視点」に関するステークホルダーは株主が中心とはいうものの債権者や顧客も含まれ、「財務の視点」と「顧客の視点」を厳格に区別できていない。また、地方自治体の業績や成果を財務情報で表すことは困難であり、それだけでは不十分であることが、米国の公会計基準審議会の報告<sup>20</sup>からもうかがえること、さらに、顧客志向を理論の柱の1つとするNPMの立場からすれば、顧客である住民の福祉の向上のために業績を上げ、成果を出すのが地方自治体の役割である。このようなことから両者を一体と見るのが現実的である。また、練馬区の事例のように「財務の視点」を採用していないものもある。したがって本稿では「財務の視点」と「顧客の視点」を統合して視点を3つとすることが妥当ではないかと結論づける。また、その名称を「成果の視点」としたい（図表8）。視点の数が少ないほどマネジメントに活用し易いという観点からも望ましいものであろう。

図表8 本稿仮説によるBSCのフレームワーク



(出所) 筆者作成

6. 参考（米国等での視点の捉え方）

ここでは米国等でのBSCのフレームワークのアウトライン（視点の配置）を表にして比較しておく。まず左から民間企業のもの、英国ソリハル市のもの、次に米国での成功例といわれているシャーロット市の2000年当時のものと現在のものを並べると図表9のようになる。これを見ると「財務の視点」に相当する視点が徐々に下位に配置されている。

図表9 英国・米国自治体等の視点の比較

営利企業の視点(当初)	ソリハル市の現在の視点	シャーロット市の以前の視点	シャーロット市の現在の視点
Financial	Customer	Customer	Serve The Customer
Customer	Financial	Financial	Run the Business
Internal Business	Business Process	Internal	Manage Resources
Innovation and Learning	Learning and Growth	Learning and Growth	Develop Employees

(出所) 各市のWEBサイト等から筆者作成

また、練馬区役所の事例であったように、個別業務のBSCのフレームワークを検討する一例として内部監査部門のフレームワークのアウトラインの比較を示す（図表10）。

Wong（2000）とKerr（2006）は個別業務としてのものというより、組織全体のマネジメント向上のための内部監査部門のツールとしてフレームワークを提示しているようであり、シャーロット

図表10 内部監査業務の視点の比較

Wong(2000)	Kerr(2006)	Frigo(2002)	シャーロット市
Financial	Mission	Audit Committee	
Customer	Customer Focus	Management and Auditees	Run the Business
Internal/Business	Internal Processes	Internal Auditing Processes	Manage Resources
Learning and Growth	Learning & Knowledge	Innovation and Capabilities	Develop Employees
	Financial		

(出所) 筆者作成

市の内部監査部門のものには「顧客の視点」に相当するものがないのが特徴である。

## 7. おわりに

今後の課題の1つとして適用範囲に関するものがある。横浜市役所、姫路市役所でのBSCの適用上の特徴は、組織のマネジメントではなく地域へのマネジメントを対象としていることである。市役所の目的はその行政区域の福祉の向上・地域の発展であることから地域を適用範囲とするのは当然のことのように思われるが、民間企業での適用範囲とは大きく異なるものである。

また、シャーロット市はマネジメント・システム及び戦略策定のツールとしてBSCを適用したとされるが、ここにおいても組織のマネジメントを超えて、地域へのマネジメントを対象としており適用範囲が広い。今後その理由を検討したいと考えている。

ただし、現時点での筆者の見解としてはBSCを適用するにあたっては、組織のマネジメントの活用にとどめ、その限界を把握したうえで、将来的に地域へのマネジメントへと活用の範囲を広げることが妥当であるように思っている。この意味では民間企業での適用と同様に、まず、自治体組織へのマネジメントに適用した千代田区役所の取組は当を得たものであったといえよう。今後は米国等の事例を含めてさらに調査研究を行いたい。

## 謝辞

本稿に対し貴重なコメントをいただいた査読者の方に深く感謝の意を表す次第である。

## 注記

- 1 本稿でいう行政評価とは事務事業評価を典型とする業績測定のことを意味する。
- 2 BSCのためのツールであり、戦略を記述し組織的に共有するための論理的・包括的なフレームワークのことである。
- 3 吉川 (2004)、pp.132-133。
- 4 吉川 (1997)、p.81。

- 5 櫻井 (2008)、pp.220-221。
- 6 松原 (2004)、pp.156-159。
- 7 稲生 (2004)、p.27。
- 8 福田 (2004)、p.102。
- 9 <http://www.city.yokohama.jp/me/stat/bsc/> (2009年4月29日現在)
- 10 [http://www.city.himeji.lg.jp/s20/2212947/\\_1805.html](http://www.city.himeji.lg.jp/s20/2212947/_1805.html) (2009年4月29日現在)
- 11 <http://www.esri.go.jp/jp/workshop/050318/050318maekawa.pdf> (2009年4月29日現在) の「財務の視点」に関する部分の要約。
- 12 [http://www.soumu.go.jp/iken/pdf/070328\\_5\\_16.pdf](http://www.soumu.go.jp/iken/pdf/070328_5_16.pdf) (2009年4月29日現在)
- 13 松尾 (2006)、pp.148-169。
- 14 松尾 (2009)、pp.191-200。
- 15 櫻井 (2001)、p.176。
- 16 松尾 (2009)、pp.186-189。
- 17 片田 (2003)、pp.125-128。
- 18 NPMの影響を受けて導入された指定管理者制度や「競争の導入による公共サービスの改革に関する法律」に基づく、民間企業等へのアウトソーシング可能な事務事業等には適用できる。
- 19 山本 (2001)、p.9。
- 20 米国公会計基準審議会 (Governmental Accounting Standards Board) の概念報告書第2号「サービス提供努力と成果 (Service Efforts and Accomplishment) 報告」

## 参考文献

- 石田晴美 (2006) 『地方自治体会計改革論』、森山書店
- 石原俊彦・INPMバランス・スコアカード研究会 (2004) 『自治体バランス・スコアカード』、東洋経済新報社
- 稲生信男 (2004) 「行政経営とガバナンス型Balanced Scorecard (BSC) に関する一考察」、『会計検査研究』、30：11-30
- 片田保 (2003) 「ポスト電子自治体の戦略行政」、『地方財務』、6月号：119-129
- 加登豊 (2000) 『管理会計入門』、日本経済新聞社
- 権永錫 (2006) 「行政におけるBSCの適合性と成果主義に関する考察」、『同志社政策科学研究』、93-110
- キャプラン, R.S.・ノートン, D.P. (吉川武男訳) (1997)

- 『バランス・スコアカード～新しい経営指標による企業変革～』、生産性出版
- 小林麻里 (2002) 『政府管理会計—政府マネジメントへの挑戦—』、敬文社
- 櫻井通晴 (2001) 『キャプランとノートンの戦略バランス・スコアカード』、東洋経済新報社
- 櫻井通晴 (2002) 「行政評価へのバランス・スコアカードの適用—シャーロット市のバランス・スコアカードの適用」、『企業会計』、中央経済社、54 (5) : 4-11
- 櫻井通晴 (2007) 「わが国の公的機関における効率性と有効性の必要性—管理会計による効率性と有効性追求の方法論—」、『会計検査研究』、36 : 9-17
- 櫻井通晴 (2008) 『バランス・スコアカード (改訂版) —理論とケース・スタディー—』、同文館出版
- 志村正・石田晴美 (2009) 『パブリック・セクター組織におけるBSCの適用』、文教大学大学院情報ジャーナル4 (2)
- 谷武幸編著 (2004) 『成功する管理会計システム—その導入と進化—』、中央経済社
- 福田寛也 (2004) 「バランス・スコアカード (BSC) による行政経営—地方自治体の事例からの考察—」、『Kwansei Gakuin policy studies review』、3 : 93-106
- 藤野雅史 (2003) 『政府における管理会計の歴史的展開』、一橋大学博士論文 (甲第174号)
- 松尾敏充 (2006) 『地域デザインフォーラム・ブックレットNo.13 政策評価制度 (総集編)』、大東文化大学 国際比較政治研究所 市域連携班、148-169
- 松尾貴巳 (2009) 『自治体の業績管理システム』、中央経済社
- 森岡一憲 (2006) 「政府部門の評価システムへのBSCの導入」、『経済科学研究』、9 (2) : 185-197
- 山本清 (2000) 『自治体経営と政策評価—消極的顧客主義を超えるNPMを一』、公人の友社
- 山本清 (2001) 『政府会計の改革—国・自治体・独立行政法人会計のゆくえ』、中央経済社
- 吉川武男 (2004) 『バランス・スコアカード構築』、生産性出版
- 吉川武男 (2006) 『バランス・スコアカードの知識』、日本経済新聞社
- Frigo, Mark L. (2002). *Balanced Scorecard Framework for Internal Auditing Departments*, The IIA Research Foundation.
- Kaplan, Robert S. and Norton, David P. (1992). The Balanced Scorecard —Measures That Drive Performance. *Harvard Business Review*, January-February, 71-79.
- Kaplan, Robert S. and Norton, David P. (2004). *Strategy Maps: Converting Intangible Assets into Tangible Outcomes*, Harvard Business School Press, 1st ed., 00-00.
- Kaplan, Robert S. and Norton, David P. (2000). *The Strategy-Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment*, Harvard Business School Press; 1st ed., (櫻井通晴訳『キャプランとノートンの戦略バランス・スコアカード』東洋経済新報社、2001年)
- Kerr, Deborah L. (2006). Using the Balanced Scorecard Planning and Conducting Performance Audits, Mid-America Intergovernmental Audit Forum, Spring Meeting, 2006. <http://www.auditforum.org/speaker%20presentations/mamiaf/mamiaf%2005%202006/kerr.pdf> (2009年9月5日現在)
- Niven, Paul R. (2008). *Balanced Scorecard: Step-by-Step for Government and Nonprofit Agencies*, 2nd Edition, Wiley.
- Niven, Paul R. (2002). *Balanced Scorecard Step-by-Step: Maximizing Performance and Maintaining Results*, New York: John Wiley & Sons, Inc., (松原恭司郎訳『ステップ・バイ・ステップ バランス・スコアカード経営』中央経済社 2004年)
- Niven, Paul R. (2003). *Balanced Scorecard: Step-by-Step for Government and Nonprofit Agencies*, New York: John Wiley & Sons, Inc. (吉川武男、柿崎平訳『行政・非営利組織のバランス・スコアカード—卓越した組織へのロードマップ』、生産性出版、2006年)
- Wong, Jeff, The Role of Balanced Scorecard in Operational Auditing, *Internal Auditing*, July/August, 2000, 33-36.

(2010.2.2受理)

## **The application of Balanced Scorecard to the local government － taking the financial perspective －**

Tsuyoshi Sato

Hiroshima University Graduate School of Social Sciences Management Program  
d092567@hiroshima-u.ac.jp

### **Abstract**

Balanced Scorecard (BSC) is well utilized by local governments in the U.S.A. However, there are a few good examples in Japan as well. The financial perspective, which takes an approach similar to a company, is in relation to BSC. From the financial perspective, indicators express achievements and results for stakeholder satisfaction in financial terms. Measuring achievements monetarily is difficult, especially for local governments. Furthermore, the aim of local governments differs from the aim of for-profit companies, which the financial perspective advocates. Customer satisfaction, which is not a financial indicator, must also be taken into consideration by local governments. Therefore, it is necessary to integrate the financial perspective with the customer perspective.

### **Keywords**

Financial Perspective, Balanced Scorecard, Management, Management Accounting, Local Government



## 日本評価学会第10回全国大会開催のご報告とお礼

第10回全国大会

実行委員長 源 由理子

プログラム委員長 牟田 博光

2009年11月28-29日開催（於：明治大学）の日本評価学会第10回全国大会には約130以上の方々にご出席いただき誠にありがとうございました。会員各位の日頃の研究や実践活動の報告を基に、評価研究者、実務者の間の経験、情報、知識を共有化する場として、有益な機会となりましたら幸甚に存じます。多くの方のご尽力により本大会を開催することができましたこと、心より感謝申し上げます。今後とも当学会の活動にご高配賜りますようお願い申し上げます。

### 第1日：2009年11月28日（土）

9：30-10：00	受 付		
午前の部 10：00-11：30		自由論題 1 学校評価／学力分析 (橋本昭彦) Rm.1094	自由論題 2 政策評価 (湊直信) Rm.1096
お昼休み 11：30-12：45	お昼休み (理事会11：35-12：30) Rm.1092		
午後の部 I 12：45-14：45	共通論題 1 日本のODA評価の現在 (和田義郎) Rm.1093	自由論題 3 評価者 (長尾真文) Rm.1094	自由論題 4 自治体評価／地域開発 (梅田次郎) Rm.1096
午後の部 II 15：00-17：00	シンポジウム 「協働社会と評価～人びとの生活世界から評価の役割を考える」 (源由理子／大島巖) Rm.1093		
17：15-18：00	総 会 Rm.1093		
18：15-20：00	懇親会 リバティタワー23F		

### 第2日：2009年11月29日（日）

9：30-10：00	受 付		
午前の部 10：00-12：00		自由論題 5 評価手法 (三好皓一) Rm.1084	自由論題 6 実践報告 (林薫) Rm.1085
お昼休み 12：00-13：15	お昼休み		
午後の部 I 13：15-15：15	共通論題 2 政権交代とマニフェスト評価 ～政治主導体制下の政策評価のあり方を考える～ (田中弥生) Rm.1083	共通論題 3 女性政策をめぐるわが国の現状 －「評価」の視点から－ (山谷清志) Rm.1084	
午後の部 II 15：30-17：30	共通論題 4 地方自治体における公民パートナーシップの評価 (源由理子) Rm.1083		

( ) は座長です。

自由論題報告の時間は原則として、発表20分、質疑8分、交替2分（計30分）です。

日本評価学会第10回全国大会  
「協働社会と評価」プログラム詳細

11月28日（土）午前の部 10：00-11：30

自由論題 1 学校評価／学力分析				座長	橋本 昭彦	国立教育政策研究所	Rm.1094
学校経営の評価に関する一考察		石田 謙豪	広島県尾道市立栗原北小学校				
高等学校の学校評価における適切な評価指標の在り方		高月 淳	広島県立忠海高等学校				
数学・読解・科学の学力に影響を与える要因の分析		福田 優一	東京工業大学				
自由論題 2 政策評価				座長	湊 直信	(財)国際開発高等教育機構	Rm.1096
				コメンテータ	林 薫	文教大学	
日本の対太平洋島嶼国協力の政策評価と太平洋・島サミット		野田 真里	中部大学				
途上国のモニタリング・評価制度の新たな課題に関する考察： ガーナを事例として		三輪 徳子	茨城大学				
モンゴルと東欧移行諸国の公共財政問題の比較評価		上野 宏	南山大学				

11月28日（土）午後の部Ⅰ 12：45-14：45

共通論題 1 日本のODA評価の現在				座長	和田 義郎	政策研究大学院大学	Rm.1093
				コメンテータ	牟田 博光	東京工業大学	
政策レベルのODA第三者評価		畠 薫	外務省				
財務省におけるODA評価		高見 博	財務省				
国際協力機構の評価制度と取り組み		遠山 慶	(独)国際協力機構				
自由論題 3 評価者				座長	長尾 眞文	国際基督教大学	Rm.1094
川崎市の政策評価と外部評価者		高千穂安長	玉川大学				
評価者倫理・スタンダードに関する評価学会会員向け質問紙調査の分析結果	○	服部 浩昌	(株)VSOC				
指定管理者評価における民間第三者評価機関の先駆的実践		間中 恵子	(財)日本ユニセフ協会				
福祉サービス第三者評価における評価者のサービス改善支援の現状と保護について～事後評価における課題や目標の設定とその支援過程の展開～		柳原真理子	神奈川大学				
		村田 道彦	芦屋女子短期大学				
自由論題 4 自治体評価／地域開発				座長	梅田 次郎	JMAC構造改革推進セクター	Rm.1096
自治体経営力の実態に関する調査	○	田中 康人	(社)日本経営協会経営研究センター				
		星野 敦子	十文字学園女子大学				
保護者団体による自治体学童保育施策の評価活動の経緯 ～東京都小金井市における保護者と行政との「協働」の現実～		橋本 昭彦	国立教育政策研究所				
事業仕分けから見た行政評価の課題		佐藤 章	大津市役所				
地方開発のためのオンパク・アプローチの評価システム：概念と発展	○	石丸 久乃	NPO 法人ハットウ・オンパク				
		三好 皓一	立命館アジア太平洋大学				

11月28日（土）午後の部Ⅱ 15：00-17：00

シンポジウム「協働社会と評価～人びとの生活世界から評価の役割を考える」							Rm.1093
				座長	源 由理子	明治大学	
				座長	大島 巖	日本社会事業大学	
	パネリスト	入山 映	サイバー大学				
	パネリスト	荒井 英明	構想日本				
	パネリスト	長尾 眞文	国際基督教大学				
	パネリスト	大島 巖	日本社会事業大学				
	パネリスト	三好 皓一	立命館アジア太平洋大学				

11月28日（土）総会 17：15-18：00

11月28日（土）懇親会 18：15-20：00

11月29日（日）午前の部 10：00-12：00			
自由論題 5 評価手法		座長 三好 皓一 立命館アジア太平洋大学 Rm.1084	
幸せ度とその要因を評価する		大島 章嘉	市民満足学会
就職移動に関する要因分析		寺山 雅宏	東京工業大学
学力に及ばず学校要因と生徒要因の階層構造分析－REDIPの事例－	○	村田 維沙	東京工業大学
		牟田 博光	東京工業大学
財務指標から捉えた民間非営利組織の評価 ～持続性の構造を探る～	○	田中 弥生	(独) 大学評価・学位授与機構
		渋井 進	(独) 大学評価・学位授与機構
自由論題 6 実践報告		座長 林 薫 文教大学 Rm.1085	
中米ホンジュラス共和国初等教育における縦断的就学・追跡調査 －教科からの分析－		關谷 武司	関西学院大学
教員研修が学力向上に与える効果に関する研究 －ガーナSTMプロジェクトを事例として－	○	富田 真紀	東京工業大学
		牟田 博光	東京工業大学
アメリカのGPRA・再論 －連邦政府における業績管理システムの導入と発展－		田中 啓	静岡文化芸術大学
中国の高等教育における日本人留学生の満足度に関する要因分析と 収益性に関する研究－清華大学の例－		佐藤 広大	パークレイズ・キャピタル
11月29日（土）午後の部Ⅰ 13：15-15：15			
共通論題 2 政権交代とマニフェスト評価～政治主導体制下の政策評価のあり方を考える～		座長 田中 弥生 (独) 大学評価・学位授与機構 Rm.1083	
マニフェスト評価が示唆する政治主導の政策評価 ～規制改革関連政策から～		田中 弥生	(独) 大学評価・学位授与機構
新政権における政策評価の方向		中林美恵子	衆議院議員
官僚主導から政治主導への転換の意味するもの		高橋 進	(株) 日本総研
政策評価の現状と課題～政策評価法施行後の見直しの方向性に即して～		新井 誠一	総務省
政治主導とマニフェストサイクルの実現		工藤 泰志	言論NPO
共通論題 3 女性政策をめぐるわが国の現状－「評価」の視点から－		座長 山谷 清志 同志社大学 Rm.1084	
		コメンテータ 田中由美子	国際協力機構
男女共同参画推進のための拠点施設における評価		高橋 由紀	国立女性教育会館
指定管理者制度導入施設の評価		葛原 生子	広島市女性教育センター
男女共同参画政策を推進する評価－市民参加とエンパワメント－		林 やすこ	NPO参画プラネット
女性関連施設の労働問題と評価		渋谷 典子	名古屋大学
11月29日（日）午後の部Ⅱ 15：30-17：30			
共通論題 4 地方自治体における公民パートナーシップの評価		座長 源 由理子 明治大学 Rm.1083	
国分寺市における公共調達の適正化のための取り組み		内藤 達也	国分寺市役所
英国における公民パートナーシップの評価と日本への示唆		菊池 端夫	明治大学
民間パートナーとパフォーマンス評価に基づく調達の枠組み		北大路信郷	明治大学

## 共通論題セッション報告

### 共通論題セッション1 「日本のODA評価の現在」

座長 和田 義郎 (政策研究大学院大学・国際協力機構)  
 コメンテータ 牟田 博光 (東京工業大学)

2008年10月1日に、国際協力機構 (JICA) に国際協力銀行 (JBIC) ODA部門が加わり、新しいJICAが誕生して、すでに1年が経過した。また、民主党政権の発足、11月の事業仕分けなどにより、ODA評価については、これまでと異なる視点からの検討が加えられている。この間において、09年2月には、外務省より新しいODA評価ガイドラインが発行され、3月には、新しいJICA事業評価年次報告書が発行された。また、夏にはDAC対日援助審査も行われるなど、ODA評価における新しい課題についての学会、一般よりの関心は強まっている。本セッションではこうした状況をふまえ、日本のODA評価の現状と課題について様々な角度から報告をいただき、議論を行った。

第1報告者の畠薫氏 (外務省) からは、日本のODA評価の位置づけ、外務省とJICAの役割分担、2003年以降の評価有識者会議による第三者評価について、ODA評価のフィードバックメカニズムとその具体的事例 (目標体系図、援助協調、ノンプロ無償、数値目標、ファストトラック制度など) が報告され、DAC対日審査の結果も含め、政策レベルの評価の難しさ、フィードバックのための努力、評価専門家育成の必要性などについて報告があった。ついで、第2報告者の高見博氏 (財務省) からは、財務省における政策評価の実施状況、ODAに関する政策評価、そして、財務省における評価の今後の課題について報告があった。特に今後の課題としては、政策評価制度の下での改善、独自の評価システム、事業仕分けでの私的などについて説明があった。第3報告者の遠山慶氏 ((独) 国際協力機構) からは、JICAの評価制度について、外務省との役割分担、評価実施体制、事業評価の8つの特徴、事後評価実施方針 (原則3スキーム共通)、フィードバックのメカニズム、合同票あの例などの説明が行われた。最後に討論者の牟田博光氏 (東京工業大学) からは、外務省についてODAの外交上の効果を評価できないか、財務省を含むODA評価の取りまとめの仕組み、財務省のODAにおいて世銀などの国際機関による評価について問題提起があり、また、JICAの評価制度のスキームを超えて統一する際には、金額のみによるのではなく、プログラムで評価すべきなのではないか、コンサルタントによる評価の第三者性について、議論があった。

これらの報告に対して会場からは多数の質問・意見が出され、とくに、ODAの事業仕分けからほとんど日数が経過していなかったことから、事業仕分けとの関係についての議論があった。日本のODAの現在については、今後とも継続的に議論がなされる必要があるだろう。

### 共通論題セッション2 「政権交代とマニフェスト評価～政治主導體制化の政策評価のあり方を考える～」

座長 田中 弥生 ((独) 大学評価・学位授与機構)

2009年衆議院選挙は民主党の大勝をもって終わり歴史的な大転換であったともいわれている。政治主導とマニフェスト型政治を謳って誕生した現政権であるが、では、政策評価はどのように変わるのか。本セッションはこの点について、政治家、行政、民間シンクタンク、NPO関係者で議論した。

第1報告者で司会の田中弥生氏（（独）大学評価・学位授与機構）は現在の政策評価システムでは、政治と有権者をつなぐマニフェスト政治の仕組みは実現できないのではないかと問題提起した。そのうえで、言論NPOのマニフェスト評価活動に携わった経験から、現行の制度では内閣や内閣官房に事務局をおく政策は評価の対象にならないことや、経済財政諮問会議など政治側が評価の結果を無視するといった事例もかつて見られたことを挙げ、省庁の中に閉じてしまっている日本の政策評価を、国民に開かれたかたちに設計し直す必要があることを提示した。

第2報告者である中林美恵子氏（衆議院議員、民主党）が、アメリカ上院や日本の省庁で政策評価などに携わった経験をもとに、大統領制のもとアメリカでは政策決定プロセスの透明性がある程度確保されている一方で、日本では政策評価が政治家やその先の有権者につながっていないと指摘した。第3報告者の高橋進氏（株式会社日本総研）は、行政刷新会議の「事業仕分け」に仕分け人として参加していた経験をふまえ、「仕分け」の長所や短所を挙げつつ、議論が全て公開され、役所に説明責任が移ったことで族議員が裏で画策することもできなくなるなど、ゲームのルールが変わったことは評価できると説明した。第4報告者の新井誠一氏（総務省の行政評価局）は、当局の概要と機能を説明し、政策評価を予算・決算につなげることで、政策の体系性や客観性の確保が重要であることを述べた。第5報告者の工藤泰志（言論NPO）は、民主党内の政策形成におけるガバナンスの仕組みが整っていないことを指摘したうえで、選挙での約束を政府の約束に発展させることの必要性を強調し、選挙でのマニフェストを政府の約束に変え、それを予算に連動させるというというマニフェストのサイクルが、日本ではまだ始まっていないと指摘したうえで、政党のマニフェストを如何に政府の約束に変えるのか、その実行プロセスを自己評価する仕組みを日本でどう設計するのかという2つの課題を挙げた。

その後のディスカッションでは、日本では立法府の政策立案や評価の機能が極めて弱いことや、代案の提示や政治的な判断に踏み込むことができない「事業仕分け」の限界などについて議論した。政策評価については、衆参両議院の機能分担のあり方や、複数の省にまたがる評価の難しさなども挙げられた。最後に、有権者である市民が強くないと政治は変わらないが、その有権者の判断材料として、マニフェスト評価結果を政党自身が、また民間の複数の主体が行い提示することが重要であるとの意見でまとめられた。

### 共通論題セッション3 「女性政策をめぐるわが国の現状－「評価」の視点から－

座長 山谷 清志（同志社大学）

コメンテータ 田中 由美子（国際協力機構）

わが国においては男女共同参画社会基本法制定以来、いわゆる「ジェンダーの主流化」を目指す理念の普及、関連する法律・制度の整備は進んできた。しかし女性政策を推進する現場においては、実は女性政策そのものを衰弱させる「効率の主流化」が進行している。すなわち政策手段を実施する体制は独立行政法人化や指定管理者制度によって混乱し、政策資源（資金・人員・情報）は適正規模の議論が無いまま節約を強制され、政策に対する社会の理解はその他の困窮化（高齢者の貧困・限界集落・自治体破産・若年失業者・医療崩壊）の中で希薄化し、政策目的を達成する意志そのものが弱まっている。こうして女性政策の目的・理念は形骸化してきた。このセッションは上記課題の具体例を「評価」の視点から提示し、「実践的研究」の糸口を探った。

第1報告者の高橋由紀氏（国立女性教育会館）は全体の総論的立場から、議論の出発点を示すため「男女共同参画推進のための拠点施設における評価」と題して報告した。それによると男女共同参画社会基本法の制定以降、男女共同参画推進のための実施拠点として位置づけられる女性関連施設であるが、その政策上の位置づけと評価は必ずしも法律が予定した方向に向かっていない状況にあるという。

第2報告者、葛原生子氏（広島市女性教育センター）は指定管理者自らが施設職員と協働して行ってい

る内部評価の実践活動、すなわち「評価のあり方研究プロジェクト」で行った約2年間の取組の実態、協働活動のプロセスと内容を報告した。報告のタイトルは「指定管理者制度導入施設の評価」であり、指定管理者制度の意味するところと実態とが若干乖離している状況がうかがわれる。

3番目に林やすこ氏（NPO法人・参画プラネット）は「男女共同参画政策を推進する評価－市民参加とエンパワメント－」のタイトルで、男女共同参画の推進をミッションとするNPO法人（参画プラネット）が、指定管理者としての場を活用し、取り組んでいる「市民の視点」による評価の内容とその成果および課題について報告した。精力的にミッションに取り組んでいるが、それはNPO関係者の「志」によるもので、行政との距離が感じられた。

最後の第4報告者、渋谷典子（名古屋大学）は「女性関連施設の労働問題と評価」の中で女性関連施設における労働（活動）状況については多様な課題が表出していると問題提起され、その課題解決に向けて、労働法の視点をふまえた評価アプローチの必要性について報告した。

討論者の田中由美子氏（JICA）からの鋭いコメントの後に、論点整理を行いつつ、評価の「意図的な誤用」が目立つ現場の問題について充実した議論が行われた。

そのなかで、わが国では評価の定着がある程度進んだ反面、評価の‘political’な活用事例が福祉や雇用、教育、ジェンダーなど、必ずしも政治的な立場が強くない人びとを対象とする政策領域で目立ち始めていることが明らかになった。今後の懸念材料であり、日本評価学会としても注視しなければならないところである。

## 共通論題セッション4 「地方自治体における公民パートナーシップの評価」

座長 源由理子（明治大学）

約30年前に英国から広がった新公共経営の考え方が現在も日本の地方行政に強い影響力を持ち、指定管理者に代表される代理人制度が使われているが、昨今、この種の手法が真に公的サービスの質向上とコスト軽減に役立っているのか疑問が呈されている。特に、市場競争原理の活用を重視して民間サービス供給主体の選定に競争的調達法をとるべきという考えと、民間主体との効果的な協働を進めるために行政が民間主体と持続的に強い信頼関係を築いていくべきという考えが混在している。本セッションでは、この問題を整理することをねらいとし、公民パートナーシップが持つ多様な目的を見分け、目的と実施形態の整合性、公民協働の効率性や有効性などを評価するための枠組みについて複数の角度から報告をいただき、議論を行った。

第1報告者の内藤達也氏（国分寺市役所）からは、公共調達における「適正な調達」を実現するための制度化の検討事例が報告された。適正な調達には、「公正性、透明性、及び競争性の発揮を通じて、地域社会や地域経済の向上に寄与する機能と役割」が期待されており、そのためにコスト重視の競争入札から品質や地域経済への貢献など、多面的な評価への取り組みが具体的に始まっていることが示された。ついで第2報告者の菊池端夫氏（明治大学）からは、英国の事例や近年の研究動向を基に、公民パートナーシップにおける委託や評価のための論点が提示され、自治体側のイニシアティブによる指定管理者の制度設計の在り方、官民競争入札の導入における民間側の主体性の取り入れ、業績情報やコストの開示等による校正で透明な公民パートナーシップ形成など新たな枠組みに対しての提言がなされた。また、第3報告者の北大路信郷氏（明治大学）からは、NPMの理念を構成する「市場競争原理の活用」と「権限移譲による自律的経営の重視」といったしばしば相反する二つのアプローチのはざまに指定管理者制度が運用されていることによる弊害が指摘され、短期的な競争入札をしなくても透明性の高い公共調達を行うための評価の仕組みが必要となるとし、公民パートナーの間の「信頼」の評価について問題提起がなされた。

これらの報告に対して会場からは質問・意見が出され、信頼性の評価の枠組み、英国の事例の日本への

適用可能性、自治体の取組の現状と課題などについて活発な議論が行われた。公共調達と評価について取り上げたセッションはこれまで限られていたが、協働社会における公民パートナーシップの評価をひとつの切り口として、公共経営分野における評価のあり方について議論が深められたことは、本セッションの成果であったと言える。

## シンポジウム報告

### シンポジウム「協働社会と評価～人びとの生活世界から評価の役割を考える」

座長 源 由理子（明治大学）

座長 大島巖（日本社会事業大学）

現代は、公的セクター（政府・自治体）、私的セクター（企業）、共的セクター（民間非営利団体）等の様々なアクターによる協働の時代である。協働社会のめざすところは、おそらく、私達ひとりひとりが生活する日常的な「生活世界」や地域社会がより良いものになることであろう。「評価」が人々の生活に近い、さまざまな分野の現場で実践されることにより、どのように協働社会の実現に貢献できるのか。シンポジウムでは協働社会と評価の関係を、市民の「生活世界」の視点からとらえ、市民を巻き込んだ評価がより良い社会の実現にどのように貢献できるのかを探ることを目的に、異なるフィールドからの報告が行われた。

第1報告者の入山映氏（サイバー大学）と荒井英明氏（構想日本）からは、地方自治体における評価の切り口から、地域社会における公的社会サービスの担い手である自治体の事業評価について、「事業仕分け」の手法による実践事例の報告があった。第2報告者の長尾眞文氏（国際基督教大学）からは、学校評価と地域社会という切り口で事例に基づく報告があり、地域との連携による学校づくりにおけるツールとしての評価への期待が示された。また第3報告者の大島巖氏（日本社会事業大学）は、保健・医療・福祉分野におけるサービス利用者の評価活動を取り上げ、サービス利用者たる市民が評価活動に参加することが、社会に評価文化を根付かせる原動力になることが指摘された。最後に第4報告者の三好皓一氏（立命館アジア太平洋大学）からは地域活性化における評価の活用について、各地域の文脈にそった価値観や規範を基準とした評価が重要で、そのためには住民の視点からの評価が不可欠であることが示された。

これらの報告に対して討論者の山谷清志氏（同志社大学）から、評価の多様性について、そもそも政府の改革を目的とする行政評価と協働やエンパワメントを背景に展開される評価研究やプログラム評価とは相容れないものがあるのではないかと問題提起がなされた。それに対し各報告者から、①評価は目的によって異なるので使い分けが必要である、②事業仕分けの手法はコストカットだけではなく地域のガバナンスを再検討することを目的とすべき、③地域社会の問題に関わる中で行政と関わりのない評価は意味があるのか、④当事者評価における評価者の役割が重要である等の様々な指摘と見解が述べられた。

本来であれば、この後フロアとの議論がなされるべきであったが、時間の不足により十分に行えなかったことをこの場を借りてお詫びしたい。本シンポジウムでの報告と議論が、人びとの生活世界に密着した評価の役割についての研究に一石を投じることができたならば幸いである。

## 自由論題セッション報告

### 自由論題セッション1 「学校評価／学力分析」

座長 橋本 昭彦（国立教育政策研究所）

本分科会では、3報告が行われ、フロアとのやり取りを経て理解を深めた。

石田謙豪会員（広島県尾道市立栗原北小学校）による「学校経営の評価に関する一考察」は、ここ数年、全国的に拡大・深化しつつある学校評価の取組みの傾向として、教育活動の状況のほうにばかり目が行きすぎて、学校運営の状況の評価が手薄になっている点に着目した。「組織」「予算」「研修」「費用対効果」などの視点から、「教育関係者でなくても経営の概要を把握することのできる評価項目」及び「人材育成」の在り方を考察した。新人教員の大量着任が近い将来に予想される中、学校評価の未来をみつめる報告である。

高月淳会員（広島県立忠海高等学校）による「高等学校の学校評価における適切な評価指標の在り方」では、小中学校に比べて学校評価の導入が進んでいない高等学校の状況について注目した着眼の良い報告である。たくさん提示された「仮説」に対する「検証結果」は、必ずしも明らかではないが、県内の85公立高校から回答を得たアンケートによる各校の学校評価についての実情調査は稀少であり、オリジナルな価値がある。

福田優一会員（東京工業大学大学院）による「数学・読解・科学の学力に影響を与える要因の分析」は、学力に影響を与える要因についての諸研究の中でこれまで注目されていなかった「月齢の影響」について分析しようとした意欲的な報告であった。PISA2006のデータを使用して、月齢と学力に一定の相関傾向があることは貴重な発見である。ただ、「影響」という語は「相関関係」を超えて「因果関係」に言い及ぶ表現であるので、慎重な立論が必要である。

3報告とも、時宜にかなったテーマ設定であり、将来の発展がさらに期待できる貴重な報告である。

### 自由論題セッション2 「政策評価」

座長 湊 直信（(財)国際開発高等教育機構）

コメンテータ 林 薫（文教大学）

このセッションでは3名の発表の後に、コメンテーター（林薫）から丁寧なコメントがあり、参加者を交えた活発な質疑応答があった。

「日本の対太平洋島嶼国協力の政策評価と太平洋・島サミット」（野田真里）では、国別援助計画がない太平洋島嶼国の国別評価を事例に、評価方法、結果、提言、政策への反映（過程を含む）を論じている。日本の対太平洋ODA政策は、同じ太平洋のパートナーとして、島嶼国の経済および環境における脆弱性に有効な支援を行っている結論付け、提言は太平洋・島サミットにおける新たな政策立案過程においてインプットされ、対太平洋ODA政策における有効なPDCAサイクルの実現が可能となったとして、評価結果の政策への活用の具体的な事例として報告された。

「途上国のモニタリング・評価制度の新たな課題に関する考察：ガーナを事例として」（三輪徳子）では、近年のプログラム・アプローチの主流化、財政支援を含めた援助モダリティの多様化の中で、途上国の

M&Eの課題が変化している点に焦点を当てた。ガーナを事例に、PRS（貧困削減戦略）M&Eの制度化の枠組みにおいて、M&E情報の質、政策・予算へのフィードバック、M&E制度間の整合性について報告している。国レベルでの政策・戦略から事業に至る統合されたM&E制度の整備、計画・予算とのリンケージの強化に向けた支援において、包括的な視点からの制度作りと人材育成の必要性を論じている。

「モンゴルと東欧移行諸国の公共財政問題の比較評価」（上野宏）では、PEFA（国際組織）が開発した枠組みを活用して、モンゴルを他の途上国と公共財政制度と執行の視点で比較している。結果は、モンゴルは東欧3国平均、中央アジア3国平均、途上国17国平均よりも良く、全体的にみれば、財政経営評価データが存在する途上国・移行国の中では、比較的、良い財政制度を持ち財政執行を行っている。しかし、最大の問題は予算の信頼性であり、その原因は、モンゴル経済が資源輸出経済へ変貌し、資源の国際価格の高騰、さらに政策の誤りにある旨、報告された。

以上の3つの発表はいずれも開発援助に関わる政策評価に関するテーマであった。プロジェクト評価に比べて、評価の目的、手法、評価結果の活用、評価体制といった点で困難が多い政策評価に関して、それぞれの視点からの分析や経験は示唆に富むものであった。

### 自由論題セッション3「評価者」

座長 長尾 眞文（国際基督教大学）

本セッションは以下の4件の発表の後、質疑応答があった。

「川崎市の政策評価と外部評価者」（高千穂安長）では、川崎市で外部の政策評価委員がメタ評価を行うことにより、市職員の一次評価に「市民の目線」が加わり、結果として川崎市の内部体制、組織文化、戦略が一体的に機能するようになったことが報告された。

「評価者倫理・スタンダードに関する評価学会会員向け質問紙調査の分析結果」（服部浩昌、間中恵子）では、本学会「評価者倫理・スタンダード策定分科会」における作業の一環として、本年6月から7月にかけて評価学会会員を対象に実施した質問紙調査の集計・分析結果の概要報告があり、今後の追加作業計画に関する説明があった。

「指定管理者評価における民間第三者評価機関の先駆的実践」（柳原真理子）では、2002年に指定管理者制度が施行されて以来、ほとんどの自治体で行政が指定管理制度におけるモニタリングを実施してきた中で、2005年に横浜市が先駆的に開始した民間第三者評価機関による指定管理者第三者評価の試みを取り上げて、そのメリット・デメリットを検討し、客観性・専門性を生かした民間評価機関が果たし得る役割や実践上の課題について報告した。

「福祉サービス第三者評価における評価者のサービス改善支援の現状と保護について～事後評価における課題や目標の設定とその支援過程の展開～」(村田道彦)では、全国的に進められている福祉サービスの第三者評価で問題化しているサービス事業所からの受信率の低下に焦点を絞り、質的手法を用いて第三者評価作業から評価の専門性に関する項目を抽出して、その妥当性を検討した結果を報告した。

以上の4報告の内、2番目の評価者倫理・スタンダードに関する学会分科会の報告を除く3報告は、内部評価が中心となっている行政サービスの評価における外部評価者と第三者評価者の役割や資質について論じている。特に3番目と4番目の報告は、いずれも福祉サービスの第三者評価者の潜在的な社会的有用性とその発揮のために乗り越えるべき課題について制度論と機能論の異なる視点からの分析が提示するもので、参加者の強い関心を呼んだ。最初の報告の外部評価者の果たす役割もそうであるが、評価者の中立性が強調される時、そこで想定されている評価の専門性が実際にどこまで担保されているかという課題が残されているように感じられた。それと同時に、行政サービスの評価で圧倒的な比重を占める内部評価慣行の充実抜きには、評価の貢献に限りがあることは明らかであることから、評価者倫理・スタンダードに関

する学会分科会の今後の作業では、内部評価者と外部評価者双方の倫理・資質をバランスよく取り上げることが期待される。

## 自由論題セッション4 「自治体評価／地域開発」

座長 梅田 次郎（JMAC構造改革推進セクター）

本セッションでは「自治体評価／地域開発」というタイトルのもと以下の4件の発表が行われた。

「自治体経営力の実態に関する調査」（田中康人、星野敦子）では、全国の自治体を対象に、組織体としての程度透明性と効率性のあるプロセスを持って整合性のある業務遂行や意思決定を行っているかという行政マネジメント領域とどのくらい財政的な力量を有しているかという財政面領域について実態調査を行い、この2つの領域を総合する「行政経営力」を算出した結果が発表され、今後の活かし方などについての報告があった。

「保護者団体による自治体学童保育施策の評価活動の経緯～東京都小金井市における保護者と行政との「協働」の現実～」（橋本昭彦）では、小金井市における公営学童保育所の民営化論議の中での保護者団体による「学童保育研究」の事例を追いながら、その中で、「保育の質」の評価活動のありようや、自治体の審議会と保護者（団体）との関係のありようについて得たヒントなどについての報告があった。

「事業仕分けから見た行政評価の課題」（佐藤章）では、「事業仕分け」で当該事業に対するニーズ、あるいは政策デザイン等について説明者側が十分に説明できない状況にありがちであることをとらえて、このことが現在の業績測定型行政評価の欠陥の一つであるとし、今後評価者自身が「評価クエスチョン能力」を磨くことが業績測定型行政評価が抱える課題の解決方策の一つではないかなどの報告があった。

「地方開発のためのオンパク・アプローチの評価システム：概念と発展」（石丸久乃、三好皓一）では、発祥地大分県別府市から全国10か所以上に広がっている地方開発のための新しい手法であるオンパク・アプローチについて紹介され、活動の成果を示すために評価が欠かせない取り組みになっていることや参加型ワークショップの実践のなかで発展しその形を変化させている状況の報告があった。

これらの報告に対して会場からは時間いっぱい多数の質問・意見が出され、これに対する報告者からの応答と再質問など自治体での評価活動や行政経営また地方開発のあり方について議論が深められたことは、本セッションの成果であった。

## 自由論題セッション5 「評価手法」

座長 三好 皓一（立命館アジア太平洋大学）

本セッションでは、4つの報告がなされ、報告者の報告に基づき有意義な議論が行われた。

まず、大島章嘉会員（市民満足学会）から、「幸せ度とその要因を評価する」について報告がなされた。本報告では、日本人が直接的に幸せになれる方法を見出すことができればとのことで、幸せ要因を把握するために実施した自由回答式予備調査と、これを基に行った幸せ度とその形成要因を定量的に把握しようとした本格調査について紹介された。フロアーからは、調査方法等に関連して質問、コメントがなされた。

次に、寺山雅宏会員（東京工業大学）から、「就職移動に関する要因分析」について報告が行われた。本報告は、若年労働者数の不均衡に焦点を当てるものであり、「学生が新規に職に就く際に行う県間移動」について、その移動の傾向とその要因について分析結果を紹介するものであった。若年労働者数の不均衡は、主にある地域で育った若者が進学や就職を機にその地域を出て行ってしまうことで発生するものであるが、本研究によって「距離」や各県の「卒業生数」及び「就職者数」が説明力を持つという進学移動と

同様の結果が得られたこと、また、各県民の「所得」を変数として加えて分析を行った結果、さらに説明力の高いモデルを得られることが提示された。フロアーからは、分析方法、また、本研究の評価論における位置づけなどに関して質問、コメントがなされた。

3番目に、村田維沙・牟田博光会員（東京工業大学）から、「学力に及ぼす学校要因と生徒要因の階層構造分析－REDIPの事例－」について報告が行なわれた。本報告は、インドネシアにおけるSchool-based Managementによる教育開発事業である地方教育行政改善計画（REDIP）を事例とし、途上国における学力成果に影響する生徒レベルの要因と学校レベルの要因の関係性について階層線形モデルを用いた検証結果を紹介するものであった。分析の結果、教員及び保護者の当事者意識、仕事の専門性、教育への参加によって、生徒の学力を説明するモデルを構築することができたこと、また保護者の仕事の程度により、学力にもたらされる影響が学校間で異なることが明らかとなったことが報告された。フロアーからは、分析方法、また、本研究の評価論における位置づけなどについて質問、コメントがなされた。

最後に、田中弥生・渋井進会員（(独)大学評価・学位授与機構）から、「財務指標から捉えた民間非営利組織の評価～持続性の構造を探る～」について報告が行なわれた。本報告は、NPO法人の財務パネルデータ分析を基に、NPOの持続性にかかる実態と促進・阻害要因を明らかにするものであり、大多数のNPO法人が事業収入という単一の収入に依存し規模を拡大しているが、それが財務的な持続性に負の影響を与える可能性があることが指摘された。フロアーからは、分析方法、また、行政の契約形態の影響などについて質問、コメントがなされた。

## 自由論題セッション6「実践報告」

座長 林 薫（文教大学）

本報告では教育に関する3件、および業績管理システムに関する1件の報告が行われた。

第1報告の關谷武司会員（関西学院大学）による「中米ホンジュラス共和国初等教育における縦断的就学・追跡調査－教科からの分析－」は春の大会の報告に引き続き、1986年から2002年の間にホンジュラスの地方都市に入学した1,377人の追跡調査を行ったものである。今回の報告では、分析に結果として教員の側の問題に焦点が当てられ、落第、退学者が多い理由が指導力の不足にあることが、明らかになっていることが報告された。膨大なデータを分析した労作であり、春の大会に比べ具体的な改善策への道筋が見えてきたことが評価される。

第2報告の富田真紀会員および牟田博光会員（いずれも東京工業大学）による「教員研修が学力向上に与える効果に関する研究－ガーナSTMプロジェクトを事例として－」はJICAの支援で2000年も開始された小中学校理数科教育改善計画の中間評価データを再分析したものであり、研修効果を確認するとともに。優秀な学校かどうかによる学力差、性別による学力差が生じていることが報告された。今後のさらなる展開に期待できる。

第3報告の田中啓会員（静岡文化芸術大学）による「アメリカのGPRA・再論－連邦政府における業績管理システムの導入と発展－」はGPRAの発展および展開をフォローしたもので、オバマ政権の今後の業績評価の方向性や日本の政策評価制度の今後の発展のあり方に関して示唆に富むものであった。行政刷新会議による事業仕分けが行われていたこともあり、これに関する質疑応答もなされ、時宜を得た報告であった。

第4報告の佐藤広大会員（パークレイズ・キャピタル）による「中国の高等教育における日本人留学生の満足度に関する要因分析と収益性に関する研究－清華大学の例－」は中国で学ぶ日本人留学生の留学における満足度を調査したデータに基づいたもので、満足度の説明要因を分析したほか、留学生のパーセプションレベルでの内部収益率による分析を行ったものである。日本人留学生の満足度というテーマ設定はユニークであり、満足度説明要因や収益率分析のさらなる発展が期待される報告であった。

## 企画委員会

企画委員会は、日本評価学会が行う各種の活動を企画、実行することをおして、わが国の評価学の普及と発展に寄与することを目的としている。最近の活動としては以下の活動がある。

### 1. 研究受託事業

前年度に引き続き、2009年度も国際協力機構から「課題別研修第三者検証委員会作業部会」業務も受託した。国際協力機構が2010年度以降に更新または新設する課題別研修案件の検討において、第三者検討委員会を設置し、案件内容の妥当性および効率性を検証するにあたり、検証プロセスの効率化と客観性のさらなる担保を図るべく、作業部会で審査原案を作成した。また、今後の事後評価をどのように行うかに関する調査研究を行った。

このような受託事業は学会としての社会的貢献を具体的な形で示すと共に、会員に評価作業を経験してもらう場を提供することにもなる。さらに学会に対しては財政的な貢献もできるなど利点が多い。企画委員会としてはこのような受託事業を積極的に拡大していきたいと考えており、他の事業の受託についても検討する。

なお、学会として事業を受託することが適切であるかどうかの判断基準、及び受託事業の品質確保については学会が定めた「受託評価関連事業実施要領」に基づいている。

### 2. 顕彰事業

学会賞には論文賞、奨励賞、功績賞の3種がある。2009年度は学会賞受賞者選考細則第6条により、2009年8月5日付けで学会ホームページ上に学会賞候補者の公募を行い、10月30日まで自薦、他薦を受け付けた。11月13日に学会賞審査委員会を開催した。論文賞、奨励賞、功績賞のそれぞれについて、申し合わせに基づき、功績、論文、発表原稿等公表された資料、アンケート結果、及び、日頃の学会活動などを総合的に検討し、さらに、受賞候補者の意思確認のためメールにて議論を継続し、11月26日、学会賞受賞候補者選考について審査員全員一致で下記の結論に達した。

論文賞は、選考当該年度の前年度に発行された本学会誌に掲載された原著論文等から推薦されるものとする、と定められており、受賞候補者として西出順郎会員を選考した。

奨励賞は、選考当該年度の前年度に開催された全国大会での発表等から推薦されるものとする、と定められている。第9回全国大会、及び春季第6回全国大会の場で発表された原稿等公表された資料、及び、それぞれの大会でのアンケート結果などを総合的に検討し、受賞候補者として田中博、米原あき会員を選考した。

功績賞は、前年度までの学会活動への貢献に基づき、推薦されるものとする、と定められているが、今年度は該当なしとした。

これらの選考結果は、企画委員長を通じて理事会に諮られ、原案通り承認された。また、受賞者には、11月28日明治大学で開催された総会の場で、学会から賞状及び副賞が授与された。

企画委員会委員長 牟田 博光

## 国際交流委員会

2009年10月—2010年3月期の国際交流委員会の活動は、以下の通りです。

①2009年10月—2010年2月

2009年度の外務省主催、日本評価学会協賛のODA評価ワークショップ開催準備の打ち合わせを外務省、FASID、国際フォーラムで開催。なお、その間、アジア地域の各国評価学会（協会）、OECD/DACを初めとする国際機関とのメールを通じて、APEANET（アジア太平洋評価協会ネットワーク）立ち上げのため協議を続けた。また、2009年10月にはシンガポール、韓国にて関連大学専門家と協議し、当該国の評価学会設立への働きかけを実施。11月初旬には、UNDP本部（ニューヨーク）と世界銀行本部（ワシントンD. C）にて、APEANET設立準備について協議。また12月には北京にて中国科学技術省（開発政策・プロジェクトの評価担当）の評価班と中国評価学会設立への働きかけを実施。さらに、同月にバンコックにてタイ開発研究所幹部とタイ評価学会設立の可能性の打診。本年2月9日-11日には、北九州市でのアジア16カ国から参集した専門家国際会議でAPEANET設立準備状況を披露し、各国評価学会・協会の参加へ勧誘。

②2010年2月18日

外務省主催、日本評価学会協賛の第9回ODAQ評価ワークショップを国際協力機構の東京研修センター（渋谷区西原町）にて開催。

③2010年2月18日

第9回ODA評価ワークショップでの国際機関とアジア諸国からのスピーカーおよびアジア諸国からのワークショップ参加者とAPEANET設立第1回準備会合を開催し、日本評価学会国際交流委員会APEANET設立検討会が予め用意した素案に基づき討議する予定。

④3月のブルネイとインドネシアへの廣野委員長の出張時に、当該国での評価学会設立とAPEANETへの参加の打診と協議を継続。

国際交流委員会委員長 廣野 良吉

## 広報委員会

### 1. 活動方針

広報委員会は、その目的である、「評価の意義を広く普及し、学会活動の強化を支援するために、日本評価学会および学会の行う種々の事業を広報し、学会の情報活動に寄与する」ことを目的とします。評価学会の活動のみならず広く評価に関わる情報も会員の皆さんと共有すべくあらゆる機会を捉えてその実現を目指します。

### 2. 活動状況

#### (1) 「学会活動の強化と学会の健全な運営を目指して会員獲得に努めます」

会員獲得のための学会説明などのツールについては、ホームページを第一に考えていますが、国、自治体、企業などの説明ツールとしては、マス・メディアを使ったパブリシティも利用する方向で検討を進めています。

#### (2) 「広報の手段は主として学会ホームページを活用します。従って、学会からのおしらせはホームページに随時掲載します」。

「ホームページのより魅力的かつ効果的な広報」のために、専門家を使ってホームページの刷新に費用をかけて取り組んでいます。4月までに更新の予定です。

### 3. 今後の方針

広報委員会は、メンバーの交代も視野に入れ、進捗状況管理、新企画の検討などを行い広報面での有効性、効率性を高めて行く所存です。

広報委員会委員長 高千穂 安長

## 日本評価学会春季第7回全国大会及び 10周年記念シンポジウムのご案内

春季第7回全国大会  
10周年記念シンポジウム  
実行委員会委員長 和田 義郎

来たる2010年6月19日（土）に日本評価学会春季第7回全国大会が、翌20日（日）には10周年記念シンポジウムが政策研究大学院大学（東京都港区六本木）にて開催されます。

プログラム等詳細につきましては、学会メーリングリストおよびホームページにて随時ご案内いたします。

会員各位におかれましては奮ってご参加くださいますようお願い申し上げます。

### 記

1. 日時： 春季第7回全国大会 2010年6月19日（土）  
10周年記念シンポジウム 2010年6月20日（日）
2. 場所： 政策研究大学院大学（東京都港区六本木7-22-1）

以上

## 日本評価学会誌刊行規定

2005.2.15改訂

2002.9.18改訂

2001.9.9改訂

### (目的および名称)

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

### (編集委員会)

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

### (編集方針)

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

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

13. 論文等は2名の査読者により査読することとし、その人選は編集委員会が行う。「研究論文」については、査読結果と編集委員会が査読者とは別に指名する担当編集委員1名の参考意見をもとに、編集委員会が掲載に関する決定を行う。「総説」、「研究ノート」、「実践・調査報告」および「その他」の論文については、査読結果にもとづき編集委員会が掲載に関する決定を行う。
14. 編集委員が「評価研究」に投稿した場合には、当該委員はその投稿に係わる常任編集委員会あるいは編集委員会の議事に一切参加しないものとする。
15. 上記いずれのカテゴリ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。

(抜刷の配付)

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

(インターネット上の公開)

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

(著作権)

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

(事務局)

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

(以上)

## 『日本評価研究』投稿規定

2008.9.29改訂

2003.4.18改訂

2002.3.25改訂

2001.9.9改訂

1. 『日本評価研究』(The Japanese Journal of Evaluation Studies)は、評価に関する論文、論考、調査報告等を掲載する。
2. 『日本評価研究』は、会員間の研究成果交流の場を提供し、内外における評価研究の一層の発展に資することを主目的として発行されており、原則として会員による寄稿を掲載する。なお、依頼原稿を除き、ファーストオーサーは学会員でなければならない。また、投稿は、一時に一原稿に限るとともに、他学会誌などへ二重に投稿などのない未発表のものとする。
3. 投稿された原稿は、編集委員会の責任において審査を行ない、採否を決定する。審査にあたっては、1原稿毎に2名の査読者を選定し、査読結果を参考にする。(査読者には、投稿者名を伏せて査読を依頼する。)
4. 原稿料は支払わない。
5. 『日本評価研究』に掲載された論文等は、その全文をインターネット上の本学会のホームページに掲載する。
6. 投稿にあたっては、投稿原稿が、①研究論文、②総説、③研究ノート、④実践・調査報告、⑤その他のうち、どのカテゴリーに入るかを明記する。ただし、カテゴリーについての最終判断は、編集委員会で行なう。「研究論文」は評価の理論構築あるいは慣行の理解について重要な学問的貢献となると認められる論文、「総説」は、評価の理論あるいは慣行について概観する論文、「研究ノート」は「研究論文」作成過程での理論的あるいは経験的な研究の中間的成果物に相当する論考、「実践・調査報告」は評価事業の実践あるいは評価にかかわる調査の報告、「その他」は編集委員会が独自に企画する特集に掲載する依頼原稿等である。
7. 投稿方法
  - (1) 使用言語は日本語又は英語とする。
  - (2) 著者校正は原則として第一校までとする。
  - (3) 英文原稿については、ネイティブスピーカーによる英文チェックを済ませ、完全な英文にして投稿すること。
  - (4) ハードコピー4部(A4版)を提出する。その際、連絡先(住所、Tel、Fax、Email)と原稿の種類を明記すること。掲載可と判断された原稿については、必要なリライトを経た後に、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。①研究論文20,000字、②総説15,000字、③研究ノート15,000字、④実践・調査報告20,000字、⑤その他適宜。それぞれ和文要旨を400字程度、英文要旨を150words程度、及び和文・英文でキーワード（5つ以内）を別に添付する。印刷は1ページ、20字×43行×2段（1,720字）とする。20,000字の原稿の場合、単純計算では英文要旨1ページを加えて合計13ページとなるが、図表の量によっては、それ以上のページ数となり得るので、注意すること。
- (7) 英文ではA4版用紙に左右マージン30mmをとり、10ポイントフォントを使用し、1ページ43行のレイアウトとする（1ページ約500words）。論文冒頭に150words程度のAbstractをつける。14ページでは、7,000words相当になるが、タイトルヘッド等を考慮して、最大語数を約6,000words（図表、注、文献込み）とする。図表の量によっては、ページ数が予想以上に増える場合もあり得るので、注意すること。

#### 8. 送付先

〒140-0002 東京都品川区東品川4-12-6 日立ソフトタワーB 22F  
財団法人 国際開発センター内 日本評価学会事務局  
TEL: 03-6718-5931  
FAX: 03-6718-1651  
E-mail: jes.info@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)

①

(2)

(3)

例2（英文）

1.

1.1

1.1.1

1.1.2

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

例1：日本語原稿の場合

図1 ○○州における生徒数の推移



(注)

(出所)

表1 ○○州における事故件数

--

(注)  
(出所)

例2：英文原稿の場合

Figure 1 Number of Students in the State of ○○

--

Note:  
Source:

Table 1 Number of Accidents in the State of ○○

--

Note:  
Source:

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

(5) 本文における注記の付け方は、(…である<sup>1</sup>)とする。英文の場合は、(…<sup>1</sup>)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 6<sup>th</sup> edition*. Beverly Hills, Calif: Sage Publications.

英語雑誌論文：著者（発行年）. 題名. 雑誌名, 巻（号）, 頁－頁.

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

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

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

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

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

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

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

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

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

## 『日本評価研究』査読要領

日本評価学会 『日本評価研究』編集委員会

2005年9月10日決定

### 1. 本査読要領の趣旨

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

### 2. 査読の目的と投稿者の責任

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

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

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

### 3. 査読の視点

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

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

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

### 4. 投稿に当たっての留意点

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

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

など、内容及び形式に関する留意点があります。

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

## 5. 査読にあたっての判断事例

### (1) 完成度において不十分であるが掲載を考慮できる場合

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

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

### (2) 掲載を考慮するのが困難と判断される事例

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

## 6. 判定

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

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

別添 (1)

## **Publication Policy of the Japanese Journal of Evaluation Studies**

Last revised on 15<sup>th</sup> February 2005

### The Purpose and the Name

1. The Japan Evaluation Society (hereinafter referred to as “evaluation society”) publishes “The Japan Journal of Evaluation Studies (hereinafter referred to as “evaluation study”) in order to widely release evaluation studies and outputs of practical activities to domestic and international academic societies, interested individual and institutions, and contribute to the advancement and prevalence of evaluation practice.

### Editorial Board

2. The editorial board administrates editing of evaluation study based on the editorial policy stated below.
3. The editorial board is formed with less than 20 members of the evaluation society who are assigned by the board of directors. Terms of editors are two years but can be extended.
4. The editorial board assigns one editor-in-chief, two vice-editors-in-chief, and a certain number of standing editors among the members.
5. The editorial board may hold at least one meeting to discuss the editing policy, plans of editorial board, and others.
6. The editorial board reports activities to the board of directors as needed and receives approval. Also it is required to report the progress of the past year and an activity plan for the following year at the annual conference.
7. The editor-in-chief, the vice-editors-in-chief and the standing editors organize the standing committee and administrate editing on a regular basis.

### Editorial Policy

8. The evaluation study, as a principle, is published twice a year.
9. The evaluation study is printed on B5 paper, and either in Japanese or English.
10. Papers published in the evaluation study are categorized as five types;
  - 10.1. Review
  - 10.2. Article
  - 10.3. Research note
  - 10.4. Report
  - 10.5. Others
11. The qualified contributors are members of the evaluation society (hereinafter referred to as “members”) and persons whose contribution is requested by the standing editors. Joint submission of members and joint submission of non-members with a member as the first author are accepted. Submission by the editors is accepted.
12. Submitted manuscripts are treated as the above categories, however, the standing editors will decide based on the application of the contributors and the following guidelines;
  - 12.1. “Review” is a paper, which provides an overview of evaluation theory or practice. The editorial board will make the decision regarding publication.
  - 12.2. “Article” is considered as a significant academic contribution to the theoretical development of evaluation or understanding of evaluation practice. The standing editors committee makes adoption judgments following the referee-reading process described in the next section.
  - 12.3. “Research note” is a discussion equivalent to the intermediate outputs of a theoretical or empirical enquiry. The standing editors committee makes adoption judgments following the referee-reading process described in the next section.

- 12.4. "Report" is the study report related to a practical evaluation project or evaluation. The standing editors committee makes adoption judgments following the referee-reading process described in the next section.
- 12.5. "Others" includes requested papers for special editions organized by the editorial board and announcements from editorial board to members regarding publication.
13. The editorial board selects two referee readers. For the "article", the editorial board makes adoption judgments referring to the results from referee readings and comments provided by one editor assigned by the editorial board. For "review", "research note", "report" and "others", the editorial board makes adoption judgments referring to the results from referee readings.
14. When editors submit a manuscript, the editors are not allowed to attend any of the standing editors committee meetings or editorial board meetings regarding the manuscript.
15. The standing editors have alternative of approval or not-approval for adoption judgment of manuscripts submitted to any categories. However exception is permitted if the editorial board approves the publication after minor rewrite. Even if the manuscripts are considered insufficient as an "article", standing editors can decide whether the manuscripts are published as a "research note" or "report" if the authors wish to publish.

#### Formulation and Release of Submission Procedure

16. The editorial board formulates the submission procedure based on the editorial policy described above and release after approval from the board of directors.

#### Distribution

17. The evaluation study is distributed to all members for free and distributed to non-members for a charge.

#### Distribution of the Printed Manuscript

18. 30 copies of the respective paper are reprinted and distributed to the authors. The authors must cover any costs incurred by author's requests for printing more than 30 copies.

#### Release on the Internet

19. The papers published in the evaluation study are released on the internet with approval from the authors.

#### Copyright

20. Copyright of papers which appear in the evaluation study is attributed to the respective authors. Editorial right is attributed to the evaluation society.

#### Office

21. The office is in charge of administrative works for editing, distribution, and accounting.

別添 (2)

## **Information for Contributors (For English Papers)**

Last revised on 29<sup>th</sup> September 2008

1. “The Japanese Journal of Evaluation Studies” is the publication for reviews, articles, research notes, and reports relating to evaluation.
2. “The Japanese Journal of Evaluation Studies” is primarily published to provide opportunities for members of the Japan Evaluation Society (hereinafter referred to as “members”) to exchange findings, and to contribute to further development of the study of evaluation both domestically and internationally. As a principle, this journal publishes the contributions submitted by the members. With the exception of requested papers, the first author must be a member. A submission (as the first author) is limited to one manuscript that has not been published or submitted in any form for another journal of academic association etcetera.
3. Adoption judgments of the manuscript are made at the discretion of the editorial board. Comments from two referee readers who are appointed for every manuscript are referred to in the screening process (the editorial board requests referee readers without notifying the author of manuscript).
4. Payment for the manuscript is not provided.
5. Papers published in “The Japanese Journal of Evaluation Studies” are released on the Internet at homepage of this academic society.
6. Regarding submission, manuscripts must be identified as one of the following categories: 1) article, 2) review, 3) research note, 4) report, and 5) others. However, the final decision of the category is made by the editorial board.  
 “Article” is considered as a significant academic contribution to the theoretical development of evaluation or understanding of evaluation practice.  
 “Review” is a paper which provides an overview of evaluation theory or practice.  
 “Research note” is a discussion equivalent to the intermediate outputs of a theoretical or empirical study in the process of producing an “article”.  
 “Report” is the study report related to a practical evaluation project or evaluation.  
 “Others” are manuscripts for special editions requested by the editing committee.
7. Manuscript Submission
  - (1) Manuscripts may be written in either Japanese or English.
  - (2) Correction by the author is only for the first correction.
  - (3) English manuscripts should be submitted only after the English has been checked by a native speaker.
  - (4) Submit four hard copies (A4 size) of the manuscript. Contact information including mailing address, telephone number, fax number, and e-mail address, and the category of the manuscript should be clearly stated.  
 For approved manuscripts, after necessary rewriting, the author needs to submit two hard copies of the final paper as well as a text file saved on a DOS/V formatted floppy disk. Original figures, charts, and maps should be provided.
  - (5) Total printed pages should not exceed 14 pages. Any cost incurred by printing more than 14 pages must be covered by the author.
  - (6) The layout for English papers should be 30 mm of margin at left and right side, 10pt for font size, 43

lines on A4 paper (about 500 words per page). An abstract of 150 words should be attached to the front. 14 pages are equivalent to 7,000 words but the body should not exceed 6,000 words to allow for the title, header, figure, chart, footnotes, and references. Please note that the number of pages may be more than expected depending on the number of figures included.

8. Mailing address

Office of Japan Evaluation Society at International Development Center of Japan  
Hitachi Soft Tower B 22nd Floor, 4-12-6, Higashi-Shinagawa, Shinagawa, Tokyo,  
140-0002, Japan

Phone: +81-3-6718-5931, Facsimile: +81-3-6718-1651

E-mail: [jes.info@idcj.or.jp](mailto:jes.info@idcj.or.jp)

別添 (3)

## Writing Manual of the Japanese Journal of Evaluation Studies (For English Papers)

Revised on 18<sup>th</sup> September 2002

### 1. Text, Charts, Figures, Graphs, Diagrams, Notes, and References

(1) The paper should be written in the follow order:

First page: Title; the author's name; Affiliation; E-mail address; Abstract (150 words); Keywords (5 words)

Second page: The main text; acknowledgement; notes; references

(2) Section of the text should be as follow:

- 1.
- 1.1
- 1.1.1
- 1.1.2

(3) Source of the charts, figures, graphs, and diagrams should be clarified. Submitted charts and others will be photoengraved, therefore it is important that the original chart is clear. Pictures shall be treated as figures.

Figure 1 Number of Students in the State of ○○



Note:

Source:

Table 1 Number of Accidents in the State of ○○

Note:

Source:

(4) Citation of literature in the text should be, (Abe 1995, p.36) or (Abe 1995).

- (5) Note in the text should be, (-----<sup>1</sup>)  
 (6) Note and references should be written all together in the end.

Note

1 -----.

2 -----.

- (7) Reference should list the literature in alphabet order, and arranged in chronological order. Follow the examples:

Book: author (year of publication). *Title of the book*. Published location: publishing house.

(e.g.) Rossi, P. H. (1999). *Evaluation: A Systematic Approach 6<sup>th</sup> edition*. Beverly Hills, Calif: Sage Publication.

Article from magazine: author (year of publication). Title. *Title of the magazine*, volume (number), page-page.

(e.g.) Rossi, P. H. (1999). Measuring social judgments. *American Journal of Evaluation*, 15(2), 35-37.

Article in Book: author (year of publication). Title. In editor (Eds.), *Title of the book*. Published location: publishing house, page-page.

(e.g.) 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 Fransisco, Calif: Jossey-Bass, 177-196.

Book by two authors: surname, first name, and surname, first name. (year of publication). *Title of the book*. Published location: publishing house.

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

Book by more than three authors: surname, first name, surname, first name, and surname, first name. (year of publication). *Title of the book*. Published location: publishing house.

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

(note 1) If some references are from the same author with the same publication year, differentiate by adding a,b,c as (1999a), (1999b).

(note 2) If the reference is more than a single line, each line from the second should be indented by three spaces.

(e.g.) 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 Fransisco, Calif: Jossey-Bass, 177-196.

別添 (4)

## Referee-Reading Guideline

The Japanese Journal of Evaluation Studies Editorial Board,  
The Japan Evaluation Society  
Approved on 10<sup>th</sup> September 2005

### 1. Content of the Referee-Reading Guideline

This Referee-Reading Guideline is to provide explanation of the main publication judgment, procedure of the referee-reading, to the members who submit the manuscript and for the members who are requested to conduct referee-reading in order to carry out the procedure efficiently and effectively.

### 2. Purpose of Referee-Reading and the Responsibility of the Author

Referee-reading is necessary for the editorial board to make decisions of whether submitted manuscripts are appropriate to publish in the Japanese Journal of Evaluation Studies or not.

If there is doubt or obscurity identified in manuscripts during the referee-reading corrections may be required. Therefore, referee-reading also contributes to the improvement of the submitted manuscripts. However, although the manuscripts are requested corrections, the author is still solely responsible in regards to the contents and it is not attributed to the referee-readers.

Referee-readers are two persons who are requested by the editorial board depending on the specialty or the field of the submitted manuscript. People who are not members of this academic society also may be requested.

### 3. Items of Consideration in Referee-Reading

Five points are considered in referee-reading, however, the importance of each may be different depending on the type of manuscript.

- (1) Importance and utility of the theme
  - (2) Originality of the study
  - (3) Structure of the logic
  - (4) Validity of verification and methodology
  - (5) Contribution to evaluation theory and practice
- For the article, all of above five are considered.
  - For the research note, especially (1), (2), (3), and (4) are considered.
  - For the report, especially (1), (3), and (5) are considered.
  - For the review, especially (3) and (5) are considered.

#### 4. Attentions in submission of manuscript

Besides above five viewpoints, basic completeness as a paper is also considered, for example;

- appearance of the paper is organized
- written according to the writing manual
- described simply and distinctive
- verification data is appropriately used
- notes and references are corresponding with the text
- terminology is appropriately used
- no wording and grammatical mistakes
- no errors and omission
- no punctuation mistakes
- expression in English abstract is appropriate
- word count is according to the manual

The above mentioned forms and contents are also considered. There have been cases in which graduate students and practitioners posted without organizing the manuscripts as a paper. On those occasions, referee-reading was not conducted. Necessary consultation is strongly recommended prior to submission.

#### 5. Judgment Cases in Referee-Reading

(1) In the case of the manuscript which is considered acceptable for the publication but is not yet complete:

The referee reader should evaluate carefully whether the paper can contribute to the development of evaluation theory or evaluation studies.

- Verification is lacking but the theory and formulation are useful for academic development.
- Analysis lacking but useful for formation and promotion of new theory.
- The literature review is not of a high standard but, the overall study is meaningful.
- Comparative study is not up to standard but is meaningful as an example of application.
- Analysis is lacking but it is meaningful as an evaluation of socially and historically important cases.
- Analysis is lacking but it is meaningful as an evaluation of particular social activities.
- Organization and expression are not up to standard as a paper but the contents are worthy to evaluate.
- Logic is not strong enough but useful in practice.
- The paper has significance as a report.

(2) In case of the manuscript which is considered as difficult for publication:

- Awareness of the issue or setting of the problem is indecisive.
- Understanding or analytical framework of notion of basic terminology is indecisive or inappropriate.
- There is a lack in credibility of data for the grounds of an argument.
- There is no clear point of an argument or appropriateness of proof.
- Organization of the paper and presentation (terminology, citation, chart, etc) are inappropriate (or not consistent).

## 6. Judgment

The final decision will be made on publication at the standing editors committee following one of four patterns (listed below). However, these judgments are not based on the number of errors but on the strength of the overall report. In the case of (3) and (4), there is a possibility to be published as a different type of paper. If it is published as a different type of paper, major rewrite concerning the number of words may be required.

- (1) The paper will be published.
- (2) The paper will be published with minor rewrite.
- (3) The paper will be published with major rewrite, however as a different type of paper (review, article, research note, or report).
- (4) The paper will not be published; however there is the possibility that it will be published as a different type of paper (review, article, research note, or report).



～投稿案内～

日本評価学会では、「日本評価研究」掲載のための投稿原稿を募集しております。1年に2回（3月末／9月末）の投稿締切を設けておりますので、ご興味をお持ちの方は投稿規定・執筆要領をご参照のうえ、奮ってご投稿ください。

投稿先： 特定非営利活動法人日本評価学会 事務局  
〒140-0002 東京都品川区東品川4-12-6 日立ソフトタワーB 22F  
財団法人国際開発センター内  
TEL: 03-6718-5931 FAX: 03-6718-1651

『日本評価研究』第10巻第1号  
2010年3月31日

編集・発行 特定非営利活動法人 日本評価学会  
〒140-0002 東京都品川区東品川4-12-6  
日立ソフトタワーB 22F  
財団法人国際開発センター内  
TEL: 03-6718-5931 FAX: 03-6718-1651

印 刷 株式会社 研恒社

©日本評価学会

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

# Japanese Journal of Evaluation Studies

Vol. 10, No. 1, March 2010

## CONTENTS

### Special Issue: The world trend of Evidence-based practice (EBP) and its trials in Japan

Editor's Note Ryo Sasaki, Iwao Oshima

Lessons Learned from the Worldwide Spread of Evidence-Based Medicine (EBM)  
Tomoya Masaki, Kiichiro Tsutani

Evidence-based Policies in Education: New Directions and Issues  
Kumiko Iwasaki

Contribution to methodology of program evaluation from the perspectives of  
development of Evidence-Based Practice (EBP) research: Program model  
constructions, fidelity assessments, and methods for their technology transfers  
Iwao Oshima

Dissecting the process of using evidence as a process of technological transfer  
and a process of replication Hiroshi Tsutomi

Challenges with the Implementation of RCT in International Development Aid Evaluation  
— from the RCT without theory to RCT with theory — Keitaro Aoyagi

Evidence-based development aid evaluation: The history of aid evaluation,  
the origin of RCT, and the comparison of the thoughts of Scriven and Banerjee  
Ryo Sasaki

#### Report

Participatory Evaluation of NGO Educational Project Facilitator Roles and  
Empowerment of Stakeholders Hiroshi Tanaka

#### Research Note

The application of Balanced Scorecard to the local government  
— taking the financial perspective — Tsuyoshi Sato