

# 日本評価研究

## Japanese Journal of Evaluation Studies

Vol. 17, No. 1, November 2016

### 特集：エビデンスの実践的活用とその方向性

巻頭言 特集「エビデンスの実践的活用とその方向性」に寄せて  
佐々木 亮 正木 朋也

エビデンスに基づくがん検診はなぜ実現しないのか  
—アイデア理論を用いた—考察— 田辺 智子

国際開発分野におけるエビデンス活用の現状と課題 浅岡 浩章

SROI（社会的収益投資）に関する批判的考察 津富 宏

貧困アクションラボの最新動向：政策教訓と拡大適用の事例 佐々木 亮

### 研究論文

評価者の倫理教育におけるケース・メソッドの利用  
—監査人の倫理教育手法からの示唆— 小林 信行

春季第13回全国大会開催報告

『日本評価研究』編集委員会  
Editorial Board

編集委員長  
Editor-in-chief

山谷 清志(同志社大学)  
Kiyoshi YAMAYA

副委員長  
Vice-Editor-in-chief

西野 桂子(関西学院大学)  
Keiko NISHINO

常任編集委員  
Standing Editors

牟田 博光(東京工業大学)  
Hiromitsu MUTA

編集委員  
Editors

岩渕 公二(NPO政策21)  
Koji IWABUCHI

大島 巖(日本社会事業大学)  
Iwao OSHIMA

岡本 義朗(新日本有限責任監査法人)  
Yoshiaki OKAMOTO

小野 達也(鳥取大学)  
Tatsuya ONO

窪田 好男(京都府立大学)  
Yoshio KUBOTA

佐々木 亮(国際開発センター)  
Ryo SASAKI

渋谷 和久(内閣府)  
Kazuhisa SHIBUYA

田中 弥生(大学評価・学位授与機構)  
Yayoi TANAKA

西出 順郎(岩手県立大学)  
Junro NISHIDE

林 薫(文教大学)  
Kaoru HAYASHI

松岡 俊二(早稲田大学)  
Shunji MATSUOKA

源 由理子(明治大学)  
Yuriko MINAMOTO

事務局  
Office

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

# 日本評価研究

第17巻 第1号 2016年11月

## 目次

特集：エビデンスの実践的活用とその方向性

佐々木 亮 正木 朋也

巻頭言 「エビデンスの実践的活用とその方向性」に寄せて ..... 1

田辺 智子

エビデンスに基づくがん検診はなぜ実現しないのか

—アイデア理論を用いた—考察 ..... 3

浅岡 浩章

国際開発分野におけるエビデンス活用の現状と課題 ..... 19

津富 宏

SROI（社会的収益投資）に関する批判的考察 ..... 33

佐々木 亮

貧困アクションラボの最新動向：政策教訓と拡大適用の事例 ..... 43

研究論文

小林 信行

評価者の倫理教育におけるケース・メソッドの利用

—監査人の倫理教育手法からの示唆— ..... 55

春季第13回全国大会開催報告

開催の報告とお礼 ..... 69

プログラム詳細 ..... 70

共通論題セッション報告 ..... 72

自由論題セッション報告 ..... 80

日本評価研究刊行規定	84
日本評価研究投稿規定	86
日本評価研究執筆要領	88
日本評価研究査読要領	91
Publication Policy of the Japanese Journal of Evaluation Studies	93
Information for Contributors (For English Papers)	95
Writing Manual of the Japanese Journal of Evaluation Studies (For English Papers)	97
Referee-Reading Guideline	99

## 【巻頭言】

## 特集「エビデンスの実践的活用とその方向性」に寄せて

佐々木 亮

株式会社国際開発センター

正木 朋也

独立行政法人国際協力機構

私たち日本評価学会社会実験分科会では日本評価研究第6巻第1号（2006年）および同10巻第1号（2010年）においてエビデンスに基づく評価およびその実践の世界動向と日本における取り組みについての特集を組み、対応した企画セッションを開催してきた。1990年代に世界的な広がりをみせたEBM<sup>1</sup>の普及もあり、上記特集号の企画時点で既に、エビデンスに基づく意思決定は保健医療領域ではもはや常識として受け入れられていた。

このような世界的潮流の中で、介入との因果関係を証明できる方法論としてのランダム化比較試験（RCT）の重要性は広く認識され、各方面でその普及・浸透のための活動もなされてきた。2006年の特集の巻頭言には、保健医療外の分野においては「日本ではまだこれからの話だ」としつつも、その理想と有るべき姿を示し、世界動向も踏まえた自然な導入と進展の期待が高まっていた。2010年の巻頭言には「日本でもついに本格的な取り組みがいくつか始まった」として、精神保健福祉、国際開発援助、教育の各分野の状況が紹介され、社会実験分科会発足5年足らずで一定の進展が認められたことを報告した。

初回特集から早10年が過ぎ、日本においても上記領域に加えて犯罪学の分野における応用報告も加わり国内において一定の活動があったことは明らかではあるが、現時点で世界動向をみればその後またさらに水を開けられた感がある。例えば、国際開発領域で実施されたCluster RCT等を含む妥当なデザインを用いたインパクト評価の研究報告件数は2000年時点で年間40件に満たなかったが、それ以降健康科学領域のみならず社会科学の領域における検討が増え、2012年では年間400件近くになり、その勢いはますます加速する様相をみせている（Cameron et al. 2015）。この世界的な展開と比べれば、日本においてはその後期待されるほどにはエビデンスに基づく評価・判断が進展していないと見受けられる。その背景と課題解決の方策については学会の場を通じて議論を行っているところであるが、エビデンスに基づく評価・判断が日本において広く一般的となるには更なる時間が必要というのが現状であろう。

因みに、RCTにより得られたエビデンスをもとに評価・判断がなされることが常識とされる保健医療分野においても、全てが順風満帆に現在の状況に導かれた訳ではなかった（津谷 2011）。先行する領域で遭遇した課題とその解決策については国内外を通して学ぶべき点も多いので、それらを参考に同じ轍を踏むことなく発展させることが望まれる。

また、エビデンスに基づく実践（Evidence Based Practice）という用語はEBPと略され、当初その推進者らの間では保健医療外の領域の現場における実践（またはその普及活動）として認識され使われてきた。その後、政策形成においてEvidence Based Policy あるいはEvidence Based Policy Makingといった用語とともに、政策レベルの意思決定にエビデンス情報を活用している海外事例の紹介もあり（家子ほか 2016）、国内においても政策レベルにおけるエビデンスを参照した意思決定の重要性がにわかに脚光を浴び始めている。

このように、同じエビデンスに基づく意思決定を行うトピックを扱い、またゴールも同じくする動向ではあるが、相互にその対象や想定している範囲の違いなどに留意した議論が必要な状況も生じている。政策までを想定する広義の概念は、EBMの始祖であるカナダMcMaster大学の流れを継承する研究者の間では、Evidence Informed Policy and Practice (EIPP) あるいはEvidence Informed Decision Making (EIDM)<sup>2</sup>として、従来のEBPも包括した概念とした議論がなされており、こうした流れについても把握しておく必要がある。

このような背景を踏まえ、この特集では、各分野におけるEBP普及の現状と課題に関して議論した。田辺智子氏の論文では、がん検診を事例に、日本において、エビデンスに基づくがん検診がなぜ実現しにくい状況となっているかについて、政治学で発展したアイデア理論を用いた分析を行っている。浅岡浩章氏の論文では、国際開発分野におけるエビデンスに基づく実践の最新の世界的動向を手際よく解説し、それを踏まえて今後エビデンスに基づく実践を推進するために必要な取り組みを述べている。津富宏氏の論文では、やや視点を変えて、社会的投資のための評価ツールのひとつであるSROI (social return on investment) について、インパクト評価の観点も踏まえた批判的考察を行っている。最後の佐々木亮の論文では、過去2回の特集号でも扱った貧困アクションラボ (J-PAL) の実績を対象に、いったいどのような政策教訓が産出され、どのような大規模適用につながったのかを解説するとともに、日本が学ぶべき政策案を解説している。

今後、本特集を含むこれまでの特集で取り上げたトピックに加えて、倫理性や経済面を含む議論も深める必要がある。また、エビデンスを情報として活用するためのインフラ整備や利害関係者を含む推進母体をいかに組織するかといった、より具体的かつ実践的な行動と対策も必要となる。その際、EIPP/EIDMを解する評価専門家の育成とエビデンスに基づく政策形成への関わりと役割についても明らかにしてゆく必要もあり、本特集を含めたこれまでの一連の特集がその一助となることを願ってやまない。

## 注記

- 1 Evidence-Based Medicineの頭文字を取ったもので「根拠に基づく医療」と訳すこともある。利用可能な信頼できる情報をもとに患者に最善の医療を施すこと、およびそれら実践活動の全体を示す。
- 2 エビデンス情報を有効に活用した政策および政策形成にかかわる活動全般を示す。Evidence basedな意思決定は「エビデンスを中心に行われるとの誤解」を与え易い点に配慮して、政策形成の場合には特に「エビデンスを有益な情報のひとつとして活用すること」が有益であるとの概念をより明確にする意図で用いられる。

## 参考文献

- 家子直幸・小林庸平・松岡夏子・西尾真治 (2016) 「エビデンスで変わる政策形成～イギリスにおける「エビデンスに基づく政策」の動向、ランダム化比較試験による実証、及び日本への示唆～」、『MURC政策研究レポート』、Retrieved from [http://www.murc.jp/thinktank/rc/politics/politics\\_detail/seiken\\_160212.pdf](http://www.murc.jp/thinktank/rc/politics/politics_detail/seiken_160212.pdf) (参照日 2016/10/03)
- 佐々木亮 他 (2006) 「特集：エビデンスに基づく評価の試み」、『日本評価研究』、6 (1) :1-2
- 佐々木亮・大島巖 他 (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」、『日本評価研究』、10 (1) :1-2
- 津谷喜一郎 (2011) 「日本のEBMの動きからのレッスンー前車の轍を踏まないために (特集 教育研究におけるエビデンス)」、『国立教育政策研究所紀要』、140:45-54
- Cameron, D., Mishra, A. and Brown, A. (2015). The growth of impact evaluation for international development: how much have we learned? *Journal of Development Effectiveness*, 1-21.

## 【研究論文】

エビデンスに基づくがん検診はなぜ実現しないのか  
— アイディア理論を用いた一考察 —

田辺 智子

国立国会図書館

stanabe0509@gmail.com

## 要 約

日本において、エビデンスに基づくがん検診がなぜ実現しにくい状況となっているかについて、政治学で発展したアイディア理論を用いて分析を行った。日本のがん検診は世界的に見ても早い時期に導入されたが、その後、死亡率減少のエビデンスがあるがん検診を行うべきという新たなアイディアが海外から輸入され、既存のがん検診を見直す政策変容が進められた。分析の結果、この政策変容が不徹底となっており、エビデンスが確立したがん検診に加え、エビデンスが不十分ながん検診が広く実施されている状況が明らかとなった。

その原因としては、死亡率減少という観点で有効性を評価すべきというアイディアが市町村レベルでは十分受容されていないこと、過去の政策が次の政策選択に影響を与える政策遺産が存在することが挙げられ、政策決定は必ずしもエビデンスのみに基づいて行われるわけではないという現実が浮き彫りとなった。

今後も、他の政策分野を含め、エビデンスに基づく政策を阻害する要因について、さらなる分析が求められる。

## キーワード

エビデンス、エビデンスに基づく政策・実践、エビデンスに基づく医療（EBM）、がん検診、アイディア

## はじめに

エビデンスに基づく政策・実践の必要性が議論されている。しかし、日本における取組はまだ始まったばかりであり、比較的先行しているかに見える保健医療の分野においても、エビデンスが必ずしも活用できていない状況が報告されている（田辺 2015、大島 2014）。日本においてエビデンスに基づく政策・実践が実現しにくい状況にあるとすれば、次なる段階として、なぜそのような状態が生じているのか、またエビ

デンスの活用を阻害する要因があるとするれば何なのかを明らかにする必要がある。本稿ではそうした問題意識のもと、がん検診を事例に分析と考察を行う。

わが国のがん検診は、1960年代という世界的に見ても早い時期から取組が開始されたが、当初は現在のような有効性評価の考え方はなく、1990年代後半からランダム化比較試験（Randomized Controlled Trial: RCT）によるエビデンスを重視する海外の有効性評価の考え方が導入された。しかし、現在でも市町村の行うがん検診では、有

効性が証明されていない検診方法が数多く実施されている。本稿では、このような現状が生じている背景や要因について、政治学で発展してきた“アイデア”の理論を用いて分析する。政策のアイデアという概念は、1970年代以降に欧米で起こった規制緩和等の大規模な政策変容が、従来の利益や制度といった要因だけでは説明できなかったことから提起されたものである。日本では、死亡率<sup>1</sup>の減少効果が証明されたがん検診を行うというアイデアが海外から輸入され、すでに行われていたがん検診の有効性を見直し、エビデンスのあるがん検診に置き換えるという政策変容が意図されたものの、十分な実現に至っていない。本稿では、この政策変容がどのように生じ、なぜ貫徹できなかったのかを、アイデア理論の枠組みを用いて分析する。

本稿の構成は以下のとおりである。まず分析の前提として、第1章でエビデンスに基づくがん検診の基本的な考え方、第2章で日本のがん検診の経緯と現状を整理する。その上で、第3章でアイデア理論を用いた本稿の分析視角を提示する。続く第4章では、がん検診における政策変容がどのように生じているかをアイデア理論の枠組みを用いて分析する。第5章では、分析を踏まえ、まとめと考察を行う。

## 1. エビデンスに基づくがん検診の考え方

がん対策は先進国に共通する政策課題であり、有効な対策としては、がん検診と喫煙対策があ

げられる。がん検診については、これまでに、子宮頸がん検診（細胞診）と乳がん検診（マンモグラフィ検査）のエビデンスがほぼ確立しており、欧米では公費で積極的に実施された結果、国レベルで死亡率が減少する成果が上がっている（斎藤ほか 2015、p.225）。大腸がん検診（便潜血検査）についても有効性が確認され、施策として導入が進んでいる。

### (1) がん検診の有効性評価

医療におけるエビデンスとは、治療や予防の有効性、つまり効果があるかどうかについての信頼性の高い研究結果を意味する。がん検診は、対象集団のがんによる死亡率を減少させることを目的に実施されるため、有効性を評価する際は死亡率を指標とする必要がある。つまり、がん検診の場合の有効性評価とは、死亡率減少効果を信頼性の高い研究で検証することと理解できる。

一般には、がん検診の目的は早期発見にあると誤解されやすいが、早期発見がそのまま死亡率の減少につながるとは限らない。がんの中には進行が非常に遅く、放置しても無症状のまま死亡につながらないものもあるためである。がん検診ではそうした無害のがん（過剰診断がん）も発見され、がんの発見率が増加しても死亡率が減少しない場合があるため、発見率をがん検診の有効性の指標とすることはできない。また、がんと診断された患者のその後の生存率<sup>2</sup>についても、表1に示すような各種バイアスがあるため、がん検診の有効性の指標とすることはできない。

表1 生存率を有効性の評価指標とした場合に影響するバイアス

バイアスの種類	性 格
リードタイム・バイアス (lead time bias)	検診で発見されるがんは、早期に発見された期間の分だけ生存率の計算の始点が前にずれるために生ずるバイアス。
レングス・バイアス (length bias)	検診では、進行の早いがんより進行の遅いがんが発見されやすく、そのために検診で発見されたがんで生存期間が長くなることによるバイアス。
過剰診断バイアス (overdiagnosis bias)	検診では、受診者が余命をまっとうするまで無症状で、臨床的に問題にならないがんも発見してしまうために生ずるバイアス。
セルフセレクション・バイアス (self-selection bias)	一般に、検診受診者は非受診者より健康意識が高く、健康管理に注意を払っているためにリスクが低く、生存期間が長いために生ずるバイアス。

(出所) 斎藤ほか「がん検診」佐藤隆美ほか編『What's New in Oncology—がん治療エッセンシャルガイド—』南山堂、2015、p.227を基に作成。

有効性を評価する研究デザインとしてはRCTが最上位であり、国際標準となっている（祖父江2012、p.845）。がん検診がすでに普及している場合などはRCTの実施が難しいため、症例対照研究などの観察研究も用いられるが、観察研究は各種バイアスの影響を受けやすいため結果の解釈に注意が必要となる。乳がん検診（マンモグラフィ検査）と大腸がん検診（便潜血検査）についてはRCTによるエビデンスが確立しており、子宮頸がん検診（細胞診）についてはRCTはないものの症例対照研究等で死亡率の減少効果が確認されている（菅野・勝保 2015、p.454）。

なお、日本ではこれら3種類に加え、諸外国では一般的でない胃がんと肺がんの検診が、症例対照研究によるエビデンスを根拠に実施されている。胃がん検診については、海外でもRCTの実施例はない。肺がん検診については、海外で行われたRCTで否定的な結果が出ているが、日本の研究者による反論も見られ、有効性について議論が残っている（佐川ほか 2012、pp.940-941）。

## (2) 検診の不利益と費用対効果

がん検診には、前述の過剰診断に伴い検診がなければ不要だった治療や精神的苦痛が生じる不利益に加え、実際はがんでないにも関わらず陽性となって精密検査等の負担が生じる偽陽性、検診に伴う偶発症、放射線被ばく等の不利益も存在する。このため、がん検診の導入を考える際には、有効性のエビデンスが存在することに加え、利益が不利益を十分上回っていることが重視される。前立腺がんや甲状腺がんは生命予後に影響を与えないものの割合が高く、米国や日本の前立腺がん、韓国の甲状腺がんなどで、検診の導入後に罹患率が急増する現象が見られたことから、過剰診断の問題が認識されるようになった。乳がん検診についても過剰診断が問題視されており、施策としての実施の可否について議論が続いている。不利益を定量的に把握するのは難しいため、現状では検診ごとに個別に判断が行われる。

がん検診を政策的に公費で実施するには多額の費用がかかるため、かかった費用に見合う効果が上がっているかの検証も必要となる。この

ために用いられるのが医療技術評価（Health Technology Assessment: HTA）であり、英国で大腸がん検診の対象年齢の決定、米国で子宮頸がん検診の対象年齢・受診頻度の決定に用いられた例が報告されている（濱島 2014）。

## (3) がん検診のマネジメント

がん検診によって死亡率を減少させるためには、有効性の高い検診方法を選択することに加え、対象者の多くが検診を受け、要精密検査とされた人が精密検査を受診し、がんが発見された場合は治療を行うという各段階について、十分なカバー率を持って高い精度で実施される必要がある。このためには、欧米で組織型検診（organized screening）と呼ばれている方法が有効だとされる（斎藤2014、p.648）。対象者を名簿化して管理し、未受診者に個別に受診勧奨して受診率を高め、その後の精密検査や診断・治療の効果を把握するとともに、対象集団での最終的ながん死亡率を評価する体制を整備するものである。

## 2. 日本のがん検診の経緯と現状

日本では、諸外国で実施されている子宮頸がん、乳がん、大腸がんの検診に加え、胃がんと肺がんの検診が施策として実施されている。

### (1) がん検診の導入

日本のがん検診は、悲惨な末期がん患者を診ていた臨床医が、早期発見で一人でも救えないものかと考え、検診技術を開発して地域住民に応用したことから始まっている。最も早い時期から取組が始まったのは胃がんであり、結核検診のために普及していたX線撮影装置を応用する形で1950年代から始められた。1960年には、検診車による最初の集団検診が宮城県で実施された。当初は精度の高いものではなかったが、1960年代前半にバリウムを用いた二重造影法が開発されたことで精度が高まった。この時期はまだ国際的に見てもがん検診は実施されておらず、世界初という「バイオニア的な意味合い」（細川ほか 2015、p.991）があったとされる。子宮頸がん

についても1950年代後半から取組が始まり、1962年には宮城県で本格的な集団検診が行われた。

がん検診への国の関与は、各地で行われていたこうした先駆的な取組を補助する形で始まり、1966年には胃がん検診、1967年には子宮頸がん検診への国庫補助が開始された。1983年には、これらのがん検診が老人保健法に基づく国の事業として実施されるようになり、全国的な体制が整備された。1987年には肺がん（X線検査）、乳がん（視触診）、子宮体がん（細胞診）、1992年には大腸がん（便潜血検査）の検診が対象に加えられた。

## (2) がん検診の有効性評価

日本では当初、診療現場で発見されるがんよりも検診で発見されるがんのほうが生存率が格段によかったことから、がん検診が有効であると認識されていた（久道 2009、p.29）。

諸外国でもがん検診の取組が始まると、検診の有効性や評価の考え方に対する関心が高まり、国際対がん連合（Union for International Cancer Control: UICC）による国際会議が1978年から不定期で開催されるようになった。後に厚生省の「がん検診の有効性評価に関する研究班」の総括班長となる久道茂は、第1回の会議に出席し、早期発見で生存率が延長するよう見えるのは見せかけに過ぎないというリードタイム・バイアスの考え方に触れて非常に驚いたと述べている（久道 2005、p.230）。1985年の会議では、大腸がん検診について当時すでに有効性を示唆する研究が出ていたものの、RCTによる結果を見るまでは対策に取り入れるべきではないという議論が行

われている（大島 2013、p.24）。そしてこの1985年の会議の結果、胃がんや大腸がんの検診はコストと手間がかかるが死亡率の減少効果が示されていないとして、すでに導入済みの日本を除き、公衆衛生施策として推奨はできないという結論が示された（久道 2005、p.232）。

こうした議論を受け、日本のがん検診の有効性を諸外国に説明する必要が生じ、そのための研究が行われるようになった。胃がん、子宮頸がん、肺がん検診については有効性を示す結果が得られたが、これは国の施策として導入された後に、後追いで検証されたものである。乳がんの視触診については、有効性を示す証拠が不十分とされ、現在では単独では推奨されていない。大腸がん検診については、唯一、事前に有効性の評価を行い、その結果をもって導入の判断が行われた。これら日本で行われた有効性評価は、いずれも症例対照研究によるもので、RCTは実施されなかった。前述の久道は、1985年に胃がん検診のRCTを試みているが、コンプライアンスとコンタミネーションの問題から途中で断念したことを報告している（久道 2011a、pp.174-175）。

厚生労働省は1996年以降、数次にわたって研究班を設置し、がん検診の有効性評価について学術文献をレビューし、有効性評価の基本的考え方を整理するとともに、各がん種について検診ガイドラインを作成している。その上で、検診ガイドラインを踏まえ、市町村に対しエビデンスに基づく正しいがん検診の実施を推奨するために、「がん予防重点健康教育及びがん検診実施のための指針」（以下「予防指針」）を定めている。現在推奨されている検診は表2のとおりで

表2 日本で推奨されているがん検診

がん種	推奨されている検診方法	対象	受診間隔
胃がん	胃部X線検査または内視鏡検査	50歳以上の男女	2年に1回
子宮頸がん	細胞診	20歳以上の女性	2年に1回
肺がん	胸部X線検査 (喫煙者は喀痰検査と併用)	40歳以上の男女	年に1回
乳がん	マンモグラフィ (視触診を実施する場合はマンモグラフィと併用)	40歳以上の女性	2年に1回
大腸がん	便潜血検査	40歳以上の男女	年に1回

(出所)「がん予防重点健康教育及びがん検診実施のための指針」（平成20年3月31日 健発第0331058号通知別添、平成28年2月4日最終改正）を基に作成。

ある。諸外国でも一般的な子宮頸がん、乳がん、大腸がんの検診に加え、胃がんと肺がんの検診が推奨されている。

予防指針は、その後の研究の進展等を反映するため適時に更新されており、厚生労働省に設置された「がん検診のあり方に関する検討会」でそのための議論が行われている。近年では、2016年に胃がんと乳がんの検診方法が改定された。胃がんについては、新たに内視鏡検査が推奨に加えられ話題となったが、その過程では、エビデンスとなる質の高い国内研究が少ないことが課題として指摘された（『日経メディカル』2015.7）。

### (3) がん検診の実施体制

日本のがん検診は、集団全体の死亡率を下げることを目的に、公費負担による公的なサービスとして提供される「対策型検診」（市町村の住民健診等）と、個人の死亡リスクを下げることを目的に、受診者の自己負担で実施される「任意型検診」（人間ドック等）があると整理されている。本稿の議論は、前者の対策型検診を対象としている。

対策型検診として行われる市町村のがん検診は、当初は老人保健法に基づき、国が主体となって行われていた。しかし、1998年に地方分権の一環として一般財源化され、現在のがん検診は、市町村を実施主体に健康増進法に基づく努力義務として行われている。検診方法については厚生労働省が予防指針を示しているが、実際に採用されている検診方法や精度管理体制は市町村によって様々である。

この市町村によるがん検診は、職場で健康診断を受けるサラリーマン等の被雇用者は対象となっていない。被雇用者は、労働安全衛生法に基づき雇用者が提供する検診（職域検診）の対象となるが、同法で定める項目には、がん検診は含まれていない。職域検診や人間ドック助成という形でがん検診を実施する雇用者もあるが、その実施状況は十分把握されていない。また、がん検診をめぐることは、受診率が30%台程度と低いこと、職域検診を含めた正確な受診率が測定できていないこと等も課題となっている。

こうした状況から日本のがん検診に対しては、海外と比較してほとんど成果を上げていないとの厳しい見方が存在している（斎藤ほか 2015、p.225）。

## 3. 分析の視角

以上で見たような日本のがん検診について、本稿ではアイデア理論の枠組みを用いて分析を行う。本章では、政治学において政策変容におけるアイデアの影響がどのように理論化されてきたかを確認した上で、エビデンスに基づくがん検診がアイデアとして理解できることを示し、本稿の分析対象を明示する。

### (1) 政策変容とアイデア

政治学では、ある政策変容がなぜ起こったかを説明する変数として、「利益」、「アイデア」、「制度」という三つの要素に着目してきた（徳久 2008、p.53）。伝統的な多元主義の分析枠組みでは、当該政策に関わる政治家、官僚、利益団体等の各アクターが、自らの利益を最大化するために行動した結果として政策が実現すると分析されてきた。しかし、1970年代以降に欧米で起こった各種の大規模な政策変容が、この従来の枠組みで十分説明できなかったことから、「アイデア」という要素への関心が高まった。

アイデアの概念は様々に定義・整理されているが、代表的な文献であるGoldstein and Keohane (1993, p.3)では、アイデアを個々人の持つ信念と定義し、以下の3種類があると整理している。すなわち、宗教のように幅広く人々の行動に影響を与える「世界観」(world views)、物事の善悪や正義・不正義などの規範的な判断基準を提供する「道義的信念」(principled beliefs)、エリート層の合意に根拠を持つ因果関係の理解であり目的達成の指針となる「因果的信念」(causal beliefs)である。

先に述べた三つの要素（利益、アイデア、制度）のうち、利益とアイデアがアクターの行動に目標を与えるものであるのに対し、三点目の制度は、各アクターの行動を制約する性格

を持つ。つまり、あるアクターが政策決定にどのように関与できるかは、制度のあり方によって決まってくる。

アイデアの概念を用いた初期の研究である Derthick & Quirk (1985) は、規制緩和の事例を分析し、既得権益を維持するために政治家、官僚、利益集団によって形成されていた強固なコミュニティが、経済学者らが生み出した規制緩和というアイデアによって打破されたことを説明した。この過程では、信頼性の高い経済分析に裏付けられたアイデアの説得力と、アイデアを推進するアクターとしての大統領や有力議員の活動が重要な役割を果たした。その後、アイデア理論の枠組みによって、米国の貿易政策や英国の経済政策などが分析されたほか、日本においても、規制緩和、地方分権改革、教育政策等を対象とした分析が蓄積されている (Haas 1993、Jacobsen 1995、秋吉 2007、木寺 2012、徳久 2008)。これらの分析は、多くの場合、政治的な意思決定を伴う大規模な改革を対象としているが、本稿では、がん検診という個別施策の政策変容にアイデア理論の適用を試みる。

## (2) 「エビデンスに基づくがん検診」というアイデア

本稿で取り上げるがん検診という施策は、前章でみたように、1960年代という早い時期から試行錯誤の中で実施されてきたものである。当初は死亡率減少という観点での有効性は重視されていなかったが、2000年前後に境に認識が変化し、エビデンスに基づく検診ガイドラインの作成や、それを踏まえた予防指針の発出が行われ、検診方法の見直しが進んだ。そこでは、「がん検診を実施する際は、信頼性の高い研究で死亡率減少の有効性が示された検診方法を採用すべき」というアイデアが、検診方法の見直しという政策変容において重要な役割を果たしている。

このがん検診を方向付けたアイデアは、先に触れた Goldstein and Keohane (1993) によるアイデアの3類型でいえば、因果的信念に当たるものといえる。つまり、がん検診の専門家というエリート層が科学的検討の中で生み出した、

がん検診が効果を上げるためには信頼性の高い研究で死亡率減少効果が確認された検診方法を用いるべき、という信念である。以降では、このアイデアを “Evidence-Based Cancer Screening” の頭文字をとって EBCS アイデアと表記する。

## (3) 本稿における分析

続く第4章では、この EBCS アイデアがどのように政策変容につながっているかを分析する。最初に、がん検診をめぐる主要なアクターを確認し、政策変容プロセスが三つの段階に整理できることを明らかにした上で、各段階ごとに政策変容の様態と、そこにどのアクターが関わるかを分析する。

なお、近年では科学的根拠のあるがん検診と言った場合、死亡率減少のエビデンスがあることに加え、検診の利益が不利益を上回るエビデンスがあることも求められる (祖父江 2014、p.340)。また、がん検診が効果を上げるためには、有効性が確立した検診方法を採用することに加え、十分な受診率・精度管理の下で検診を実施する検診マネジメントも必要となる。本稿では紙幅の制約があるため、このうちの有効性が確立した検診方法を採用するという部分に焦点を当てる。また、分析の対象とするのは、日本のがん検診のうち、市町村が事業として実施するがん検診のみとする。

## 4. がん検診における政策変容の分析

### (1) がん検診をめぐるアクター

がん検診という施策に係わるアクターとしては、厚生労働省の官僚、がん検診を専門とする研究者、市町村の施策担当者、市町村議員、がん検診を提供する医療関係者、住民等がある。

一般に、あるアイデアが政策変容につながるためには、そのアイデアを推進するアクターが必要となる。がん検診において、EBCS アイデアを最もよく理解し、施策として推進してきたのは研究者であった。Haas (1993) は、ある政策分野の専門家が、しばしば「認識コミュニティ」(epistemic community) と呼ばれるネッ

トワークを形成し、アイディアの供給、海外との情報交換、国際的な政策協調などを担うことを指摘した。がん検診をめぐるのは、1970年代後半から国際会議を舞台に有効性の評価について議論する認識コミュニティが形成されており、そこでEBCSアイディアが形成され共有された。日本の研究者は、その認識コミュニティの一員としてEBCSアイディアに触れ、それを日本の施策に導入する上で重要な役割を果たしている。

そして、EBCSアイディアを導入する以前から、がん検診はすでに市町村で広く実施されていた。そこには、既存のがん検診を事業として実施してきた市町村の職員や、業務・ビジネスとして検診に関わってきた医療関係者や医療機器メーカー、がん検診の受診者である住民等の多様なアクターが、利害関係者として存在していることになる。

## (2) 政策変容のプロセス

秋吉（2007、pp.57-65）は、アイディアに基づく政策変容を以下の三つの段階に整理している。第一段階では、既存の政策の方向性を規定していた政策パラダイムの限界が認識され、新たなパラダイムへの転換が行われる。この契機としては、政策自体の問題点が顕在化し継続が難しくなる「内的要因」と、社会経済状況の変化や政治的事件等の「外的要因」があるとされる。第二段階では、新たな政策パラダイムの下で、具体的な政策や制度設計の中核となるアイディアが形成される。そこでは専門家集団である認識コミュニティが重要な役割を果たし、また特定のアイディアが採用されるためにはアイディア自体の説得力とそれを推進するアクターの存在が必要となる。第三段階では、新たな政策パ

ラダイムとアイディアの下で、個別具体的な政策の内容が決定される。この段階では、当該政策にかかわる諸アクターの利害が直接影響を受けるため、その調整が行われる。

これを本稿が対象とする日本のがん検診の政策変容に当てはめると、表3のように理解できる。第一段階は、既存の早期発見を目指す方向性に換わり、死亡率減少という観点での有効性が必要だと認識されるようになる検診パラダイムの転換である。検診パラダイムは主に国レベルの方針を決定づけるものであるが、それと同時に、実際のがん検診を実施する市町村でも同じ検診パラダイムが受容されていないと、後述する第三段階のプロセスが実現しないことになる。第二段階は、厚生労働省が市町村に対して有効な検診方法を示す段階であり、国レベルの施策として実施される。第三段階は、市町村が行うがん検診において、予防指針の推奨する検診方法が採用される段階である。

秋吉（2007）の整理では、第二段階において政策や制度設計の中核となるアイディアが形成されるとしているが、このがん検診のケースでは、EBCSアイディアは第一段階より前からすでに海外で形成されており、それが日本国内に輸入された。このため本稿では、EBCSアイディアの輸入による検診パラダイムの転換を第一段階、EBCSアイディアに沿った有効な検診方法の提示を第二段階と整理した。

## (3) 政策変容の第一段階：検診パラダイムの転換

### ① 第一段階の様態

日本のがん検診の研究者は、1978年から行われたUICCの国際会議に参加する中でEBCSアイディアに触れ、がん検診の有効性について理解を

表3 がん検診における政策変容のプロセス

政策変容の段階		内 容	政策のレベル
第一段階	政策パラダイムの転換	死亡率減少という観点での有効性評価の必要性が認識されるようになった検診パラダイムの転換	国レベル 市町村レベル
第二段階	アイディアの構築	有効性評価に基づく検診方法の推奨の提示	国レベル
第三段階	アイディアの制度化	検診方法の見直し・変更による有効な検診方法の採用	市町村レベル

(出所) 秋吉貴雄（2007）『公共政策変容と政策科学—日米航空輸送産業における2つの規制改革—』有斐閣、pp.57-65による三つの段階を参考に筆者作成。

深めていった。しかし、それがすぐに日本のがん検診の見直しにつながったわけではなく、彼らはこの時点では、どちらかといえば日本の従来のがん検診を擁護し、その有効性を海外に示すための研究に力を入れていたといえる。

その後、1990年代になると、マスコミ等でがん検診についての批判がしばしば取り上げられるようになった(村上 1998、p.92)。その一つの契機となったのが、1992年に医師の近藤誠が『文芸春秋』誌上で発表した「がん検診・百害あって一利なし」という論考である。そこでは、海外で実施されたRCTの結果を紹介しながら、日本のがん検診の有効性への疑問や、検診に伴う負担・不利益の問題が提起された。

そうした中、厚生省は1996年に「がん検診の有効性評価に関する研究班」を設置した。この研究班は、「がん検診が普及するとともに国民の関心が高まり、がん検診の有効性に関するだけ正しい情報を国民が共通のものとして持つことの大切さが認識されるようになった」ことを背景に、がん検診に関する内外の学術文献を批判的にレビューする目的で設置された(がん検診の有効性評価に関する研究班 1998、序文)。この報告書では、がん検診の評価においては死亡率減少効果の検討が最も重要であること、評価のための研究方法としてはRCTや観察研究がありRCTが最も妥当性が高いこと等のEBCSアイデアが示す内容が含まれ、その後のがん検診のあり方を方向付けるものとなった。

アイデア理論によれば、政策パラダイムの転換の契機としては、政策自体の問題点が顕在化し継続が難しくなる「内的要因」と、社会経済状況の変化や政治的事件等の「外的要因」があるとされる。がん検診についていえば、この時点まで、必ずしも検診の現場で課題や問題が生じていたわけではないと考えられ、マスコミ報道を契機とした国民的関心の高まりという外的要因が、政策見直しの契機となったと見ることができる。

## ②政策決定の場とアクター

この第一段階のパラダイム転換を決定づけた「がん検診の有効性評価に関する研究班」の報告

書は、「研究班」という比較的外部から閉じられた専門家の作業によって作成された。アイデア理論によれば、制度の制約により政策決定の“場”に参加できるアクターは限定され、どのアクターが参加できるかどうかが政策変容の内容に大きな影響を及ぼす(秋吉 2007、p.50)。「がん検診の有効性に関する研究班」は、国内各地の医学系の大学・研究所・医療機関に所属する研究者によって構成されていた。結果として、この研究班には、EBCSアイデアを推進するアクターであるがん検診の研究者が参加できた一方で、既存のがん検診に利害を持つ、市町村関係者や検診提供に関わる医療者等は参加していなかった。研究班の任務は学術的な観点からの文献レビューであり、この結果、海外で確立されていたEBCSアイデアが報告書で示されることとなった。

ただし、この報告書を受け、厚生省におけるがん検診の方針が直ちに変更されたわけではない。「がん検診の有効性に関する研究班」では、報告書を踏まえ、がん検診の有効性について市町村向けの手引きを作成している。その配布に当たって厚生省が添付した文書には、科学の世界での有効性評価と行政施策に求められる有効性評価では「データのレベルは自ずから異なる」ため、「生存率や地域の死亡率等のデータから効果が示唆されるものについても…受診を希望する住民に広くその機会を提供することが自治体に求められている」という記載が含まれ、「証拠に基づく保健医療とは、全く相容れない」ものであったことが指摘されている(大島 2004、p.99. 傍点は筆者による)。

また、後述のように、市町村においてはEBCSアイデアは現在に至るまで必ずしも十分受容されたとはいえない状況にある。

## (4) 政策変容の第二段階：予防指針の作成

### ①第二段階の様態

厚生省では、前述の報告書を受け、1998年以降、研究班を設け、各種がん検診について学術文献の系統的レビューをもとに検診ガイドラインの作成を行った。当初は推奨を提示しない形であったが、2004年以降は国際標準の手法によ

り、推奨レベルを提示する形式のガイドラインが作成・更新されている。検診ガイドラインは学術的な観点からの政策提言という位置付けであり、厚生労働省の見解や政策を示すものではないとされている。

これに対し、厚生労働省が政策文書として、がん検診のあり方を示しているのが予防指針である。予防指針では、検診ガイドラインで死亡率減少効果が確認された胃がん、子宮頸がん、肺がん、乳がん、大腸がんの5種類のがん検診について、有効な検診方法を示し、市町村にエビデンスに基づく正しいがん検診の実施を推奨している。予防指針の作成・更新のため、2003～2007年と2012年以降、「がん検診のあり方に関する検討会」（以下、「検討会」）が設置されており、がん種ごとに中間報告書を作成し、その内容を予防指針に反映する流れとなっている（祖父江2015、p.997）。

つまり、EBCSアイデアに基づいて有効と判断されたがん検診は、予防指針の推奨という形で市町村に示されていることになる。

## ②政策決定の場とアクター

検診ガイドラインの作成は、国立がん研究センターに置かれた研究班が行っており、参加できるアクターは研究者に限定される。これに対し、検診ガイドラインを受けて予防指針を作成する検討会では、研究者のみでなく、自治体関係者、医療保険者、日本医師会理事がメンバーに含まれる。ここでは、検診ガイドラインを踏まえてはいるものの、純粋に有効性のみを議論するわけではなく、検診の実施可能性等の他の要素も考慮される。こうした場では、過去の政策が次の政策選択にも影響を及ぼす「政策遺産」の影響が生じやすい。いったん形成された政策は、各アクターの利益を規定し、それが新たなパラダイムのもとで政策決定する際にも影響を及ぼすと考えられている。

その一例として、直近で行われた2016年の予防指針の改定時の議論を検討する。この改定では、胃がんと乳がんの検診方法の見直しが行われた。胃がんについては、従来のX線検査に加え内視鏡検査を新たに推奨に加えるとともに、対

象年齢を40歳から50歳に引き上げ、検診間隔を年1回から隔年に変更する改定が行われた。2015年7月の検討会の時点では、50歳以上、隔年実施を推奨する方向で意見の一致を見ていたにもかかわらず、同年9月の検討会で了承された最終版の中間報告書では、X線検査について、「当分の間、40歳代の者に対して…実施しても差し支えない」、「当分の間…逐年実施としても差し支えない」という、案の段階では含まれなかった経過措置が追加されていた。この経過措置の理由については、その後の学会等の意見や、内視鏡検査の体制整備に一定期間を要することを考慮したためと報じられているが（『国保実務』2015.11.16、p.31）、最終版を了承した9月の検討会の議事録は公開されておらず、なぜ最終段階になって従来どおりの実施を容認する経過措置が盛り込まれたのかは不透明といえる。

乳がんについては、視触診を推奨から外す改定が行われた。現行の2013年度版の乳がん検診ガイドラインでは、マンモグラフィが推奨されている一方、視触診はエビデンスが不十分だとされている（国立がん研究センター 2014）。しかし、旧予防指針では、市町村のマンモグラフィ実施体制が不十分であったことから、視触診とマンモグラフィの併用が推奨されていた。つまり、予防指針の推奨は、エビデンスだけでなく実施体制を考慮して決定されていたことになる。今回の予防指針の改定では、市町村でマンモグラフィの実施体制が整い視触診の必要性は薄れていること、また視触診は精度管理の面でも問題があることから、マンモグラフィによる検診を原則とし、視触診については「推奨しない」と明記された。それにもかかわらず、視触診を「仮に実施する場合は、乳房エックス線検査（マンモグラフィ）と併せて実施すること」（厚生労働省健康局長 2016、p.12。丸括弧は筆者による補記）という記載が残り、改定後も視触診を併用するという選択肢が残されている。

この予防指針の改定に関する論評を見ると、「事業あるいはビジネスとして、医療関係者・事業者そして自治体など多くのステークホルダーが絡むがん検診について、これまでの方針を大きく変える提言を行なえば混乱は必至」である

一方、最新のエビデンスを考慮することも必要であるため、「新しい方向を目指して、慎重な中にも一歩踏み出す内容となった」（大橋 2015、p.57）との見方もされている。つまり、利害関係者と大きな摩擦・調整を生じかねない大きな変更を避ける配慮が働いたと見ることが可能である。

以上からは、検診ガイドラインの作成まではEBCSアイディアが貫徹していたとしても、予防指針については、必ずしもエビデンスだけでは決まっていないことがわかる。その背景には、第一段階より幅広いアクターが直接・間接の影響を与えている状況があるものと推測される。

#### (5) 政策変容の第三段階：エビデンスに基づいた検診方法の採用

##### ①第三段階の様態

1998年以降、がん検診の実施主体は市町村となっており、厚生労働省の示す予防指針には強制力はない。つまり、政策変容の第三段階は、実際にかん検診を事業として行う市町村が、予防指針の推奨通りの検診を採用して初めて実現することになる。

厚生労働省が2015年度に実施した「市区町村におけるがん検診の実施状況調査」によれば、全1,738市区町村のうち、予防指針が推奨する胃がん、肺がん、大腸がん、乳がん、子宮頸がん以外のがん種について検診を実施している市区町村は1,477に上り、全体の実に85%の市区町村でエビデンスが不十分な検診がおこなわれている。最も多いのが前立腺がんのPSA検査で1,355市区町村が実施、次いで肝臓がんの超音波検査で534の市区町村が実施している。予防指針に含まれるがん種であっても、推奨されていない検診方法を実施しているケースも多く、乳がんの超音波検査が554市区町村、胃がんの内視鏡検査（調査時点ではまだ推奨に含まれていない）が353市区町村で実施されている。ここで問題となるのは、予防指針が推奨するがん検診が実施されていないということではなく、推奨されているがん検診を実施した上でさらに、エビデンスが不十分ながん検診が多く実施されているということである。

こうした状況となっている背景として、日本では、発見率や早期発見が有効性の指標とはならないことが「著しく誤解されて」おり、がん検診の専門家にも誤った判断が多いことが指摘されている（斎藤ほか 2015、p.226、229）。つまり、市町村レベルにおいては、政策変容の第一段階である検診パラダイムの転換が実現できていない状況にあるといえる。予防指針でも、冒頭で「指針以外のがん種の検診を実施している市区町村及び指針以外の検診項目を実施している市区町村の数はそれぞれ1,000を超え、科学的根拠に基づくがん検診の実施について十分でない」（p.1）という問題意識が述べられている。

個々の市区町村がどのように検診方法を決定しているか、またなぜある検診方法が採用・非採用となっているかを調査するのは容易ではないが、その一端が伺える資料として、千葉県のがん検診についての報告がある（橋本 2015）。この報告によれば、千葉県では、「できるだけ多くの乳がんを発見することを目的に」県として乳がん検診ガイドラインを作成している（橋本 2015、p.11。傍点は筆者による）。また、日本人に多い若年層の乳がんはマンモグラフィで発見が難しく、40歳代のマンモグラフィでは約3割が見逃されていることから、そうした年代には超音波検査が最も適しているとしている。千葉県のガイドラインでは、40歳代は予防指針の推奨する隔年実施のマンモグラフィの間の年に、超音波検査を実施することとしている。超音波検査による死亡率減少の有効性は証明されていないため、あくまでも各市町村の判断で施行しているとしているが、2013年度で、千葉県内54市町村のうち9割以上の50市町村が超音波検査を実施している。

この事例から伺えるのは、自治体はがん検診で少しでも多くのがんを発見しようと努力しており、そうした中で、エビデンスがまだ確立していないがん検診を積極的に採用している状況である。無理解や怠慢等ではなく、現場の熱意が、エビデンスに基づかないがん検診の採用につながっている。しかし、EBCSアイディアの観点からは、エビデンスが不十分ながん検診の実施は望ましいことではない。たしかに、日本人

の乳がんは欧米に比べ若い年代に多く、しかもその年代は乳腺密度が高くマンモグラフィでは乳がんの発見が難しいという問題があることが指摘されている。このため、マンモグラフィと超音波検査を併用することで検診精度を高めることが検討されており、その有効性を評価する大規模なRCTが国内で進められているが、死亡率減少を確認するには長期間の観察が必要であり現時点ではまだ結果が得られていない（鈴木・大内 2013）。このRCTを実施している東北大学のグループ自身、「これだけ多くの超音波検診がエビデンスの裏付けがないままに実施されていることは憂慮される事態である」との見解を示している（鈴木ほか 2014、p.59）。

## ②政策決定の場とアクター

市町村レベルのがん検診の政策決定過程は、国レベルと比較して以下の二点の特徴がある。これらは、いずれも、推奨どおりの検診方法の採用を阻害する方向に作用するものとなっている。

一点目として、政策決定の場が国レベルと比較して開かれており、幅広いアクターの影響を受けやすいことである。実際の決定過程のあり方は市町村によって異なると考えられるものの、市町村には、検診に利害を持つ様々なアクターが存在する。市町村の担当者は、市民・議会・医師会等から「受診率向上」を要請され、「最先端の検診方法で、様々ながん種の検診を、広い対象に提供すること」（菅野 2013、p.619）が求められる状況にある。

二点目として、先にも触れた「政策遺産」の存在がある。がん検診が新たに導入されるのであれば問題は少ないが、市町村はすでに長きにわたりがん検診を実施している。このため、事業やビジネスとしてその検診に関わる、自治体職員、医療関係者、医療機器メーカー等の多くのステイクホルダーが存在している。これらステイクホルダーの存在は、既存のがん検診の見直しを阻害する方向に作用すると考えられる。日本におけるエビデンスに基づく医療（Evidence-Based Medicine: EBM）の代表的な提唱者の一人である名郷（2014、p.71）は、エビデンスに基づいたがん検診の議論が進まない現状について、

「がん検診は科学的な検討のもとに行われるべきだと思うが、世の中全体からすれば、そんなことよりも検診が事業化されて、そこに雇用が生まれ、経済が回るというようなことのほうが、はるかに重要なかもしれない。そうでも考えないと、いまのがん検診の現状を理解することは困難である」と述べている。政策変容の第三段階は、現場で政策変容の影響を受けるアクターとの利害調整を伴うものであり、それが進まない検診方法の見直しは実現できないことになる。

検診ガイドラインの作成において主導的役割を果たしてきた祖父江友孝（2015、p.999）は、「早期発見につながるがん検診は常に“よいこと”、と刷り込まれてきた政策決定者や一般住民にとって（おそらく、医療関係者にとっても）、[がん検診ガイドラインによる] 推奨の内容が現実の普及の程度と大きく乖離し、受け入れがたい場合もありうる」（丸括弧は原文どおり、角括弧は筆者による補記）と述べている。つまり、前項で見たように現在でもEBCSアイデアが十分受容されていない市町村では、早期発見の観点からすでに実施してきたがん検診があり、それが政策遺産となって、推奨が変更されても検診方法の見直しが難しい状況があるものと見られる。

## 5. まとめと考察

### (1) まとめ：エビデンスに基づくがん検診の現状

以上では、アイデア理論の枠組みを用いて、日本のがん検診の現状を分析した。この結果、日本ではEBCSアイデアに基づく政策変容が不徹底となっていることが明らかとなった。具体的には、死亡率減少のエビデンスがあるがん検診を実施することに加え、死亡率減少のエビデンスがないがん検診が広く実施されている。

この政策変容は、以下のようにまとめることができる。日本では、死亡率減少効果が証明されたがん検診を行うというEBCSアイデアが海外から輸入され、すでに行われていたがん検診の有効性を見直し、エビデンスのあるがん検診

に置き換えるという形の政策変容が意図された。この政策変容は、第一段階の検診パラダイムの転換、第二段階の有効な検診方法の提示、第三段階の有効な検診方法の採用というステップに整理できる。第一段階の検診パラダイムの転換は、国レベルでは実現しているものの、市町村レベルでは必ずしも徹底されていない。第二段階の有効な検診方法の提示は国レベルの施策として実現しているが、その過程ではエビデンス以外の要因も考慮に入れられていることが明らかとなった。第三段階の有効な検診方法の採用は個々の市町村において行われるが、この段階が十分実現しておらず、エビデンスの不十分ながん検診が多く実施されている。

つまり、第一段階のパラダイム転換が市町村までは徹底されておらず、このために、第三段階の有効な検診方法の採用も不徹底となっているといえる。

## (2) 考察：なぜエビデンスに基づくがん検診が実現しないのか

このようにエビデンスに基づく検診方法の採用が不徹底となっている原因、言い換えれば、エビデンスの活用を阻害している要因としては、本稿の分析から以下の二点が指摘できる。

一点目は、発見率など早期発見の指標ではなく、死亡率減少という指標で有効性を評価すべきというEBCSアイデアが、市町村レベルでは十分受容されていないことである。その背景として、がん検診の目的は早期発見であるという理解が広く定着している一方で、死亡率減少効果が重要だという点は直感的に理解が難しいことがあるものと推測される。がん検診関係者の、できるだけ多くのがんを発見しようという熱意が、死亡率減少効果は証明されていないが早期発見に有効と考えられるがん検診、すなわちエビデンスの不十分ながん検診の採用につながっている。つまり、根本的な問題として、がん検診という施策の目標やあるべき姿について、関係者間の合意ができていないという点を挙げるができる。

二点目は、すでに実施しているがん検診という政策遺産の存在である。政策変容の第二段階

である厚生労働省の予防指針を見ると、改定を行う際に、既存の推奨を経過措置として残す激変緩和的な配慮がなされていることが観察される。第三段階の市町村の検診方法の決定においては、すでに実施しているがん検診が政策遺産となっており、検診方法の見直しを阻害している可能性が示唆される。検診方法を見直す際は、市町村の既存の施策や制度を変更する必要があり、利害関係者との調整が生じる。政策変容の説明変数として、利益、制度、アイデアの三つがあるとするならば、がん検診についていえば、アイデアが政策変容を促進する方向に作用するのに対し、利益と制度が政策変容を阻害する方向に作用しているといえる。

以上の分析から浮かび上がるのは、政策決定は必ずしもエビデンスだけに基づいて行われるわけではないという現実である。がん検診のように、医学や公衆衛生学のエビデンスに基づいて決めることが合理的と考えられる施策であっても、市町村における検診の実施体制や、住民の要望等の様々な要素が作用する中で意思決定がなされる。「エビデンスは政策プロセスに影響を及ぼす多くの要素の一つにすぎない」（ナトリー 2015、p.39）という理解は、このがん検診の事例にも当てはまるものといえる。

## (3) 残された課題

本稿では、がん検診を取りまく問題のうち、なぜ有効な検診方法が採用されないのかという部分に焦点を当てて分析を行なった。しかし、がん検診とエビデンスをめぐっては、これ以外にも多くの疑問や問題点が存在する。本稿で分析対象としなかったが、今後、さらなる研究や解明が必要な課題として以下がある。

第一に、なぜ日本では、有効性についてのエビデンスが確立するより前に、がん検診を国の施策として導入してしまったのかという問題がある。日本のがん検診は1980年代前半には国の施策となっており、この時期にはまだ死亡率減少による有効性評価が必要だという認識が国内では定着していなかったという事情はある。しかし、海外では1985年の時点でも、大腸がん検診の導入はRCTの結果を見てからにすべきと議論

されており、十分なエビデンスを確認するまでは施策として導入すべきでないと考えられていた。がん検診に限らず、日本の保健医療政策では、効果が確認される前に本格的な事業を開始することが繰り返される傾向にあると指摘されている（福田・今井 2008、p.25、大島 2012、p.851）。なぜ日本では有効性を確認してから施策として導入するという発想が希薄になりがちなのか、またその背景として日本特有の問題構造があるのかどうかについて、明らかにしていく必要がある。

第二に、国内において質の高いエビデンス、特にRCTが不足しているのはなぜかという問題がある。がん検診の有効性評価が必ずRCTによらなければいけないわけではないが、RCTが行われていないことで、特に日本独自の方式となっている胃がんと肺がんの検診について、有効性に議論の余地が生じている。この状況は、EBCSアイデアについて最も理解の深いアクターであるがん検診の研究者が、RCTの必要性を認識しながらも、それを実施できていない状況と理解できる。がん検診に限らず、日本の医学研究においては、2000年代半ばまでRCTがほとんど実施されてこなかった（福井 2011、2006）。近年ではこの点は改善されつつあるが、医学以外の政策領域についても、RCTの実施が進まない要因を分析する必要があるだろう。

第三に、エビデンスに基づいた検診マネジメントが実現しない現状をどうしたら改善できるのかという問題がある。がん検診が成果を上げるためには、死亡率減少のエビデンスのある検診方法を採用するだけでは不十分であり、受診率、精度管理等の検診マネジメントもエビデンスに基づいて改善する必要がある。日本では、国民の大多数を占めるサラリーマンを対象とした職域健診でがん検診の実施義務がなく、がん検診が国民全体をカバーするように提供されていない。また、受診率が低く精度管理体制も不十分で、欧米で成果を上げている組織型検診の条件を満たしていない。この問題に関しては、厚生労働省のがん検診のあり方に関する検討会でも議題として取り上げられており、今後の検討の行方が注目される。

## おわりに

本稿では、がん検診を事例に、エビデンスに基づく政策・実践が実現しにくい状況について考察を行った。現在、様々な分野でエビデンスの活用が議論されるようになったが、その源流は医療におけるEBMにあり、保健医療の分野は、エビデンスに基づく議論や政策が最も受け入れられやすい分野といえる。しかし、その保健医療においてもなお、政策を取り巻く様々な状況や制約のために、エビデンスの活用が阻害される場合があることがわかる。

エビデンスに基づく政策・実践を推進するということは、本稿で見たアイデア理論に即しているならば、研究による最良のエビデンスを意思決定に活用するというアイデアを導入し、各種分野で政策変容を起こすということにほかならない。今後は、他の政策分野についても掘り下げた分析を行い、エビデンス利用を促進あるいは阻害する要因を明らかにし、エビデンス活用に資する知見を蓄積することが求められる。

## 付記

文中の意見は筆者個人のものであり、所属する組織を代表するものではない。

## 注記

- 1 死亡率とは、ある集団で一定期間中に死亡した者の割合。総死亡または死亡原因別に計算できる。（がん検診の有効性評価に関する研究班 1998、p.5）
- 2 生存率とは、ある病気（本稿の場合はがん）をもつ患者集団において、ある期間（診断から5年経過後など）までに生存している者の割合。（がん検診の有効性評価に関する研究班 1998、p.6）

## 参考文献

- 秋吉貴雄（2007）『公共政策変容と政策科学—日米航空輸送産業における2つの規制改革—』、有斐閣
- 秋吉貴雄（2006）「政策変容とアイデアの因果関係に関する研究」、『熊本大学社会分館研究』、4、1-15

- 大島明 (2013) 「わが国のNCD (非感染性疾患) 対策への警告」、『社会医学研究』、30(2)、23-29
- 大島明 (2012) 「NCD対策におけるスクリーニングの限界と展望」、『公衆衛生』、76(11)、849-852
- 大島明 (2004) 「がん検診の考え方」、『治療』、86(1)、2004.1、97-103
- 大島巖 (2014) 「科学的根拠に基づく実践とその形成評価アプローチが日本社会に定着しない現状と要因—改善への示唆—」、『日本評価研究』、14(2)、17-28
- 大橋靖雄 (2015) 「統計学から医療を斬る (18) 乳がん検診はどこに行くのか?」、『メディカル朝日』、44(9)、55-57
- がん検診のあり方に関する検討会 (2015) 「がん検診のあり方に関する検討会中間報告書—乳がん検診及び胃がん検診の検診項目等について—」、2015.9
- がん検診の有効性評価に関する研究班 (1998) 「がん検診の有効性評価に関する研究班報告書」、日本公衆衛生協会
- 菅野匡彦 (2013) 「「地方行政における費用対効果評価の活用」がん検診の場合—予算獲得競争の現場から—」、『保健医療科学』、62(6)、617-624
- 菅野哲平・勝俣範之 (2015) 「がん検診は必要か?」、『medicina』、52(3)、452-455
- 木寺元 (2012) 『地方分権改革の政治学—制度・アイデア・官僚制—』、有斐閣
- 厚生労働省健康局長 (2016) 「がん予防重点健康教育及びがん検診実施のための指針」、健発第0331058号通知別添、平成20年3月31日、(平成28年2月4日最終改正)
- 厚生労働省 (2016) 「平成27年度 市区町村におけるがん検診の実施状況調査 集計結果」、(第16回がん検診のあり方に関する検討会 参考資料)、2016.2.18
- 国立がん研究センター がん予防・検診研究センター (2015) 「有効性評価に基づく胃がん検診ガイドライン 2014年度版」、2015.3.31
- 国立がん研究センター がん予防・検診研究センター (2014) 「有効性評価に基づく乳がん検診ガイドライン 2013年度版」、2014.3.31
- 近藤誠 (1992) 「がん検診・百害あって一利なし」、『文藝春秋』、70(9)、302-313
- 今野良編著 (2012) 『知っておきたい 子宮がん診療ハンドブック』、中外医学社
- 斎藤博 (2014) 「大腸がん検診：成果を上げるには」、『成人病と生活習慣病』、44(6)、647-651
- 斎藤博ほか (2015) 「がん検診」、佐藤隆美ほか『What's New in Oncology—がん治療エッセンシャルガイド—』、南山堂、225-233
- 佐川元保ほか (2012) 「日本肺癌学会編纂の肺癌診療ガイドラインにおける肺がん検診の推奨度に関する2010年版改定 (追記:PLCO研究結果に関するコメント)」、『肺癌』、52(6)、938-942
- 鈴木昭彦ほか (2014) 「乳がん検診における超音波診断の役割」、『総合健診』、41(2)、57-63
- 鈴木昭彦・大内憲明 (2013) 「乳がん検診の薦め—受診者にとって利益のある検診とは—」、『臨床と研究』、90(10)、15-19
- 祖父江友孝 (2015) 「胃がん検診ガイドラインの考え方」、『胃と腸』、50(8)、995-999
- 祖父江友孝 (2014) 「わが国のがん検診の現状と展望—諸外国の動向との比較—」、『公衆衛生』、78(5)、2014.5、337-341
- 祖父江友孝 (2012) 「進化するスクリーニング検査—その適用をめぐる課題と展望—」、『公衆衛生』、76(11)、844-848
- 高見元敏 (2015) 「日本における胃がん検診の歴史とこれからの展望」、『癌と人』、42、20-22
- 立道昌幸 (2012) 「職域でのがん検診—胃がん検診の転換期—」、『健康開発』、16(3)、8-14
- 田辺智子 「エビデンスに基づく医療政策の必要性—医療の質と費用対効果—」、『調査と情報 -ISSUE BRIEF-』、907、2016
- 徳久恭子 (2008) 『日本型教育システムの誕生』、木鐸社
- 名郷直樹 (2014) 「がん検診の光と影—影の部分に光を当てて—」、『月刊薬事』、56(6)、67-71
- 中山健夫 (2010) 「エビデンス—つくる・伝える・使う—」、『体力科学』、59(3)、259-268
- ナトリー、サンドラ・M.ほか著、惣脇宏ほか訳 (2015) 『研究活用の政策学—社会研究とエビデンス—』、明石書店
- 橋本秀行 「千葉県における超音波併用乳がん検診の現状と課題」、『INNERVISION』、30(8)、2015、11-14
- 濱島ちさと (2014) 「医療経済評価の方法論と事例1—がん検診の費用対効果—」、一般財団法人医療品医療機器レギュラトリーサイエンス財団、『基礎から学ぶ医療経済評価—費用対効果を正しく理解するた

- めに一』、じほう、77-90
- 久道茂 (2011a) 「がん検診判断学について」、『日本がん検診・診断学会誌』、18(3)、172-180
- 久道茂 (2011b) 「これからの消化器がん検診についての提言」、『日本消化器がん検診学会雑誌』、49(1)、5-10
- 久道茂 (2009) 「がん検診の科学的検証」、『日本消化器がん検診学会雑誌』、47(1)、28-34
- 久道茂 (2005) 「がん検診の有効性評価に関する研究」『癌の臨床』、51(3)、229-236
- 久道茂 (2003) 「科学的視点からみたがん検診の再評価と将来展望」、『日本消化器集団検診学会雑誌』、41(6)、537-545
- 福井次矢 (2011) 「我が国の臨床研究の特徴と改善の方向性」、『最新医学』、66(7)、1676-1679
- 福井次矢 (2006) 「わが国の臨床研究の現状と課題」、『学術の動向』、11(8)、12-17
- 福田吉治・今井博久 (2008) 「地域における特定健診・保健指導の進め方—エビデンスと理論に基づく実践とスキルアップに向けて—」、『保健医療科学』、57(1)、22-27
- 細川治ほか (2015) 「対策型検診と任意型検診における胃X線、内視鏡、血清学的検査の位置づけと問題点」、『胃と腸』、50(8)、2015.7、991-994
- 村上明宏 (1998) 「検診反対論者の主張に反論する」、『ばんぼう』、204、92-93
- 「がん検診検討会中間報告書 エックス線検査の対象・間隔で経過措置」、『国保実務』、2015.11.16、30-31
- 「胃癌検診に内視鏡推奨も課題山積—国立がん研究センターが9年ぶりにガイドライン改訂—」、『日経メディカル』、44(7)、2015.7、22-23
- 「無駄で有害な健康診断」、『選択』、2014.6、110-113
- Derthick, M., and Quirk, P. J. (1985). *The politics of deregulation*. Brookings Institution Press.
- Goldstein, J. (1993). *Ideas, interests, and American trade policy*. Cornell University Press.
- Goldstein, J. and Keohane, R. O. (1993) "Ideas and foreign policy: an analytical framework." *Ideas and foreign policy: Beliefs, institutions, and political change*, 3-30.
- Haas, P. (1993). "Policy paradigms, social learning, and the State." *Comparative Politics*, 25, 275-297.
- Jacobsen, J. K. (1995). "Much ado about ideas: the cognitive factor in economic policy." *World politics*, 283-310.

(2016.9.13 受理)

## **What inhibits evidence-based cancer screening?: An analysis using the concept of policy ideas**

Satoko Tanabe

National Diet Library  
stanabe0509@gmail.com

### **Abstract**

The author analyzed the factors that inhibit evidence-based cancer screenings based on the concept of policy ideas. In Japan, the idea of evidence-based cancer screenings was introduced from abroad and it was intended to re-examine existing cancer screening programs. This policy change is incomplete so far at the municipal level, which actually conduct screening programs. It became apparent that many screenings without enough evidence are provided in addition to screenings that passed effectiveness test.

It can be pointed out that the factors inhibiting evidence-based cancer screenings are the lack of understanding on effectiveness at the municipal level and the existence of “policy legacies” that previous policies affect the next policy making.

### **Keywords**

Evidence, Evidence-based policy and practice, Evidence-Based Medicine (EBM), Cancer screening, Policy ideas

## 【研究論文】

## 国際開発分野におけるエビデンス活用の現状と課題

浅岡 浩章<sup>1</sup>

国際協力機構

Asaoka.Hiroaki@jica.go.jp

## 要 約

国際開発分野で、開発効果に関する科学的に信頼性の高いエビデンスの欠如が著しいとの問題意識から、国際的に多数のインパクト評価が実施されてきた。特に過去10年間で評価や研究の実施数が急増し、信頼性の高いエビデンスは相当数整備されてきたと言える。その一方、評価結果の活用という点では改善の余地がある。

活用が十分でない背景として、これまで取り組まれた評価の多くは研究者によるものが多く政策判断への活用意図が弱かった、外的妥当性の課題を克服できていない、システムティックレビューも援助実務者のニーズに応えられていない、加えて、途上国の政策決定者や援助機関関係者の意識醸成も十分でないといったことが挙げられる。

国際開発分野において、エビデンスに基づく事業実施を推進するためには、JICAを含む援助機関はエビデンスの活用を実践し、好例を示していくことが必要である。

## キーワード

インパクト評価、国際開発、国際協力機構、エビデンス、評価の活用

## 1. はじめに

2006年にアメリカのシンクタンクであるCenter for Global Development (CGD) により「When Will We Ever Learn? Improving Lives Through Impact Evaluation」が発刊されてから10年の年月が経った。同報告書では、CGDのほか、大学、財団、世界銀行などの研究者やエコノミストから成るEvaluation Gap Working Groupにより、開発援助事業を進める上で信頼性の高いエビデンス<sup>2</sup>が著しく不足しているという問題点が指摘され、インパクト評価の更なる実践が提言された。これにより、援助機関等の開発コミュニティの中

でも、援助事業の介入効果に対するエビデンスの不足が一層意識されるようになった（青柳 2006）。

同報告書には、「10年後に、社会開発事業の実践において、今日のように何が有効な手立てかが分からないと知見の欠如を嘆いているか、エビデンスに基づいたリソース活用に向けて改善がなされているか、国際社会は二つの状況のどちらかにあるだろう」と記述されている。現在、開発援助コミュニティは、いずれの状況にあると言えるだろうか。

CGDの報告書が示したエビデンスに基づく開発事業へのリソース活用の実現には、まず第一に信頼性の高いエビデンスを産出すること、第

二にそのエビデンスを活用できることが必要となる。本稿の目的は、同報告書で挙げられた問いに対する回答として、国際開発分野でのエビデンスの産出と活用状況を確認し、その課題を整理することである。そのために、援助機関や研究機関などによるエビデンスの産出や活用について、公開されている報告書や学術論文をレビューすることで現状を把握する。そして、活用における課題と改善に向けた方向性を提示する。それら国際的な状況を踏まえた上で、日本の援助実施機関である国際協力機構（JICA）の取り組みと方向性について示す。

## 2. エビデンスの産出と発信

### (1) 国際開発分野でのインパクト評価への取り組み

CGDの報告書によって、国際開発分野においてエビデンスの不足が広く認識されることとなったが、インパクト評価の実施自体はそれ以前から開始されていた。援助関係者の中で、開発分野での大規模なランダム化比較試験（Randomized Controlled Trial:RCT）の端緒として認知されているのは、メキシコの条件付き現金給付（Conditional Cash Transfer）プログラムであ

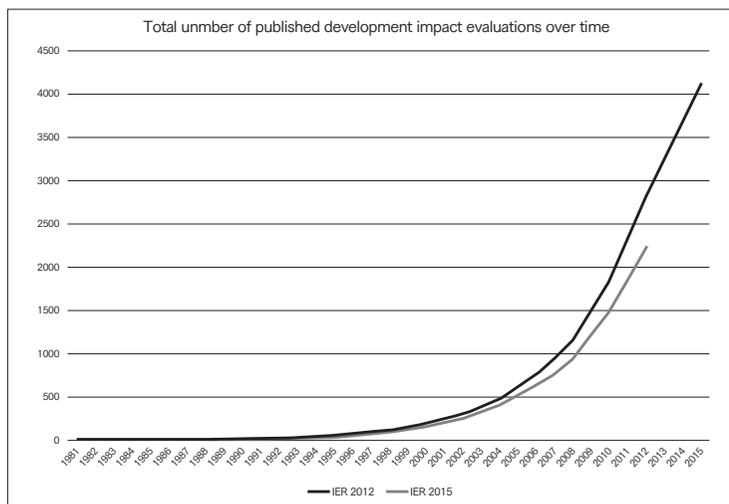
ったPROGRESAのインパクト評価であり、1997年から評価が開始された。2000年代に入ってから徐々に評価実施件数が増え、この15年ほどで実施件数が大きく伸びた。以下、現在に至るまでのインパクト評価の実践に係る変遷を概観する。

開発分野でのインパクト評価やエビデンス活用を促進する国際的なNGOであるInternational Initiative for Impact Evaluation（3ie）<sup>3</sup>は、2013年から2014年に掛けて、途上国における介入を対象とするインパクト評価の大規模なレビューを行った。検索プロトコルに基づき、45にもものぼるアカデミックや援助機関などのオンラインデータベースを中心にレビューを行ったものである。

1981年から2012年に出版された開発分野のインパクト評価の論文は2,259本に上る。その内訳をみると、2000年以前のものでは僅か132本に留まり、多くが保健医療分野の論文であった。2000年代に入り評価数が増加していくが、特に2008年以降、出版論文数は加速度的に増えた。2008年は年間173本であったものが、2009年には274本となり、2012年に350本を超えと現在までその勢いは続いている。分野としては、保健分野が全体の65%を占め、教育23%、社会保障15%、農業農村開発10%と続く（Cameron et al. 2015）。

3ieのインパクト評価データベース（Impact

図1 Impact evaluations published per year (1981–2015) in low and middle income countries



\*IER2012は2012年時点での蓄積、IER2015は2015年時点での再レビューを踏まえたもの。

(出所) (Miranda et al. 2016) より抜粋

Evaluation Repository:IER)には、図1の通り、2015年9月までに4,100件を超える評価<sup>4</sup>が蓄積されるまでになった (Miranda et al. 2016)。

他にもShiらが、3ieのデータベースに基づき、保健や教育分野などのインパクト評価件数の変遷を纏めている (Y.Shi et al., 2015)。2004年にCGDがEvaluation Gap Working Groupを形成したことから、その年を境としてインパクト評価の実施件数を比較している。例えば、同データベースで「保健／栄養／人口」に区分された途上国でのインパクト評価は、1995年から2004年までには272件しかなかったのが、2005年から2014年の間には1,599件となった。この増加件数の多くはRCTによるものである。教育分野も同様に、2004年までは92件だったものが、その後から2014年までには512件に急増している。

こうしたインパクト評価の量産、関心の高まりは、2006年のCGD報告書に加えて、2002年のIPA (Innovations for Poverty Action)<sup>5</sup>設立、2003年のJ-PAL (Abdul Latif Jameel Poverty Action Lab) 設立<sup>6</sup>、2005年の世界銀行DIME (Development Impact Evaluation Initiative) 設立、2008年の3ie設立といった一連の動きが後押しとなってきた。なお、3ie設立は、CGD報告書での提言が結実したものである。また、J-PALの共同設立者であるEster Duflo教授は、CGDのEvaluation Gap Working Groupに参加しており、互いに影響、作用し合いながら大きな潮流を生み出してきたと言える。

個別の援助機関に目を向けると、援助機関の中でインパクト評価普及の牽引役を担ってきた世界銀行では、2005年から2010年の間に、年間平均57件のインパクト評価を完了している (IEG 2012)。また、2008年には10件以下の実績に留まっていた米州開発銀行は、2013年には60件を超え、2014年には145件を実施中と急増させている (Gray 2014)。援助機関のみならず、途上国政府も評価業務への関与を強めている。メキシコ、コロンビア、チリ、南アフリカ、インドなどは政策の評価やインパクト評価実施促進を担う部署を設置し、ウガンダやフィリピンといった国でもインパクト評価によるエビデンス獲得に関心を払っている (Levine and Savedoff 2015)。

上述の経過を辿って、インパクト評価は開発

援助の表舞台で認知された存在となり、3ieや世界銀行等による継続的な評価資金の供与もあり、その勢いは続いている。

## (2) 国際開発分野のエビデンスの発信状況

援助機関にとっては、インパクト評価から得られる結果をエビデンスとして、それを事業に反映していくこと、つまりEvidence-based Practice (EBP: 根拠に基づく事業実施) に繋げることが一義的な実施目的となる。エビデンスの使い方として、まずは評価対象案件で実施した介入の効果を示すという説明責任の観点で活用することが考えられる。加えて、その案件の中で、または、案件終了後にスケールアップを行うことを想定した場合、その判断根拠として使うことが考えられる。評価対象案件以外では、類似の環境下で類似の介入内容による事業を計画する際にエビデンスとして活用することが想定される。

3ieやJ-PALといった国際組織や、世界銀行や英国国際開発省といった援助機関は、エビデンスを作り、公共知として世に広く知らしめ、活用を促す取り組みを進めているが、主要機関の発信状況について以下紹介する。

3ieは、そのウェブサイトで、政策ブリーフ、インパクト評価、システムティックレビュー、エビデンス・ギャップ・マップ<sup>7</sup>を提示している。他にも、Replication Reportという、既存の評価結果を再現することで、提示されているエビデンス内容の検証を行うなど、開発分野でのエビデンス整備と活用促進に注力している。J-PALなどは異なり、RCTによる実験法のみならず、準実験法によるものや質的なエビデンスも扱っている。既述したインパクト評価のデータベースに加えて、システムティックレビューのデータベースもあり、2015年時点で303件<sup>8</sup>が掲載されている。また、水・衛生、初等・中等教育、平和構築など5分野において、エビデンス・ギャップ・マップを整えている。加えて、「3ie Breifs」、 「Policy Highlights」という媒体で、政策決定者などが使いやすいよう簡潔に纏めた政策ブリーフも作成しており、2016年5月時点で73点、作成されている。3ieでは、インパクト評価実施に係る研究資金の提供もしており、2008年の発足から

これまで63件の評価が完了し、また、2015年には新たに28件の資金提供を決めるなど精力的にエビデンス産出を進めている。3ie Annual Report 2015では、これまで47の低中所得国において、3ieが資金提供した評価からのエビデンスにより活用されているという記載があり、また、掲載のある表からは63件の政策インパクトを生み出していると読み取れる。例えば、スケールアップに繋がったのが7件、効果がないため事業停止したものが3件、政策や事業の変更には15件があった(3ie, 2015)。

J-PALについては、そのウェブサイトで、計67か国における729件の評価結果を掲載しており、また、Brief Caseという数件の評価結果から纏められたエビデンスを示した文書を49件掲載している。研究者サイドから、過去10年にわたるインパクト評価の実施促進を推し進めてきた一大勢力と言える<sup>9</sup>。なお、J-PALは、政策実施者、実務者に対するインパクト評価の研修機会を設けており、また、オンラインコースも開設するなど評価手法の浸透にも尽力してきた。

IPAについては、その2014年年報によると、275件の評価を完了し245件が実施中とある。IPAのウェブサイト<sup>10</sup>には327件<sup>11</sup>の評価内容(実施中を含む)が掲載されているが、評価内容や結果を簡潔に示す努力が見られる。IPAもスケールアップにエビデンスを活用することを標榜している。

世界銀行は、援助機関の中で最も多くのインパクト評価を実施しており、2000年から2010年にかけて、計460件のインパクト評価を実施した(IEG 2012)。2005年のDIME設立により、実施件数は更に増えた。DIMEが中心とはいえ、事業部門独自によるもの、独立評価局(IEG)によるものなど評価実施部門が分かれており、以前はインパクト評価のデータベースがあったものの、それが廃止されたことで世界銀行によるインパクト評価の取り組み全容を一元的に把握するのは難しくなった。アウトプットもIEG、DIME等それぞれの部門のサイトに掲載されている。評価結果は、評価報告書、ワーキングペーパーなどの媒体で発表されるほか、「from Evidence to Policy」という政策ブリーフとして簡潔に公開されている。多数の評価を実施する一方、その結

果の活用状況については明確に伝わってこない。一例ではあるが、IFC(国際金融公社)がハーバード大学に発注した研究結果は、公共知としての貢献はある一方、IFCスタッフには十分に共有されていなかった(IEG 2012)。最近では、DIMEが、研究者と事業部門(世界銀行スタッフ及び途上国政府側)とのマッチングの場を設け評価案件形成を進める、途上国関係者の評価能力強化を進めるなど、実施促進と結果活用に向けた取り組みがなされている。

英国国際開発省は、その2015年度評価年報によると、2012年から開発分野でのエビデンスギャップを埋めるという方針を掲げ、これまで8,770万ポンドをインパクト評価等の実施に投入した。その資金により、外部機関などが299案件の評価をこれまで実施している。3ieへの活動資金拠出や、世界銀行のThe Strategic Impact Evaluation Fund(SIEF)への拠出など、多額の資金を提供しており、開発分野の公共知の産出促進を行っている(DFID 2016)。

### (3) 小括

上述のように、過去10年近くで、インパクト評価からのエビデンスの産出数は急増し、また、それらの発信も関連機関の努力により相当進んできたと言える。しかし、あくまで過去と比較してのことであり、また、保健医療分野や教育分野等に偏っていることから、継続的に産出していくことが重要となる。

## 3. エビデンス活用の現状

上述の通り、インパクト評価からのエビデンスを得る情報環境はここ数年で整ってきており、また、評価結果を政策決定者に平易に伝えるための各機関の努力も払われてきた。スケールアップに繋がった事例も出てきている。

一方、開発事業を実施する途上国政府の政策決定者や援助機関の実務家は、インパクト評価結果を活用している、EBPを十分に実践出来ているといった実感は未だ薄いものと思われる。昨今、エビデンスの活用状況について問題提起さ

れる場面が見られるようになってきた。Langerら(2015)は、開発分野でのエビデンス活用が不十分であるとの認識の下、その理由として、1) 政策決定者、実務者ともに研究エビデンス活用の訓練や動機づけがなされていない、2) 開発政策の中で研究エビデンスを取り入れるための組織的なメカニズムや動機づけがない、3) 開発事業においてエビデンスの活用を確かめる組織的なメカニズムがない、4) そもそも開発政策は漸進的な変化に基づき社会的なインパクトを引き起こしているが、研究のエビデンス対象はその一部に限られている、といった点を挙げている。

エビデンスの産出の主要アクターであるアカデミア側からの視点と、援助に携わる実務家側からの視点から、その現状への認識を見ていきたい。

(1) アカデミアから見たエビデンス活用における課題

J-PALのDeputy Directorを務めるIqbal Dhaliwalらが執筆した「From Research to Policy」(2012)では、研究結果を政策決定者による活用に繋げる難しさについて言及がなされている。同文書は、J-PAL関係者、パートナー機関関係者に対する聞き取りに基づき纏められたものである。表1にて、活用されない理由として挙げられている点を示す。

J-PALは、上述の課題への対応策として、1) 研修コースの提供によって、政策決定者(途上

国、援助機関等)にエビデンスの使用者としての能力獲得を促す、2) 政策決定者がエビデンスの産出者となるよう働きかけ、また、その組織の評価能力の向上を支援する、3) 評価実施を通じて研究者と政策決定者間で長期的な関係を築く、といった点を挙げている。J-PALから政策決定者への働きかけをする一方で、研究者が政策決定に関与していく困難さへの認識、更にはそのプロセスに巻き込まれることへの躊躇や恐れが見て取れる。研究者としての関心を追求することに注力するか、実世界を変えていくことに向き合うか、研究者の葛藤が伺える。

(2) 援助実務者のエビデンス活用に関する意識

開発分野事業の政策決定者として、途上国の政策決定者と、援助機関の政策決定者とに大別できる。後者である援助機関の実務者の中でも、エビデンスの活用の意識が十分とは言えないのが現状である。

Ravallion(2011)は、世界銀行の事業部門のスタッフが、研究からのエビデンスを活用できているか分析した。世界銀行職員の約4分の1を占めるシニアスタッフと言える2,900名を対象に行われ、555名分の回答があった。アカデミックな研究への精通度、エビデンス活用の動機、世界銀行の研究に対する馴染み、研究に対する価値といった複数の項目に対して、1(とても低い)から10(とても高い)の10段階で聞いている。そ

表1 エビデンスが政策決定に活用されない理由

① 政策決定は、エビデンスだけでなく、様々な要因(イデオロギー、無知、慣習、直感)によって決まる。
② 政策決定者がエビデンスを欲していても、得ることが難しい。
③ 政策決定者がエビデンスの解釈、活用することが難しい。 <sup>12)</sup>
<p>【③の理由】</p> <ul style="list-style-type: none"> <li>▶ 研究結果はアカデミック向けに示されている。</li> <li>▶ 政策決定者は、エビデンスの質の見極めが難しい。</li> <li>▶ 政策決定者は、違う条件下で実施されたエビデンスの比較が難しい。</li> <li>▶ 政策決定者は、同じ政策目標を目指した異なるプログラムのエビデンスを比較することが難しい。</li> <li>▶ 政策判断とエビデンスの産出タイミングが合わない。研究結果が出るのに時間が掛かる。</li> <li>▶ 研究者が、継続的に政策決定者に関与していくことが難しい。研究者が政策決定に関わることは、アカデミアの世界では評価されない。研究者にとって、政策決定者に関わり影響を与えることは、長期にわたり、労力がかかり、リスクなこと。</li> </ul>

(出所) (Iqbal Dhaliwal et al. 2012) から筆者まとめ。

の集計結果の一例を表2に示す。

表2の通り、「(エ) 活用し、よく知っている」と回答したのは42%と高い一方、「(ア) 知らされなくて結構」との回答も20%を超えている。部門による違いも大きくあり、(エ) の割合が高いのが、貧困削減部門 (83%)、経済政策 (58%)、教育部門 (51%)、保健部門 (47%) であり、エコノミストの数や、インパクト評価をはじめとする定量的な分析への馴染みの深さが影響していると考えられる。

一方、(ア) の割合が高いセクターは、エネルギー・工業 (41%)、都市開発 (37%)、環境 (36%)、農業・農村開発 (34%) など、インフラ分野を中心にエビデンス活用の意識が低い。インフラ案件は一般的にインパクト評価が実施しにくいと言われており (Estache, 2010)、その影響もあるやもしれないが、加えて、アカデミックな研究への精通度や、伝統的な支援方法の名残といった側面が影響していると考えられる。これら世界銀行の研究結果への活用が低いセク

ター部門 (エネルギー・工業、都市開発、農業・農村開発、交通) は、貸付額全体の45%を占める一方、スタッフの15%しか研究に馴染みがないという結果であり、特にこれらの分野での研究成果の活用が必要とされる。

ドイツでも、PEGNetという開発分野のネットワークに関わる個人や組織に対するウェブや質問票送付による調査によって、類似の調査がなされた (Kleeman & Bohem, 2013)。回答者は105名で、その45%がNGO関係者、37%が援助機関関係者を占めている。ドイツの開発分野全体の傾向を示すものではないと考えられるが、研究からのエビデンス活用への意識が伺える一例である。表3の通り、40%が「Happily uninformed」を占めた。また、研究結果へのアクセスの問題も挙げられており、回答者の3分の1から、有料の学術論文購読のための資金的障壁が大きいとの回答があった。開発分野の実務家による研究エビデンスの活用にあたり、意識面での大きな課題に加えて、資金面でも障壁になっている。

表2 世界銀行の研究に対する価値とスタッフによる精通度に係る調査結果

		(2) 業務に対する世界銀行研究の価値	
		低	高
(1) 知識: スタッフの世界銀行研究に対する精通度	低	(ア) Happily uninformed 「知らされなくて結構」 人数:117 (22.54%) 平均精通度: 3.35 平均価値 : 2.72	(イ) Frustrated uninformed 「知らされないと嫌だ」 人数: 123 (23.70%) 平均精通度: 4.00 平均価値 : 6.35
	高	(ウ) Independently well-informed 「仕事には不要だが、関心あり」 人数: 62 (11.95%) 平均精通度: 7.26 平均価値 : 2.90	(エ) Functionally well-informed 「活用し、よく知っている」 人数: 217 (41.81%) 平均精通度: 7.58 平均価値 : 7.57

(出所) Ravallion (2011) より筆者訳

(回答者数:519)

表3 PEGNetによる調査結果

		日常業務に対する研究の価値	
		低	高
活用度合	低	(ア) Happily uninformed 「知らされなくて結構」 40%	(イ) Frustrated uninformed 「知らされないと嫌だ」 17.14%
	高	(ウ) Independently well-informed 「仕事には不要だが、関心あり」 9.52%	(エ) Functionally well-informed 「活用し、よく知っている」 33%

(出所) Kleeman & Bohem (2013) から筆者まとめ。

(回答者数:105)

(3) 小括

インパクト評価からのエビデンスを量産してきたアカデミアと、その結果の活用が期待される援助実務家を含む政策決定者との間では、評価実施にあたり、関心事、実施目的、活用への意識といった様々な点で未だ隔たりがあると認識される。援助のプロと言える援助実務家ですら、エビデンス活用自体への意識が十分には醸成できていないことが見てとれる。

4. エビデンス活用を進めるための方向性

(1) 政策判断に活用される評価の促進

それではインパクト評価からのエビデンスが活用されるには何が必要であろうか。Shahら(2015)<sup>13</sup>は、従来のインパクト評価の多くは研究者を中心に新たな知識を生み出す研究目的でなされてきたが、政策への反映が十分でなかったとの見解を示している。その上で、研究志向の強いKnowledge-focused evaluation (KFE) に対し、政策判断を目的とするDecision-focused evaluation (DFE) という考えを提示した(表4)。なお、評価には両方の目的を包含して実施されることもあり、KFEとDFEの違いは必ずしも明確に二分できるものではないが、一義的な目的が理論形成か、政策判断への活用かという点で区分したものと認識される。

同ペーパーでは、KFEの活用が不十分な論拠として、J-PALが過去に626件の評価を支援しつつも、15のスケールアップ例しかWEBでの掲載がないことや、IPAの創始者であるDean Karlan氏の

言葉として、「J-PALとIPAが政策へ大きな影響を与えた評価結果は5から10のみであり、想定していたほどの結果を示せていない」といった認識を示している。そして、KFEが政策判断に活用されづらい理由を以下のように挙げている。

KFEと政策との弱い繋がり／使われない理由	
【関心のずれ】	KFE評価者の関心と、政策実施者との関心の不一致
【時間】	学術論文として世に出るまでに掛かる時間が長い
【費用】	評価費用が高く、政策実施者の関心ある評価ができていない。
【外的妥当性 <sup>14</sup> 】	別のコンテキストでも使える介入内容かが検証できていない。
【難解さ】	政策実施者にとって、評価結果の解釈が困難。
【政策環境】	エビデンスはあっても、それを使う政治的なバリアが存在する。

(出所) Shah et al. (2015) から筆者まとめ

これらの問題意識を踏まえて、開発理論を掘り下げの意義は保持されるとしてKFEの実施意義には言及しつつも、開発援助の現場で政策・施策への反映に活かされるDFEの一層の実施を提案している。

DFEの有すべき特徴として以下の4つを挙げている。

- ①Demand-driven (政策実施者の要望に沿っている)
- ②Tailored (状況・環境に合わせられる。より安価に実施しうる)
- ③Embedded (政策決定プロセスに組み込まれている)
- ④Cost-effectiveness (費用対効果の視点がある)

(出所) Shah et al. (2015) から筆者まとめ

表4 KFEとDFEの実施目的と成功の定義

	Knowledge-focused evaluation (KFE)	Decision-focused evaluation (DFE)
実施の目的	研究者を中心に開発の理論や介入に関する新たな知識を生み出す研究目的で実施する評価	特定の対象者・対象地域や時間軸を有する政策実行者に対して、政策決定への判断を目的に実施する評価
成功の定義(活用例)	<ul style="list-style-type: none"> <li>・開発理論への貢献</li> <li>・高いレベルの政策討議への貢献</li> <li>・一般的な介入のスケールアップ</li> </ul>	政策実施者のコンテキストにおける意思決定の情報提供 <ul style="list-style-type: none"> <li>・政策判断</li> <li>・スケールアップ判断</li> <li>・事業中止判断</li> </ul>

(出所) Shah et al. (2015) から筆者まとめ

調査に係るコストは実施方法次第であるものの、常に大規模かつ検証に長期間を有するような精緻な分析を行わずとも、目的に応じて既存データの活用や、短期でアウトカムを検証するといった、より簡易なやり方の検討を提示している。研究としての質よりも、如何に政策判断に使っていくかという観点から評価設計を行うことの必要性に言及している。また、外的妥当性に関し、DFEは同じコンテキストを持つ環境でのスケールアップ判断などに活用することを目的に実施することから、KFEよりもDFEの方が、外的妥当性に関し問題となる脅威は小さいとしている。なお、DFEの実施促進のためには、実施意義の啓発、予算配布、評価者の育成<sup>15</sup>といった対応が必要とされている。

KFEは、言い換えれば、研究目的が先行し、政策決定者の関心をカバーできていない評価と言えよう。それなりの数のインパクト評価が産出された一方、結果の活用については不十分という認識があるならば、まずは地道にDFEを実施して、結果を活用していくという方向性は現実的であり、エビデンス活用の意義を示すためにも有効と言える。

## (2) エビデンスの汎用性 ～外的妥当性への対処～

DFEは、その評価結果を当該事業の実施や展開において直截的に用いることが想定されている。しかしながら、DFEをすべての事業において実施することは不可能である。また、特定の介入にDFEを実施したとしても、その介入をより広範囲に適用して同様の効果が発生しうるかという外的妥当性の観点は、DFEにも求められる。つまり、インパクト評価から得られる結果には、外的妥当性の問題が常についてまわると言える。この点、エビデンスの活用に向けて留意が必要であるため、ここで触れておく。

Petersら(2015)は、2009年から2014年に掛けて主要な経済学学術誌<sup>16</sup>に掲載された92のRCTや実験研究を対象に、研究の実施方法や、外的妥当性の説明が含まれているかを検証した。分析者による論文内容の検証結果を、論文の筆者に直接確認し、回答があったものは必要に応じ

て分析者の誤認を修正している。質問として、1) ホーソン効果<sup>17</sup>、ジョン・ヘンリー効果<sup>18</sup>に関する説明があるか、2) 調査対象者が調査に参加していることを認識しているか、3) 一般均衡効果<sup>19</sup>についての説明があるか、4) スケールアップされる場合についての記載があるか、5) 時間が経過した場合の効果発現について言及があるか、結果の一般化や外的妥当性について言及があるか、6) 調査の母集団の代表性について言及があるか等の、計10個の問いを投げかけている。結果として、まず調査の実施方法の説明が十分に記載されていないことに加えて、調査対象者が実験に参加しているかの認識有無については半数以上で明らかにされていなかった。また、一般均衡効果についてはほとんど言及がない。外的妥当性に関する記載はケースによって異なるが、調査対象が政策裨益者の中でかなり限定されている、研究者と調査対象者の間で密な接触があったといったケースも見られたとのことである。研究者からの反論として、評価設計段階で外的妥当性について検討をしているものの論文には書いていないだけという説明が時になされるものの、政策決定者にはそれら含まれなかった情報も必要としている。RCTは内的妥当性には応えているものの、外的妥当性についての説明も努力が払われるべきとの見解が示されている。

J-PALの共同創始者であるAbhijit Banerjeeも、実験手法からのエビデンスにおける外的妥当性について課題を認識している(Banerjee et al. 2016)。外的妥当性に関する重要事項として、1) その介入はスケールアップ可能か、2) 異なる母集団に対してどう効果が表れるか、3) 同じ母集団に対しても異なる状況下に置かれた際にはどうなるか、といった点を挙げている。外的妥当性の課題を乗り越え、異なる条件や、異なる母集団に対しても、介入効果を予測していくことを、“Structured Speculation”と呼んでいる。Structured Speculationのガイドラインとして、実験研究者は、1) 研究結果の外的妥当性についてシステマティックに推測すべき、2) “Speculation”は論文の中で分けて明示すべき、3) “Speculation”は正確かつ再現可能であるべき、と主張してい

る。実験結果を政策に反映していくには、外的妥当性の課題に応える必要があるという考えに基づき提言した内容と認識される。なお、外的妥当性に関する仮説の検討にあたり、仮説が明瞭かつ再現可能である限り、間違ったやり方というものはないとし、また、質的な内容が必ずしも分析的な内容より精緻さに欠けるものではないといった指摘もなされている。RCTを推進してきたBanerjeeの見解として興味深く、外的妥当性への対処の重要性と共にその難しさが伺われる内容である。

外的妥当性については、唯一の解が示されるわけではないが、どのような状況、条件、コンテキストの下で事業の介入が行われ、また、評価については、どのようなデザイン、対象範囲、時間軸で行われたかということ意識し、その設定からの結果を、他の設定に対して適用するかを熟慮すべきというメッセージと捉えられる。

### (3) システマティックレビューへの期待

Shahらが示したKFEは、「開発の理論や介入に関する新たな知識を生み出す研究目的で実施する評価」と定義されていたが、個々のインパクト評価等を統合して整理するシステマティックレビューも、開発の理論や知識の産出に貢献するものである。また、上述した外的妥当性を直接説明するものとはならないが、検討にあたり参考情報となる。加えて、システマティックレビュー自体がエビデンスの汎用性を高める取り組みと言えることから、システマティックレビューの取り組みと今後の期待を以下に示す。

国際開発分野でシステマティックレビューへの取り組みが本格化してからまだ10年も経っていないが、3ie、英国国際開発省、豪国国際開発庁などの支援でレビューの数が、昨今増えている。特に3ieは、社会科学分野でのシステマティックレビューの産出、活用を推進するCampbell Collaborationの国際開発グループで中心的な役割を果たし、また、多数のレビュー実施を外注や内部作業で進めるなど、この取り組みを牽引してきた。3ieのシステマティックレビューのデータベースには、303件のレビュー情報が格納され

ている<sup>20</sup>。システマティックレビューには、通常、外的妥当性を分析に含まないものの、ユーザーが外的妥当性に対処するためのコンテキスト、母集団、介入内容に関する情報を提供することが可能としている(White et al. 2012)。

一方、国際開発分野のシステマティックレビューは、未だ課題が多い。まず、レビューの対象となる個別の評価や研究数にまだ限りがあるため、数少ない案件に基づいたエビデンスの提示になるケースが見られる。また、3ieによると、システマティックレビュー推進の初期段階では、レビューの設問内容を広く設定しすぎてしまったが、最近では、より絞った設問を置くようになったということである(Waddington et al. 2012)。政策決定者がシステマティックレビューに期待する内容と、レビュー経験者による実務的な視点に基づくレビュー対応範囲とに差異があると思われ、埋めていく必要がある。

国際開発分野でのシステマティックレビューの改善点として、設問を絞ることでより実用的な結果を示すことに加えて、Snijlsvit (2012)は、従来のレビューにおけるプログラムセオリー<sup>21</sup>の欠如を指摘し、精緻に分析からの定量的なエビデンスのみに固執せず、定性的な分析や文書(プロセス評価、プロジェクト文書等)も併せてレビューに加えることによって一層使えるレビュー内容になるものとしている。過去のレビューでは欠けていた、事業実施の情報などが入ることで、他の状況、環境下に置かれた介入の検討においても、より具体的な検討ができるようになる旨を指摘している。

以上のように課題はまだ多いが、今後レビュー数が増えることによって、また、インパクト評価の数が増えてより質の高いレビュー結果を示せるようになると、将来的には異なる条件下での介入策を当てるかといった判断材料が増えることとなる。システマティックレビューは政策判断において、また、開発分野の理論形成に向けて、有効なツールの一つになると期待される。

### (4) 国際開発分野特有の課題

上述のように、国際開発分野において生み出

したエビデンスの活用を進める上で乗り越えていかなければならない課題が多い。そもそも、EBPは、欧米におけるエビデンスに基づく医療（Evidence-based Medicine:EBM）の活性化に追随する形で社会政策全般に浸透してきた。EBMは介入に対するアウトカムを生体反応として計測することから、社会、文化、経済状況、個人の嗜好など多様な要因が影響してくる社会政策ほどの複雑さはない。また、評価の目的、対象にもよるものの、一般的に先進国で行われてきた医療分野の評価は、医療施設などよく管理された状況で行うことが可能な場合が多く、既存の行政データやカルテなどの医療データを活用できるケースも多い。

一方、国際開発分野で扱う事業は、最終的に多数かつ広範囲に居住する受益者を対象としたものが多い。例えば、農村開発事業であれば農村部に点在する農家が調査対象となり、それに掛かる手間やコストは小さくない。加えて、学校、医療施設、その他行政機構によって取られる既存データの不備や不足、政府統計の不備やデータの信頼度が低いといった問題もあり、途上国特有の課題があると言える。

また、アメリカやイギリスで推進されてきたEBPでは、自国の社会政策の実践のためにエビデンスが整えられてきた。それら得られたエビデンスから、他国への単純な政策移植には注意が必要となるが、類似した社会政策を取りうる国において大いに参考になるものと考えられる。一般論として、途上国においても同様に参考になるとも言える一方、予算や人員の制約による政策実施能力の低さ、行政システムや社会文化の多様性から、各国や地域間での違いが大きいと言える。よって、得られたエビデンスの活用には、それを適用する社会や行政体制といった各国特有のコンテキストを踏まえる必要がある。

これら国際開発分野特有の課題は筆者の見解の域を出ないものであるが、外的妥当性の問題やシステムティックレビューが有する課題に直結する事項でもあり、また、エビデンス活用時に留意すべき重要なものと思料する。

## 5. まとめ 一国際開発分野における更なるエビデンス活用への期待一

「When Will We Ever Learn?」の発刊から10年、国際開発の分野においてエビデンスは確実に増えたと言える。勿論、引き続きエビデンスが不十分な分野や介入内容は残されており、その産出を続けていく必要があるものの、援助機関、アカデミア、NGOなどの国際的な取り組みにより大きな進歩があった。これら取り組みから得られたエビデンスの活用についても、3ieやJ-PAL等のホームページでスケールアップ事例が示されるなど、具体例を目にする機会は増えてきた。活用ケースを示すことは、国際開発分野でのEBPを推進する上で重要な取り組みである。ただし、その多くは個別案件に対する評価を用いたスケールアップであり、いわばDFEからの活用事例と言える。

他方、その評価対象案件を超えたエビデンスの活用という観点では、援助機関やアカデミア側から具体的な活用の方策は示せていない。昨今、国際的にエビデンスの活用について一層注目が払われるようになってきたのは、その問題意識の表れと言えよう。システムティックレビューの結果からもその有効な手立てが示せていない。研究者を中心に整えられた研究や評価の結果の多くが開発事業の改善に向けて十分に使われていないとすると、これまでのエビデンス産出の大きなうねりが減退する可能性もあろう。

現状は、エビデンスに基づいた事業が多く実践されているという段階には至っておらず、まだその使い方が分からないと嘆いている段階にあると認識される。勿論、政策決定や事業実施判断の場面では、介入効果に関する信頼性の高いエビデンスのみならず、財務面、実施体制、政治的な判断など様々な要素を踏まえて判断していくものである。よって、インパクト評価やシステムティックレビューから唯一の処方箋が示せるものではない。しかし、そのことはエビデンスを軽視して政策判断をしても良いことではないのは自明であろう。国際開発分野にてエビデンスに基づく実践を定着させていくためにも、援助機関、研究機関などは早々にエビデン

ス活用の良例を得て、広く共有していく努力が求められる。今から10年後にはエビデンスを活用した意思決定が当然のこととして、援助事業の実践において取り入れられていることを強く願うばかりである。

## 6. JICAによるインパクト評価の取り組みと方向性

以上のように、国際開発分野のエビデンスの産出、発信、活用といった点について、国際的な取り組みを見てきた。翻って、日本の援助実施機関であるJICAの取り組み状況はどうであろうか。

ODA分野でインパクト評価の試行が始まったのは2006年であり、当時の国際協力銀行海外経済協力部門によるものである。より本格的な取り組みは、統合して現在の国際協力機構となった2008年からである。評価部、JICA研究所、保健や教育分野を担う人間開発部が、主に進めてきた。JICAの事業評価のウェブサイトでは、17件の完了したインパクト評価結果を掲載<sup>22</sup>している。評価の実施方法は様々で、評価部やJICA研究所の内部関係者による直接実施、プロジェクト本体の契約に含めた実施、外部研究機関などへの評価業務の委託など様々である。また、そのアウトプットも用途によって評価報告書、研究論文やワーキングペーパー、プロジェクト業務文書への反映などの形態をとる。

インパクト評価導入の初期段階は、実践を通じてインパクト評価の実施経験を培っていった段階と言える。この時期は案件開始当初からインパクト評価を計画したものはなかったため、評価部は事後的に評価対象案件を発掘していた。2008年頃に開始したインパクト評価では、データの入手可否という点も選定判断の一つであったが、当時インフラ事業のエビデンスを得るという志向もあり、インドネシアやフィリピンなどの灌漑案件などを対象案件として選定した。その後、インパクト評価の実施と並行し、JICA内部関係者に向けた研修実施などを通じ、組織内でインパクト評価の意義、信頼性の高いエビ

デンスに基づく事業実施への意識の醸成を図った。その結果、徐々に、評価部や研究所以外の部署によってインパクト評価が実施されるようになった。事業部門が示す協力の実施方針としても、例えば、2012年には、人間開発部が「保健協力分野におけるインパクト評価の導入に係る指針」を作成した。また、2015年に示された教育分野のポジションペーパーでは、エビデンスに基づく政策提言と実施の強化が実施アプローチの一つとして掲げられ、エビデンスと調査が不足している領域への対応、広域的な事業展開を図る事業に対するタイムリーなインパクト評価の実施、世界の教育協力におけるエビデンスの蓄積と発信への貢献が謳われた。

以前は事業効果をより精緻に示すこと自体を目的に実施されることが多かったが、事業の中でエビデンスを活用することを目的に行われるケースが徐々に増えてきている。また、キャパシティデベロップメント、エンパワーメント、信頼度の醸成など、これまで測定が困難と考えられていた項目の検証が試みられる等、インパクト評価の適用の場が広がってきたと言える。評価結果とエビデンス活用の連動については、今後より多くの事例を生み出していくことが期待される。

他の援助機関と同様に、JICAにおいても、エビデンス活用の重要性を認識しているものと、意識できていないものとに分かれる。そのため、組織内での意識啓発、内外関係者への能力強化を引き続き進める必要がある。この動きを押し進めるためにも、エビデンスの活用策を事業の関係部門へ具体的に示していくことが重要となる。JICA内外でインパクト評価を担う人材が限られている現状から、開発分野の理論構築を目指すことを目的とするKFEの取り組みや、膨大な人的リソースを必要とするシステムティックレビューを独自で行うことの優先度は下がる。当面、DFEを目的として評価を実施し、当該案件の中でその結果を活用すること、当該事業の広域展開に繋げること、また、類似の案件に使っていくことを評価計画時から明確に定め、産出から活用に繋げていくことが重要である。これに加えて、すでに世に出されているエビデンスを、

協力案件の計画や実施のサイクルにおいてどのように取り込めばよいかを分かりやすく提示していくことが併せて必要となる。JICAは援助機関として開発事業の実施を所掌することから、エビデンスと事業での意思決定とを繋げることができることが強みとしてある。その強みを生かして、JICAが評価結果を事業に繋げる事例を示すことで、より意義のあるインパクト評価の実施に繋がり、ひいては、国際開発分野におけるエビデンスに基づく事業実施の普及に寄与できるものとする。

## 注記

- 1 本論において示す考えは筆者の個人的見解であり、所属する組織を代表するものではない。
- 2 エビデンスには様々なグレードがあり、専門家の経験知と言えるものから、精緻な統計技法を用いることで得られた分析結果に到るまで様々である。本論では、後者に該当するようなものを「信頼性の高いエビデンス」として示している。
- 3 エビデンスによった開発政策やプログラムの実施促進を行うため資金を提供する国際NGOで、英国開発協力省、ドイツ財団などが大口の資金提供者である。
- 4 2012年までには3ieが見つけられていなかった評価が、新たな大規模サーチを踏まえてデータベースに含められたことも、急増した理由の一つである。
- 5 2002年に、Dean Karlan（現Yale大学教授）により設立された。J-PALの姉妹組織ともいえ、開発分野でのRCT実践に尽力してきた。IPAには400名以上の研究者が関与しているが、J-PALとIPA両方に関与している研究者は多く、共同で研究するケースも多々ある。
- 6 2003年にマサチューセッツ工科大学で教鞭をとっていたAbhijit Banerjee、Ester Duflo、Sendhil Mullainathanの3名によってPoverty Action Labとして立ち上げられた。RCTの実施を通じた開発効果向上を目指しており、2016年5月時点で、136名の研究者が関与している。
- 7 対象セクターにおいて、介入内容とアウトカムから成る表の中で既存のエビデンス有無、多寡を示したもの。
- 8 2015年5月時点。
- 9 J-PALが生み出したエビデンスによるスケールアップについて、本特集号の別稿（佐々木）を参照されたい。
- 10 <http://www.poverty-action.org/search-studies>からカウントした。
- 11 2016年5月時点。
- 12 一般的に、途上国政府等の政策決定者は多忙のため、数十ページにわたる評価報告書や学術論文を読む時間がなく、エビデンスの活用には繋がらないという言説がある。3ieにより、政策ブリーフによつて的確に内容を伝えられるか、変化を起こせるかの検証がなされた（E. Masset et al. 2013）。本調査では、75,000人を対象にメールで調査が依頼され、807名がベースラインに回答するも、半数近くがフォローアップ調査ではサンプルから落ちたことをはじめ、調査として高い質を有しているとは言い難いものの、難しさを前提の上で政策と研究とのリンク付けについて定量的な分析を試みた例である。結果として、扱った内容に対して意見が無い者に考えを得させたケースは見られたが、元々の考えを変えさせるまでのエビデンスは得られなかった。制約が大きい調査であり、結果は一般化できないものの、このような定量的な分析を積み上げることで、どのようにエビデンスを広め、政策決定に繋げるかの普及方法についても検討が必要とされている。
- 13 IDinsightというアメリカの開発コンサルタント社所属
- 14 ある研究から得られた結果を、違った母集団、状況、条件へ一般化し得る程度を指す。
- 15 大学での評価実施者育成コース開設の提案が示されるなど、アメリカにおいても実務に寄り添ったインパクト評価の担い手はまだ十分ではないことが伺われる。
- 16 American Economic Review, Econometrica, The Quarterly Journal of Economics, The Journal of Political Economy, The Review of Economic Studies, the Economic Journal, The Journal of Public Economics, The American Economic Journal
- 17 介入を受けた被調査者が期待感に応えようとし、結果、行動が変容してしまうこと。
- 18 介入を受けなかった被調査者が、調査や実験で比較対象となることで意識が変化し、結果、行動が変容

してしまふこと。

19 市場における需要と供給が価格の調整機能によって同時に均衡した状態を一般均衡というが、インパクト評価において分析した介入をより広く適用した時に均衡が移り、異なる効果を示す可能性がある。

20 2016年5月時点。

21 「プログラムセオリー」を、「ロジックモデル」とも言い換えている。

22 2016年5月時点。

## 参考文献

- 青柳恵太郎 (2006) 『開発援助の新しい潮流:文献紹介』 No.64, FASID ([https://www.fasid.or.jp/\\_files/library/BriefingReview/BriefingReviewNo64.pdf](https://www.fasid.or.jp/_files/library/BriefingReview/BriefingReviewNo64.pdf))
- 国際協力機構人間開発部 (2012) 『保健協力分野におけるインパクト評価の導入に係る指針』、国際協力機構
- 国際協力機構人間開発部 (2015) 『JICA教育協力ポジションペーパー』、国際協力機構
- Banerjee, A., Chassang, S. and Snowberg, E. (2016). *Decision Theoretic Approaches to Experiment Design and External Validity, draft of Handbook of Field Experiments.* (<https://www.povertyactionlab.org/handbook-field-experiments>)
- Cameron, D., Mishra, A. and Brown, A. (2015). *The growth of impact evaluation for international development: how much have we learned? Journal of Development Effectiveness*, 8(1), 1-21.
- Center for Global Development (2006). *When Will We Ever Learn? Improving Lives Through Impact Evaluation.*
- Dhaliwal, I. and Tulloch, C. (2012). *From research to policy: using evidence from impact evaluations to inform development policy, Journal of Development Effectiveness*, 4(4), 515-536.
- The Department for International Development (2016). *Annual Evaluation Report 2015.*
- Estache, A. (2010). *A Survey of Impact Evaluation of Infrastructure Projects, Programs and Policies.*
- Gray, C. (2014). *Finding Out What Works: Tracking Results in the Inter-American Development Bank, Journal of Development Effectiveness* 6: 480-489.
- Independent Evaluation Group (2012). *World Bank Group Impact Evaluations: Relevance and Effectiveness.* Washington: World Bank.
- International Initiative for Impact Evaluation (3ie) (2016). *Evidence, influence, impact: annual report 2015.*
- Kleemann, L. and Böhme, M. (2013). *Happily Uninformed? The Demand for Research among Development Practitioners in Germany, PEGNet Survey on the Demand for Research* (<http://www.pegnet.ifw-kiel.de/research/Survey>)
- Langer, L., Stewart, R. Erasmus, Y. and Wet, T, D. (2015). *Walking the last mile on the long road to evidence-informed development: building capacity to use research evidence, Journal of Development Effectiveness*, 7(4), 462-470.
- Levine, R. and Savedoff, W. (2015). *Aid at the frontier: building knowledge collectively, Journal of Development Effectiveness*, 7(3), 275-289.
- Miranda, J., Sabet, S., and Brown, A. (2016). *Is impact evaluation still on the rise?* 3ie Blog (<http://blogs.3ieimpact.org/is-impact-evaluation-still-on-the-rise/>)
- Masset, E., Gaarder, M., Beynon, P. and Chapoy, C. (2013). *What is the impact of a policy brief? Results of an experiment in research dissemination, Journal of Development Effectiveness*, 5(1), 50-63.
- Ravallion, M. (2011). *Knowledgeable Bankers? The Demand for Research in World Bank Operations, Policy Research Working Paper 5892.*
- Peters, J., Langbein, J. and Roberts, G. (2015). *Policy Evaluation, Randomized Controlled Trials, and External Validity – A Systematic Review, Ruhr Economic Papers #589.*
- Shah, NB, Wang, P, Fraker, A and Gastfriend, D. (2015). *Evaluations with impact: decision-focused impact evaluation as a practical policymaking tool. 3ie Working Paper 25.*
- Shi, Y., Zhang, L. and Rozelle, S. (2015). *When will we ever learn ...to change policy: current state of impact evaluation, Journal of Development Effectiveness*, 5(1), 402-422.
- Snilstveit, B. (2012). *Systematic reviews: from 'bare*

- bones' reviews to policy relevance, Journal of Development Effectiveness*, 4(3), 388-408.
- USAID U.S. Agency for International Development (2016). *Strengthening evidence-based Development. Five years of better evaluation evaluation at USAID, (2011–2016)*.
- Waddington, H., White, H., Snilstveit, B., Hombrados, J, G., Vojtkova, M., Davies, P., Bhavsar, A., Evers, J., Koehlmoos, T, P., Petticrew, M., Valentine, J, C. and Tugwell, P. (2012). *How to do a good systematic review of effects in international development: a tool kit, Journal of Development Effectiveness*, 4(3), 359-387.
- White, H. and Waddington, H. (2012). *Why do we care about evidence synthesis? An introduction to the special issue on systematic reviews, Journal of Development Effectiveness*, 4(3), 351-358.
- (2016.9.29 受理)

## **Current situation and challenges of utilization of evidences in the international development field**

Hiroaki Asaoka

Japan International Cooperation Agency  
Asaoka.Hiroaki@jica.go.jp

### **Abstract**

In the international development field, it has often been pointed out the significant lack of scientifically sophisticated and rigorous evidences related to development impact. With the recognition, the large number of rigorous impact evaluations and researches has been carried out for last decade. On the other hand, there is room for improvement in terms of the use of those results.

There are several reasons of the less utilization. The evaluations and surveys by researchers might pay less attention to influence on the decision-making. The decision-makers have little awareness on the utilization as well. The evaluation results could not overcome the external validity issues and the systematic reviews have the room of improvement for directly using on decision making.

In order to promote the Evidence-based Practice in the international development field, aid agencies including Japan International Cooperation Agency (JICA) need to apply the evidences on the implementation of projects or programs and show the good examples toward further evidence-based practice.

### **Keywords**

Impact Evaluation, International Development, Japan International Cooperation Agency, Evidence, Evaluation Use

## 【研究論文】

## SROI（社会的収益投資）に関する批判的考察

津富 宏

静岡県立大学

tsutomi@u-shizuoka-ken.ac.jp

## 要 約

本稿では、エビデンスを「つくる・つたえる・つかう」運動であるEBP（Evidence-based practice）の観点から、社会的投資のための評価ツールのひとつであるSROI（Social Return On Investment: 社会的収益投資）について批判的検討を行った。まず、SROIの普及状況について概説した後、SROIがCBA（Cost-Benefit Analysis: 費用便益分析）の一種であることを確認し、Nicholls et al.（2009）に従って、SROIの原則、SROIの手順について概観した。これを踏まえて、Arvidson et al.（2010, 2013）によるSROIに対する、的を得た8つの批判を紹介した。その後、SROIに関する具体例の検討を行い、SROI比率算定における恣意性やSROI比率がインフレートされる可能性を見出した。最後に、福祉国家論における社会的投資の役割についての考察を踏まえ、SROIは、投資対象としての事業や組織を評価するためではなく、EBPが長年にわたり行ってきたように、社会的共通資本としてのセクターの漸進的改善を支援するために用いられるべきであると主張した。

## キーワード

SROI（社会的収益投資）、社会的投資、福祉国家、社会的共通資本、社会権

## 1 はじめに：問題意識

本研究は、昨今注目を浴びている、社会的投資／社会的責任投資のためのツールのひとつであるSROI（social return on investment: 社会的収益投資、社会的投資収益分析、社会的投資収益率など）について批判的に検討し、その建設的な活用を提案することを目的とする。

この問題意識は、評価研究者としてのものであり、かつ、NPO経営者としてのものである。評価研究者として、筆者は、代表的なEBPプロジェクトである、キャンベル共同計画<sup>1</sup>へ参画してきた。キャンベル共同計画は、医学分野にお

けるEBPプロジェクトであるコクラン共同計画にならって1999年に発足した、社会科学分野（同計画が現在対象としているのは、教育、社会福祉、刑事司法、国際協力の四分野）で、エビデンスをつくり、つたえ、つかうための活動を行う国際プロジェクトで、世界各国の研究者・実務家が協力して運営している。EBPは、エビデンスを科学的に産出し、それを共有して、実務に反映する活動であるが、SROIは、EBPが積み上げてきた、エビデンスを「つくる・つたえる・つかう」プロセスを軽視している。すなわち、EBPは、個別事業の評価ではなく、その分野全体で活用しうる科学的エビデンスの蓄積と提供を重視し

てきたが、SROIをはじめとする社会的投資のための評価は、EBPが発達させてきた、慎重なエビデンス活用のあり方を十分に尊重していない。

また、NPO経営者として、筆者は、青少年の就労支援分野のNPO法人（特定非営利活動法人青少年就労支援ネットワーク静岡<sup>2</sup>）を10年余り運営してきた。この団体は、静岡県内で500名ほどのボランティアを組織化し、行政などの外部資金の提供に頼らずに、働きたいけれども働けない人を支える相互扶助の市民社会をつくることを目指して活動している。すなわち、同団体は就労支援に携わりつつ、地域社会を市民社会として編みなおすことを目的としている。しかし、SROIは、非営利活動の直接的な成果（たとえば、就労及びそれに伴う便益）に着目するあまり、市民セクターを形成していくというより根幹的な成果（たとえば、地域コミュニティに相互扶助が根付くこと）を見逃しがちである。その結果、地域社会としてみればより重要な意義をもたらしている事業体であっても、費用便益比率に劣るという表面的な理由で退場を迫られるおそれがある。

非営利活動に対する評価のあり方は、その分野・事業体に対して大きな影響をもたらす。一例を挙げよう。青少年の余暇活動を担うユースワークは、北欧諸国（スウェーデン、フィンランド、ノルウェーなど）では基本的には公的セクターが運営しているが、英国では民間団体に事業委託されている。北欧諸国では、仕事を「振り返る」ための専門性を問われる評価は厳しく行われているが、スタッフの身分は守られており、現場には自由でまったりした空気が流れている。一方、英国では民間団体が競争で受託しているため、スタッフは仕事を「取る／取り続ける」ための評価に追われ疲弊している。社会的投資の概念はいずれの国家にもあるが、前者では基本的にセクター全体の効率を上げるために用いられ、後者では事業体単位の評価に用いられている。このように文脈が異なれば、社会的インパクト評価自体がもたらす「社会的インパクト」は正反対となる。

SROIは、単に評価をするだけでなく、評価と投資（investment）を結び付ける概念であり、

現実における資源配分を変更しようという意図を持って発展してきた。しかし、SROIは、①EBPが積み上げてきたエビデンスを「つくる・つたえる・つかう」プロセスを軽視している、②非営利組織が担う市民セクターを形成していくという成果を見逃してしまうという二つの欠点がある。とすると、SROIは、その活用の結果、良質の活動（体）を退場させてしまうおそれがある。すなわち、SROIには、負の「社会的インパクト」をもたらす可能性がある。

なお、SROIは、企業内においては、他の一般収益投資事業と比較して、社会的投資の価値を示すために用いられる側面もある。つまり、SROIは、社会的投資に資源誘導をする可能性を持ちつつ、その過程において、特定分野の社会活動の質を低下させてしまう可能性がある。

## 2 SROI

### 2.1 SROIの普及

SROIとは、1990年代にアメリカで開発され、2000年代になってからイギリスに導入された、非営利セクターの活動を投資の観点から評価するための代表的ツールである。英国では、2009年、Cabinet Office（内閣府）のOffice for the Third Sector（サードセクター局）が、公的なマニュアルとして、A guide to social return on investment（『投資の社会的収益についての手引き』（Nicholls et al. 2009））を発刊したことがひとつの節目となった（さらに、2012年には、その改訂版（Nicholls et al. 2012）が、SROI Network から発刊されている）。現在、イギリスでは、非営利活動に対する政府助成プログラム評価の基本手法として位置付けられると評価され、非営利活動のさまざまな分野（児童養護、就労支援、再犯予防、貧困対策など）で活用されている。

SROIは、Social Value UK<sup>3</sup>を通じて、国際的な普及が促進され、世界各国で急速に普及が図られている。同団体のHPを見ると、2015年5月現在、SROIのガイドブックは、英語に加え、イタリア語、チェコ語、韓国語、フランス語、スペイン語、中国語、そして、日本語に翻訳されている。

わが国では、SROIを紹介する章を含む書籍（塚本・関 2012）の発刊、特定非営利活動法人SROIネットワークジャパン<sup>4</sup>の活動などを通じて普及が図られ、同団体は実際の適用事例の報告も行っている。

## 2.2 CBAとしてのSROI

SROIとは、CBA（cost-benefit analysis: 費用便益分析）の一種であり、CBAを、社会的投資（social investment）の観点から、非営利活動を評価するツールとして発展させたものである。その活動が生み出す社会的価値（social value）を貨幣換算し、それをもとに、費用便益比率であるSROI比率を求めることが目的となっている。SROIでは非貨幣的な指標、定性的な評価も併用されるが、投資の判断基準として注目されるのは、やはり、貨幣換算された指標であるSROI比率である。なお、CBAと比べると、評価過程におけるステークホルダーの関与が強調される。

ところで、筆者の専門である犯罪学においては、CBAは研究としては取り込まれるものの、政策判断への応用については従来慎重になされてきた。CBAは特定かつ多くの前提条件を基にした試算であるので、政策判断における参考情報のひとつとして参照するがそれによって直ちに資源の再配分（投資など）を行わないということである。

たとえば、再犯抑止に関して、施設内処遇（刑務所）は社会内処遇（保護観察）に比べて、コストパフォーマンスが劣ることは実証的に示されている。だからといって、施設内処遇を廃止するわけではない。そのひとつの理由は、施設内処遇を廃止すれば、負の社会的インパクトが生じるからである。一次的には、雇用の乏しい地域にあえて立地し地域経済を支える役割を果たしている刑務所の雇用が減るからであり、二次的には、犯罪が減れば、警察、司法など刑事司法関係の雇用、民間警備業の雇用が減り、損害・生命保険の需要が下がり、メディアなど報道機関への需要も減少するであろうからである。すなわち、資源再配分には負の社会的インパクトが伴うことが想定されるがゆえに、事業単体での費用対効果さえよければよいというCBAの

安易な利用は避けられてきた。

## 2.3 SROIの原則

SROIは、以下の7つの原則を掲げている（Nicholls et al. 2009）。

### 1 ステークホルダーを巻き込む

ステークホルダーを巻き込むことで、何を測定し、それをどのように測定し価値づけるかを共有する

### 2 何が変化するかを理解する

どのように変化が作り出されるかを明示し、意図的・非意図的な変化およびプラス・マイナスの変化があることを認識しつつ、収集したエビデンスを用いて評価する

### 3 意義のあるものを価値づける

アウトカムの価値が認識できるように代替的な金銭指標を用いる。多くのアウトカムは市場で取引されていないため、その価値は認識されていない

### 4 重要なものだけを含む

ステークホルダーが、インパクトについて十分に合理的な結論を導けるよう、真実かつ公平な評価となるように、どの情報とエビデンスを計算に含めるのかを決定する

### 5 過大な主張をしない

組織が創り出したといえる価値のみを主張する

### 6 透明性を保つ

分析が正確かつ廉直であるとみなしうる根拠を示し、それをステークホルダーに報告し議論する

### 7 結果を点検する

適切な独立した確認を行う

これらの原則は、さまざまな疑問点を引き起こす。たとえば、「1」は、ステークホルダーの関与を、CBAにはない、SROI固有の特徴として強調するが、権力差のある複数のステークホルダーの関与を認めるとき介入の対象である当事者の参加は十全に保障されるのか、「3」は、金銭換算することをもって非営利組織の活動を測ることができるという前提を置いているが、なぜそのような前提を置くのか、また置けると考え

るのか、「5」は、組織の存在／活動が正の価値を生み出しているとは限らないにもかかわらず、なぜ「価値」を生み出しているという表現になっているのかといった点である。これらの疑問点は、以下の検討とも関連する。

## 2.4 SROIの手順

SROIは、以下の6つの段階からなる手順を進めることとされている (Nicholls et al. 2009)。

### 1 評価対象を決め、ステークホルダーを確定する

SROI分析がカバーすべき明確な領域を確定し、過程と進め方に誰を巻き込むかを決める

### 2 アウトカムをマッピングする

インパクトマップ (ロジックモデル) (インプット→アウトプット→アウトカム→インパクト)をつくる

### 3 アウトカムに関するエビデンスを手に入れ、価値づける

アウトカム指標を決めデータを集める。アウトカムを価値づける

### 4 インパクトを確定する

死荷重 (deadweight: インプットなしでも生じうるアウトカム)、寄与率 (attribution: アウトカムに対してインプットが寄与する割合)、置換効果 (displacement: インプットのもたらすアウトカムが、他のアウトカムを置き換えてしまう割合)、ドロップ・オフ (drop-off: アウトカムの持続性)を考慮する

### 5 SROI比率を計算する

割引率を計算したうえで総便益を総費用で割り、SROI比率を計算する。感度分析を行う

### 6 結果を報告し活用し内製化する

ステークホルダーに報告する、結果を活用する、結果を点検する

評価研究の伝統におけるインパクト評価では「4」のためのデザインが重視されるが、その点を比較的軽視した上で、「5」「6」に進むのがSROIの特徴である。

## 3 Arvidson et al. (2010, 2013) によるSROIに対する批判

Arvidson et al. (2010, 2013) はSROIに対し、以下のとおり、的を得た批判を行ってきた。私見を加えつつ、それらを紹介したい。

### 3.1 便益の価値の貨幣換算

収入の増加など貨幣換算されやすい (tangible) 便益が優先され、主観的な満足度など見えにくい (intangible) 便益は、見逃されがちである。公的セクターの支出削減の想定額を求めることによって貨幣換算をすることが多いが、そもそも、公共セクターの支出削減は、受益者本人に対する便益とは異なる。また、公的セクターの支出 (いわば公共投資) が削減されたとしても、それが社会全体にとってよいとは限らない (たとえば、第三セクターによって費用が内部化されるだけだったりする)。また、公的セクターの支出削減額を算定するに当たっては、①固定費用と (それよりはるかに小さい) 変動費用を混同する、② (直ちに回収されるのではなく) 長期的に回収されるしかない額を算定するといった問題もある。

### 3.2 ボランティアの価値づけ

非営利組織においては、ボランティアが多く活用されている。しかし、それを貨幣換算したインプット費用として (のみ) 価値づけることの困難性については、以下の3点が指摘できる。第一に、インプット費用としてのボランティアの価値づけ (貨幣換算) には、最低賃金で換算するアプローチ、雇用の機会費用あるいは余暇の機会費用で換算するアプローチがあるが、ボランティア自身は労働の市場化を好ましくないと考えており、自らの貢献を金銭評価することを不適切であると考えている可能性がある。第二に、ボランティア活動はインプット (コスト) であると同時に、アウトプット (喜び、すなわち、ベネフィット) であるので、SROIの算定に当たっては後者を見落としてはならない。第三に、私見だが、ボランティアが担い手となること自体が相互扶助に基づく社会のインフラの形

成であり、よって、ボランティア活動は、個人にとってだけでなく社会にとっても費用ではなく成果である。

### 3.3 死荷重、寄与率、置換効果、ドロップ・オフの扱い

死荷重と寄与率を算定するには、本来は、反事実（counterfactual）を入手する必要がある。残念なことに、SROIにおいては、反事実の入手に最も近い状況を可能とする、ランダム化比較試験をはじめ、質の高いデザインを利用することは十分に強調されず、安易に反事実に代わる想定がおかれる。また、置換効果の推定も困難で、たとえば、進路支援サービスについて、80%～90%の死荷重を想定した例（NatCen et al. 2011）も、0%を想定した例（Wright et al. 2009）も報告されるなど、恣意的になりがちである。ドロップ・オフの想定も評価者の判断に任されている。

### 3.4 指標設定における恣意性

SROIに、どのインパクトを含め、どの指標を用いるかを判断するにあたっては、ステークホルダー間に力関係があること、指標の入手可能性・時間的制約・評価に投入しうる資源が制約されていること、指標の測定期間があらかじめ決められているわけではないことなどの要素により、恣意性が入り込む。また、これらの判断要素はすべて、政策動向によって影響を受けると同時に、政策決定者の意向に影響を与えることを意識して考慮される。すなわち、SROIを構成する指標は、客観的なものではなく「社会的な構成物」である。また、SROIが社会的に設定されるということは、SROI比率が低いからといって活動が不成功であるわけではないことを意味している（たとえば、ターミナルケアは余命の低い対象に対するケアであり、ターミナルケアのSROIは低くなるが、だからといって価値がないわけではない）。

### 3.5 SROIの活用

非営利組織がSROIを用いるのもっとも一般的な理由は資金調達のためである。競争環境におかれた非営利組織は、自らのイメージを向上させ、

正当性を獲得するため、マーケティングの手段としてSROIを用いる。SROI比率を比較することは好ましくないとされているが、公的資金の削減を背景として導入されるSROIは、結局のところ「勝者を選ぶ」「勝者となる」ために用いられる。その結果、非営利組織は、SROI比率を誇張したり、インパクトをひいき目に表現したりする誘惑に誘われる。

### 3.6 プロセスの軽視

SROIは、成果を要約した指標であるため、どのようにその変化がもたらされたかというプロセスを十分に明らかにせず、その結果、SROI率が高くても、介入を改善したり普及したりするには役立たない。たとえば、保育サービスについてのSROIは、どれだけ社会的インパクトをもたらしたかは明らかにするが、社会的インパクトが、スタッフの質、施設の設備、親に対する支援のいずれによってもたらされるかは明らかにしない。すなわちその介入が有効であるメカニズムの理解に役立たず、結果として、その拡散・普及に役立ちにくい<sup>5</sup>。

### 3.7 目標の（非意図的な）すり替え

数量化されやすい目標に焦点があてられるため、数量化されにくい目標の達成が軽視されがちになる。その結果、組織の方向性が、数量化されやすい目標の達成へと（意図せずに）ずれてしまい、組織本来のミッションとのあいだに矛盾が拡大してしまう。なお、この批判は、量的なインパクト評価全般にあてはまる批判である。

### 3.8 SROIにかかる費用

SROIを行うための研修と時間には相当の費用が掛かる（数千ポンドから数十万ポンドまで）ため、規模の小さな非営利組織では負担できない。その結果、「3.5」で指摘したように、SROIが「勝者となる」ために用いられるような競争環境においては、小さな組織は不利な位置におかれる。

## 4 SROIの具体例の検討

### 4.1 RooP (Routes out of Prison Project: 刑務所からの脱出路プロジェクト) に関するSROI

RooPは、スコットランドで行われた、ライフコーチといわれるメンターによる、刑務所入所中から始まる出所者の支援である。RCTによらず、比較群と比して介入群において、再犯率が減少(4%)した。Jardine and Whyte (2013)は、4%の差をもとに、公的費用の削減想定額についての4つの推計(刑事司法に掛かる費用、被害者に掛かる費用、加害者本人にかかる費用を算定した162,255ポンドから、刑務所収容費用のみの33,244ポンドにわたる)と、4つの寄与率の推定(100%から50%まで)を組み合わせてSROI比率を計算したところ、SROI比率は、(介入期間の追加費用を考慮すると)4.6から0.4まで、(平均費用に基づく)6.7から0.7まで変動した。これは2変数のみを変動させた感度分析であるが、SROI比率に影響する変数はもちろんこれらだけに限られず、それらも考慮すれば、上限と下限がさらに離れることになる。

また、それぞれの推計値が推定誤差を持つので、これを考慮することも必要である。すなわち、SROIは、さまざまな要素を組合せて感度分析を行い、いくつかのSROI比率を求めるだけでなく、さらに、感度分析において変動させたすべての推定値の誤差を考慮して、それぞれのSROI比率の推定値の上限と下限(信頼区間)を示すのが適当であるということである。

### 4.2 「若者UPプロジェクト」に関するSROI

「若者UPプロジェクト」は、日本マイクロソフト社が地域若者サポートステーションを運営する就労支援団体等に提供したITスキル講習である。このプロジェクトに関する、SROIが公表されている(株式会社公共経営・社会戦略研究所2014)。このSROIについて、以下の点を指摘したい。

- ・(RCTを用いず)コントロール・グループを非受講者群としているにも関わらず、死荷重を考慮していない。また、クリーム・スキミングが生じていると思われる。

- ・就職決定者の多くが1年間勤務できるとは想定できないにもかかわらず、賃金を年額換算し、ドロップ・オフも考慮していない。
- ・本講習によって、市場における就労可能人口が増えた想定できないのに、労働市場における置換効果を0%と仮定している。
- ・このほか、受講者のみを対象にアンケートを取り、受講者のデータのみを用いて、賃金増分を計算している。

上記で指摘した特徴は、いずれもSROI比率をインフレートする方向で働いていると思われるが、実際このSROIのSROI比率は13.18である。Arvidson et al. (2010, 2013)の指摘とも符合するが、これは、事業者、投資者(そして、いずれかから付託されて評価を行っている評価者)のすべてにとって、SROI比率がより高いことが好ましいことの反映であるように思われる。

そもそも、事業者間には、SROI算定のコストの負担能力やSROI評価を行うニーズに差がある。SROIを行うことに積極的な事業者の事業のSROI比率が押し上げられる結果、SROIを行えない/行わない事業者への資源配分が不当に縮小してしまう可能性がある。

## 5 討論

SROIは、主としてイギリスで発展して来たが、そもそも社会的投資について考える際、問われなければならないのは「社会的投資は何のためにあるのか」という問いである。20世紀末よりヨーロッパにおいて福祉国家再編のキーワードとなっている社会的投資の理念的な側面と実際の政策としての展開パターンについて検討した濱田(2014)によれば、欧州における潮流は大きく分けて二つに類型化できる。濱田の議論を表1にまとめてみた。北欧諸国を代表とする社会民主主義レジームの国々と、イギリスを代表とする自由主義レジームの国々では社会的投資の位置付けが大きく異なっていることがわかる。

前者の一つであるフィンランドは、競争モデルを用いずに、1990年代以降学力ランキングを

表1 欧州の社会的投資の整理

	社会民主主義レジーム 代表は、北欧諸国	自由主義レジーム（第三の道） 代表は、イギリス
経緯	もともと社会民主主義勢力の強い国であり、1970年代の福祉国家の黄金期から女性の就労支援や子育て支援に取り組んでいた	福祉国家の縮減期にあたる1990年代半ばから女性の雇用の活発化に取り組み始めた
社会的投資	21世紀において社会権を重層的に再構築する取り組み	個人を投資対象として位置付け労働市場における商品価値を最大化し国家の生産性を高めるための福祉の効率化の一環
社会的投資のステークホルダー	当事者	投資者

（出所）濱田（2014）の内容の一部を要約し、筆者が作成したもの

急上昇させた。フィンランドの学力が高い理由は、学力の学校間格差が世界で最小である（すなわち、平等度が高い）ことで、学力急上昇の背景には、競争ではなく学校間でアイデアを共有し問題を一緒に解くことを重視した学校間ネットワークを形成する政策（Aquarium Project）があった（Sahlberg 2011）。この政策は、（競争を用いずに）社会全体にとってのインパクトを最大化したものである。本稿で取り扱った、社会的投資／SROIは、この意味で「社会」全体のためになっているのだろうか。

すなわち、私たちが問わなければならないのは、SROIは、はたして「社会」のためになっているのかということである。ドイツ憲法における「社会的」（国家）は「福祉」（国家）を意味（市野川2006）し、社会的包摂というときの「社会」は「連帯」を意味する。SROIは、はたして「連帯」を強化するものになっているのだろうか。

このように考えるとき、社会的投資の本来の対象として想起されるのは、宇沢が提唱した社会的共通資本（social common capital）である。社会的共通資本は、宇沢がコモنزから発展させた概念であるが、宇沢（2015）によれば「ゆたかな経済生活を営み、すぐれた文化を展開し、人間的に魅力ある社会を持続的、安定的に維持することを可能にするような自然環境や社会的装置」（p. 45）であって、「それぞれの分野の職業的専門家集団によって、専門的知見と職業的規律にしたがって管理される」（p. 46）、「すべての人々の人間的尊厳が守られ、魂の自立を保たれ、市民的権利が最大限に享受できるような、リベラリズムの理念」（p. i）を具現化するもので

ある。フィンランドの例で取り上げた教育も社会的共通資本のひとつである。

しかし、現状においては、SROIは、事業体間の競争的な環境を促進しており、結果として、社会的共通資本は棄損されている。藤井ほか（2013）は、社会的企業の評価としてのSROIが建設的に利用されるには、活用環境を、投資者優位の競争環境から、当事者優位の協同環境に変えていくことが必要であると主張している。

## 6 おわりに

SROIをはじめとする社会的活動の評価は、事業や組織を評価するためではなく（投資対象の選別に役立つ評価としてではなく）、社会的共通資本を高めるために、社会的共通資本の要素となるセクター（たとえば、教育）全体を改善するために行われるべきである。具体的には、SROIをはじめとする社会的活動評価の生み出すエビデンスは可能な限りメタ分析を行えるまで蓄積して、セクター全体の効率を高めるための情報として「つくられ、つたえられ、つかわれる」ことが望ましい。これは、EBPにおいて、コクラン共同計画やキャンベル共同計画が行ってきた、エビデンス活用の仕方である。エビデンスは、事業レベルの評価として拙速に用いられるようなものではなく、社会全体の共有財を漸進的に積み上げるために用いられるべきである。

社会的投資における投資対象は、そもそも事業ではなく社会なのではないか。投資効果（リターン）の低い個人の社会権は守られなくてよ

いのだろうか。「社会的」な国家とは、人びとの社会権をできる限り「平等」に保障する国家である。とすると、私たちは、この市民社会において、なんのためにSROIを用いているか、「社会」的投資の評価を行っているかを、自分たち自身に問い直さなければならない。

SROIは、EBPが慎重に積み上げてきた、社会のために、エビデンスを「つくる、つたえる、つかう」ための作法を十分に尊重することなく、次のステージへ向かおうとしているように思われる。私たちはここで立ち止まり、SROIが社会にどんなインパクトをもたらそうとしているかを理解する必要がある。

## 注記

- 1 <http://www.campbellcollaboration.org/>を参照。
- 2 <http://www.ssns.org/> (2016年8月20日アクセス)
- 3 <http://socialvalueuk.org/> (2016年8月20日アクセス)
- 4 <http://www.sroi-japan.org/> (2016年8月20日アクセス)
- 5 Arvidson et al. (2010, 2013) はこのように指摘するものの、SROIはロジックモデルに基づいているので、単純にプロセスを軽視しているわけではない。しかしながら、残念なことに、SROIの結果が利用される際には、SROI比率のみに着目が集まることが多く、また、パス解析などの計量モデルを構築してロジックモデル自体を検証しているSROIはほとんどないので、彼らの指摘を否定することは困難である。

## 引用文献

- 市野川容孝 (2006) 『社会 (思考のフロンティア)』、岩波書店
- 宇沢弘文 (2015) 『宇沢弘文の経済学：社会的共通資本の論理』、日本経済新聞出版社
- 株式会社公共経営・社会戦略研究所 (2014) 『マイクロソフトコミュニティITスキルプログラム「若者UPプロジェクト」(第4年次：2013年度) (ITを活用した若者支援プロジェクト) SROIによる第三者評価報告書』、株式会社公共経営・社会戦略研究所 (公社研)  
<http://koshaken.pmssi.co.jp/upfile/MSYR4.pdf> (2016年5月5日アクセス)

- 塚本一郎・関正雄 (編著) (2012) 『社会貢献によるビジネス・イノベーション』、丸善出版
- 濱田江里子 (2014) 「21世紀における福祉国家のあり方と社会政策の役割：社会的投資アプローチ (social investment strategy) の検討を通じて」、『上智法学論集』、58 (1) : 137-58
- 藤井敦史・原田晃樹・大高研道 (共編) (2013) 『闘う社会的企業』、勁草書房
- Arvidson, M., Lyon, F., Mckay, S., and Moro, D. (2010). *The ambitions and challenges of SROI (Third Sector Research Centre Working Paper 49)*. Third Sector Research Centre.  
<http://www.birmingham.ac.uk/generic/tsrc/documents/tsrc/working-papers/working-paper-49.pdf> (2016年5月5日アクセス)
- Arvidson, M., Lyon, F., Mckay, S., and Moro, D. (2013). Valuing the social? The nature and controversies of measuring social return on investment (SROI). *Voluntary Sector Review*, 4(1), 3-18.
- Jardine, C., and Whyte, B. (2013). Valuing Desistance? A Social Return on Investment Case Study of a Throughcare Project for Short-Term Prisoners. *Social and Environmental Accountability Journal*, 33(1), 20-32.
- NatCen (National Centre for Social Research), Institute for Volunteering Research, University of Southampton, University of Birmingham, and Public Zone. (2011). *Formative evaluation of v: The National Young Volunteers' Service: Final Report*. London, NatCen.  
<http://www.natcen.ac.uk/media/23287/formative-evaluation-final-report.pdf> (2016年5月5日アクセス)
- Nicholls, J., Lawlor, E., Neitzert, E., and Goodspeed, T. (2009). *A guide to Social Return on Investment*. London: Office of the Third Sector, Cabinet Office.  
[http://b.3cdn.net/nefoundation/aff3779953c5b88d53\\_cpm6v3v71.pdf](http://b.3cdn.net/nefoundation/aff3779953c5b88d53_cpm6v3v71.pdf) (2016年8月20日アクセス)
- Nicholls, J., Lawlor, E., Neitzert, E., and Goodspeed, T. (2012). *A guide to Social Return on Investment*. SROI Network.  
[http://socialvalueuk.org/publications/publications/cat\\_view/29-the-guide-to-social-return-on-investment/223-the-guide-in-english-2012-edition](http://socialvalueuk.org/publications/publications/cat_view/29-the-guide-to-social-return-on-investment/223-the-guide-in-english-2012-edition) (2016年5月5日アクセス)
- Sahlberg, P. (2011). *Finnish Lessons: What Can the World*

*Learn from Educational Change in Finland? (The Series on School Reform)*. New York, New York: Teachers College Press.

Wright, S., Nelson, J. D., Cooper, J. M., and Murphy, S. (2009). An evaluation of the transport to employment

(T2E) scheme in Highland Scotland using social return on investment (SROI). *Journal of Transport Geography*, 17, 457-67.

(2016.9.26 受理)

## Critical Reflections on SROI (social return on investment)

Hiroshi Tsutomi

University of Shizuoka

tsutomi@u-shizuoka-ken.ac.jp

### Abstract

This paper critically reflects upon SROI (social return on investment), one of the evaluation tools for social investment from the viewpoint of EBP, an initiative to produce, communicate, and use evidence. Firstly, I describe the history and spread of SROI worldwide and point out that SROI is indeed a type of CBA (cost-benefit analysis), and then, following Nicholls et al. (2009), review the SROI principles and stages. Next, I introduce Arvidson et al. (2010, 2013)'s vivid eight criticisms against SROI. Further, I examine two cases of SROI to find the arbitrariness in calculating the SROI ratio and the possibility of inflating the ratio. Lastly, based on the research on the role of social investment in welfare states, I contend that SROI should not be used to evaluate a project and/or an organization as a target of investment, but should be used to support the gradual investment in a sector as social common capital as EBP has done for years.

### Keywords

SROI (social return on investment), social investment, welfare state, social common capital, social rights



## 【研究論文】

## 貧困アクションラボの最新動向：政策教訓と拡大適用の事例

佐々木 亮

株式会社国際開発センター<sup>1</sup>

sasaki.ryo@idcj.or.jp

## 要 約

2003年に設立されたアメリカの貧困アクションラボ（Poverty Action Lab）は現在までに770件に及ぶRCTを適用したインパクト評価を実施してきた。その実績を踏まえて、教育分野、保健分野、政治経済・ガバナンス分野のインパクト評価の総合的なレビューを行い、抽出された複数の教訓から構成される合計4本の「政策教訓」（Policy Lessons）を産出・公表してきた。さらに合計7例の拡大適用（Scale-Ups）を実現してきた。本報告では、いったいどのような政策教訓が産出され、どのような拡大適用の事例があったのかを解説するとともに、今後の日本の開発援助への示唆を得た。

## キーワード

RCT、インパクト評価、政策教訓、拡大適用、系統的レビュー

## 1. 本実践・調査報告の背景

開発援助（ODA）分野における厳格な評価の必要性の高まりを背景として、2003年に貧困アクションラボ（正式名称：The Abdul Ratif Jameel Poverty Action Lab、J-PALとも呼称）が設立された。設立趣旨は、もっとも厳格なデザインである無作為化比較試験（RCT:Randomized Controlled Trial）を適用したインパクト評価の実施により、確かに効果がある介入を明らかにすることであった。設立したのは、米ハーバード大学とマサチューセッツ工科大学（MIT）の経済学部研究者たちである。その後、RCTを用いたインパクト評価の本数を着実に増加させてきた。

筆者は、貧困アクションラボが活動を開始した直後にその設立の経緯と動向を解説して論じる論文を執筆した（佐々木、2006）。その後、同

ラボのその後の活動をレビューしつつ、創設者でRCT普及の推進を唱えるAbijit Banerjeeの主張と、評価研究の大御所でRCTの急激な普及に関して批判的な立場を取るMichael Scrivenの主張を比較する論文を執筆した（佐々木、2010）。その後、貧困アクションラボは、RCTによるインパクト評価の実績が蓄積されたことを受けて、総合的なレビューを開始した。言い換えれば、単一分野の個別のインパクト評価の結果をレビューして教訓をまとめ上げて、「政策教訓」（Policy Lessons）として産出・公表し始めた。それに加えて、政策教訓（および個別の評価結果の提言・教訓）が具体的に途上国政府によって全国あるいは全州的な拡大適用が行われた事例を特定して公表し始めた。つまり、正木・津谷（2006）が提唱する「エビデンスに関わる3つの立場」である「つくる」（=個別の評価結果の産出）、「つ

たえる」(=系統的レビューの実施と公表)、「つかう」(=政策への反映)の3つの“場”のうち、「つたえる」にあたる活動にも着手し、さらに「つかう」に関する情報も収集し始めたと見ることができる。本論文では、過去2回にわたりその動向を整理して論じてきた貧困アクションラボにおける政策教訓と拡大適用事例の産出・公表の動向に焦点を絞る。

## 2. 本論文の目的

本論文では、貧困アクションラボがとりまとめて公表している「政策教訓」(4本)と「拡大適用事例」(7本)の解説をレビューして、以下の項目に関して報告して論じることを目的とする。

- (1) 貧困アクションラボが国際開発の分野で何が本当に効果があり、何が効果があるとは言えないと主張しているのかを明らかにする。
- (2) 今後の日本の開発援助への示唆を得る。

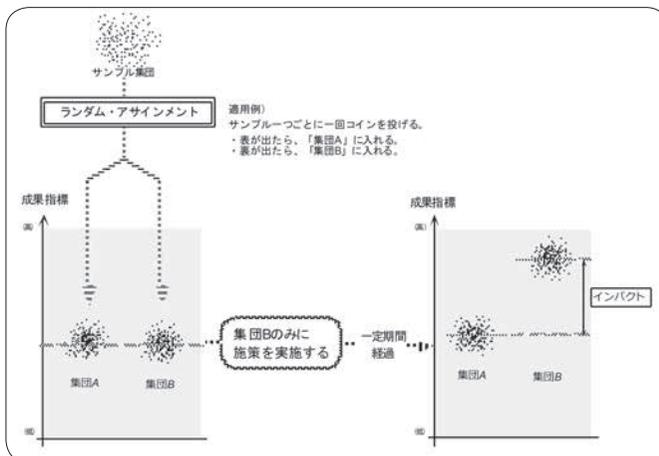
## 3. 状況設定

最初に本論文の全ての議論の前提となるラン

ダム化比較試験(RCT)に関して簡単に解説する。すでに過去2本の筆者の論文でも解説しているが次の通りである。RCTでは、介入行為(開発援助事業)を適用する事前の段階でサンプル集団を無作為割当(ランダム・アサインメント)によって二つのグループに分ける。コインの裏表で分けて考えると分かりやすい。そして片方のグループには介入行為を適用する一方で、もう片方のグループには何もしない。二つのグループはあらゆる特徴や背景の平均値(それは測れるものと測れないものがある)が事前段階で一致していることが無作為割当によって保証されていることから、介入行為の実施後に二つのグループの成果指標値の間に表れた差は、純粹に、途中の唯一の違いである介入行為によって引き起こされたと判断できる(龍・佐々木、2000)。またその差を「インパクト」(Impact)と呼称する。これにより、もっとも厳格に因果関係を明らかにすることができる。下図に示したので確認されたい。

RCTはそもそも実験デザイン(Experimental design)と呼ばれ、1920年代にフィッシャー(Fischer, R.A., 1925, 1935)によって農業分野で提案されたアプローチである。それが1960年代にキャンベル(Campbell, D.T., 1966, 1969)によって社会科学分野に紹介され、その後普及していったものである。したがって、新しい方法とい

図1 RCT (無作為化比較試験) の概念図



(出所) 龍・佐々木 (2000)

うわけでもないし革新的な方法というわけでもない、長い歴史を有するアプローチである。ただし、評価研究の世界で一時は「クラシック」（古典的）なデザインと言われたアプローチが息を吹き返したわけで、その功績は貧困アクションラボにあると言える。筆者は認識している。

貧困アクションラボは、インパクト評価の方法としてRCTのみを用いると設立時に宣言している点が特徴的である。2003年の設立から12年経過したことになるが、同ラボのウェブサイトによると現在までに合計770本（2016年7月19日時点）のインパクト評価を実施済みあるいは実施中としている<sup>2</sup>。なお、貧困アクションラボは最近ではインパクト評価と呼称せずにランダム化評価（Randomized evaluation）という用語を普及させようとしているようである。その内訳は、分野別に見るとファイナンス（227本）、教育（185本）、政治経済・ガバナンス（172本）、保健（159本）、労働市場（100本）、農業（72本）、環境・エネルギー（34本）の順となっている。ファイナンスという見慣れない分野が最多となっているが、内容は、貧困層への貸付サービス、起業支援の貸付サービス、農業金融サービスなどで構成される。地域的には、アフリカ（227本）、南アジア（159本）、北米（147本）、中南米（124本）、東南アジア（43本）といった順番となっている。

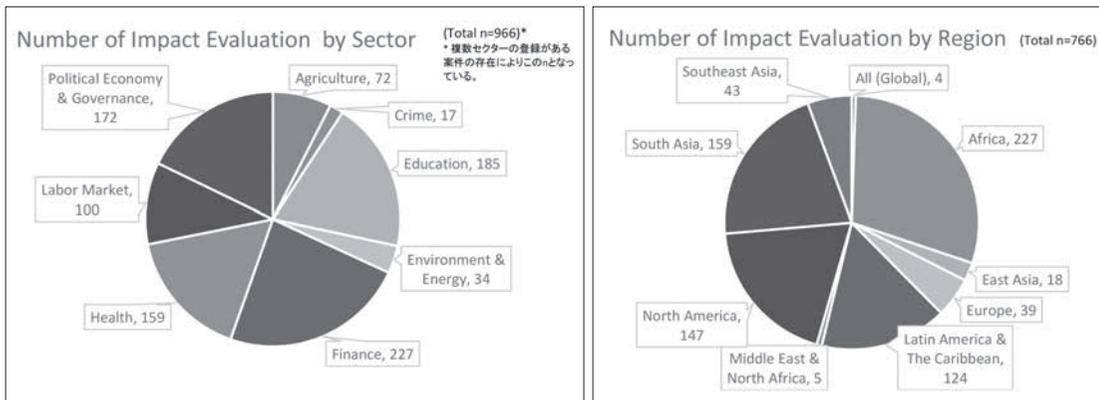
#### 4. 本研究の対象となる介入行為

本論文は、貧困アクションラボが実施した770件の個別のインパクト評価を直接の対象とするのではなく、それらのインパクト評価をレビューして産出・公表された文書である「政策教訓」（Policy Lessons）という文書、および実際に実現した「拡大適用事例」（Scale-Ups）の報告を分析の対象とした。なお、「拡大適用事例」の報告は、単純に政府による拡大適用の事例を報告しているのではなく、その前段階で、政策教訓にあたる分析を行っている場合があることが多いため、分析対象に加えることにしたものである。

ただしここで留意すべきは、レビューを行ったとは言っても、個別介入の効果サイズ(Effect size)とサンプルサイズから平均的な効果サイズを計算するメタ分析（Meta-analysis）を用いた正式な「系統的レビュー」（Systematic Review）のことを指しているわけではないという点である。「政策教訓」は、貧困アクションラボの研究者が、費用－効果分析による比較も併用しつつ、複数のインパクト評価報告書に関して専門家の視点から総合的なレビューを行って得られた知見をまとめた文書のことを指している。

なお、通常の系統的レビューでは事例間でほぼ同一の介入行為を対象とするが、貧困アクションラボのレビューは、比較の括りがかなり広いようである。通常の系統的レビューとはアブ

図2 貧困アクションラボによるRCTを用いたインパクト評価の構成（セクター別、地域別）



(出所) 貧困アクションラボのデータベースから筆者作成

ローチがかなり違うので、本論文では総合的なレビューと記載している。これにより、複数のタイプの介入行為を比較して（そしてそれは費用-効果分析の比較も含む）、複数の政策選択肢の中からより良い政策選択肢の提示に役立てようとしていると理解される。

また、たとえば貧困アクションラボのデータベースではPolicy Issueを選択できるが、「生徒の出席の改善」(Improving Student Participation)を選択すると46件のインパクト評価がヒットし、そのうち37件が終了したことが分かる(2016年7月24日時点)。しかし、後述する政策教訓1「生徒の出席の改善」では13件のみが選択されているが、その理由は明示されておらず、網羅的な検索を行い対象を漏れ無く選び出すという系統的レビューの通常の手続きとは違う手続きが取られていることに留意が必要である。

## 5. データ収集

データは貧困アクションラボのウェブサイトから入手した。「政策教訓」(Policy Lessons)として産出・公表されている4文書、「拡大適用事例」(Scale-Ups)の事例報告として産出・公表されている7文書をそれぞれのサイトから入手した(<http://www.povertyactionlab.org/policy-lessons>、<http://www.povertyactionlab.org/scale-ups>)。

以下が分析対象とした「政策教訓」と「拡大適用事例」の一覧である。なお、原題の英語名には統一性に書ける部分があったが、日本語訳はなるべく統一した。例えば、拡大適用事例の10には国名が入っているが、日本語訳では省略している。

表1 「政策教訓」と「拡大適用事例」の一覧 (n=11)

#	セクター	文書のタイトル
政策教訓 (Policy Lessons)		
1	教育 Education	生徒の出席の改善 Improving Student Participation
2	教育 Education	生徒のテストスコアの増加 Increasing Test Score Performance
3	保健 Health	予防目的の保健製品の価格付け Pricing Preventive Health Products
4	政治経済・ガバナンス Political Econ.& Govern.	コミュニティ参加 Community Participation
拡大適用事例 (Scale-Ups)		
5	教育 Education	レベル分けによる教育 Teaching at the right level
6	教育、保健 Education, Health	学校での回虫駆除薬投与 Deworming in schools improves attendance and benefits communities over the long term
7	保健 Health	無料の防虫蚊帳 Free Insecticidal Bednets
8	政治経済・ガバナンス Political Econ.& Govern.	警察のスキル向上研修 Police Skills Training
9	政治経済・ガバナンス Political Econ.& Govern.	援助効果向上のインセンティブ:コミュニティ向けブロックグラント Incentives improve aid efficacy: The case of the Generasi community block grants
10	保健、政治経済・ガバナンス Health, Political Econ.& Govern.	コミュニティの水供給場所への塩素剤デイスペンサーの設置 Chlorine dispensers at community sources provide safe water in Kenya, Malawi, and Uganda
11	政治経済・ガバナンス Political Econ.& Govern.	社会サービスの受給のためのIDカード Identification cards improve national social assistance in Indonesia

(出所) 貧困アクションラボのウェブサイトの情報から筆者作成。

## 6. 分析方法

収集されたデータにあたる文書を筆者が丹念に読み込み、主要な主張を「政策教訓」(Policy Lessons)の一覧としてとりまとめた。また、その政策教訓の中で「拡大適用事例」(Scale-Ups)につながった事例を特定してさらに詳細に概要を記載するという方法を採用した。その結果を踏まえて、筆者が知り得る日本の開発援助の現状を踏まえた日本への示唆を整理して提示した。

## 7. 分析結果

### 7-1. 貧困アクションラボ (J-PAL) が国際開発の分野で何が本当に効果があり何が効果があるとは言えないと主張しているのか。

表1に掲載された「政策教訓」と「拡大適用事

例」の概要を解説する。これらは、貧困アクションラボが明らかにした確かに効果がある開発援助事業のリストであるとも言える<sup>3</sup>。ゴシック体で示してある政策教訓は、この後で解説する「拡大適用」が実現している政策教訓を示している。

第1番目に、「生徒の出席の改善」のための政策教訓が産出・公表されている。この「生徒の出席の改善」は、教育分野の政策・事業の目標として第1番目に挙げられることが多い。この目標を実現するためのエビデンスに基づく確かな政策教訓を得るために、貧困アクションラボでは、RCTを用いたインパクト評価13本の結果に関して総合的なレビューを行い、さらに費用-効果の分析も行ったとしている(表2)。

第2番目に、「生徒のテストスコアの向上」のための政策教訓が産出・公表されている。指標名の「テストスコア」がそのまま用いられてい

表2 政策教訓1：教育分野：生徒の出席の改善

種類/番号	得られた政策教訓の内容
政策教訓1-1	親は費用負担に敏感であり、少額の補助金であっても子供の出席を増加させる。
政策教訓1-2	学校を修了した子供がより高額の収入を得るといった情報を与えることは子供の出席を増加させる。
政策教訓1-3	回虫と慢性的貧血といった保健上の問題を解決することが、子供の出席を増加させるし、(教育分野の伝統的な介入と比べて)たいへん費用対効果が高い。
政策教訓1-4	教育の質を高めること(例:教員の質を高める等)が子供の出席を増加させるというエビデンス(RCTから得られた科学的根拠)はほぼない。

(出所) <https://www.povertyactionlab.org/policy-lessons/education/improving-student-participation>

表3 政策教訓2：教育分野：生徒のテストスコアの向上

種類/番号	得られた政策教訓の内容
政策教訓2-1	教育へのアクセスが極めて限られている場合には(例:アフガニスタンなど)、とにかく学校に来させることが大きな学力習得につながる。
政策教訓2-2	学校に行って学ぶことに関するインセンティブ(例:出席を条件とする現金給付、奨学金等)の設定は、たいへん費用-効果が高い。
政策教訓2-3	教員の増員や教材(例:フリップチャートやテキスト)の増加が、生徒の学力を改善するというエビデンスはほぼない。
政策教訓2-4	生徒の学習レベルによるクラス分けは、生徒の学力習得のために一貫性をもっとも効果的であるし、たいへん費用-効果が高い。
政策教訓2-5	教員に対するインセンティブの設定は大きな学力習得につながる。ただし、その仕組みが客観的に運用された場合には。
政策教訓2-6	短期契約の教員の増員は、比較的安い費用で生徒の学力を改善する。
政策教訓2-7	コミュニティの能力強化プログラムの一環としてコミュニティに補助金を与えることは、生徒の学力向上を導く。

(出所) <https://www.povertyactionlab.org/policy-lessons/education/increasing-test-score-performance>

るが、これは評価学上はやや問題があり、より適切には「生徒の学力の向上」とすべきである。ともあれ、貧困アクションラボでは、RCTを用いたインパクト評価29本に関して総合的なレビューを行い、政策教訓を産出・公表している（表3）。

第3番目に、「予防目的の保健製品の価格付け」と題した政策教訓が産出・公表されている。保健分野では、使用者の一部自己負担の導入が長年にわたり開発援助業界の常識となっていたが、外部経済の効果の大きさ等からその導入を真っ向から否定する内容である。貧困アクションラボでは、RCTを用いたインパクト評価10本に関して総合的なレビューを行い、政策教訓を産出・公表している（表4）。

第4番目に、「コミュニティ参加」と題した政策教訓が産出・公表されている。コミュニティ参加の導入によって、住民のニーズをより良くみ取って反映させることができるはずだと想定されてきたとともに、住民による事業のモニタリングがしっかりなされるだろうとも想定されてきた。貧困アクションラボでは、RCTを用いたインパクト評価4本に関して総合的なレビューを行い、政策教訓を産出・公表している（表5）。

次に「拡大適用事例」（Scale-Ups）の7本のうち、各分野の政策教訓と関連があると思われる拡大適用の事例を合計4つ解説する。RCTによるインパクト評価は、新しい介入案を試す場合には、本当に効果があると言えるかを確認することを目的とするが、その確認によって、政府の政策に取り入れられて全国展開（あるいは全地方展開）されることを究極の目標とすると理解することができる。一方で、すでに長い期間に渡って広範囲に実施されてきた場合には、RCTによるインパクト評価により本当に効果があると言えるかを確かめて、効果がある場合には継続するし、効果があるとは言えないと出た場合には思い切って中止という政策判断に用いることができる。

第1番目は教育分野の政策教訓である「生徒の出席の改善」に沿った拡大適用の事例である。教育分野の伝統的な介入ではなく、回虫駆除薬投与プログラムという保健分野から提案された介入である。RCTによるインパクト評価を実施して、大幅な改善効果があることを示して複数の途上国での拡大適用が実現している（表6）。

第2番目は教育分野の政策教訓である「生徒の

表4 政策教訓3：保健分野：予防目的の保健製品の価格付け

種類/番号	得られた政策教訓の内容
政策教訓3-1	極めて少額であっても使用者料金を導入することは、アクセスを（つまり利用者数を）シャープに減少させてしまう。
政策教訓3-2	ほとんどのケースでは、使用者料金の導入により、保健製品を必要としていた人たちの一部の集団が受け取れなくなる。
政策教訓3-3	政府が保健製品や保健サービスの無料化を検討する場合には、正の波及効果（spillover effects）が大きい製品やサービスに優先順位を置くべきである。

（出所） <https://www.povertyactionlab.org/policy-lessons/health/pricing-preventive-health-products>

表5 政策教訓4：政治経済・ガバナンス：コミュニティ参加

種類/番号	得られた政策教訓の内容
政策教訓4-1	コミュニティを訓練して能力強化することは、確かに公共サービスを改善する傾向がある。
政策教訓4-2	コミュニティで行われている事業やサービスは何かを正確に知ることが第一番目のステップである。
政策教訓4-3	活動計画を作成して保持したり、その活動計画の実施過程を監督しているコミュニティは、公共サービスのデリバリーをより効率的に改善する傾向がある。
政策教訓4-4	ある一つのケースでは、コミュニティ参加型よりも中央政府によるモニタリングの方がより効果的であった。そのかわり、実施にかかる費用はより高価となるが。

（出所） <https://www.povertyactionlab.org/policy-lessons/governance/community-participation>

テストスコアの向上」に沿った拡大適用の事例である。レベル分けによる教育の事例である。複数の教授法に関してRCTによるインパクト評価を実施して、一貫して効果がある教授法をインド政府が拡大適用した事例である（表7）。

第3番目は保健分野の政策教訓である「予防目的の保健製品の価格付け」に沿った拡大適用の事例である。正の外部効果が大きい場合には無

料配布が利用者の増加と外部効果の発現を同時に実現して、公衆衛生（Public health）を改善することを確認した。それが各国政府、開発援助機関、NGOの政策転換につながった（表8）。

第4番目は政治経済・ガバナンス分野の政策教訓である「コミュニティ参加」に沿った拡大適用の事例である。援助と開発成果（パフォーマンス）をリンクさせて、開発成果（パフォーマンス）

表6 拡大適用事例1：学校での回虫駆除薬投与

拡大適用した受益者	合計9,500万人の児童に拡大適用された（2009年以来現在まで）。ケニア政府、インドのビハール州政府などが採用した。これに加えて、2015年にインド中央政府は、14の州で1億5,500万人の生徒を対象とした適用を決定した。さらに最近、エチオピア政府、ベトナム政府が国レベルの採用を決定した。
介入の内容	学校ベースの回虫駆除プログラムは、学校がすでに存在している利点を利用して、全生徒を対象として一年に一度、回虫駆除プログラム（安全で噛み砕くことができる錠剤を配布する）を教員・職員のリードのもとに実施する低コストの介入である。
インパクト評価から拡大適用への道筋	貧困アクションラボのフラッグシップ的なインパクト評価がこれである。1998年以来実施されているRCTによるインパクト評価によると、回虫感染率の低下を実現し、生徒の出席率を向上させるし、さらに最近のケニアでの追跡調査によると、1) 女子生徒の学校での成績（School performance）を向上させ、2) 男子の学校卒業後の所得を向上させるという結果が出ている。具体的には、女子の初等教育の卒業時国家試験の合格率が41%から9.5%ポイント向上した。また、プログラムを適用された男子は（適用されなかった男子と比較して）、週当たり3.4時間多く働かし、起業活動により多くの時間を割くし、高賃金の製造業の職を得ていることが確認された。なお、一人当たり1年間の必要費用はわずかUS0.5ドル（=50円）であり、伝統的な教育プログラムよりも格段に安い。これらのインパクト評価の結果を受けて、2009年にケニア政府が国レベルのプログラム導入を決定した。2011年にインドのビハール政府が州全体での事業実施を決定した（1,700万人対象）。2015年にはインドの14の州で1億5,500万人の生徒を対象とした導入が決定された。さらに、エチオピア政府とベトナム政府でも国レベルの導入を決定した。

（出所） <https://www.povertyactionlab.org/scale-ups/deworming-schools-improves-attendance-and-benefits-communities-over-long-term>

表7 拡大適用事例2：レベル分けによる教育

拡大適用した受益者	合計4,700万人のインドの児童に拡大適用された（2008年以来）。インド政府により実施されている試験的実施の対象人口の人数がこれにあたる（インドであるから試験的実施の段階でこのように大きな対象人数となっている）。
介入の内容	簡単なテストによって生徒の学力レベル（Competency level）を査定して、クラス分けし、そのレベルにあった教育方法（level-appropriate learning activities）で授業を行う。簡単なテストは、読み、書き、算数、理解力に関して継続的に実施する。
インパクト評価から拡大適用への道筋	2001年以来、貧困アクションラボが現地NGOのPrathamをパートナーとして複数の教授法に関する試験をした。それらは以下のとおりである。 ① “Child’s Friend”：同級生から遅れを取ったと認定された児童を対象に、日中の授業時間のうち2時間を（正規の授業からはずれて）補習授業に充てる。雇用された地元の若年女性が教える。 ② Reading Camps：夏休みの間2-3ヵ月にわたって、地元のコミュニティのボランティアが、リーディングの技能を教える。 ③ Read India：4つの違う介入を実施・比較した。1) 教員を夏季休暇中に鍛える、2) 生徒に読み書きの教材を供与する、3) 前2者のミックス、4) 教員をサポートするコミュニティのボランティアを養成する。 ④ Learning Enhancement Program (LEP)：学年の最初にヒンディ語のテスト（2分間の簡単なテスト）を実施してクラス分けしてレベルにあった教授法を適用する。Haryana州政府の協力を得て実施。それぞれのインパクト評価報告書が作成され、その結果、④クラス分けが一貫して効果的で、テストスコアを0.07-0.28標準偏差分向上させることが示された。これが、「レベル分けによる教育」（Teaching at the right level）としてPrathamによって確立され、その後インド政府によって現在実施されている試験的導入につながった。

（出所） <https://www.povertyactionlab.org/scale-ups/teaching-right-level>

ンス)を向上させるためにコミュニティが自由に用途を決めるブロックグラントの導入は効果があることをRCTを用いたインパクト評価で明らかにした。アメリカの援助機関がその評価結果に呼応して莫大な資金拠出を約束している(表9)。

以上、正木・津谷(2006)が提唱する「エビデンスに関わる3つの立場」のうち、「つたえる」にあたる政策教訓の産出・公表と、「つかう」にあたる拡大適用の段階を解説して論じてきた。一連の解説に基づいて、RCTによるインパクト評価の産出・公表(「つくる」)、政策教訓の産出・公表(「つたえる」)、拡大適用の実施(「つかう」)の3つの“場”の関連を整理して図示した(図3参照)。これによると、「つたえる」と「つかう」の関連が見られるものもあれば見られないものがあったことは明記されねばならない。関連が

見られるものとして、①「つたえる」(政策教訓の産出・公表)から「つかう」(拡大適用実施)への直接的な影響があったと考えられるもの(例:学校での介入駆除薬投与)、②「つたえる」(政策教訓の産出・公表)の段階ですでに拡大適用の実施が始まっていたもの(例:レベル分けによる教育)、③「つたえる」(政策教訓の産出・公表)の段階は特に観察されず、RCTによる個別のインパクト評価結果がそのまま拡大適用の実施につながったと思われるもの(例:警察のスキル向上研修)などがあつた。言うまでもなく、政策教訓の産出・公表が自動的に拡大適用の実施を導くわけではなく、いろいろな要因が影響することが表れていると言える。

なお、「つかう」(拡大適用実施)に関しては、以前から外部妥当性(External Validity)の議論が

表8 拡大適用事例3:無料の防虫蚊帳の配布

拡大適用した受益者	人口数に関する情報の記載なし。ただし、英国国際開発庁(DFID)、セーブ・ザ・チルドレン、国連ミレニアムプロジェクトなどが防虫蚊帳の無料配布を支持している。これに関して、ブルンジ、ネパール、マラウイ、ザンビア、シェラレオネ、ガーナ、リベリアの中央政府が無料配布への動きを進めている。
介入の内容	無料の防虫蚊帳(insecticide-treated nets (ITNs or bednets))を配布することで、防虫蚊帳を普及させ、マラリアの被害を減少させる。
インパクト評価から拡大適用への道筋	ケニアにおける貧困アクションラボのRCTによるインパクト評価(2006年開始)により、1)極めて少額であっても使用者料金を導入することは、アクセスを(つまり利用者数を)シャープに減少させてしまうこと、2)マラリアの村落内感染を防ぐという防虫蚊帳の外部効果が大きいこと、が確認された。この評価結果に基づいて、上記の援助機関が無料配布への支持を表明し、上記の被援助国の中央政府が無料配布に向けた着実な動きを進めている。

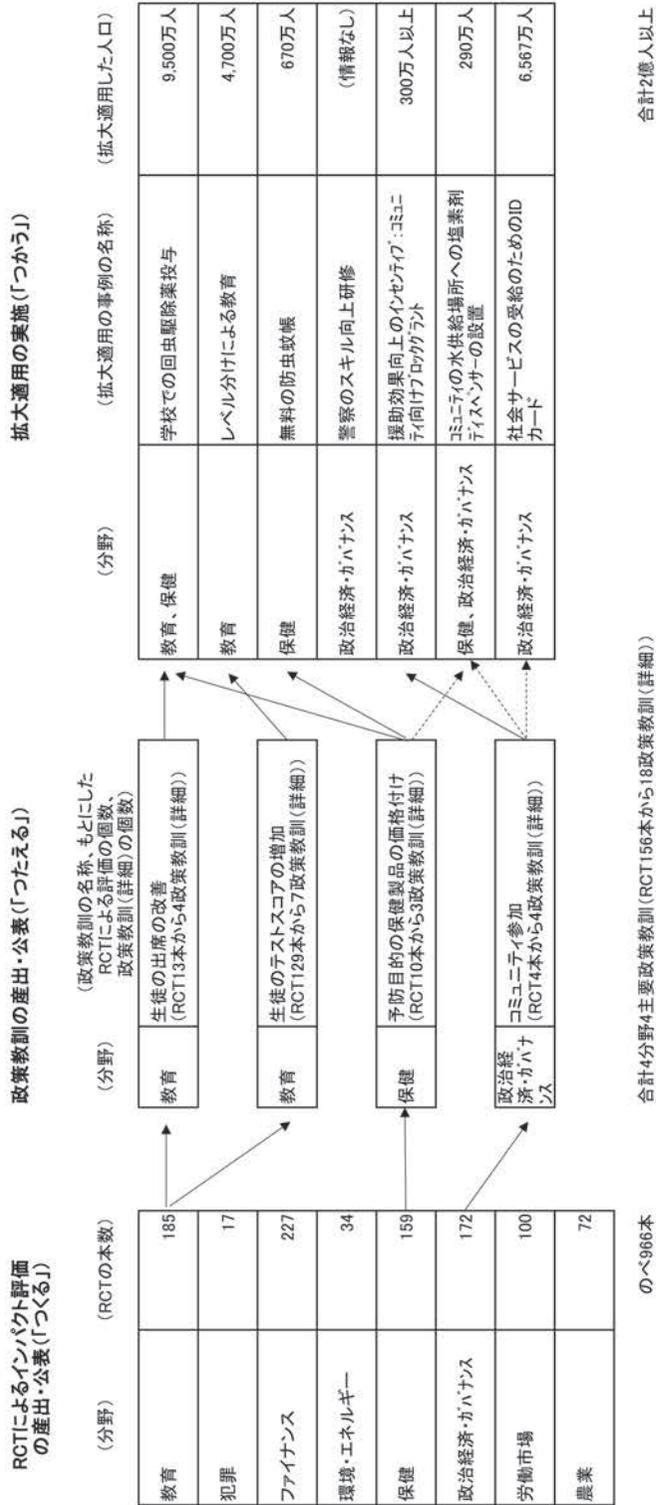
(出所) <https://www.povertyactionlab.org/scale-ups/free-insecticidal-bednets>

表9 拡大適用事例4:援助効果向上のインセンティブ:コミュニティ向けブロックグラント

拡大適用した受益者	合計670万人のインドネシアの女性と子供が受益している。インドネシア政府がこのプログラムを実施しているほか、アメリカのMillennium Challenge Corporation (MCC)が、プログラムの次期フェーズで拡大適用のためにUS\$6億(=600億円)を支出することを約束している。
介入の内容	近年、政府・国際機関・NGOの間で、開発援助と開発成果(パフォーマンス)を結びつけることが重視されている。この結びつけが開発援助の効果を増大させるという意見がある一方で、適切に予算が配分されないリスクも指摘されている。2007年に、インドネシア政府は、保健状況と教育状況を改善するための新しいアプローチとして、コミュニティ向けブロックグラントの試験的实施を開始した。このアプローチ(“Generasi”と呼称)のもとでは、保健・栄養・教育の12の指標を改善するために村落自身が選んだ活動に自由に予算配分できるブロックグラントが設定された。なお、指標は、長期的な成果指標(例:乳児死亡率やテストスコア)ではなく、予防接種を受けた子供の数、妊娠時のケアを受けた人数、学校に入学した人数と出席した人数などのコミュニティがコントロール可能な直接的な指標を選定した。
インパクト評価から拡大適用への道筋	貧困アクションラボとインドネシア内務省が協力してRCTを用いたインパクト評価を実施した。その結果、ブロックグラントでインセンティブ付けされたコミュニティの方が、インセンティブ付けされていないコミュニティよりも、保健指標が改善したことが確認された。一方で教育指標には効果が見られなかった。総じて、保健指標の改善の50-75%はブロックグラントの貢献であると判断された。Generasiは、すでにインドネシア政府によって全国展開が開始されていたが、その政策の正当性が確認されたと言える。

(出所) <https://www.povertyactionlab.org/scale-ups/incentives-improve-aid-efficacy-case-generasi-community-block-grants>

図3 貧困アクションラボのRCTによるインパクト評価、政策教訓の産出・公表、政策教訓の産出・公表、拡大適用の実施



ある。RCTによるインパクト評価は、他のアプローチによるインパクト評価に比べて、高い内部妥当性（Internal Validity）を有すると言える。一方で、他の事業や他の場面への拡大適用可能性である外部妥当性に関しては、他の手法（例：定性的アプローチなど）による評価の方が問題がより少ないという見解もある（Rodrik, D., 2008）。したがって、今回解説した拡大適用の事例をもって、他のアプローチに比べてRCTは外部妥当性もより高いと一概に判断することはできない。

今回の論文では、政策教訓が拡大適用に直接的に影響したと判断された事例が確認されたが、明確に宣言されなくても当該国や援助機関の政策に間接的に反映されている例はあると思われる。その間接的な反映を明らかにする方法の検討は今後の課題である。

## 7-2. 今後の日本の開発援助への示唆

今回検討した貧困アクションラボの知見の蓄積から、日本の開発援助で利用できる主な知見として、すでに解説しているが、少なくとも次の4項目を挙げることができる。なお、筆者が知る日本の開発援助の状況との比較を踏まえている。

- (1) 教育以外の分野からの教育改善の提案の真摯な検討。出席日数の向上という目的の実現のためには、インフラ建設や教員研修よりも、単純な回虫駆除薬の配布（年に一度の配布で十分）が効果があることが確認されたが、日本の開発援助では導入例があるとは聞いておらず、導入を選択肢のひとつとして検討すべきである。
- (2) 生徒の学習レベルのクラス分けによる、レベル別の教育。生徒の心理面に細心の注意を払いつつ導入すれば、教育効果の向上が期待できる。やはり日本の開発援助で導入例があるとは聞いておらず、対応を検討すべきである。
- (3) 外部効果が高い物品やサービスに関して安易な住民負担を求めることは、住民のアクセスを劇的に減らすかもしれない。逆に、適切な運用を伴うそれら物品・商

品の無料化は、事業の目的の達成に貢献するかも知れない。なお、日本で最近議論されている相手国政府の一部負担を求める新しいタイプの援助も同様の考え方ができるわけで、外部経済を考慮に入れて慎重に検討すべきである。

- (4) コミュニティの能力強化は、確かに公共サービスの提供の質を改善する傾向がある。日本の開発援助では継続的に支援がなされてきたアプローチであり、今後も十分に力を置くべきであると言える。

貧困アクションラボによる、RCTを用いたインパクト評価の継続的な実施によって、世界で合計2億人以上にその影響が及んでいることが確認された。総額が継続的に減少している日本の開発援助が、より効率的に効果を発揮するためには、この実績は十分に参考にされるべきである。

## 8. 本論文の制約および今後の研究への示唆

本論文は、RCTによるインパクト評価の実施機関である貧困アクションラボのウェブサイトから得られた報告書や情報に基づいており、情報源が限られていることは認めねばならない。一方で、3ie（International Initiative for Impact Evaluation）のウェブサイトでは、すでに303件の系統的レビューが行われているとの記載がある（2016年7月23日時点）。また同様に英国国際開発庁（Department for International Development）のウェブサイトでは、3ieと協働して系統的レビューを行っているとしている（101件の登録がある（2016年7月23日時点））。これらの系統的レビューに関しても今後分析が必要である。

### 謝辞およびお断り

本稿の草稿にあたっては匿名の査読者の方から丁寧なコメントを頂戴した。この場を借りて御礼申し上げます。なお、本論文は筆者の所属機関の考え方を示すものではない旨、明記します。

## 注記

- 1 〒108-0075 東京都港区港南1-6-41 品川クリスタルスクエア12階
- 2 セクター別の件数の合計が966件となっているが、複数のセクターに登録がある案件が存在することが理由である。また、地域別の件数の合計は766件となり、同ラボのウェブサイトに表示のあった770件より少ないことになるが、理由は不明である。
- 3 これらのリストを見ると、個別のインパクト評価の実施数が最多となったファイナンス分野では政策教訓の抽出のための総合的なレビューが行われていないということになる。ただし、「政策教訓」「拡大適用事例」の形をとっていないが、貧困アクションラボの研究者がレビューを行っている論文はファイナンス分野でも多数あることは指摘されねばならない。

## 参考文献

- 佐々木亮 (2010, 2014) 『政策評価の理論と技法』、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)、日本評価学会
- 佐々木亮 (2010) 「エビデンスに基づく開発援助評価－援助評価の歴史、ランダム化比較実験の起源、スクリヴェンとバナージェの考え方の比較－」、『日本評価研究』10 (1)、日本評価学会
- 佐々木亮 (2010) 『評価論理：評価学の基礎』、多賀出版
- 佐々木亮 (2011) 「MDGs達成のための「エビデンスに基づく援助評価」－7つのベストバイとマイクロファイナンスの評価－」、『21世紀社会デザイン研究』、2011.10号

- 津谷喜一郎 (編著) (2015) 『いろいろな分野のエビデンス 温泉から国際援助までの多岐にわたるRCTやシステムティック・レビュー』、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)、日本評価学会
- 龍慶昭・佐々木亮 (2000) 『政策評価の理論と技法』、多賀出版
- Campbell, D.T. and Stanley, J.C., (1966). *Experimental and quasi-experimental designs for research*. Stokie, IL: Rand McNally.
- Campbell, D.T., (1969). "Reform as Experiments" *American Psychologist*.
- Fischer, R.A., (1925). *Statistical Methods for Research Workers*. NY: Hafner Publishing Company Inc.
- Fischer, R.A., (1935). *The Design of Experiments*. NY: Hafner Publishing Company Inc.
- Glass, G.V; McGaw, B; and Smith, M.L. (1981). *Meta-analysis in Social Research*. CA: SAGE Publication.
- J-PAL. (2016). *Policy-Lessons*. Retrieved from <https://www.povertyactionlab.org/policy-lessons> (retrieved at 27 April, 2016)
- J-PAL. (2016). *Scale-Ups*. Retrieved from <https://www.povertyactionlab.org/scale-ups> (retrieved at 27 April, 2016)
- Rodrik, D., (2008). *The New Development Economics: We Shall Experiment, but How Shall We Learn?*. Faculty Research Working Papers Series, Harvard Kennedy School.  
<https://research.hks.harvard.edu/publications/getFile.aspx?Id=317>

(2016.8.25 受理)

## **Communicate and Use of Evidences in International Development Field: Movement of J-PAL**

Ryo Sasaki

International Development Center of Japan Inc. (IDCJ)

sasaki.ryo@idcj.or.jp

### **Abstract**

Abdul Latif Jameel Poverty Action Lab (J-PAL) was established at 2003 and it has conducted over 770 impact evaluations using randomized controlled trial (RCT) since its establishment. Based on the accumulation of the impact evaluation using RCT, J-PAL conducted comprehensive review of a set of impact evaluations at education sector, health sector, political economy & governance sector and prepared four “Policy Lessons” consisting of a set of detailed lessons generated by those reviews. In addition, J-PAL achieved 7 cases of scale-up. This research report focuses what lessons have been actually generated and what scale-up cases have been realized. In addition, this thesis discusses what Japan should learn from the systematic review results and how Japan can conduct more the practice of systematic review and scale-up cases.

### **Keywords**

RCT, impact evaluation, policy lessons, scale-up, comprehensive review

## 【研究論文】

## 評価者の倫理教育におけるケース・メソッドの利用 — 監査人の倫理教育手法からの示唆 —

小林 信行

OPMAC株式会社<sup>1</sup>

kobayashi@opmac.co.jp

### 要 約

本稿は、監査人が倫理教育手法を考察し、評価者の倫理教育に向けた示唆を導出する。監査人の倫理教育では、ケース・メソッドに基づく参加型手法が重視されており、幾つかの体系化された意思決定プロセスがこの倫理教育手法で使用されている。また、倫理上の問題を理解するための分析枠組みや選択肢を比較考量するための手法も準備されている。

評価者の置かれた環境で特徴的な点は、評価者はしばしば同一業務でアカウントビリティと業務改善という異なる評価目的が与えられ、局面毎に異なる利害関係モデルに直面することである。評価者は状況を読み解く能力がより求められ、また「線引き問題」や「相反問題」への対処を学ぶ必要があり、ケース・メソッドはその教育上のニーズに合致する手法である。ケース・メソッドによる倫理教育は「実践」から「制度」への道筋を整えるものであり、参加者による評価倫理ガイドラインの検討を通じて有意義なフィードバックを生み出すことが期待される。

### キーワード

評価倫理、職業倫理、監査人、倫理教育、ケース・メソッド

### 1. はじめに

評価者の職業倫理の特徴として、強制力を伴う仕組みが構築しにくい点が指摘されている(Fitzpatrick 1999, p.12)。その理由として、評価手法が依拠するパラダイムの多様さやプロフェッションとしての歴史の短さから、倫理規程の解釈や適用についてプロフェッション内でコンセンサスが形成されにくく、その結果として、プロフェッション内部でも罰則といった強制力のあるメカニズムが整備されていない点が挙げられている。評価者の倫理面での逸脱を防止する

にあたり、罰則を通じて倫理を順守させることは困難である。そのため、評価者というプロフェッションにおいては、内発的動機の涵養が職業倫理を担保する上で肝要となり、個々人の倫理高揚に向け倫理教育の果たす役割が大きい。日本評価学会の主催する評価士養成講座では評価倫理の講義が行われているが、評価者向けの倫理教育手法の開発はまだ緒についたばかりである。倫理教育プログラムを確立する上で、教育手法の開発に長年取り組んできた他プロフェッションから学ぶことは多い。本稿は、監査人の倫理教育手法、特にケース・メソッドに焦点

を置いて、評価者に同様の教育手法を適用するに際して、留意すべき点を考察する。

## 2. 本研究の視点

### (1) 用語の定義

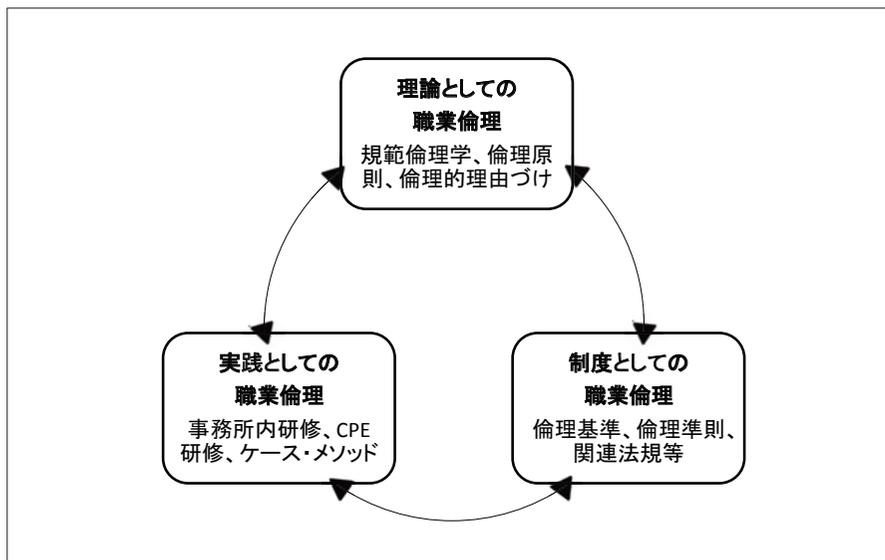
評価者というプロフェッションの職業倫理を論考の対象とするに際し、本稿では一義的に規定することが難しい倫理という単語に作業定義を設定する。訳語としての倫理 (ethics) と道徳 (moral) はそれぞれギリシア語とラテン語の「慣習・習俗」を受け継ぐものであり、語源にさかのばれば両者は同義である (星野ら1997, p.1)。両者の意味するところが重なるため、職業倫理を扱う応用倫理学では、しばしば倫理と道徳は明確に区分されずに用いられている<sup>2</sup>。本稿でも応用倫理学の考え方に沿って両者を区分せず、倫理は道徳と重複する意味を持つものとし、杉本・高城 (2008, pp.4-5) を参考に倫理に「コミュニティ内の人々が順守するよう期待される自律的な規範」との作業定義を与える。依拠する参考文献に用いられた用語についても、この定義に合致する範囲内で極力転記する。この定義

に基づく、職業倫理は特定のプロフェッションのコミュニティに属する人たちが自ら守る規範と理解される。倫理が守られるためには、理解できる文章にまとめられる必要があり、倫理を成文化したものには倫理規程の用語を充てる。倫理規程のうち、個々の規則を根拠づける抽象的な倫理上の価値を基本原則と呼ぶ (Newman & Brown 1996, p.23)。

### (2) 本研究の視点

本稿では、職業倫理を具体的、かつ包括的に理解する枠組みとして、梅津 (2002, p.7) の提唱したビジネス倫理のフレームワークを援用する。このフレームワークでは「理論」、「制度」、「実践」の三点から成り立っており、倫理的なジレンマに対処する「実践」、意思決定の理由づけを考察し根拠を与える「理論」、加えて個々の事例を統合的に扱えるよう統合する「制度」が相互にフィードバックを繰り返し、現実のニーズに答えるべく倫理が深化するモデルとなっている。八田・町田 (2003, pp.42-43) では、ビジネス全般を対象とした上記フレームワークを特定分野に特化させ、会計プロフェッションの職業倫理フレームワークが作成されている (図1を参照)。

図1 会計プロフェッションの職業倫理フレームワーク



(出所) 八田・町田 (2003) p.43より抜粋

また、本稿は各プロフェッションの有する利害関係にも着目する。利害関係には、サービス提供者とその利用者のみのものであり、さらに第三者が介在するモデルがある。両者では倫理的なジレンマの発生する原因が異なり、三者間モデルではインセンティブの捻じれが発生する可能性がある。インセンティブの捻じれとは、クライアントの持つ三つの側面（サービスの選任者、費用負担者、利用者）が一致しない現象を指す（岡崎 2012、p.20）。インセンティブの捻じれがある場合、サービスの選任者や費用負担者がプロフェッションのサービス提供に介入し、利用者の利益を損ねる可能性がある。そのため、弁護士や医師といった原則的に二者間の利害関係の基にあるプロフェッションに比べて、保証業務を行う公認会計士は利益相反を起こしやすい状況にあると見なされる。この捻じれへの対応が公認会計士の職業倫理を特徴的なものとしている。

### 3. 共通点と相違点

監査人の倫理教育手法を考察する前に、監査人と評価者の職業倫理上の共通点と相違点について詳述したい。

#### (1) 共通点

監査の伝統的な定義は「当事者以外のものが、当事者の行動や成果について分析、批判して、その適否や正否に関して判断を行うこと」（日本経済新聞社 1992、p.45）であり、監査主体に基づく、組織体外部の専門家（主に公認会計士）が実施する外部監査、組織体内部の監査人（内部監査人）が実施する内部監査、に大別される。監査人は第三者として行為、成果の適否や正否に保証を提供する位置にあるが、公認会計士、内部監査人ともに伝統的な監査の定義に合致する保証業務ばかりでなく、専門知識を活用できる周辺業務にもその職域を広げている。

日本公認会計士協会の公表した「公認会計士等が行う保証業務等に関する研究報告」では、公認会計士の業務を保証業務とそれ以外の業務

（非保証業務）に大別している（日本公認会計士協会 2009、p.3）。前者は財務諸表監査が代表的であり、後者には税務や会計分野におけるコンサルティング業務が含まれる。保証業務は、企業が提供する情報の適切さを、公認会計士が第三者（主に市場参加者）に対して保証することになる。つまり、利害関係は公認会計士、企業、サービス利用者の三者間のものとなる。非保証業務を一つの利害関係モデルに定型化することは難しいが、コンサルティング業務の多くは依頼者が公認会計士に業務を委託する二者間のものである。

内部監査人の業務は、保証業務とコンサルティング業務に大別される。米国に本部を持つ内部監査人協会（Institute of Internal Auditors : IIA）が策定した「内部監査の専門職的実施の国際基準」では、保証業務を監査対象について「独立した監査の意見ないし結論を得る基礎として、内部監査人が入手した証拠の客観的な評価」と定義され、一般的に三者（監査対象のプロセスに責任を持つ部署、内部監査人、監査結果の利用者）が関与すると説明する（IIA 2012、p.2）。保証業務は、監査対象の特定のプロセス（例えば、購買や在庫管理等の個々の活動）のリスク・マネジメントや内部統制の有効性を内部監査人が評価し、サービスの利用者たる経営者がその意見を評価判断に用いるもの、となる。他方、同基準はコンサルティング業務を「助言の提供」と定義し、「個々のコンサルティング業務の内容と範囲は、依頼部門との合意による」と見做し、同業務は一般的に二者間（内部監査人、依頼部門）のものとして説明している。

評価は「プログラムや政策の改善に寄与するための手段として、明示的または黙示的な基準と比較しながらプログラムや政策の実施あるいはアウトカムを体系的に査定する」（Weiss 1997、p.5）のものであり、実務上の用途からプログラム・政策に関するアカウントビリティ、または業務改善を目的とする2種類の業務に大別される<sup>3</sup>。評価主体としては、組織外部の専門家による外部評価、組織の職員が実施する内部評価に区分され、外部評価、内部評価ともに2種類の業務を行う。評価目的の違う業務は異なる利害関係を有

表1 利害関係モデル

公認会計士		内部監査人	
保証業務	非保証業務	保証業務	コンサルティング
三者（公認会計士、対象企業、サービス利用者）	二者（公認会計士、サービス利用者）	三者（内部監査人、監査対象部門、経営者）	二者（内部監査人、組織内のサービス利用者）

外部評価者		内部評価者	
説明責任を目的とする評価	業務改善を目的とする評価	説明責任を目的とする評価	業務改善を目的とする評価
三者（外部評価者、評価委託者、一般国民）	二者（外部評価者、評価委託者）	三者（外部評価者、所属先の組織、一般国民）	二者（内部監査人、組織内の評価結果の利用者）

(出所) 筆者作成

表2 インセンティブの捻じれ

	公認会計士： 保証業務 (財務諸表監査)	外部評価者： アカウンタビリティ 目的の評価	内部評価者： アカウンタビリティ 目的の評価
1. 選任権者	クライアント企業	評価委託者	所属組織
2. 費用負担者	クライアント企業	評価委託者	所属組織
3. 利用者	主に市場参加者	主に一般国民	主に一般国民
インセンティブの捻じれ	あり	あり	あり

(出所) 筆者作成

し、監査人と同様に二者間と三者間の利害関係モデルが存在する（表1を参照）。

アカウンタビリティを目的とする評価では、介入セオリーの適切さ、介入の目標達成度、介入効果の大きさ等を明示し、評価者が評価対象となる社会介入に対し価値判断を行う。評価結果の利用者は一般国民であり、利害関係モデルは評価者、（外部評価者の場合）評価委託者／（内部評価者の場合）所属先の組織、一般国民の三者間となる。評価の利用者たる一般国民は評価者の選任には関与せず、費用も直接負担しない。保証業務を行う公認会計士と同様に、評価者によるアカウンタビリティ目的の業務にもインセンティブの捻じれが生じうる（表2を参照）。

業務改善を目的とする評価では、評価者は課題を有する組織や部門に対して、今後の活動に対する教訓や提言を提示する。評価結果の利用者は、外部評価者の場合には評価委託者であり、内部評価者の場合には所属先の組織となる。利害関係モデルは評価者と評価結果の利用者との二者間で、公認会計士の非保証業務や内部監査

人のコンサルティングと同じモデルとなり、インセンティブの捻じれは生じない。

## (2) 相違点

最初の相違点として、監査人に対しては法による処罰が存在し、職業倫理から著しく逸脱し、社会一般に損害を与える行動を抑制する強力な外発的動機が存在する。わが国においては、金融商品取引法、公認会計士法が監査人及びその所属する組織に対して罰則を有する法として挙げられる。監督官庁が、それらの法律に抵触する行動に対し、個々の監査人や彼らの所属する組織に対し処分を行っている。金融商品取引法に基づく処分は、金融機関に加え、当該法律の適用を受ける一般事業会社、それらを監査する会計監査人にも及ぶため、不適切な行動が外部により摘発されうる体制が公認会計士の所属する組織の多くを網羅する点が指摘できる。また、同法に基づき内部監査体制の不備が指摘され、内部監査人の所属する組織が処分を受けることもある。公認会計士及び監査法人に対しては、

公認会計士法の罰則も適用され、問題のある行動を抑制する「制度」が強固に構築されている。

冒頭で言及したように、評価者には職業倫理からの著しい逸脱を抑制する外発的動機は弱く、その点は監査人との比較でも改めて明確となる。金融商品取引法、公認会計士法と異なり、行政機関が行う政策の評価に関する法律（通称、政策評価法）には評価者やその所属組織に対する罰則は設けられていない。職業倫理からの逸脱を明確に定義し、倫理規程への違反を理由に評価者に対し深刻な影響のある罰を与えることには困難を伴う。

第二の相違点として指摘できるのは、監査人においては、保証業務とそれ以外の業務では異なった規程が設けられ、プロフェッション内部では2種類の業務は明確に区分される点である。監査人、評価者ともに利害関係モデルの異なる業務を実施するが、監査人に関しては保証業務とそれ以外の業務が同一業務内に含まれることはない。例えば、東芝不正会計事件<sup>4</sup>では、内部監査部門のコンサルティング業務への過度の集中が不十分な保証業務につながった点を第三者委員会が指摘している（株式会社東芝 第三者委員会 2015、p.283）。この指摘は、2種類の業務が区分されていることにより可能となるものである。

監査人が業務を区分し、それぞれに異なる規程を持つのは、保証の有無が利害関係の違いに繋がり、両業務で利益相反の原因や生じる損害の程度が異なるためである。公認会計士や内部監査人の保証業務には三者が関与し、監査人が監査対象となる部門や組織に有する利害が保証業務の利用者の利益を脅かす可能性が生じる。とりわけ公認会計士の場合、財務諸表監査の利用者は一般の株主や債権者であり、財務諸表上の重大な虚偽表示は、市場の信頼性及び債権者の保護の目的を損ね、ひいては公共の利益を損ねる深刻なものとなる。一方、公認会計士の非保証業務や内部監査人のコンサルティング業務は二者間のものである。監査人が自己利益の過度の追求によりサービスの質を下げる可能性はあるが、低い品質は主としてサービスの利用者に損害をもたらす。

評価分野においても、総括評価と形成評価で

は答えるべき評価設問が異なり、必然的に異なる評価情報を必要とするため、両者の区分が望ましく、複数の評価目的が設定される場合でも主たる目的を設定すべきことは理解されている（Patton 1997, pp.49-51）。総括評価と形成評価は異なる利害関係モデルを有するため、倫理面でのジレンマやその原因や対応も異なり、その観点からも両者を区分する意義がある。しかし、同一業務に総括評価が目指すアカウンタビリティと形成評価がゴールとする業務改善を同程度に重視する評価目的が設定されることも多く、評価の局面毎に評価者は異なる利害関係に直面する（小林 2014、p.32）。局面で変化する利害関係は内部評価者、外部評価者ともに生じる状況であり、評価者は倫理的に望ましい行動を選択するにあたり、その都度自らの状況を的確に把握する必要に迫られる。

さらなる監査人との相違点として、評価者の倫理では情報提供者や政策／プログラムの受益者の安全や尊厳が基本原則に含められ、重視される価値となっている点が指摘できる（小林・一寸木 2010、p.154）。その一方、監査人の職業倫理においては、情報提供者の保護は基本原則で明示的に言及されるほどの重きは置かれていない。評価者のこの特徴はむしろ研究倫理の被験者保護と通底し、評価者がクライアントのニーズに答えるべく科学的な調査手法を使う点に起因している。とりわけランダム化比較実験を用いるような評価では、評価デザインがプログラム実施と密接な関係を持ち、プログラム受益者の保護が倫理の焦点となる。加えて、政策／プログラムの受益者には多くの社会的な弱者が含まれ、保護をより必要とする点も看過できない。

#### 4. 監査人における倫理教育の考察

監査人の倫理教育としては、実務に携わる前の教育機関での講義、実務従事後の継続教育に大別できる。実務前の倫理教育は個々人の規範意識の基礎形成に関与するため、多くの論考がなされ、評価者の倫理教育にも貴重な示唆が存在している。そのため、本稿では、前者に含ま

れる大学や大学院における会計倫理の教育を中心に論考する。

### (1) ケース・メソッド

大学や大学院における会計倫理クラスでは、倫理規程の解釈を講義するばかりでなく、ケース・メソッドも導入されている。国際会計士連盟（International Federation of Accountants: IFAC）の下部組織である国際会計教育基準審議会（International Accounting Education Standards Board）では、プロフェッション初期段階の倫理教育として、ロールプレイ、ディスカッションといった参加型手法を勧めている（IFAC 2014, p.10）。ケース・メソッドと意思決定プロセスを組み合わせ、一定の手順で倫理的なジレンマへの対処を教授する手法は、講義で取り組みやすく、初学者が主体的に参加する点が評価されている（Boyce 2014, p.544）。米国の大学における会計倫理クラスでは、会計倫理の理論と事例（ケース）をまとめた教科書が使用されている<sup>5</sup>。わが国の会計専門職大学院の倫理教育においては倫理規程の解釈に関する講義が主であるが、識者により参加型手法による倫理教育が提唱されている（八田 2011, pp.130-131）。

倫理教育におけるケース・メソッドでは具体

的なケースに基づき討論を行い、ケースの当事者の立場で意思決定の実践的な訓練を行う。参加者は他のメンバーとの議論を通じて、意思決定やその元となる価値観を検討する。ケース・メソッド用の教材では、論点は明示されないこともあり、参加者は状況を読み解く必要にも迫られる。明確な結論を想定せず、オープンエンドの議論を前提とした題材も多い。会計倫理におけるケース・メソッドは、参加者が同一の意思決定に到達することを目的とはせず、多様な倫理的視点でジレンマを分析し、実際にジレンマを抱える局面で自らの意見を周囲に提示する能力を涵養することをゴールとしている（原田 2012, p.135）。

会計倫理では、道徳的論点検査（The Defining Issues Test: DIT）を用いて道徳性発達を測定する研究があり、倫理教育の効果推計が取り組まれている。DITはKohlbergの道徳性発達理論（表3を参照）を下敷きに道徳性発達の程度を測定するものであり、倫理的ジレンマを含んだケースと道徳性発達理論の6つの段階に対応した選択項目を配した質問紙を使う。具体的には、ケースの読後、特定の行動をとる上で重視する観点を選択して、評点を付けるという手順となる。

DITを用いて倫理教育が会計専攻の学生の道徳

表3 Kohlberg理論に基づく道徳性の発展段階

道徳の発展段階		定義
慣習以前レベル	ステージ 1： 罪と罰への服従の志向	行為の物理的結果が、善悪を決定する。罰の回避と権威への服従が、自らの視点で決定される。
道徳的価値は人や規範にあるのではなく、外的、物理的な結果や力にある。	ステージ 2： 個人主義的、道具主義的志向	自分の欲求、時に他者の欲求を道具的に満たすことが正しい行為で、自らの利益や欲求に合致するように行動することが正しい。
慣習的レベル	ステージ 3： 対人的同調、相互関係への志向	他者から期待されるよい役割をすることが正しい。
道徳的価値は良いあるいは正しい役割を遂行すること、慣習的な秩序や他者からの期待を維持することにある。	ステージ 4： 社会的秩序への志向	正しい行為は、自らの義務を果たして、権威を重んじて、社会秩序を維持することを目的とすることから成る。
慣習以後レベル	ステージ 5： 社会契約的、遵法的志向	一般的な個人の権利と幸福を守るために社会全体によって吟味され一致したものとしての規準に従うことが正しい。
現実の社会や規範を超えて、妥当性と普遍性をもつ原則を志向し、自己の原則を維持することに道徳的価値をおく。	ステージ 6： 普遍的倫理原則への志向	自ら選択した倫理的原則に従うことが正しい。特定の法や社会的合意は、この原則にもとづいている場合に妥当と考える。

（出所）原田・矢部（2011）p.95より抜粋。

発達性に与える効果を計測する研究が進められており、ケース・メソッドを組み込んだ倫理教育の研究も多い。Ponemon (1993) では、米国の会計専攻の大学生（介入群39名、比較群34名）と大学院生（介入群26名、比較群27名）への講義、ビデオ視聴、ケース・ディスカッションを用いた10週間の倫理コースでは、4つの群（大学生、大学院生のそれぞれのコース参加者と非参加）のいずれもDITの評点にはコースの事前事後で有意な差が生じなかった。原田・矢部 (2011) では、日本で会計関連科目を受講中もしくは受講した学生（661名）を対象に、無作為にグループA（会計教育をまったく行わない）、グループB（倫理規程のみを渡す）、グループC（倫理規程を渡して書面で意見をまとめさせる）、グループD（倫理規程を渡しケースを議論する）の4群に割り当て、実験後にDIT検査で道徳性発達を計測した。実験の結果、各群で有意差はなかったものの、グループDが最も高い評点となった。研究結果として、短時間の倫理教育が道徳水準に与える効果は限定的である一方、教育手法としてはケース・メソッドの有効性が示唆された。一方、上記の研究とは逆に倫理教育が道徳性発達水準に与える効果が有意であると結論する研究もある。倫理規程の学習、ケース分析とディスカッションを含めた講義（5週間）を受けた米国の会計大学院の学生（45名）の事前事後での会計道徳的論点調査（Accounting Defining Issues Test: ADIT）を計測したところ、有意な差があったとの結論を得ている（Welton & Guffey 2009）。ADITはDITでは会計や監査分野に特徴的な倫理面の課題を適切に計測できないとの理由から、DITを会計分野に適合させた計測手法である。

過去の研究では、ケース・メソッドが道徳性発達水準に与える効果については明確な結論に至っていない。各研究で異なった結果が出ている理由としては、ケース・メソッドによる倫理教育の期間、ケース・メソッドの進め方、他の教育手法との組み合わせ、倫理教育への参加者数、測定手法が研究毎に異なる点が挙げられる。

ケース・メソッドの進め方として、組織全般、社会的・政治的な文脈をケースの議論から切り離すことの問題性が指摘され、また倫理規程の

適用に絞って教育を行う限り、専門職としての責任を狭い範囲に限定する理解にしか到達しないとの意見がある（Boyce 2014, pp.544-545）。さらには、合規性を重視する会計分野の専門職の特性から、伝統的な会計教育カリキュラムが道徳性発達、特に道徳のステージ5以上（既存の規範を超える普遍的な原則への志向）への発達を妨げ、倫理教育の効果がでにくい点を主張する研究がある（原田 2012, pp.194-195）。

ケース・メソッドの道徳性発達水準に与える効果は明確ではないが、監査人の教育にケース・メソッドを用いる意義は認識されている。将来の監査人が、ケース・メソッドを通じて社会の要請を再考し、既存の倫理規程がプロフェッション全体の規範として十分であることを検討することは貴重な機会と考えられている（八田 2011, pp.130-131）。つまり、監査分野では「実践」において活用されるケース・メソッドが将来的に「制度」に有意義なフィードバックを与える点が重視されていると思料される。

前述のとおり、評価分野では方法論の多様さから、倫理規程の解釈や適用にコンセンサスが形成されにくく、外発的動機となる罰則の適用が困難である。しかし、その反面、倫理規程の解釈や適用に共通点を見出し難いため、評価者は監査人に比べ、倫理規程を教条的に順守する傾向は弱いと推察される。つまり、外発的動機の弱さがデメリットとなる一方、合規性への志向が弱い点はメリットとなりうると考えられる。そのため、ケース・メソッドが有効な教育手法となる可能性はあるが、監査人に関するこれまでの研究結果を踏まえると、期待される効果を得るには十分な条件を整える必要がある。

## (2) 意思決定プロセス

倫理的ジレンマの局面における適切な意思決定を学ぶために、ケース・メソッドは体系的な意思決定プロセスと組み合わせられる。ケース・メソッドを通じて学生が意思決定プロセスを内面化した後で、実務者として実際に意思決定に活用する企図があり、教育の場における学びから業務における意思決定へのスムーズな移行が想定されている。幾つかの意思決定プロセスが

提案されているが、本稿では7段階法、8段階法を採り上げる<sup>6</sup>。

7段階法はMintz (1997, pp.35-36) により、8段階法はLagendefer & Rockness (1989, pp.66-68) により提案された意思決定プロセスである(表4を参照)。7段階法は問題分析のステップを8段階法に比べてより多く設けており、業務上の問題(企業組織が抱えている問題)、会計上の課題(会計処理上の問題)を分けて問題を抽出する。また、7段階法ではステップ⑤の選択肢の分析において、法規や専門職の基準との整合性ばかりでなく、社会契約論や義務論に根拠を持つ他者の権利の尊重、功利主義の観点に立つ選択肢の社会的有益さ等を問うとしており、複数の倫理思想を基準に選択肢を精査する手順がある(原田2012, p.135)。つまり、「理論」における規範倫理学上の価値と「実践」の意思決定の関連性が

表4 意思決定プロセスの例

7段階法のプロセス
① 事実関係を確認しなさい。
② 業務上の問題を確認しなさい。
③ 会計上の問題を確認しなさい。
④ 利害関係者を識別し、その責務を確認しなさい。
⑤ 選択肢に対して倫理的な分析を加えなさい。
⑥ 行動を決定しなさい。
⑦ 意思決定の再確認をしなさい。- 自問しなさい。
8段階法のプロセス
① 事実関係を確認しなさい。
② 倫理的な問題点と利害関係者を確認しなさい。
③ その状況と関係する規範、原則、価値観を明らかにしなさい。
④ 代替の行動方針を確認しなさい。
⑤ 規範、原則、価値観に調和する最適な行動方針を決定しなさい。
⑥ 各々の可能な行動方針について、結果を評価しなさい。
⑦ もしその選択肢が適切と思うなら、より客観性を高めるために信頼できる人とその選択肢について議論しなさい。
⑧ 意思決定をしなさい。

(出所) 原田 (2012) pp.132-133 より抜粋。

明確なものとなっている。一方、8段階法では選択肢の再検討に際し、他者の意見を聞く手順があり、意思決定の客観性を当事者以外での確認で確保する点が特徴的である。

上記のような差異はあるが、両モデルにおいて、i. 事実関係の確認、ii. 問題点の抽出、iii. 利害関係者とその利害の確認、iv. 意思決定の選択肢の列举、v. 分析と選択、vi. 選択肢の再検討、vii. 最終的な意思決定、という一連のプロセスは共通している(原田2012, p.134)。評価者に適用できる意思決定プロセスを策定する上でも、両手法に共通する上記 i から vii までの手順は参考となるだろう。

前述のとおり、評価者は同一業務内に利害関係モデルが混在し、局面毎に適切な状況把握が求められる点で、監査人とは異なった環境にある。監査人の手法を参考に評価者の意思決定プロセスを策定する場合、三者間の利害関係モデルとなるアカウントビリティを重視するか、二者間のモデルとなる業務改善に徹するか、といった2つの評価目的の相対的な重要性が、ステップ iv 以降の選択肢の決定に大きく影響すると考えられる。いずれの評価目的を重視するかにより、適切な行為は異なるからである。状況把握に関連するステップ i から iii までが、利害関係者と利害の程度を明らかにする作業となり、2つの評価目的のいずれを優先するか判断に直結するため、監査人と比べて重点的にトレーニングが行われるべきステップになると思料される。

### (3) 脅威カテゴリー

上記した7段階法と8段階法に共通するステップの i から iii までは、意思決定に向けて事実関係を整理し、意思決定に必要な情報を洗い出す段階となる。この段階において、倫理的問題のある状況を類型化できれば、自らのおかれた状況を理解し、倫理的なジレンマの原因を特定し、対策をとることがより容易になる。本来、倫理教育のために開発されたものではないが、この段階に援用できるアプローチとして、脅威カテゴリーという概念がある。IFACの倫理規程(Code of Ethics for Professional Accountants)では、公認会計士の職業倫理上の基本原則が脅かされ

る状況を5種類に分類し、それぞれの脅威が生じる具体的な状況や緩和策が提言されている (IFAC 2015, p.12, pp.23-27)。IFACの倫理規程で言及される脅威は以下の通りである。

- ①自己利益の脅威：経済的、他の利害が会計士の判断や行動に不当に影響する脅威
- ②自己レビューの脅威：以前に会計士や所属する組織が提供したサービスや活動が現在のサービスや活動における判断の根拠となっており、その会計士が以前に下した判断を適切に評価できなくなる脅威
- ③援護の脅威：会計士の客観性が侵害される程度までクライアントや所属組織の立場を斟酌する脅威
- ④馴れ合いの脅威：会計士が持つクライアントや所属組織との長期で親密な関係により、クライアントや所属組織の利害や過去の業務を過剰に受け入れてしまう脅威
- ⑤不当なプレッシャーを受ける脅威：実際の圧力や圧力と受け取られる行為（会計士に不当な影響を及ぼす試みを含む）により、会計士が客観的な業務を妨げられる脅威

内部監査人に関しても、その客観性に対する脅威カテゴリーとして、7分野（①自己レビュー、②社会的圧力、③経済的利害関係、④個人的関係、⑤親密さ、⑥文化的・人種的・性的偏向、⑦認識上の偏向）の脅威が識別されており、それぞれのカテゴリーに対して緩和する要因や脅威を管理する具体的な手法が提案されている (IIA 2001, pp.47-66)。

脅威カテゴリーは、保証業務において監査人の客観性が著しく侵害される状況が主に想定されている。前述のとおり、保証業務は三者間のモデルを有し、サービスの利用者以外の第三者が監査人との間に持つ利害が利益相反を生み出しやすくなる。監査人については、深刻な倫理上の問題が起こりうる業務に関しては、倫理的なジレンマの理解を助けるフレームワークが存在し、問題とその原因が予め類型化されている点が特記される。分析用のフレームワークは実務家の適切かつ効率的な判断を支えるのみなら

ず、教育面でもより体系的な教授法を容易にする点でも有意義である。

評価者においても、アカウンタビリティ目的の評価は利益相反が生じやすい三者間モデルを有する。監査人と同様に、リスクの高い業務を重点的に、倫理上の判断を助ける分析枠組みを構築することは合理的であり、そのフレームワークの利用を教育段階から進めることは検討に値する。とりわけアカウンタビリティ目的の評価にはインセンティブの捻じれがあるため、倫理上発生しやすい問題の分類とその原因を特定し、評価者が取りうる対策を例示することは、今後、取り組む意義のあるタスクとなるだろう。

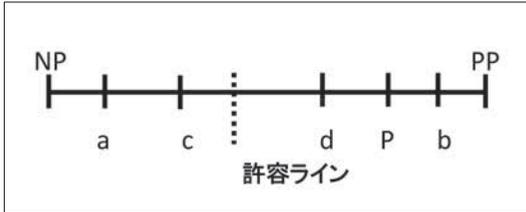
#### (4) 選択肢の分析手法

上記した7段階法と8段階法に共通する手順では、ステップivからviまでが選択肢の是非を検討する段階にあたる。監査人の倫理教育では、この検討を体系的に行うべく、意思決定プロセスと親和性の高い分析手法を導入する試みがなされている。

倫理上の問題として、行為者は「線引き問題」や「相反問題」にしばしば直面する。「線引き問題」とは、何らかの価値に基づき、正しい行為と不正な行為は定義できるものの、明らかに正しい行為と明らかに不正な行為の間にはグレーな領域があり、幾つかの選択肢がその領域に入るため、どこまでが正しい行為か判断に迷うことを指す。判断基準が外部から与えられる法と異なり、倫理はこの線引きを行為者自らが行う。一方、「相反問題」とは、2つ（またはそれ以上）の相反する倫理上の責務があり、いずれかに基づく行為を選択しなければならぬ状況を目指す (杉本・高城 2008, pp.107-110)。

線引き法は「線引き問題」への対処法であり、複数ある選択肢を何らかの基準に沿って一線上に配置できる場合には「相反問題」にも適用できる。線引き法は明らかに正しい行動案（肯定的模範事例、positive paradigm: PP）と明らかに正しくない行為案（否定的模範事例、negative paradigm: NP）を直線の両端に配し、その中間に幾つかの仮定的な行動案（図2では行動案a~d、アルファベットの順序は何らかの尺度に基づくも

図2 線引き法の例



(出所) 原田 (2012) p.147を参考に筆者作成。

のではない)を導出し、線上に配置する。そして、直線のどこかで許容できる線(許容ライン)を設定して、最後に判断しようとする行動案(疑問事例、problematic case: P)が許容ラインのいずれの側に配置されるかを決定する(原田 2012、pp.146-147)。図2の例では、Pは許容される選択肢となる。

ケース・メソッドでは、参加者が議論の上、模範的事例の設定、仮定的な行動案の配置、許容ラインの設定、疑問事例の位置について合意形成を図る。合意形成は特定の行動案を支持する／しない理由に焦点を充てるのではなく、行動案の配置や許容ラインの位置を中心に進められる。つまり、参加者間で行為の是非を判断する価値観が異なっていたとしても、行為の相対的な正しさは合意できることを想定している。

ケース・メソッドでは、選択肢の分析にディシジョンツリー法が使用される場合もある(Mintz 1997, p.41)。ディシジョンツリー法は行為の結果を判断基準とする功利主義に立脚する手法であり、選択した行動案が次にどのような結果になるかを想起して、特定の行為の結果を複数検討するものである。相反する複数の責務を一線上に配置することが困難な場合にも、結果を基準に行為を選択することが可能となる。

監査人と評価者の持つ三者間の利害関係の基では、行為者はサービス利用者と費用負担者のそれぞれに対して責務があるため、相反問題が生じやすい。また、評価者については、同一業務内にアカウントビリティと業務改善という異なる目的が同時に設定されるため、それぞれの評価目的に合致するが相反する複数の行為が存在する状況も想定される。そのため、評価者の倫理教育にて選択肢を比較考量する手法を学ぶ

ことは有意義である。

留意すべき点としては、評価者の場合、利害関係者とその価値観がより多様である点が指摘できる。監査人と比較した場合、評価者の業務では、①評価の利用者がさまざまな意見を持つ一般国民となるケースがあること、②評価対象の政策／プログラムの受益者や情報提供者も重視される利害関係者であり、彼らの意見も尊重すべきこと、の2点が特徴的である。その結果として、問題点や事実関係の認識、望ましい行為が多様な利害関係者間で大きく異なるケースも想定され、トレーニングでもその点への対処を学ぶ必要がある。実際の評価の現場を完全に模倣することは難しいものの、ディスカッションの参加者に異なる立場を割り当てるロールプレイ方式を用いて、多様な価値観を有する利害関係者が存在する状況を意図的に作りだす取り組みが必要となるだろう。

## 5. 結論

本稿では、監査人と評価者の共通点と相違点を導出した上で、監査人の倫理教育手法を紹介し、評価者に同様の手法を適用する際の留意点についても考察した。

監査人の倫理教育には、ケース・メソッドに基づいた意思決定プロセスの習得が含まれ、その過程で習得された手法は実務への適用が可能なものである。局面毎に利害関係モデルが変転し、状況を的確に読み解く能力が求められる評価者にとっては、ケース・メソッドは意義のある教育手法である。特に監査人の脅威カテゴリーのような分析枠組みに習熟することは、状況を把握する能力の育成に寄与するだろう。また、評価者は、同一業務内で異なる2つの評価目的を追求することがあり、評価目的の違いから倫理的に望ましい行為が複数提示されることが想定され、そのような状況への対処を学ぶため、線引き法やディシジョンツリー法を使ったトレーニングは検討に値する。

監査人に関しては、ケース・メソッドが道徳性発達水準に与える効果は有意ではないと結論

する研究がある。評価者については、倫理規程の解釈に幅があり、監査人と比べて合規性への志向は強くないと考えられ、ケース・メソッドによる倫理教育が道徳性発達水準に与える効果はあるものと推察される。しかし、評価者の倫理教育でも条件次第で効果を得られない可能性も考慮すべきである。そのため、ケース・メソッドの導入にあたり、様々な条件での試行と効果測定を通じ講義手法の継続的な改善が必須と言えよう。

変化する社会の要請に答えるべく、倫理規程の内容は継続的に見直される必要があり、日本評価学会の評価倫理ガイドラインもその例外ではない。監査人においては、ケース・メソッドによる倫理教育は「実践」から「制度」に至る1つの経路となっていると思料される。評価者の倫理教育にケース・メソッドを導入する際にも、参加者が評価倫理ガイドラインの適切さを検討することが望ましいだろう。

本稿はクライアントとの関わりに重点を置き、監査人と評価者を比較し、倫理教育手法への示唆を導出したが、政策／プログラムの受益者、情報提供者との関係性も評価倫理を考える視点として欠くことはできない。この視点は研究倫理で扱う被験者保護とも重なる議論となるため、研究倫理とも照らし合わせて評価倫理のあり方を検討することも有意義であろう。

## 謝辞

本稿は日本評価学会 第16回全国大会の発表要旨に論考を加えて、大幅に加筆・修正を行ったものである。発表要旨及び本稿の執筆にあたり、有益なコメントを頂いた方々、特に国際協力機構の正木朋也氏、公認会計士の標夏樹氏、評価者倫理・スタンダード策定分科会のメンバーに感謝を申し上げたい。

## 注記

1 本稿にて述べられている見解は、筆者個人としてのものであり、筆者が所属する組織の見解とは必ずしも一致しない。

- 2 本稿で言及した道徳発達性理論や道徳的論点検査を扱う論文でも、倫理と道徳に関して明確に区分せず、職業倫理を論考しており、両者を同義とする暗黙の前提が立てられている。
- 3 評価目的としては、アカウントビリティ、業務改善、知識の創出が挙げられる (Patton 1997, p.41) が、日常的な評価業務では、アカウントビリティと業務改善が評価目的として設定されることが多いため、本稿ではこの2つの目的に焦点を置き論考した。
- 4 2015年9月7日の株式会社 東芝のプレスリリースに基づく、不適切会計処理により過年度決算の修正が必要となり、2008年度から2014年度の期間で累計2000億円を超える損失が計上された (株式会社 東芝 2015)。
- 5 米国においてケース・メソッドに利用される教材としては、Mintz (1997)、Mintz & Morris (2014) が挙げられる。
- 6 他の意思決定プロセスの例として、Reisch & Seese (2005) による5段階法がある。また、Mintzも現在は7段階法を発展させた10段階からなる意思決定プロセスを提案している (Mintz & Morris 2014, p.70)。

## 参考文献

- 梅津光広 (2002) 『ビジネスの倫理学』、丸善出版
- 岡崎一浩 (2012) 「第2章 倫理研究の基本的視角」、藤沼亜起『会計プロフェッションの職業倫理』、同文館書店:13-46
- 株式会社 東芝 (2015) 『過年度決算の修正、2014年度決算の概要及び第176期有価証券報告書の提出並びに再発防止策の骨子等についてのお知らせ』、[http://www.toshiba.co.jp/about/ir/jp/news/20150907\\_1.pdf](http://www.toshiba.co.jp/about/ir/jp/news/20150907_1.pdf) (2016/3/3アクセス)
- 株式会社 東芝 第三者委員会 (2015) 『調査報告書』、[www11.toshiba.co.jp/about/ir/jp/news/20150721\\_1.pdf](http://www11.toshiba.co.jp/about/ir/jp/news/20150721_1.pdf) (2016/2/29アクセス)
- 小林信行・一寸木英多良 (2010) 「『評価者倫理・スタンダード策定分科会』中間報告－基本原則 (Principles) 案について－」、『日本評価学会第11回全国大会発表要旨集録』:147-154
- 小林信行 (2014) 「証券アナリストと評価者による倫理向上への取り組み－職業倫理の比較研究－」、『日本評価研究』14 (2) :29-41

- 杉本泰治・高城重厚 (2008) 『第四版 大学講義 技術者の倫理 入門』、丸善出版
- 日本経済新聞社 (1992) 『会計用語辞典 (第4刷)』、日本経済新聞社
- 日本公認会計士協会 (2009) 『監査・保証実務委員会研究報告第20号「公認会計士等が行う保証業務等に関する研究報告」』、[http://www.hp.jicpa.or.jp/specialized\\_field/files/2-8-20-2-20090703.pdf](http://www.hp.jicpa.or.jp/specialized_field/files/2-8-20-2-20090703.pdf) (2016/3/22 アクセス)
- 八田進二・町田祥弘 (2003) 「会計プロフェッションの自主規制における職業倫理の位置づけ」、『企業会計』、55(3):39-48
- 八田進二 (2011) 「会計専門職大学院における倫理教育の現状と課題」『会計プロフェッション』2011年3月 第6号:125-134
- 原田保秀・矢部孝太郎 (2011) 「会計教育における倫理規程の意義－コールバーグ理論に基づく実証研究－」、『四天王寺大学紀要』、第51号 (2011年3月) :93-106
- 原田保秀 (2012) 『会計倫理の視座』、千倉書房
- 星野勉・三嶋輝夫・関根清三 (1997) 『倫理思想辞典』、山川出版
- Boyce, G. (2014). Ethics and accounting education. In Wilson, R.M.S. (Eds.), *The Routledge Companion to Accounting Education*. Abington, Oxon: Routledge, 533-557.
- Fitzpatrick, J. L. (1999). Ethics in Discipline and Professions Related to Evaluation. In Fitzpatrick, J. L. and Morris, M. (Eds.), *Current and Emerging Ethical Challenges in Evaluation*. San Francisco, CA: Jossey-Bass, 5-14.
- IFAC (2014). *International Education Standard (IES) 4: Initial Professional Development – Professional Values, Ethics, and Attitudes (Revised)*. New York, NY: IFAC.
- IFAC (2015). *Handbook of the Code of Ethics for Professional Accountants (2015 Edition)*. New York, NY: IFAC.
- IIA (2001). *Independence and Objectivity: A Framework for Internal Auditors*. Altamonte Springs, FL: IIA. (松井隆幸訳「独立性と客観性－内部監査人のためのフレームワーク」、日本内部監査協会) .
- IIA (2012). *International Standards for the Professional Practice of Internal Auditing*. Altamonte Springs, FL: IIA. (檜田信男監訳『内部監査の専門職的実施の国際基準』、日本内部監査協会)
- Langenderfer, H.Q. and Rockness, J.W. (1989). Integrating Ethics into the Accounting Curriculum: Issues, Problems, and Solutions. *Issues in Accounting Education*, Vol.4 No.1, 58-69.
- Mintz, S.M. (1997). *Cases in Accounting Ethics & Professionalism (3<sup>rd</sup> Edition)*. New York, NY: Irwin/McGraw Hill.
- Mintz, S.M. and Morris, R. E. (2014). *Ethical Obligations and Decision Making in Accounting (3<sup>rd</sup> Edition)*. New York, NY: McGraw Hill Education.
- Newman, D.L. and Brown, R.D. (1996). *Applied Ethics for Program Evaluation*. Thousand Oaks, CA: SAGE Publications
- Patton, M.Q. (1997). *Utilization-Focused Evaluation (3<sup>rd</sup> Edition)*. SAGE Publications. (大森彌監修、山本泰・長尾眞文編(2001)『実用重視の事業評価入門』、清水弘文堂書房)
- Ponemon, L.A. (1993). Can Ethics be taught in accounting?. *Journal of Accounting Education*, Vol.11, 185-209.
- Reisch, J.T. and Seese, L.P. (2005). Compliance with Title IX at Kingston State University: A Case Study on Cost Allocation and Ethical Decision Making. *Issues in Accounting Education*, Vol. 20, No. 1, 81-97.
- Weiss, C.H. (1997). *Evaluation (2<sup>nd</sup> Edition)*. Prentice Hall (佐々木亮監修、前川美湖・池田満監訳 (2014)『入門 評価学 政策・プログラム研究の方法』、日本評論社)
- Welton, R.E. and Guffey, D.M. (2009). Transitory or Persistent?: The Effects of Classroom Ethics Interventions : A Longitudinal Study. *Accounting Education: An International Journal*, 18(3), 273-289.

(2016.8.10 受理)

## **Use of Case Method in Ethics Education for Evaluators: Implications from the Teaching Method in Ethics Education for Auditors**

Nobuyuki Kobayashi

OPMAC Corporation  
kobayashi@opmac.co.jp

### **Abstract**

This paper examined teaching method in the ethics education of auditors and withdrew implications for that of evaluators. The ethical education of the auditors emphasized the use of a participatory approach based on case method and applied several systematic decision-making models to the teaching method. Also, an analytical framework to understand ethical issues and methods to compare choices for ethical decision-making had been developed.

It was notable that evaluators were often given two different evaluation purposes, accountability and improvement, under the same assignment. As a result, evaluators dealt with different relationships with clients on each occasion. Evaluators were required to acquire the abilities to interpret their situations and to solve line-drawing problems and conflict problems. Case method is an appropriate technique to satisfy their educational needs. The ethics education based on the case method will establish a path to link ethical practices to institutional arrangement. Thus, it is expected to withdraw meaningful feedbacks on the Guidelines for the Ethical Conduct of Evaluations, which the Japan Evaluation Society approved in 2012, from participants in the ethics education for evaluators.

### **Keywords**

Evaluation Ethics, Professional Ethics, Auditors, Ethics Education, Case Method



**日本評価学会春季第13回全国大会**  
**「SDGsの幕開けにあたり評価を考える」**  
**開催の報告とお礼**

2016年5月28日開催（於：JICA横浜）の日本評価学会春季第13回全国大会には129名の方々にご出席いただきました。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、皆様にとっ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。

実行委員長 和田義郎（国際協力機構）  
 プログラム委員長 牟田博光（東京工業大学）

2016年5月28日（土）

9：30 - 10：00	受 付			
午前の部 10：00 - 12：00	共通論題1 政策決定のための系 統レビューにおける 国際的潮流 (湊直信) かもめ	共通論題2 ODA評価の現状分析と その展望 (山谷清志) いちよう	自由論題1 自治体評価 (西出順郎) やまゆり	
お昼休み 12：00 - 13：15	お昼休み (12：05 - 13：00 理事会)			
午後の部 I 13：15 - 15：15  ※自由論題2 14：15 - 15：15	共通論題3 「持続可能な開発のため の教育国連10年：2005-14」 の評価：特に日本における その成果とプロセスについて (廣野良吉) かもめ	共通論題4 「行政評価士」の役割 とそのカリキュラム (山谷清志) いちよう	共通論題5 テストに拠らない学 校評価の試み－体験 学習の評価を中心に (石田健一) やまゆり	自由論題2 大学評価 (牟田博光) 会議室1
午後の部 II 15：30 - 17：30  ※自由論題3・4 15：30 - 18：00	共通論題6 政策形成のためのエ ビデンス活用 (正木朋也) かもめ	共通論題7 米国におけるGPRAMA と日本への示唆 (南島和久) いちよう	自由論題3 評価の国際潮流と評価 手法 (源由理子) やまゆり	自由論題4 教育評価 (橋本昭彦) 会議室1

( ) は座長です。

日本評価学会春季第13回全国大会  
「SDGsの幕開けにあたり評価を考える」プログラム詳細

5月28日(土) 受付 9:30 - 10:00

5月28日(土) 午前の部 10:00 - 12:00

共通論題1 政策決定のための系統レビューにおける国際的潮流				かもめ
		座長・モデレーター	湊 直信	国際大学
		コメンテーター	米原 あき	東洋大学
			佐分利 応貴	総務省
			APEA参加者	(skypeによる)
K1-1	Enabling the use of research: relevant methods of primary research and research synthesis and the science of research use		Gough, David	EPPI-Centre, University College London
K1-2	Evidence-based decision for policymaking in Japan		正木 朋也	国際協力機構
共通論題2 ODA評価の現状分析とその展望				いちよう
		座長	山谷 清志	同志社大学
		コメンテーター	南島 和久	新潟大学
K2-1	ODAにおけるPDCAサイクルの評価結果と今後の課題		村岡 敬一	外務省
K2-2	JICAの事後評価-外部有識者によるレビュー結果と対応-		鳴谷 哲	国際協力機構
K2-3	ODA評価の現状と課題-第三者評価は日本の官僚文化になじまない?-		佐藤 寛	アジア経済研究所
自由論題1 自治体評価				やまゆり
		座長	西出 順郎	岩手県立大学
J1-1	参加型評価を通じた市民教育の可能性		橋本 圭多	同志社大学
J1-2	バランス・スコアカードを用いたSDGsのための事業モデルの評価に関する検討	○	加藤 郁夫	(株)国際開発センター
J1-3	地方議会議員の質問に対する重み付け設定の試行		氏川 恵次	横浜国立大学
J1-4	指定管理者制度導入施設の管理運営に関する第三者評価の実態-岩手県盛岡市の取り組みを事例に-	○	熊谷 智義	特定非営利活動法人政策21
			岩淵 公二	特定非営利活動法人政策21

5月28日(土) お昼休み 12:00 - 13:15 (12:05 - 13:00 理事会)

5月28日(土) 午後の部I 13:15 - 15:15 (※自由論題2 14:15 - 15:15)

共通論題3 「持続可能な開発のための教育国連10年：2005-14」の評価：特に日本におけるその成果とプロセスについて				かもめ
		座長	廣野 良吉	成蹊大学
K3-1	我が国におけるESD活動の特徴と課題：その総合的評価		阿部 治	立教大学/日本環境教育学会
K3-2	ESD推進施策の評価		岩本 渉	千葉大学
K3-3	企業の立場からみたDESとCSR ～進化の10年を振り返って～		関正 雄	損害保険ジャパン日本興亜(株)
K3-4	市民社会からの挑戦—ESD推進12年間の軌跡		池田 満之	持続可能な開発のための教育推進会議(ESD-J) / 岡山ユネスコ協会
共通論題4 「行政評価士」の役割とそのカリキュラム				いちよう
		座長	山谷 清志	同志社大学
		コメンテーター	小野 達也	鳥取大学
			源 由理子	明治大学
K4-1	行政評価者と評価士		西出 順郎	岩手県立大学
K4-2	「行政評価士」資格の創設をめぐる諸論点		佐藤 徹	高崎経済大学
K4-3	「評価人材」に対するニーズと「行政評価士」		田中 啓	静岡文化芸術大学
共通論題5 テストに抛らない学校評価の試み—体験学習の評価を中心に				やまゆり
		座長・司会	石田 健一	東京大学
K5-1	テストによらない学校評価の試み—体験学習の評価を中心に企画趣旨の説明		石田 健一	東京大学
K5-2	体験学習の作文分析	○	大河原 尚	大東文化大学
			西村 邦雄	東洋学園大学
			石田 楓軒	Owl Medical Service
			石田 健一	東京大学
			小澤 伊久美	国際基督教大学
			石田 洋子	広島大学
K5-3	インタビュー、フォーカスグループ・ディスカッション	○	石田 健一	東京大学
			大河原 尚	大東文化大学
			石田 楓軒	Owl Medical Service
			西村 邦雄	東洋学園大学
			石井 徹弥	コーエイ総合研究所
			伊藤 美保	インテムコンサルティング
K5-4	作文分析による評価	○	西村 邦雄	東洋学園大学
			石田 健一	東京大学
			大河原 尚	大東文化大学
			石田 楓軒	Owl Medical Service
			小澤 伊久美	国際基督教大学
			石田 洋子	広島大学

K5-5	学校現場への提言	○	石田 楓軒 大河原 尚 石田 健一 西村 邦雄	Owl Medical Service 大東文化大学 東京大学 東洋学園大学
K5-6	第三者評価に対する現場教員の評価－アンケート結果から－	○	大河原 尚 小澤 伊久美 西村 邦雄 石田 健一 石田 楓軒 石井 徹弥	大東文化大学 国際基督教大学 東洋学園大学 東京大学 Owl Medical Service コーエイ総合研究所
K5-7	総合討論のための話題提供：学校評価士の共同作業による学校評価、その課題と可能性	○	小澤 伊久美 石田 健一	国際基督教大学 東京大学
<b>自由論題2 大学評価</b> (14:15 - 15:15)				<b>会議室 1</b>
		<b>座長 牟田 博光 (一財) 国際開発センター</b>		
J2-1	大学評価支援へ向けた指標設定のチェックリストの開発	○	洪井 進 田中 弥生	(独) 大学改革支援・学位授与機構 (独) 大学改革支援・学位授与機構
J2-2	教育と医療の第三者評価の比較検討－大学機関別認証評価と病院機能評価を中心に－		高池 宣彦	筑波大学大学院大学院生 (博士後期課程)
<b>5月28日(土) 午後の部Ⅱ 15:30 - 17:30 (※自由論題3・4 15:30 - 18:00)</b>				
<b>共通論題6 政策形成のためのエビデンス活用</b>				<b>かもめ</b>
		<b>座長 正木 朋也 国際協力機構／北里大学</b>		
K6-1	ODAプロジェクトに関するエビデンス形成の一事例：インドネシア・スラウェシ島の稲作振興プロジェクトから		高橋 和志	上智大学
K6-2	エビデンスに基づく犯罪対策：犯罪者治療プログラムの開発と評価		原田 隆之	目白大学
K6-3	イギリスにおけるEvidence-Based Policyのエコシステムと日本への示唆		小林 庸平	三菱UFJリサーチ&コンサルティング
K6-4	「エビデンスの活用」を日本の政策形成過程に導入するための自治体での試み		家子 直幸	三菱UFJリサーチ&コンサルティング
<b>共通論題7 米国におけるGPRAMAと日本への示唆</b>				<b>いちよう</b>
		<b>座長・コメンテーター 南島 和久 新潟大学 司会 岡本 義朗 新日本有限責任監査法人</b>		
K7-1	米国連邦政府GPRAMAの成立過程及びそれに対するGAOの関与		三浦 雅央	新日本有限責任監査法人
K7-2	業績達成度評価の進展：米国連邦政府のGPRAMAの取組を参考に		左近 靖博	新日本有限責任監査法人
K7-3	スタット・モデルの背景にあるデータ活用型行政経営の進展		高木 麻美	新日本有限責任監査法人／国際大学
<b>自由論題3 評価の国際潮流と評価手法</b> (15:30 - 18:00)				<b>やまゆり</b>
		<b>座長 源 由理子 明治大学</b>		
J3-1	グローバル評価アジェンダと日本の市民社会における評価キャパシティの強化に向けて	○	黒田 かをり 今田 克司 長谷川 雅子 高木 晶弘 今田 克司	(一財) CSOネットワーク (一財) CSOネットワーク (一財) CSOネットワーク (一財) CSOネットワーク (一財) CSOネットワーク
J3-2	発展型評価～概念と可能性		林 薫	文教大学
J3-3	SDGsの評価：第7目標「エネルギーと持続性の指標」		中嶋 学	ニューヨーク州立大学 (博士課程)
J3-4	協働型ネットワークにおける業績情報の拡散型探索と深耕型探索のための活用：試験			
J3-5	実存論的評価手法を用いた気候変動適応プロジェクト事後評価のメタ分析	○	宮口 貴彰 Uitto, Juha	立命館大学 GEF独立評価局
<b>自由論題4 教育評価</b> (15:30 - 18:00)				<b>会議室 1</b>
		<b>座長 橋本 昭彦 国立教育政策研究所</b>		
J4-1	教育協力に関する政策レベル評価3件からの学び		石田 洋子	広島大学
J4-2	国際教育協力プロジェクトは公正的吗		田中 紳一郎	国際協力機構
J4-3	Improving Educational outcomes at Primary Schools in Lao: Anevaluation using a randomised controlled trial	○	吉川 香葉子 神谷 祐介 野村 真利香 荻野 妃那	大阪大学大学院生 (博士後期課程) 龍谷大学 国立保健医療科学院 四日市看護医療大学
J4-4	修学パターンによるミャンマー初等教育における進級制度の検討	○	關谷 武司 吉田 夏帆	関西学院大学 関西学院大学大学院生 (博士前期課程)
J4-5	ミャンマー国基礎教育学校の教育条件が学力試験結果に及ぼす影響に関する分析		牟田 博光	国際開発センター

○印は、共同研究の代表者

## 共通論題セッション報告

### 共通論題セッション1 「政策決定のための系統レビューにおける国際的潮流」

座長・モデレーター 湊直信（国際大学）  
 コメンテーター 米原あき（東洋大学）  
 佐分利応貴（総務省）

国際開発の分野では、SDGs推進等の政策決定のため、ゴールの達成度を示すための具体的指標を用いたアプローチが求められている。その際、介入効果のエビデンスを用いることは有益であり、これは世界的な潮流となっている。保健医療、社会・行動・教育分野のほか、国際開発の領域においても、エビデンスの系統レビュー（Systematic Review）が始まっている。英国の公共政策におけるEvidence Based Practiceの普及とその経験、およびエビデンスの系統レビューと政策形成につながるプロセス構築の黎明期にある日本の現状について共有すると共に、多様な視点からエビデンスと政策形成の関係について論じた。尚、本セッションは英語で実施され、APEA（アジア太平洋評価協会）との連携プログラムとして、APEAの会員もSkypeで参加した。

David Gough氏（ロンドン大学）による研究報告1では、研究の有効活用について、その目的に合致した手法や、研究における知見の政策決定や実践への効果的な活用プロセスについて論じられた。調査結果と活用のシステムに関して、研究結果とその活用は統合されたシステムの一部であるために、裨益者の意向をも含むシステムの構成要素が相互に関係し合っていることを理解する必要がある。一次研究、二次研究（マッピング、統合）、メタ評価、およびエビデンス情報に基づく政策形成において研究結果をつかうための研究の重要性も指摘された。

正木朋也氏（国際協力機構）による研究報告2では、日本におけるエビデンスに基づく政策決定に関する課題について、エビデンスに基づく医療の歴史的背景を参考にしながら論じた。第一に、日本のエビデンスに基づく政策決定の世界的潮流と日本の現状、第二に、グローバルに比較した場合のエビデンスに基づく政策決定のしくみづくりの遅れとその背景分析、第三に、因果関係の証明に最小限必要な要因にも言及して、現状の課題と今後の方策について論じた。

以上の報告に対して、2名のコメンテーター及びSkypeで参加したアジア太平洋評価協会（APEA）会員から多様な視点でコメントを得た。会場の参加者からも多くの質問、コメントが寄せられた。医療、環境、教育、国際開発等の分野ごとにエビデンス活用ニーズと産出状況および適用の難易度が異なることや、調査研究と政策決定者との間の橋渡しを行うこととその動機付けの重要性等が論じられた。

### 共通論題セッション2 「ODA評価の現状分析とその展望」

座長 山谷清志（同志社大学）  
 コメンテーター 南島和久（新潟大学）

2015年は日本のODA大綱が開発協力大綱へと再編成され、また国連MDGs（Millennium Development Goals）が終了してSDGs（Sustainable Development Goals）が新しく登場した年であったが、ODA評価についても大きな議論がなされていた年でもある。しかし、このことについて、おそらく一般国民はよく承知

していないと思われる。この視点から、本セッションは開始された。

第1報告者の村岡敬一氏（外務省）のテーマは「ODAにおけるPDCAサイクルの評価結果と今後の課題」である。2015年2月に閣議決定された開発協力大綱は、ODA評価結果を政策決定過程や事業実施過程に適切にフィードバックするよう求めている。外務省ではこれを受けて、ODAのマネジメント改善と国民への説明責任強化を図るべく第三者評価「ODAにおけるPDCAサイクルの評価」を行った。外務省は、①説明責任のさらなる確保、②マネジメント改善のために評価結果を政策企画・実施部門へのフィードバックする機能を強化、あわせて対話に基づく戦略的・計画的な評価案件の選定を進める、③JICAとの連携を深め、JICA協力プログラム評価および事業評価結果の活用を図りつつ、目標体系図を意識した政策・施策・事業の指標設定の工夫を促すこととした。その際には2030アジェンダで合意されたSDGs評価の国際的議論に十分留意していく。以上の三つを確認した。

第2報告は鳴谷哲氏（国際協力機構）の「JICAの事後評価－外部有識者によるレビュー結果と対応－」である。2015年6月の行政事業レビューにおいてJICA評価事業の点検が行われ、評価における透明性の確保等の提言がなされた。この提言を踏まえJICAは説明責任の観点から、外部有識者による事後評価制度のレビューを行なった。2016年2月の事業評価外部有識者委員会は、3つの提言をした。①JICAの事業評価の評価活動が非常に細かく丁寧に行われているものの、一般国民にもさらにわかりやすいように工夫する必要がある、②JICAの事業評価は今後も外部評価者を確保し高い質を維持する必要がある、③有識者の知見を活かしたメタ「分析」によって実務者にも有用な分析を行うべき、である。これらは、JICA評価部が考える方向性とも合致しており、今後実施を予定している。また、評価機能の学習面での強化は同委員会でも累次提言されており、横断・詳細分析、定量分析、手法開発（プロセスの評価、SDGsへの貢献）、フィードバック機能をさらに強化する予定であると報告された。

第3報告者の佐藤寛氏（アジア経済研究所）は「ODA評価の現状と課題－第三者評価は日本の官僚文化になじまない？」と題して報告された。日本のODA研究を代表する国際開発学会は、過去に何度か評価をめぐる研究テーマをとりあげ、あるいは『国際開発研究』にも関連論文が掲載されてきた。ただ評価の研究を主とする日本評価学会とはODA評価の取り上げ方に微妙な違いが見られる。この点について、佐藤氏はこれまでの経験と知見から評価学会に対して有益な提言をされた。すなわち外務省のODA評価は第三者評価の形を取ることで客観性、説明責任を担保しようとしている。そのため報告書作成に先立って、評価受注者（コンサルタント会社）と外務省の担当部局との間のやり取りが行われるが、その目的は「事実経過の確認」であると説明される。しかし実際には、組織防衛的な観点から報告書の書きぶりの修正が求められることが少なくない。これは「第三者評価」の趣旨に反するのではないか。このような有益な提言であった。

コメンテーターの南島和久氏（新潟大学）はODAそのものの研究・実践と、ODA「評価」の研究・実践との違いを、評価の専門的視点、あるいは行政学的な視点からコメントされ、上記の問題提起を総合的にまとめた。

### 共通論題セッション3 「「持続可能な開発のための教育国連10年：2005-14」の評価：特に日本におけるその成果とプロセスについて」

座長 廣野 良吉（成蹊大学／（公財）地球環境戦略研究機関）

地球とその生態系は我々の故郷である。「母なる地球」は、世界の大多数の国々や地域で共通の表現、合言葉となっている。この母なる地球に住む我々が、経済・社会・環境・文化の基本的ニーズの正しいバランスを達成しなければ、自然の一部である人類と、其処に生息するあらゆる動植物に未来はない。しか

し、現実の世界では、目先の経済成長や快適な生活を求める中で、企業も個人も地球が数十億年の間に育んできた自然資源を取奪し、大自然を破壊し、政府は「国益」優先の旗印の下で宗教的対立や人種的偏見、国家間の不信・衝突を招き、国内外難民を続出させ、子ども・女性・障害者等社会的弱者の基本的人権を蹂躪するという地球社会の平和と安定に相反する行動が各地域で見られる。このような国内外情勢の下で国連は、2000年に「新世紀開発目標（MDGs）、2001-15年」を導入し、特に開発途上国における貧困削減、教育・保健・衛生の改善、環境保全等への自助努力と国際協力を促した。さらに、教育・学習の重要性に鑑み、日本政府は2002年「持続可能な開発世界首脳会議（WSSD）」で、NGOとの共同提案で、「持続可能な開発のための教育・国連の10年（UNDESD）、2005-14」を提唱し、国連総会での採択を主導した。一昨年UNDESDが、昨年はMDGsが最終年を迎えたが、これらの国際的連帯行動に引き続き、昨年9月に国連総会での首脳会議で「持続可能な開発目標（SDGs）2016-30年」が採択され、あらゆる地域社会、国、地球的規模で「持続可能な社会」の構築こそが、現代に住む我々人間の共同責務と宣言された。

筆者は、このポスト2015開発アジェンダであるSDGsは、先のMDGsとUNDESDの理念・目標を引き継いだものという国際社会の共通認識の下で、我が国のみならず、国際社会がSDGsを成功裏に遂行するためにも、日本評価学会の協力を得て、特に我が国に於けるUNDESDの成果とプロセスを客観的に評価することが時機を得たものと考えた。そのために、我が国のみならず、国連を含む国際社会で長年、ESD活動に幅広く積極的に従事してこられた数少ない貴重な専門家に登壇を願い、我が国におけるUNDESD活動の全貌と政府、地方自治体、企業、NGO等各主体の活動成果をその内容とプロセスについて、率直に反省・評価していただくことにした。幸いに全員がご快諾くださり、限られた時間内であったが、自己批判を含めて、活発な議論が展開された。

本セッションの趣旨・目的についての筆者による簡潔な報告に引き続いて、最初の報告は、UNDESDの国際的な旗振り役である国連教育科学文化専門機関（UNESCO）でUNDESD委員会に参与し、UNDESDの実施機関中に日本環境教育学会会長を務め、尚且つ、ESD活動を推進するためのNGOであるESD-Jの代表理事を務めた立教大学教授・同大学ESD研究センター長である阿部治氏によるものであった。その論点は、我が国におけるESD活動の特徴の整理と課題であった。前者については、①ESDの基本的概念は市民社会組織による発案、②政府、地方自治体、企業、学界、NGO等あらゆる活動主体を巻き込んだオールジャパン的な活動であり、そのネットワークづくり、③学校教育だけでなく、地域に根差した社会教育のネットワークづくり、④ESD教材づくり、⑤ESDコーディネーターの育成と仕組みづくり、⑥UNDESDの各節目での反省と政策提言づくり、⑦ESD 10年後の体制づくり、⑧ESDアジアネットワークづくりであった。我が国のESD活動の課題としては、①ESDでは、環境教育が中心、②SDの推進に従事している他のNGOとの接点・連携が希薄、③文科省と環境省との連携活動は良好だったが、政府内に設置されたのは、単なる情報交換に終わった各省庁間連絡報告機関であり、ESD活動を統合的に推進する本部機能はなし、④国際的なESD政策立案・実施プロセスへの参画は限定的、⑤義務初等・中等教育課程におけるSDに資する開発教育・学習方法の主流化は困難、⑥企業との連携によるESD活動が低調等の課題が指摘された。最後に、2014年以降は、ユネスコ総会で採択されたグローバル・アクション・プログラム（GAP）の推進とESD活動のさらなる強化のために、ESD活動支援センターの設置が決まったことは歓迎するが、その具体的活動は今後の検討課題となっている。

第2番目の登壇者である千葉大学エクゼクティブ・アドバイザーを務めている岩本渉氏は、ユネスコ本部にあってUNDESDプログラムに参与した後に、文科省国際統括官として我が国のESD活動計画の策定と全国都道府県、市町村におけるESD活動の実施状況を監察する立場から、我が国の学校教育課程におけるESD活動について、数多くの示唆に富む報告があった。我が国に於けるESD活動の特徴と課題については、阿部先生の報告要旨に概ねの賛同を示したが、特に世界の学校教育の潮流に言及し、我が国に於けるESD推進施策の特徴として、①各省庁間連絡会議の設置、②日本ユネスコ国内委員会による日本各地域におけるユネスコ学校の普及、③幼稚園から高等学校までの学習指導要領におけるESDへの積極的な取り組

み、④文科省によるESDの国際的展開について詳細な報告があった。中でも、ユネスコ加盟校の急増（2000年の15校から2015年の939校）は顕著であり、現在では世界のユネスコスクール10%を占めている。ESDの導入によって、2000年の「万人のための教育」目標を高く掲げ、教育の質の改善を唄ったダカール宣言が、持続可能な社会の担い手を教育するという指針を与えられたことを高く評価していた。しかし、報告者が指摘しているように、①ESDの内容についての「教育方法の固定化は避け、学習者中心のESD、学習者の変容を狙う」ことの諸困難の打破が政府、地方自治体、教育委員会、現場の先生等に強く求められるという警告は注目に値する。また、②県教育委員会レベルのESDへの関心の低度と③グローバル化時代に不可欠な「問題解決能力やコミュニケーション能力等の育成が求められる大学こそESD」を強化する必要があるという岩本氏の指摘は、本セッション参加者の共感を呼んだことを指摘しておきたい。最後に④「異なるネットワークが共同する重層的なネットワークを構築する必要があるだろう」という報告者の指摘は、地域社会でも、中央でも、どこでも求められているというのが、筆者の考えでもある。

第三番目の報告は、長年損害保険ジャパン（後に損害保険ジャパン日本興亜株式会社）にて、CSRを担当し、現在明治大学にて特任教授をなさっている関正雄氏によるものであった。最初に、①企業による社会的責任（CSR）に関する伝統的な定義を披露し、この概念規定に近年変化が生まれてきていることを指摘し、企業の社会的責任の国際基準であるISO26000の策定過程に参加された国際的経験から、②持続可能な企業経営やかかる社会の構築への関心が我が国でも高まっており、その認識と行動の必要性が叫ばれるようになったとのことである。2000年以来、国連グローバル・コンパクト企業連合に加盟する欧米諸国の多国籍企業は数千社にのぼり、グローバル・コンパクト企業日本連合によると、我が国の大手企業147社が加盟している。我が国では「企業の社会的責任」というと、古くから企業による地域社会の諸活動や大震災や台風による被害地域への寄付行為や企業財団による環境保全や社会的弱者への救済活動を意味するとの理解であったが、③近年本来の企業活動（コアビジネス）を通じて、環境保全や持続可能な社会の構築等社会的公共財への配慮を企業の社会的責任とする企業が生まれてきていることは歓迎すべきことである。そのためには、④経営陣トップだけでなく、社員一同がその企業内訓練を通じ、また現場での生産活動やその部品調達活動等を通じて、ESDを主流化することが期待されている。そのためには、⑤政府、地方自治体による法規制だけでは不十分であり、社内からの声が不可欠であり、NGO等との協働による社員教育の徹底が望まれる。関氏によれば、特に⑥ISO26000には、ESDは「必要な価値観を根付かせ、積極的な行動をし、新しい方向性を定めることで、社会的責任に関する課題に取り組む力を人々に与える。」しかし、現実の社会では、⑦昨年、ESD賛同企業が採択した「企業によるESD宣言」では、「企業内でのESDと社会に向けたESD支援活動、この二つの切り口で積極的な行動を呼びかけている」が、阿部氏の指摘にあるように、NGOからも企業からも、協働作業への呼びかけはまだ少ないのが現状である。双方による一層の努力が不可欠であろう。

最後の報告者は、2002年のWSSSDで岡山市を市長と共に代表し、岡山ユネスコ支部で長年活動し、ESD-Jの設立と共に、その副代表（政策提言担当）を兼ねている池田満之氏である。岡山市は、国連大学が提唱した地域ESD専門家集団（RCE）の我が国に於けるトップバッターである。地方自治体、小学校から大学まであらゆる教育機関、市民会館、図書館、博物館、美術館、青少年交流団体、宗教団体等を包含してESD活動を推進している国連大学認証地域であり、我が国では、岡山市以外に、北九州市、兵庫神戸、中部、横浜、仙台市広域連合がRCEとして認証されている。産業界とNGOが中心となって各主体と連携してきた特徴を持つ北九州市と並んで、岡山は公民館を活動の場としてあらゆる主体とのパートナーシップを組んで、我が国でもESD活動を最も意欲的に推進し、多くの成果をもたらしてきたモデルケースである。

その岡山市での貴重な経験に基づいて、全国的な中間市民社会組織の立場から長年「ESD推進の牽引役を担ってきた」池田氏の報告は多くの示唆に富む教訓・警句を学ぶ機会であった。

報告者本人が指摘するように、また、「ESDレポート第35号」や「ESD推進12年間の軌跡（ESD活動報告書2003-15）」が強調しているように、ESD-Jの成果は、①地域、全国的な各活動主体とのつながりの強化、

②地域、学校、部門組織等各活動現場でのモデルづくり、③ESD政策提言と推進のための仕組みづくり、④UNDESD後のESD体制づくり、⑤アジアでのESDネットワークの形成であった。これらの成果については、ESD関与者の間に異論はないであろう。しかし、池田氏が自ら牽引してきたESD-Jの広がりにも拘わらず、①2015年から始まったユネスコ学習都市グローバルネットワーク（GNLC）への参加登録都市は、我が国では現時点で岡山市が唯一であって、他の諸都市の登録が未定であるだけでなく、②ESD-Jは「出る釘は打たれる」、「縦割り社会」という我が国の伝統的な社会通念を恐れて、あらゆる個人や団体と連携して、「創造的破壊」をしなかったため、SDを希求する多くの個人、団体の信頼を得ることが出来なかった」という筆者の指摘を再確認し、「国レベルのESD活動支援センターの運営団体として、創造的破壊にどう踏み出せるかは、今後のESDの展望にもかかわる大きな課題でもある」という2つの重大な反省点は、さすが市民社会派を自他ともに認めている池田氏の発言であり、注目に値する。

セッションでの4人の登壇者の報告後、時間的制約もあって、総ての質疑やコメントを受けることはできなかったが、総ての報告者に共通した自問自答の「何のためのつながりか、その成果の評価は？」は、我が国に於ける今後のESD活動のあらゆる局面で、ESD-Jだけでなく、各地域社会においてESD活動を推進している各主体にとっても、常に厳格に検討すべき課題であろう。ESD活動の本来の受益者にとって必要な評価は、インプットに対するアウトプットよりも、アウトカムであるという筆者の本セッションの最終論考を再度強調したい。

気候変動の悪影響がますます深刻化している中で、我が国のみならず、世界のESD活動の今後に期待したいことは、自然資源の利用面では地球生態系が長期的に許容できる範囲内（再生可能原則）での生産・消費に満足する老若男女づくりであり、若者失業率が上昇し、所得格差がますます深刻化し、倫理的行動を喪失した企業経営者が続出し、立憲主義の逸脱に無批判な人々があらゆる所得・教育階層で多数化している世界にあって、包摂的・持続可能な社会の構築に不可欠な社会的平等・公正原則を厳守し、自助・共助・公助の原則に即して、互いに連携・協力意識・行動を強化する生徒・学生・社会人の養成である。特に国内にあっては社会的弱者が直面している諸々の課題の解決に真剣に対応し、対外的には後発開発途上国、内陸開発途上国、および小島嶼開発途上国に代表される脆弱な国々や紛争状態にある国々で多くの人々が直面している深刻な課題の解決にも共鳴・共感・協働する人づくりとその結果としての仕組みづくりこそ、ESD活動のアウトカムであってほしい。

## 共通論題セッション4 「行政評価士」の役割とそのカリキュラム

座長 山谷 清志（同志社大学）  
 コメンテーター 小野 達也（鳥取大学）  
 源 由理子（明治大学）

地方自治体の評価は、地方分権運動が盛り上がった1990年代末に「政策評価」として始まり、その後「行政評価」として普及した。背景には財政再建や地方行革、「夕張ショック」、平成の大合併をふまえた自治体ガバナンス健全化の意図があった。その後「行政評価」は守備範囲を拡大し、地方独立行政法人評価（大学と病院）、指定管理者制度の評価、施設評価、PFI事業評価など多様な専門分野でも使われ、重疊的に評価が存在するようになる。「評価の氾濫」である。そのため、評価の関係者であっても全体像はおろか、個々の評価がいかなる役割を持ち、どのように自治体ガバナンス健全化に貢献し、あるいは市民に対してどんなアカウンタビリティを示そうとしているのか、正確に把握できない状況にある。そこで地方自治体関係者はさまざまな評価のスキルと考え方を身につけ、こうした困難な状況に取り組む努力を求められている。こうした時代の要請に応えるべく、専門評価士の資格として「行政評価士」を設けたらどう

かと考え、本セッションを提案した。もちろん行政評価士を内向きの行政管理テクニクの専門家ではなく、地方自治体改革を市民中心の立場で進めるプロ人材として養成したいと志向している。「評価は民主主義のリテラシー」だからである。

第1報告者の西出順郎氏（岩手県立大学）は、「行政評価者と評価士」と題して、行政機関における現行評価の枠組みから評価者の特徴を整理し、その文脈から専門「評価士」の姿を考えた。以下のとおりである。地方創世の話があるのでタイミングは良いが、その一方でコンセプトが曖昧な場合需要が無い恐れがある。学会としての資格提供を大前提とするならば、誰をターゲットにするか絞り込みは必須である。行政マンを対象とした場合、「事業評価士」の方が受けはいい。仮に「行政評価士」でもほしがる行政マンは組織内のアウトサイダーであり、評価に関心があり個人の生涯学習を目的とする人のみであろう。どうしても「『行政』評価士」とするなら、ターゲットは政治家かもしれない。したがって市町村議員を対象とするのがベストと考えられていた。

第2報告「『行政評価士』資格の創設をめぐる諸論点」で、佐藤徹氏（高崎経済大学）は、以下のように可能性を述べた。行政評価士の受講対象者としては、自治体の行政職員、住民、NPOメンバー、民間コンサルタント、地方議会議員、研究者など多様な面々が想定される。そこで、まずこれら立場の異なる者たちが、行政評価士としてどのような場面で活躍が期待されるかについて検討を行った。次に、中級レベルの資格としてどのようなスキルが求められるか、そしてそのためにはどのようなカリキュラム（内容、水準、方法等）が望ましいかについて考察を加え、最後に自治体行政の実態等を踏まえながら、行政評価士の創設に向け幾つか提案された。すなわち受講資格（初級評価士がない場合の特例措置）、受講料（遠方割引・団体割引など）、大学・大学院との連携である。

第3報告「『評価人材』に対するニーズと『行政評価士』において、田中啓氏（静岡文化芸術大学）は、近年、行政機関において評価が広く普及しているだけでなく、公共部門のさまざまな分野において、評価またはそれに類する活動が実施されるようになってきている状況を指摘された。しかし田中氏によれば、日本では「評価人材」が量的に不足しているために、公共部門で評価を必要とする場面に、適切に評価機能が提供されていないという問題が生じている。この状態を放置すれば、公共部門における意思決定の劣化を通じて、公共部門の全体的なパフォーマンス低下につながる懸念がある。そこで田中氏は、社会で必要とされる評価機能を提供するために、公共部門のどのような分野においてどのような評価人材が必要とされているのかを考察した上で、公共部門で求められる「評価人材」と「行政評価士」との関係性を再検討してはどうかと提案された。

この3つの報告に対して小野達也氏（鳥取大学）は、行政評価のニーズはある、求められているのは評価の設計と運用能力、評価ツールの開発・選択・適用能力、評価作業の質を維持する能力であると指摘された。そのためには評価士養成講座のいくつかの個別科目を分割・再編し、「初級+中級」編で実践的演習を組み合わせた講座が必要であるという。

もう一人の討論者、源由理子氏（明治大学）もまた、座学よりは演習形式で、ワークショップ的な講義の重要性を指摘された。なお、源氏は「行政評価士」を構想された方であり、その可能性、将来性に期待されると同時に、慎重な対応を求められていた。

## 共通論題セッション5 「テストに拠らない学校評価の試みー体験学習の評価を中心に」

座長・司会 石田 健一（東京大学）

複数名の学校評価士が、東京都北多摩地区の中学校で実施されている「リベラルアーツ」教育（体験学習を組み合わせ教育目標の達成を狙う教育活動）を評価した。評価の性格は試行的、予備的な実践である。

7件の予稿、4名の発表者により計6題の発表を行った。

石田健一氏（東京大学）（企画趣旨から評価結果）は評価結果の概要を紹介した。体験学習の中から2年生の都内巡り、並びに、農業体験を選択した。教員と評価者が設定した評価項目と評価基準を用いて作文と日誌を分析し、生徒の力と変化について評価した。生徒、教員、保護者等に対してインタビューおよびフォーカスグループ・ディスカッション（FGD）も併用した。

西村邦雄氏（東洋学園大学）（作文の分析）は、写真を豊富に交えながら上記の分析では捕獲できない生徒の潜在的な可能性についての考察を紹介した。石田楓軒氏（Owl Medical Service）（教員研修への活用）は、作文と教師によるフィードバックで生徒の思考を更に促すことについて、本評価での分析結果と自身の過去の体験（元中学校校長）を踏まえて説明した。大河原尚氏（大東文化大学）（教員へのアンケート実施結果）は、評価のプロセス並びに評価結果をプラスであると捉える教員が多いものの、ほぼ全ての教員が学校評価の担い手とは第三者であると考えていることを示した。石田健一氏（共同作業について）は、学校評価士が共同作業を行えば多様なスキルが強みとなるが、一方で持続性に難があること、生徒、教員、保護者にとって有用な評価手法の開発が今後も必要であることに触れた。総合討論の冒頭に石田健一氏が課題を3つに絞り、学校評価手法の比較、評価の有用性について提案を行った。

総合討論における指摘と論点は次の4つに集約される。

- ◇学力測定（ペーパーテストによる評価）と「リベラルアーツ」教育による能力育成を評価すること。
- ◇学校評価と教育評価の区別。
- ◇教員の少ない時間を前提とした効果的な評価の模索（教員の通常業務の効率化と併せて考える課題でもあり、総合討論の冒頭でも石田（健）が触れている）。
- ◇体験学習実施の前後で生徒の変化（変容）を測定すること。

以上の4点である。集まっていたいただいた会場の皆様に感謝申し上げます。

## 共通論題セッション6 「政策形成のためのエビデンス活用」

座長 正木 朋也（国際協力機構／北里大学）

午前中に開催された国際セッション「政策決定のための系統レビューにおける国際的潮流（Global trend in systematic review for policy making）」も踏まえ、日本の政策形成にエビデンスを活用する方策について参加者を交えて議論した。

第1報告者の高橋和志氏（上智大学）は、国際開発の現場からの実例として、傾向スコアマッチングの手法により利用可能な既存データを用いた分析を行い、稲作の増収効果に着目した開発支援により実際に米の収量増を達成してもなお、農家の総所得向上という目標にそぐわないケースが生じる可能性があることを指摘した。

第2報告者の原田隆之氏（目白大学）は、犯罪対策・防止の立場から犯罪者の更生には厳罰化による再犯抑制効果がないとの海外エビデンスを踏まえ、犯罪者を治療する立場から日本で実施中の更生プログラムの有効性をランダム化比較試験により証明するとともに、エビデンスを欠いた対策がまだまだ広く行われている現状を指摘した。

第3報告者の小林庸平氏（三菱UFJリサーチ&コンサルティング）は、イギリスにおける独自調査に基づき、ステークホルダーのマッピングを行い、政策形成のためのエコシステムの機能とその重要性について紹介し、継続的にエビデンスが創出・伝達・活用される仕組みの中で、官民協働組織であるWhat Works Centreの役割が重要であることを報告した。

第4報告者の家子直幸氏（三菱UFJリサーチ&コンサルティング）は、社会政策領域における実証的手法

の応用経験が日本ではほとんど行われてこなかった実情を踏まえ、現在自ら推進している地方自治体での取り組みの中間報告を行った。平易なことばを用いて変革のイメージを共有するなどの工夫と支援を通じて、住民協働のエビデンス創出活動と合意形成プロセスが進展していることが共有された。

政策形成にあたりエビデンスは本当に求められるかとのフロアからの質問に応じて、同じく参加者のひとりであるロンドン大学のDavid Gough教授から、イギリスで経験した問題やその克服プロセスについて英語にて応答いただく場面もあった。

昨今、国内においても、政策形成のためのエビデンスが求められることが増えているようであるが、今後、何から取り組み、何を变えてゆくべきか、また、エビデンス産出と活用のための合意形成と住民協働活動の支援の必要性など、関連するトピックについての活発な意見交換が行われた。

一つの方向性として、産官学共同のエビデンス創出・伝達・活用を推進する中心的役割を担う組織あるいは検討グループの設立、およびそのメンバーを中心とした実践的活動の共有の場の必要性が浮上した。今後、セッション参加者有志らも含め、相互に領域横断的な意見交換を継続することを確認して閉会した。

## 共通論題セッション7 「米国におけるGPRAMAと日本への示唆」

座長・コメンテーター 南島 和久 (新潟大学)

司会 岡本 義朗 (新日本有限責任監査法人)

米国連邦政府では、2010年にGPRAが抜本改正され、GPRAMA (Government Performance and Results Modernization Act, 2010) が制定された。GPRAMAへの法改正の重要な要素は、一言でいえば、「評価の重点化」にあった。そこで注視されていたのは「リーダーシップの参画」「客観的な業績情報の活用」であった。同法改正の経緯および制度運用の実態を明らかにするため、2014年から2015年にかけて会計検査院事務総長官房調査課は、『「アメリカの政府業績成果現代化法(GPRAMA)等の運用から見た我が国の政策評価の実施及び会計検査」に関する調査研究』(委託先：新日本有限責任監査法人、研究会座長・田邊国昭東京大学教授)と題する調査研究事業を実施した。本セッションでは、本研究会メンバーを中心とし、上記報告書およびその後の追加調査について評価学会に還元するとともに、日本の評価制度への示唆を得ることを目的として実施したものである。

第1報告者の三浦雅央氏(新日本有限責任監査法人)は、「米国連邦政府GPRAMAの成立過程及びそれに対するGAOの関与」と題する報告を行い、とくにGPRAMAの改正プロセスの概説および、そこに重要な役割を果たしたGAO(立法府付属機関、会計検査院)の各種のレポートの分析を行った。GAOはGPRAの実施過程を分析し、その課題を集約するうえで重要な役割を演じていたが、本報告からこれが具体的な姿として浮かび上がった。なお、とくに三浦氏が強調していたのがGAO、あるいはそのサポートを行ったシンクタンク・コンサルタント、あるいは政府内外の学識者・有識者などの「評価コミュニティ」であったことを付け加えておきたい。

第2報告者の左近靖博氏(新日本有限責任監査法人)は、「業績達成度評価の進展：米連邦政府のGPRAMAの取組を参考に」と題する報告を行い、意思決定に活用しうるための具体的手法についての概括的なレビューを行った。とくに、GPRAMA下での具体的な業績分析の手法として注目される「ドリルダウン分析」(深堀型評価)についての紹介がその重要な内容であったと思われる。米連邦政府における政策評価の重点化はPARTの反省を含めてHPPGsとして登場し、これがGPRAMA下ではAGPsおよびCAGPsへと再編されていった。これらはいずれも特定の評価対象に特化した分析を行うものであるが、その具体的な分析に際して、左近氏は、関係者のインタビューに基づき、アウトカムの改善に向けてベストプラクティスの活用のためのベンチマーク活用(主に州政府間)、アウトカムの詳細化を行う「ドリルダウン分析」、

同一目的の対応について比較・分析する「貢献度分析」などの分析手法が駆使されていることを紹介したものである。

第3報告者の高木麻美氏（新日本有限責任監査法人）は、「スタット・モデルの背景にあるデータ活用型行政経営の進展」と題する報告を行い、とくに評価の重点化について、「スタットモデル」と称されるものが米国内をはじめとして注目されている点を詳細に紹介した。「スタットモデル」とは、高木氏によれば、行政の優先課題を解決するため、幹部の強力なリーダーシップの下、①データの活用、②定期的かつ戦略的な会議、③関係者によるコミットメントの3つの要素を機能させることであるという。さらに高木報告は、スタットモデルが、組織において活用されることにより、データ重視の文化を促進し、多様な範囲に応用され得る可能性を秘めている点を強く指摘するものであった。重点型の評価を「スタットモデル」として紹介した点に高木報告の意義はあったといえるだろう。

これらの報告に対して会場からは多数の質問・意見が出されたが、とくに日本の政策評価制度の「評価疲れ」に対する処方箋として、米国連邦政府の経験および「スタットモデル」は有効な処方箋であるという点が、本セッションの示唆としては、重要であったと思われる。

## 自由論題セッション報告

### 自由論題セッション1 「自治体評価」

座長 西出 順郎（岩手県立大学）

本セッションでは以下の4つの報告がなされた。

第1報告者の橋本圭多氏（同志社大学）の報告では、高等教育を拠点とした市民教育の在り方の一環として形成的評価手法を用いた学生と地域住民との協働による市民教育の可能性について論じられた。

第2報告者の加藤郁夫氏（国際開発センター）及び氏川恵次氏（横浜国立大学）の報告では、大阪府大阪狭山市のグリーン水素シティ構想を事例に、SDGsを視野に入れた行政と民間企業との連携、そしてその評価手法の重要性が説明され、さらにはその手法の一つとしてバランススコアカードの有用性が提示された。

第3報告者の本田正美氏（島根大学）の報告では、地方議員の議員活動を検証するひとつのアプローチとして、地方議員の議会での質問事項に対する重み付けを設定する手法の提案、及びその適用可能性について事例分析をもとに説明された。

最後の熊谷智義氏及び岩瀨公二氏（いずれもNPO法人政策21）の報告では、岩手県盛岡市における公の施設の指定管理者に対する第三者評価について、自らのアクション・ラーニングをもとに当該評価の問題点や今後の方策等が提示された。

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、地方自治体を取り巻く評価の外延が日々拡大、深化しており、その理論的・技術的対応が急務であることが改めて確認されるに至った。日本の自治体評価の今後については継続的かつ幅広く議論が展開されてくことが期待されよう。

## 自由論題セッション2 「大学評価」

座長 牟田 博光（東京工業大学）

大学評価に関しては2004年以降、文部科学大臣の認証を受けた評価機関による第三者評価が義務づけられたが、この認証評価以外にも様々な評価がなされている。そのような中で、「自由論題セッション2 「大学評価」」では、以下の2報告を中心に、貴重な質疑応答があった。

渋井進氏・田中弥生氏（大学改革支援・学位授与機構研究開発部）「大学評価支援へ向けた指標設定のチェックリストの開発」は、大学評価は成果の定量的な可視化が求められ、さらに、公的資金給付の要件として、指標の設定とそれを用いた実績評価が求められるため、指標のデザインとデータ分析が重要となっているが、基礎的な研究が不足していることを指摘した。指標をデザインする際の問題として、指標が内容を適切に捉えているか、評価者を十分に説得できる内容であるかという妥当性の問題があり、この問題解決を支援するために、信頼性・妥当性を確認するためのチェックリストを開発したことを説明した。

高池宣彦氏（筑波大学大学院図書館情報メディア研究科）「教育と医療の第三者評価の比較検討—大学機関別認証評価と病院機能評価を中心に—」は、大学と医療の第三者評価を比較分析することにより、両評価の特徴とその差異を明らかにした。病院評価は日本医療機能評価機構という第三者機関が病院機能評価を行っているが、公開文書を元に、大学認証評価と病院機能評価の導入の経緯、評価の難易度、評価結果の比較を行った。その結果、法的拘束力の違い、認定・適合率、評価結果の活用の違いを明らかにした。

各報告における質疑を通じて、大学自らが指標を設定するに際して考慮すべき点が明確になったが、実際の現場に適応してその有効性を検証することが今後の課題として示された。また、大学評価と病院評価で同じような第三者評価といっても、「認証評価」と「機能評価」という性格の違いや、最低限の質保証か、質の向上を図るかという目的の違いから来る差は大きいものの、「認証評価」でも質の向上に寄与し、「機能評価」でも結果の拘束力が強くなっているなど、相互に参考にできる点も増えていることが示された。

今後もさらに、こうした課題意識にそって、研究発表が重ねられることが期待される。

## 自由論題セッション3 「評価の国際潮流と評価手法」

座長 源 由理子（明治大学）

本セッションでは5会員による研究報告に対し、フロアー（約26名）からの質疑を得て活発な議論が行われた。

まず、第1報告者である黒田かをり氏（一般財団法人CSOネットワーク）の「グローバル評価アジェンダと日本の市民社会における評価キャパシティの強化に向けて」では、2015年9月のSDGs採択を受けて、今後5年間のグローバル評価アジェンダが採択されたことを背景とし、日本のNPO／NGOの今後の評価キャパシティ向上への示唆をまとめたものである。フロアーからは、日本の地域おこしとの関連や、市民社会のECBの具体的な内容などについての質問があった。日本の市民社会を巻き込んだ評価能力向上（ECB）の課題を浮き彫りにさせた興味深い報告であった。

第2報告者の今田克司氏（一般財団法人CSOネットワーク）からは、「発展型評価～概念と可能性」と題して、パットンの発展的評価（Developmental Evaluation）の特徴と可能性について報告があった。発展型評価は、ロジック・モデルのような単線系理論ではなく、複雑系システムを評価論に取り入れ、社会的イノベーションの取り組みを絶え間ないフィードバックサイクルに組み込んだ評価アプローチである。フロ

アーからは、そのような取り組みは既に行われているのではないかという指摘や、もともとパットン「組織開発」の文脈で実践してきたのであり、実施組織に伴走していく評価のあり方を示しているのではないかという指摘があった。同評価が内包する複雑系社会におけるより実用的な評価理論の体系化へのニーズが、現代社会においてますます高まっているのではないかということを示唆される報告であった。

第3報告者の林薫氏（文教大学）からは、「SDGsの評価：第7目標「エネルギーと持続性の指標」」と題した報告があった。エネルギー効率の改善の指標は、多くの外生的要因があり、複雑かつ困難な作業になるが、コンパクトシティがひとつの政策目標となりうるとしている。フロアーからは再生可能資源の持続性と収奪（depletion）の考え方、資本としてのエネルギーのとらえ方、原子力エネルギーのリスク評価、さらには技術革新という不確実性への対応など、多様な視点からの質問があった。今後のSDGsの評価の指標に関する研究の重要性を再認識させられた報告であった。

第4報告は、中嶋学氏（ニューヨーク州立大学アルバニー校）による「協働型ネットワークにおける業績情報の拡散型探索と深耕型探索のための活用：試論」であった。協働型ネットワークにおいてどのように業績情報を活用しているかを検討したもので、深耕型探索（exploitation）の学習のために用いられているという仮説が示された。これに対し、拡散型か深耕型かは、その組織の発展過程に応じて推移するものではないかという指摘や、パートの紐帯の理論を活用した際の組織間の関係性の強さの変数に関する質問等が寄せられた。社会ネットワーク理論と評価論を組み合わせた研究は日本においては少ないところ、今後の研究に期待したいところである。

最後に、第5報告として、宮口貴彰氏（立命館大学）より、「実存論的評価手法を用いた気候変動適応プロジェクト事後評価のメタ分析」と題した報告があった。ポーソンによって提唱された実存論的評価手法（Realist Evaluation：以下RE）は、プログラム理論を使って介入策がどのように機能するかという「説明力」を高めようとする点特徴的である。報告者は9か国の気候変動適応プロジェクトの事後評価を対象とし、REを使ってメタ評価を行った結果、将来のよりよいプロジェクト形成に関しての有用な説明を政策立案者に与えることができた結論づけている。フロアーからは、将来のプロジェクト形成においてREを活用したメタ評価活用への期待があった一方で、そもそもREそのものは外部妥当性に重きをおいていないのではないかという指摘や、メタ評価対象の介入策がREを使っていないこととの整合性など、多角的な意見、質問が出された。介入による効果を実証的に検証する評価ではなく、生成力を重視した評価の適用に関する興味深い報告であった。

本セッションでは、評価の国際潮流について主にSDGsをテーマにしたものと、新しい切り口の評価理論・手法に関する研究報告が行われた。いずれも、複雑な現代社会における評価の理論と実践についての報告であり、今後のさらなる研究の深化に期待したい。

## 自由論題セッション4「教育評価」

座長 橋本 昭彦（国立教育政策研究所）

国際教育協力に関わる5報告が工夫や独創を競い、価値ある質疑応答を共有しえた。

石田洋子氏（広島大学）「教育協力に関する政策レベル評価3件からの学び」は、ODAの教育協力の評価の結果の活用状況について明らかにする独創的かつ有用な研究の報告であった。評価から得られた提言のうち、次の施策の策定プロセスに関わるものがよく活用されたのに比べて、実施プロセスに係る提言から施策改善に至ることが難しい傾向が看取されることが報告された。フロアからは、ODA評価における施策実施部門からの調査の在り方を問う鋭い質問もあったが、本研究の丁寧な分析プロセスの一端が共有された。

田中紳一郎氏（国際協力機構）「国際教育協力プロジェクトは公正的吗」は、国際教育協力の世界で重視されてきた「公正性（Equity）」の価値観がJICAの評価文書174点の中にどのように現れるかを集計・分析する独創的な手法が注目された。フロアからは、「公正性」の定義や「公平性」との区別といった研究デザイン関係や、資料として用いる文書の性格や取り出す言葉の種類と重みなどデータ関係、さらに分析のツールやその使用法などの分析過程関係などの多様な質問が出て、特に手法の練度への期待が目立った。

吉川香菜子氏（大阪大学大学院生）・神谷祐介氏（龍谷大学）・野村真利香氏（国立保健医療科学院）・荻野妃那氏（四日市看護医療大学）の「Improving Educational Outcomes at Primary Schools in Lao: An evaluation using a randomized controlled trial」は、吉川氏から日本語で報告が行われた。ラオスの小学生数百名について教育成果を測るべく、利他的行動の程度を測るペア・ゲームを実施したこと等が報告された。キャンデーを用いたディクテーターゲームを実施するなど、経済学の実験で使われる手法を教育協力の評価に応用する工夫がみられた。この実験手法が一定の人間関係のある学級内で有効なのか、実験や分析の目指すものは何かなどの質疑応答が交わされた。

關谷武司氏（関西学院大学）、吉田夏帆氏（関西学院大学大学院生・博士前期課程）の「修学パターンによるミャンマー初等教育における進級制度の検討」は、關谷（2012）による「就学パターン」分析の研究手法を援用して、ヤンゴンの小学生について実施した調査研究である。多くの留年児童を生んでいた進級試験制度が1998年に総合評価による進級制度に改められた。その前後で、児童が卒業に至るまでの修学パターンが344から48へと激減し、卒業率や留年未経験率も急伸した。明確な数値が出たことでフロアの注目も大きく、データの解釈や考察をさらに深めることを期待するコメントが相次いだ。

牟田博光氏（国際開発センター）「ミャンマー国基礎教育学校の教育条件が学力試験結果に及ぼす影響に関する分析」は、ミャンマーの全国学力試験の結果と基礎教育学校の教育条件の幾つかを組み合わせ、教育条件が試験結果に及ぼす影響を考察した。Chin州、Mandalay管区の2地域を選び、教育条件の変数を「教員一人あたり児童生徒数」「教員の資格」「児童生徒数規模」に限定し、正統的なデータ分析を行った。教育条件の差によって全国学力試験の合格率を説明する明快な議論が提出されたので、フロアでは分析を深めるための各種の視点が交換された。時間内では到底尽くせない議論が先送りされた。

## 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

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

## 『日本評価研究』執筆要領

2002.9.18改訂

2002.3.25改訂

## 1. 本文、図表、注記、参考文献等

(1) 論文等の記載は次の順序とする。

日本語原稿の場合

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

第2ページ以下：本文、謝辞あるいは付記、注記、参考文献

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

英文原稿の場合

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

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

(2) 本文の区分は以下のようにする。

例1（日本語）

1.

(1)

①

(2)

(3)

例2（英文）

1.

1.1

1.1.1

1.1.2

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

例1：日本語原稿の場合

図1 ○○州における生徒数の推移



(注)

(出所)

表 1 ○○州における事故件数

--

(注)  
(出所)

## 例 2：英文原稿の場合

Figure 1 Number of Students in the State of ○○

--

Note:  
Source:

Table 1 Number of Accidents in the State of ○○

--

Note:  
Source:

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

- (5) 本文における注記の付け方は、(…である<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）おとしで記述する。

## 『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。

## Publication Policy of the Japanese Journal of Evaluation Studies

Last revised on 15th 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 administers 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 administer 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.

## **Information for Contributors (For English Papers)**

Last revised on 29th 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  
Shinagawa Crystal Square 12th Floor, 1-6-41 Konan, Minato-ku, Tokyo,  
108-0075, Japan  
E-mail: [jes.info@evaluationjp.org](mailto:jes.info@evaluationjp.org)

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

Revised on 18th 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 pzhotoengraved, 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.

## Referee-Reading Guideline

The Japanese Journal of Evaluation Studies Editorial Board,  
The Japan Evaluation Society  
Approved on 10th 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。随時、投稿を受け付けておりますので、ご興味をお持ちの方は投稿規定・執筆要領をご参照のうえ、奮ってご投稿ください。

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

『日本評価研究』第17巻第1号

2016年11月17日

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

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. 17, No. 1, November 2016

## CONTENTS

### Special Issue: Practical Use and its Direction of the Evidence

Editor's Note: Special Issue: Practical Use and its Direction of the Evidence  
Ryo Sasaki, Tomoya Masaki

What inhibits evidence-based cancer screening?:  
An analysis using the concept of policy ideas Satoko Tanabe

Critical Reflections on SROI (social return on investment) Hiroshi Tsutomi

Current situation and challenges of utilization of evidences  
in the international development field Hiroaki Asaoka

Communicate and Use of Evidences in International Development Field:  
Movement of J-PAL Ryo Sasaki

#### Article

Use of Case Method in Ethics Education for Evaluators:  
Implications from the Teaching Method in Ethics Education for Auditors  
Nobuyuki Kobayashi

#### Report of the 13<sup>th</sup> Spring Conference