

# 日本評価研究

## Japanese Journal of Evaluation Studies

Vol. 24, No. 1, March 2024

### 巻頭言

日本評価学会の国際協力

西野 桂子

### 特集：評価研究および評価実践におけるエビデンス 理解の多様性と多義性

特集に寄せて「評価研究および評価実践における  
エビデンス理解の多様性と多義性」

米原 あき

評価における「エビデンス」の考察  
－信頼性と活用可能性を巡る議論を中心に－

今田 克司 津崎 たから 中谷 美南子

<コメント>

西村 君平 呉 書雅

<リプライ>

今田 克司

実在論的評価の理論と日本のEBPMへの示唆

西村 君平 呉 書雅

<コメント>

津富 宏

<リプライ>

西村 君平 呉 書雅

EBPプログラムの構築を目指す形成的評価手法における  
エビデンスの捉え方

<コメント>

新藤 健太

<リプライ>

青柳 恵太郎

新藤 健太

民間非営利セクターにおけるエビデンス活用の現状と課題

－児童家庭福祉分野団体へのアンケート調査から－

<コメント>

清水 潤子

<リプライ>

新藤 健太

清水 潤子

### 研究論文

不完全一対比較表を活用した参加型評価手法の提案

－評価結果に差がつく相対評価を目指して－

飯田 洋市

### 実践報告・調査報告

少人数学級の政策決定におけるエビデンス

－エビデンスは決定に影響したのか－

荒木 進太郎

日本評価学会

Japan Evaluation Society

第24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

米原 あき (東洋大学)  
Aki YONEHARA

編集副委員長  
Deputy Editor-in-chief

村上 裕一 (北海道大学)  
Yuichi MURAKAMI

編集委員  
Editors

大島 巖 (東北福祉大学)  
Iwao OSHIMA

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

新藤 健太 (日本社会事業大学)  
Kenta SHINDO

西野 桂子 (関西学院大学)  
Keiko NISHINO

橋本 昭彦 (日本女子大学)  
Akihiko HASHIMOTO

牟田 博光 (東京工業大学)  
Hiromitsu MUTA

山谷 清志 (同志社大学)  
Kiyoshi YAMAYA

特定非営利活動法人日本評価学会  
編集委員会事務局  
E-mail: [journal@evaluationjp.org](mailto:journal@evaluationjp.org)

# 日本評価研究

第24巻 第1号 2024年3月

## 目次

### 巻頭言

西野 桂子

日本評価学会の国際協力…………… 1

### 特集：評価研究および評価実践におけるエビデンス理解の多様性と多義性

米原 あき

特集に寄せて「評価研究および評価実践におけるエビデンス理解の  
多様性と多義性」…………… 3

今田 克司 津崎 たから 中谷 美南子

評価における「エビデンス」の考察

－信頼性と活用可能性を巡る議論を中心に－…………… 7

<コメント（西村 君平 呉 書雅）>

<リプライ（今田 克司）>

西村 君平 呉 書雅

実在論的評価の理論と日本のEBPMへの示唆…………… 29

<コメント（津富 宏）>

<リプライ（西村 君平 呉 書雅）>

新藤 健太

EBPプログラムの構築を目指す形成的評価手法におけるエビデンスの捉え方…………… 49

<コメント（青柳 恵太郎）>

<リプライ（新藤 健太）>

清水 潤子

民間非営利セクターにおけるエビデンス活用の現状と課題

－児童家庭福祉分野団体へのアンケート調査から－…………… 69

<コメント（新藤 健太）>

<リプライ（清水 潤子）>

## 研究論文

飯田 洋市

不完全一対比較表を活用した参加型評価手法の提案

－評価結果に差がつく相対評価を目指して－…………… 87

## 実践報告・調査報告

荒木 進太郎

少人数学級の政策決定におけるエビデンス

－エビデンスは決定に影響したのか－…………… 103

## 第24回全国大会のご案内

開催の報告とお礼…………… 117

プログラム詳細…………… 119

シンポジウム報告…………… 123

共通論題セッション報告…………… 124

自由論題セッション報告…………… 131

## 【巻頭言】

# 日本評価学会の国際協力

西野 桂子

関西学院大学

2024年は能登地方の大地震や羽田空港の事故と波乱の年明けになりました。お亡くなりになられた方々、各地で避難生活を続けていらっしゃる方々、そして、大切な方をなくされた方々へ心からお悔やみとお見舞いを申し上げます。私が今住んでいる神戸も29年前に大震災に見舞われました。震災を経験した人も震災後に生まれた人も、今回被災した方々に共感し、何とか力になろうとしています。

話は昨年に戻りますが、2023年12月4日～7日の4日間、フィリピンのマニラでアジア・太平洋評価学会(APEA)の第4回大会が開催されました。アジア・太平洋地域を超えたすべての地域の39カ国からのべ203人が参加した一大イベントになりました。本号が発行される時には大会報告書がウェブサイトに掲載されているはずなので詳細には述べませんが、1日目はプレ・コンファレンスとして4つの参加型ワークショップ(1)Evidence-based voluntary national reviews and country-led evaluations of the SDGs. Why and how? (2)Our experiences, our data: A how to guide on facilitating locally owned evaluation processes, (3)Evaluation of humanitarian action, (4)Foundations of evaluation designが終日開催されました。また、ワークショップと並行して、若手評価者の育成を目的としたWinter Schoolも開催され、多くの若手が評価学や評価の実践を学んでいました。

2日目と3日目はAPEAの本大会と位置付けられ、InaugurationからClosingまで、アジア・太平洋評価戦略に基づく15のセッションが開催されました。主なテーマには、(1)National evaluation policies and systems for good governance and accountability: A dialogue on NEPS, (2)Engagement of young and emerging evaluators: Lessons from UN agency evaluations in the Asia Pacific, (3)Amplifying indigenous voices: Advancing equity and inclusion in evaluation practiceなどがありました。どのセッションも各国の経験や個人の研究成果を基に、非常に活発な議論が続きました。最終日は、Parliament Dayと称して、フィリピン国会に招かれ、各国の国会議員が「Institutionalization of Monitoring and Evaluation for the 2030 SDGs」というテーマで発表しました。最後に、SDGsを達成するためのコミットメント「マニラ宣言」に参加者全員で署名して式典を終了しました。

4日間を通して感じたことは、評価文化の浸透と若者の熱気、そして議員の参加でした。外務省は、2001年からアジア・大洋州諸国の政府関係者や専門家を招待して「ODA評価ワークショップ」を開催しています。その目的は、ODA評価の手法や課題についてアジア・大洋州地域における理解を増進し、特に途上国の評価能力を向上させること、途上国関係者自身のODA評価能力の向上を通じて、ドナー国側の援助効率化だけでなく、途上国側のオーナーシップ・透明性の向上や開発の効率化を目指すことにあります(外務省HP)。2022年からはAPEAと共催でワークショップが開催されており、協力体制が強化されつつあります。また、JICAも途上国行政官の評価に対する理解と知識の向上を目指した技術協力プロジェクトを

多くの国で実施してきました。このような日本の取組の中で2011年にAPEAが設立されました。10年ひと昔と言いますが、この20年の間に日本が成した多くの試みが12月のAPEA大会で結実したのではないかと密かに感動していました。それと同時に、各国の活発な発表や試みを聞いているうちに、「今や日本が遅れをとっているのではないか？」と不安を覚えたのも事実です。

2つ目は、アジア・太平洋地域の若者の熱気です。若手評価者の育成を目的としたWinter Schoolが同時開催されたこともあると思いますが、若手の参加者が多く、活発に参加し、貪欲に知識を吸収している姿勢に感銘を受けました。現在、各国でEvalYouthが設立され、若手評価者のネットワークが広がっています。日本評価学会も毎年評価研修を実施していますが、研修を受けた人を中心に日本にもEvalYouth Japanができないかなあと考えています。また、先の全国大会のラウンドテーブルにもありましたが、評価学の教育をいかに日本で推進していくかも日本評価学会に課せられた課題だと思います。最後に一言加えたいのが、国会議員の評価への関わりです。日本の議員に期待するのは無理かもしれませんが、一人でもThe Global Parliamentarians Forum for Evaluationなどの国際的なフォーラムに参加して、日本の行政評価の経験などを共有し、議論をリードしてほしいと願っています。

最後になりましたが、2024年1月からAPEAの会長を日本評価学会の石田洋子会長が務められます。マニラの大会で前会長のAsela Kalgampitiya氏から石田会長に正式にバトンが渡され、日本への期待が高まりました。今後は日本評価学会がもっと国際化して、APEAだけでなく、IPDET (International Program for Development Evaluation Training) や各国の評価学会の大会で積極的に研究発表し、ジャーナルに投稿するなどして、評価の国際協力にリーダーシップを発揮していくことを期待しています。2020年から4年間、APEAの副会長を務めさせていただいた感謝とともに、今後も「日本評価学会の国際協力」への貢献を続けていきたいと思っています。

【特集に寄せて】

## 評価研究および評価実践における エビデンス理解の多様性と多義性

米原 あき

東洋大学

エビデンスに基づく政策形成（EBPM）をはじめ、学校評価やNPO評価、あるいは国際協力やSDGsに関連するプロジェクト評価等、近年、社会課題に関するあらゆる取り組みにおいて、エビデンスが求められる傾向がある。一方で、「エビデンスとは何か」という問いに対する理解は曖昧になっていることが多く、実践現場においても、そして時には学術的な議論の場でさえ、誤解や混乱が内包されたまま議論がすすめられるようなケースも少なくない。医療、政治、行政、経済、教育、科学…様々な分野で、「エビデンス」という一つの言葉が少しずつ異なる意味で使われているという点でそれは多義的であり、同時に、それぞれの分野で重視される特徴も異なるという点でそれは多様でもある。

日本評価学会では、学会創設当初からEBPMおよびエビデンスに基づく評価に関心を払い、継続的に特集を組んで研究の蓄積を図ってきた。一連の特集では、エビデンスとは「ある介入の効果を裏付けるための情報」として理解されている。しかし、エビデンスの詳細な性質については必ずしも見解が一致しているわけではない。ひろく知られる見解としては、ランダム化比較試験およびその知見をシステムティックレビューによって統合していくことでこそ「ある介入の効果を裏付けるための情報」を純粹かつ厳格に明らかにすることができるという主張がある。一方で、このようなエビデンス観については、エビデンスを狭く限定しすぎているとか、社会プログラムの評価にはそぐわないケースがあるといった批判も存在する（大塚他 2023）。

さらに、政策形成や様々な社会プログラムの評価の実践現場では、上記のような学術上の議論はさておき、各個人・各組織・各分野が独自にエビデンスの意味を解釈して実践を展開していることもある。それがエビデンスを多様化・多義化させる妥当な要因になっていることもあれば、素朴な誤解に基づくものであることもある。本特集企画の本題に踏み込む前に、まず注意すべき素朴な誤解として、以下の3点を指摘しておきたい。

最も単純な誤解の一つは、「エビデンスとは数値情報のことである」というものである。エビデンスの提示を求められる現場では、「とにかく数字になっている情報」を評価指標にできないかと、それらしい指標を探し当てることに躍起になっている向きもある。そしてこのような指標設定の仕方が、後述する「因果性の崩壊」につながっていくケースが多い。「ある介入の効果を裏付けるための情報」という定義は、エビデンスを量的な情報のみに限定するものではなく、質的な情報でもあり得ることを担保していると考えるべきである。

次に、エビデンスと事実特定（現状把握）を混同するケースを見聞することもある。介入前のベースライン評価情報が欠落した状況で、介入後の状況のみを測定した場合、それはエビデンスではなく、ある時点における事実的な状況の測定である（ただし一時点の調査で過去の時点のデータも同時に取得して、介入前後を比較する場合は除く）。事後の状況を測定することが問題なのではなく、その情報をエビデンスであると捉えることが問題なのであって、そのようにして得られたデータは、現状把握のための情報とし

て適切に活用されるべきである。

3点目に、そもそも介入のロジックが成立していない、因果性（因果仮説）の崩壊に当たるケース——構成概念妥当性が欠落している、あるいは不十分なケース——にも注意が必要である。例えば、「Aというプログラムを実施したらBという成果が得られた」という想定があるときに、Aという介入がそもそもBという成果を生むような介入ではない（成果Bを意図していない）ケースがある。この場合、Bについての測定情報はBの事実的な情報であって、介入Aの効果を示すエビデンスとは言えない。このような因果仮説の崩壊は、指標の不適切な設定や、クライアントからの文脈に沿わない要求によって引き起こされることもある。

このような素朴な誤解に加えて、エビデンスの議論をする際に避けて通れないより本質的な論点は、科学哲学や認識論（epistemology）にルーツをもつ、社会科学の方法論におけるパラダイム論争に関わる。実証主義的な認識論に基づく量的方法論と、解釈（相対）主義的な認識論に基づく質的方法論の間には、1980年代初期から続く論争がある（Marshall 1998）。1994年にGary Kingらが“Designing Social Inquiry: Scientific Inference in Qualitative Research”を出版したことにより、北米でこの論争がピークに達した（King et al. 1994; Brady et al. 2004; Goertz et al. 2012）。冒頭に挙げたようなエビデンスをめぐる議論は、このパラダイム論争の延長に、方法論的な論争として位置付けられることが多い。

本特集の企画意図は、このパラダイム論争の枠組みの外側に、エビデンスをめぐる議論を拡張することにある。2024年の年頭は予期せぬ天災人災とともに明けることとなったが、予測不可能な事態への対応や、SDGsに示されるようなそれぞれの文脈における評価可能性の検討（米原 2021）が求められる中で、特に社会的な介入の価値を問う活動である評価という営みは、「事象の普遍性」や「因果の線形性」を前提とした従来の枠組みのなかでは、議論が十分に尽くせなくなっているのではないだろうか。あるいはそこに量・質を問わない多様なエビデンスの可能性を再考する余地があらわれるのではないか。

エビデンスに基づく評価が、多様な取り組みの評価に挑戦する実践現場を、混乱させ疲弊させるためではなく、複雑困難な状況のなかでも改善に向けての有効な手がかりを提供する活動として役立つために、評価研究にできることは何か。その一つは、実践現場から抽象理論まで多様な次元から「エビデンスとは何か」という問いに向き合い、エビデンスの捉え方を再考する契機を創出することではないか。本特集企画の主たる目的は、パラダイム論争としての議論を踏まえつつも、認識論的な拡張を視野に入れたエビデンス論議を展開することである。

また、本企画では、それぞれの特集論文に異なる評者がコメントを付し、そのコメントに著者がリプライするという、「本論—コメント小論—リプライ小論」という形態を採用した。これは大塚他（2023）に範を取ったものである。本論の執筆者は、多様な立場から各自の論考を展開しているが、エビデンスをめぐる議論は、その認識論的なレベルにおいても、方法論的な守備範囲という意味でも、深遠な広がりがあり、本企画のみをもって整理しきれものではない。それぞれの議論の複雑性や動態性を誌面上で再現し共有するための試みとして、コメント—リプライ方式に挑戦することにした。制約の厳しい出版スケジュールの中で、査読の結果がどうなるかも分からない段階から、この試みに賛同し協力して下さった本論およびコメント小論の執筆者には心から感謝を申し上げたい。

最後に、本企画を構成する4本の論文の位置付けについて概説する。まず、今田・津崎・中谷論文は、そもそもエビデンスの多様性や多義性を系統的に議論する枠組み自体が整っていないという根本的な問題を指摘し、エビデンスについて検討する際に、信頼性という水準だけでなく、活用可能性という観点も考慮する必要があるとする。この総説論文は、総説でありながらも、評価学の観点からパラダイム論争の二極化（実証主義—対—解釈主義）を乗り越えるための視点を提起する、建設的な論考を披露している。

西村論文は、RCTを代表とする実験的評価の外的妥当性に対する疑義を出発点として、実在論的評価の可能性を検討する。特に、実験的評価をプログラム評価などの文脈依存性の高い評価に応用する場合、ランダム割付やPSMなどの完全な施行は困難になる。現在の日本のEBPMでは、実験的評価による因果効果

の特定を重視する傾向があるが、この研究論文は、EBPMの限界を克服する方途として、応用可能性を重視する実在論的評価のアプローチを提唱している。また、このアプローチも、パラダイム論争の二極化をこえた、第三の立場として実在論を位置付けている点を強調しておきたい。

新藤論文は、形成的評価の観点からエビデンス概念を問い直している。一般的な用語としてのエビデンスと異なり、Evidence-Based（エビデンスに基づく）と表現される場合の「エビデンス」は「特定の介入の有効性を裏付けるもの」でなければならない。CD-TEP法などのプログラム改善を目的とする形成的評価の観点からこの意味での「エビデンス」を再検討すると、数値化し難い「有効性」を質的にも捉える必要があることが浮き彫りになる。形成的評価の実践におけるエビデンス観がこの総説によって拓かれている。

最後に清水論文は、エビデンス活用に関する調査を通して、実践者たちのエビデンス理解の現状を明らかにしている。昨今のEBPMブームの中で、予測不可能性や文脈依存性の高い社会課題への対応が求められる領域の筆頭が、民間非営利セクターの現場であろう。さらに、これまでの評価経験の蓄積とのギャップが大きいのもこの分野の特徴である。この調査報告は、そのような現場の事業実施団体を対象に調査を行い、得られたデータを分析したものである。この調査では、本稿冒頭で指摘したようなエビデンスの独自解釈やある種の誤解がひろまっていることが明らかになった。学術的な理論考察と評価実践の臨床観が極端に乖離しないよう、我々の意識を現実に戻してくれる、また同時に、そのような実践現場における評価者の役割を問い直す、貴重な調査報告となっている。

各論考に続くコメントーリプライ小論が示すように、これらのテーマは今まさに動的な議論の渦中にある。引き続きこれらの議論を深め、牽引していくことが、日本評価学会に求められているのではないだろうか。

## 参考文献

- 大塚啓二郎、樋口裕城、鈴木綾（2023）「これからの実証経済学における課題」、大塚啓二郎他編『次世代の実証経済学』、日本評論社、第3章：71-89
- 米原あき（2021）「SDG教育目標にみる理念志向ターゲットの評価に関する一考察：測定可能性（measurability）から評価可能性（evaluability）へ」、『日本評価研究』、21(2)：31-46
- King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton University Press.
- Brady, H. E., & Collier, D. (2004). *Rethinking social inquiry: Diverse tools, shared standards*. Rowman & Littlefield Publishers, Inc.
- Goertz, G., & Mahoney, J. (2012). *A tale of two cultures: Qualitative and quantitative research in the social sciences*. Princeton University Press.
- Marshall, G. (Ed.). (1998). *Oxford dictionary of sociology*. Oxford University Press.



## 【総説】

## 評価における「エビデンス」の考察 —信頼性と活用可能性を巡る議論を中心に—

今田 克司

株式会社ブルー・マーブル・  
ジャパン  
katsuji.imata@blue-marble.co.jp

津崎 たから

ウエスタン・ミシガン大学大学院  
学際的評価学博士課程  
takaratsuzaki@gmail.com

中谷 美南子

評価コンサルタント  
minakonakatani@gmail.com

### 要約

近年、医療分野、教育分野、国際開発分野等においてエビデンスに関する議論が展開され、評価研究および評価実践においてもエビデンスをめぐる議論の多層化が起こっている。しかしながら、共通理解や共通項の認識が進んでいるとは言い難い。本稿ではエビデンスの多面性に関する考察を系統的に行う方法論が確立していないという問題提起をもとに、エビデンスの有用性、特に信頼性（credibility）および活用可能性（actionability）を中心に置いた議論を展開する。前半では、科学論における基本的対立、評価学における「パラダイム・ウォー」におけるエビデンスの扱い、研究者・評価者の世界観について解説する。後半では、エビデンスの活用可能性を定める多面的要素について四つの側面を紹介し、活用可能性の高いエビデンスの抽出の鍵となる五つの文脈を提示する。さらに、活用可能性を担保する二つの実践例を挙げ、評価の妥当性や評価者の役割とエビデンスとの関連についてあらためて課題を提示する。

### キーワード

エビデンス、エビデンスの活用可能性、エビデンスの信頼性、評価の妥当性

#### 1. はじめに

近年、評価研究や評価実践においてエビデンスについての議論が多くなされるようになってきている。それはなによりも、EBPM（エビデンスに基づく政策立案）に対する注目が集まっているためと言える。日本評価研究第20巻第2号ではこれに関する現状と課題の特集が生まれ、政策形成・立案、医療分野、教育分野、国際開発分野における議論が展開された（日本評価学会2020）。

一方で、エビデンス概念や分野ごとのEBPMの

実践に関する共通理解や共通項の認識が進んでいるかといえば、必ずしもそうとは言えない。この問題意識を発端として、本稿では、エビデンスの有用性、特に活用可能性（actionability）を中心に置いた議論を展開する。すると、エビデンスにおけるドミナントな議論である信頼性（credibility）を中心に置いた議論では視野に入りにくい視点や論点が浮かび上がってくるはずである。

## 2. エビデンス：問題の所在

### (1) エビデンスの隆盛と「エビデンス」用語の不用意な使われ方

エビデンスをめぐる議論の混乱を回避するため、まず、エビデンスの用語についていくつかの側面から整理したい。「エビデンス」は英語の「evidence」を由来とする外来語であり、「証拠、根拠、形跡」という意味を持つ。英語由来であるがゆえに、「evidence」と「エビデンス」は同一の意味と理解されている場合が多いだろう。なお、「evidence」の語源は、ラテン語の「evidentia」であり、「外に (ex-) 見え (video) ている (-ens) こと (-ia)」という意味を持つ。つまり、「エビデンス」は外部から見えることや明らかなことを示す言葉という出自を持つ。

「エビデンス」という単語は多義的に用いられ、職種や業種によって意味合いが異なって使用されている。医療現場での「エビデンス」という用語は、薬や治療方法、検査方法等、それが効果があると判断を下される際の証拠を指す。医療の分野で求められる「エビデンス」は「科学的証拠」であり、実際に患者の治療段階で調査研究を実施し、薬や治療法の効果や安全性を確認している。

行政分野ではEBPMの政策方針の中で、「エビデンス」が用いられる。「エビデンスに基づく政策立案」について、内閣府は「政策の企画をその場限りのエピソードに頼るのではなく、政策目的を明確化した上で合理的根拠（エビデンス）に基づくものとする」と<sup>1</sup>と定めている。

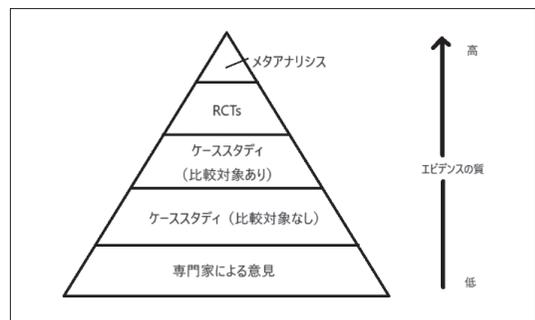
国際開発分野では、国際潮流と常に整合性を取りつつ、開発効果に関する科学的に信頼性の高いエビデンスを求めるために、多数のインパクト評価が実施されてきた。インパクト評価によって積み上げられる仮説検証の根拠がエビデンスであり、エビデンスとインパクトが相互に関連し合いながら、議論が進んでいる（浅岡2020）。

IT業界における「エビデンス」とは、結果を正しく残すための資料であり、ある試験を行った際その結果が正しかったことを示す証拠となるものや、その根拠を表すデータ・画像を残すこ

とを意味する（一般社団法人日本情報経済社会推進協会2015）。

このように多分野で多義的に「エビデンス」が定義され、理解されているが、エビデンスの議論を先導してきたのは、医療・医学領域である。この分野では「エビデンスレベル」という言葉も頻出する（図1参照）。医学の臨床研究において、信頼性をレベルで現す目的のために使用されるもので、最もエビデンスレベルが高いのは「ランダム化比較試験のメタアナリシス（同じ病態で行った複数のRCTを集めて解析したもの）」であり、逆に最もエビデンスレベルが低いのは「専門家個人の意見（経験だけの発言）」と言われている。また、「ケーススタディ」もエビデンスレベルが低いと位置付けられる（Schwandt 2015, p.262）。このようなエビデンスレベルの考え方は主に医学の文脈で活用されているが、教育学や社会学の分野でもEBPMや「エビデンスに基づく教育」（Evidence-based Education）のかけ声の中でエビデンスの信頼性を問う議論において持ち出されている。エビデンスレベルの概念含め、分野を横断して「エビデンス」用語が活用されているが、その用法が不用意であったり、定義や基本的理解が曖昧なまま使用されている場面も多い。これが、エビデンスに関する議論の混乱を招く一因となっていると考えられる。

図1 研究デザインとエビデンスレベル



（出所）Level of evidence hierarchy<sup>2</sup>より筆者（津崎）が翻訳し作成

### (2) エビデンスの質を問う

エビデンスとは、何らかの主張や判断の根拠を示す情報やデータのことであり、科学的な知

識や真理を主張するために用いられる。しかし、前述のとおり、その定義や基準は分野等によってまちまちであり、社会状況や価値観にも影響を受ける。

エビデンスの質が取り沙汰される場合、まずその組上にのぼるのは、エビデンスの信頼性(credibility)である。そして、これを評価するために、GRADEやCASPやCRAAP等、様々な手法や基準が存在する<sup>3</sup>。これらの基準やツールは、エビデンスに対して客観的な評価を行うことができることが利点であるが、社会的な文脈や価値観に十分な検討を付与することは難しい。また、エビデンスをもとに政策と成果の間の因果関係を証明する場合、多様な利害関係者の立場を踏まえて総合的にエビデンスを構築する等の構想を実践に移すのは簡単ではない。それはなぜであろうか。

筆者らの考えによれば、それは、エビデンスの多面性に関する考察を系統的に行う方法論が確立していないためである。エビデンスを使う際には、その限界や前提条件を理解し、過不足なく解釈し、意思決定の根拠とすることが重要となる。その際、エビデンスの信頼性に関しては、図1に示したように、信頼性が「高い」との判断を可能にする一連の手続きが存在する。しかし、第一に、これを政策評価や社会的介入におけるプログラム評価等にあてはめた場合、現実的な制約等でエビデンスレベルを高めていくことに困難が生じることが多い。そのような状況下で、いかにエビデンスの精度を弱めつつ一定の信頼性を確保するかを考察した分析枠組も存在する<sup>4</sup>。

そして第二に一こちらが筆者らの主たる論点になるが一、エビデンスに関しては、信頼性の精度を上げる技術や分析手法の開発が進んでいる一方で、特にその活用を視野に入れた多面性に関する考察が不十分である。シュワント(Schwandt 2015a, pp.259-263)は、介入のインパクトや有効性の論拠となるエビデンスの主たる要素には、エビデンス自体の信頼性に加え、エビデンスが仮説検証を支える力や、エビデンスと因果推論との関連性があるとし、信頼性の考察だけでは十分でないと述べている。

エビデンスの多面性に関する考察が必要であるとの論は、評価研究において一定の位置を占めている。特記すべきは、2007年にクレアモント大学で開催されたエビデンスの信頼性に関するシンポジウムとこれをまとめて翌年出版された書籍(Donaldson, Christie and Mark 2008)、そして2015年のその第二版(Donaldson, Christie and Mark 2015)であろう。注目すべきは、初版ではこの書物が「エビデンスの信頼性」をテーマにしていると掲げられているのに対し、第二版は、これが「エビデンスの信頼性と活用可能性」に関する書物だと掲げられている点である。後述するように筆者らは、エビデンス概念に関しては、信頼性とともに関与可能性の考察が少なくとも同等程度以上に必要だと考える。

上記のようなエビデンス概念のいわば「拡張」の必要性は、評価研究・実践における「問い」の性質を考えることで理解が進む。一般に社会調査(あるいは研究全般)とは、新しい知識や理解を得ることを目的として仮説を検証することであるが(Wanzer 2021)、評価の基本形は、定められた目的や基準に基づいて価値や効果を測定することである。評価でも社会調査でも、その入り口で設定する「問い」は重要であり、目的や方向性を明確にする。社会調査における「問い」が仮説検証と知識の獲得であるのに対し、評価実践における「問い」は、評価目的に従って、まず「誰が」「どのように」評価基準を定め、どのような方法でデータ収集・分析を行い価値判断に至るかを規定する。定性的手法では、「問い」は発見、説明、解明、探求に焦点を当てる場合が多く、広範な分野や具体的な領域に対して柔軟に設定される傾向にある。一方、定量的手法では、「問い」は記述、比較、関係によって仮説を証明したり反証したりする場合が多く、より具体的・限定的な文脈で利用されるものが多い。いずれにせよ、評価の「問い」は検証すべき設問の内容だけでなく、その設問にいかにか答え、その結果がいかにか活用されるかという一連の過程に関する諸作業を内包する。

研究一般においても、問いの重要性は広く認識されており、エビデンスが問いの性質に関連することを論じる論文は多数存在する。医学の

分野から一例を紹介すると、エビデンス・ベースド・メディシン（EBM）が有効性と効率性を重視する中で、アップシュールら（Upshur, VanDenKerkhof and Goel 2001）は、エビデンスは質問の性質に応じて異なる要素から構成されるという包括的なモデルを提案した。質問に適したエビデンスを探すことが重要であると主張し、エビデンスの要素は質問、方法、測定、意味の四つであり、それぞれに対応する研究デザインや分析手法、測定ツールや指標、臨床的または政策的な意味付けを行った上で、包括的に判断する必要があると論じている。

### 3. 科学論の基本的対立と評価研究への影響

エビデンスの質、特に信頼性に関する議論、すなわち取り組みや政策が有効であるという判断の論拠の信頼性はいかに確保されるのかという命題は、そもそも物事の真偽の判別は何を根拠になされるのか、科学的とはどのようなことかといった、評価や社会調査にとどまらない人文・社会科学全般の存立基盤に関わる根本的な問題を取り扱うことになる。そしてこれは、20世紀後半以降の科学論における世界観の論争が評価研究という小さな学術分野にも飛び火したものと考えられる。そこで、エビデンスの信頼性の論題に取り組むには、この論争をおさらいしておくことが必要となる。

#### (1) 科学論における基本的対立

極端に単純化して言えば、物事には私たちの知覚とは独立した客観的な事実があり、私たちはそれを認識することができる（目の前にイスがある）という考えと、そのような存在はおしなべて特定の認知・言語体系の投影にすぎない（虹は七色）という考えが存在する。前者には実在論（realism）、後者には懐疑論（skepticism）というラベルが貼られたりするが、科学史においては、前者を論理経験主義（logical empiricism）、後者を歴史主義（historicism）という呼称で表すことも多い（Papineau 1996）。また、前者を客観

主義（objectivism）、後者を相対主義（relativism）と呼ぶ等（Bernstein 1983）、論者の整理の仕方により「～イズム」の呼称や区分けも異なるが、基本的な世界観の相違としてこの対立は存在してきた。社会調査や評価研究含め、人文・社会科学の多くの分野、そして一般にも普及しているパラダイムという用語を世に出したクーン（Kuhn 1962）やファイヤーアーベント（Feyerabend 1975）は歴史主義に位置付けられるが、決して極端な懐疑論で論陣を張っているわけではなく、また、それぞれのイズムも固定的ではなく進化や分化が見られる。

評価研究において、この論争を紹介する場合、前者を実証主義（positivism）、後者を構成主義（constructivism）と呼ぶ場合が多い（Christie and Fleischer 2015, p.27; Christie and Alkin 2023, pp.17-18）<sup>5</sup>。

実証主義の基本姿勢が、物事は私たちの認識とは関係なく存在し（事象の独立性）、私たちはそれらを何らの予見や先入観なしに認知することができる（認識の客観性や価値中立性）であるのに対し、構成主義の基本姿勢は、世界の存在に関する私たちの認識は社会によって作られているというものだ（事象の非独立性、認識の間主観性や価値非中立性）。ただし、この基本形を出発点に、さらに進化した社会構成主義者は、事物の存在自体を肯定あるいは否定するのではなく、事物が実在として構築されていると考える。別言すれば、「実在している」という判断がいかに可能になっているかを問う。松村（2021 pp.37-38）は、このような流れを、「反-実在論」と呼び、実在論と反実在論の古典的な対立図式を超えるものと位置付けている。

#### (2) 評価研究における「パラダイム・ウォー」とエビデンスの扱い

評価研究においては、定量対定性手法の選択が対立化し、場合によっては先鋭化して、「パラダイム・ウォー」と呼ばれる戦闘状態になったと例えられることも少なくない。これは、上述のような科学論争が飛び火したものであり、単なる分析手法の対立ではない。そうであれば、本特集号で扱うエビデンスに関する論考も、事

象の独立性や客観性、観察者の価値中立性に関する議論を視野に入れて取り扱う必要がある。

評価理論においては、実証主義に加えて、その進化系であるポスト実証主義 (post-positivism) や実証主義と構成主義の統合系である実利主義 (pragmatism) が紹介される。ポスト実証主義では、実証主義が唱えていた科学は真理 (truth) を解き明かすものという立場に留保を入れて、真理そのものは解き明かせないかもしれないが、測定を精度を上げることによってそれとできるだけ近づくことは可能だと考える (Christie and Alkin 2023, p.17)。この伝統においては、エビデンスは客観的な現実を反映しており、人間の解釈や認識に依存しないと考える。一方、構成主義アプローチでは、エビデンスが生み出された文脈や目的を重視し、特定の利害関係者グループが現実をどのように定義し、意味付けるかを示すものとして分析される。エビデンスは、その背景や前提を明らかにすることで批判的に評価され、他の可能な解釈や視点と比較されたり、議論によって多様性や相対性を認めるものとして扱われる。

また、実利主義は、客観と主観の両方を認め、定量・定性データはどちらも有用だと考える。エビデンスの扱い含め、実利主義は実証主義と構成主義の折衷系のように受け取られている (Christie and Fleischer 2015, pp.33-34)。

さらに、評価の実在論 (realist evaluation) にも言及しておきたい。実在論的アプローチは、科学論における実在論の発展系を評価研究・実践に応用したもので、社会的介入は一定の条件のもとで機能するという文脈重視の考え方を持つ理論的なアプローチの一つである (例えば Pawson and Tilley 1997 や Marchal 2012)。このアプローチにおいては、プログラムや政策が効果をもたらすメカニズムと、そのメカニズムが作用する社会的文脈という二つの要素に注目し、「どのような文脈でどのようなメカニズムが作用してどのようなアウトカムが生じるか」という因果過程の記述に注力する。エビデンスは、現実の構造やメカニズムを発見し、説明し、予測するための手段として利用され、信頼性を検証することで客観的に評価される。

### (3) 研究者・評価者の世界観

上記の基本的対立は、当然のことながら、研究者・評価者の世界観に大きく影響し、逆に研究者・評価者の世界観が各人の科学に対する立場に影響する。そしてそれがエビデンスの捉え方を規定していく。研究者・評価者の世界観は、大きく存在論、認識論、価値論に分けることができる。

存在論 (Ontology) は、存在するものやその性質についての考え方である。研究者が実証主義者であれば、客観的に存在する真理を探求することを目指し、観察や実験等の客観的な方法で得られたデータや統計をエビデンスとして最重視する。一方、研究者が構成主義的存在論を支持していれば、真理は人間の認識や言語によって構成されるものであり、多様で相対的であると考える。

また、認識論 (Epistemology) は、知識や認識の性質、起源、根拠、限界等について考察する哲学の一分野であり、人間がどのようにして物事を知ることができるか、また、どのようにして正しい知識を得るのかの指針となる。研究者が認識論において実証主義者であれば、知識は経験から得られるものであり、科学的な方法論に基づいて検証されるべきだと考えるだろう。その場合、エビデンスは論理的かつ客観的に分析され、仮説検証や因果関係の推定等の形式で提示される。一方、研究者が構成主義者であれば、知識は人間の解釈から生まれるものであり、一般的に科学的と言われる方法論だけでは不十分と考える。その場合、エビデンスは文脈的かつ主観的に理解され、物語作りや意味付け等の形式で提示される。研究において認識論が異なると、研究対象や研究方法の選択に影響を与える可能性がある (Cresswell 2014)。

さらに、価値論 (Axiology) とは、価値や倫理についての考え方であり、価値の本質や価値と事実の関係、価値判断の基準等を扱う。研究者が価値論において価値中立主義者であれば、研究は価値判断や感情から自由であるべきだと考えるだろう。その場合、エビデンスは客観的かつ中立的に収集され、偏見や先入観を排除した形式で提示される。一方、研究者が価値中立性

に疑念を持つ価値関与主義者であれば、研究は価値判断や感情を含んでもよい、あるいは含むことは避けられないと考える。その場合、エビデンスは主観的かつ関与的に収集され、自分自身の立場や価値体系を明確にした形式で提示される。

このように、研究者の世界観は、エビデンスに関する様々な側面に影響を与えると考えられる。しかし、これらの立場は一様ではなく、研究者によって異なるほか、研究分野や目的や方法によっても変化する。

(4) 価値論の幹をたどる

2023年、アルキンとクリスティーによる『Evaluation Roots』の第三版が刊行され、下記図2のとおり「評価理論の木 (Evaluation Theory Tree)」の内容が更新された (Christie and Alkin 2023, p.13)。方法論 (Methods)、価値論 (Valuing)、実

用論 (Use) の三つの幹が並べられているのは変わらないが、第二版までの、それぞれの幹についていた人物名 (当該の評価アプローチを提唱し、流れを受け継ぐ評価理論家や実践者) が、評価アプローチや方法論の名称に変更されている。この三つの幹のなかで、実証主義からポスト実証主義への流れは方法論の幹で表されているが、本論との関連において、価値論の幹について紹介してみたい。

価値論の幹の根に近いところにある基本的考え方に、事実 (fact) と価値 (value) の二分法に対する懐疑論がある (Christie and Alkin 2023, p.27)。一般に (あるいは実証主義の伝統によれば)、客観的な事実と主観的な価値は別物で、評価作業に後者 (評価者やクライアントの価値、その他特定の評価利害関係者の価値) が混入することは客観的な価値判断を阻害するものであり避けなければならない。評価の基本図式とも

図2 評価理論の木

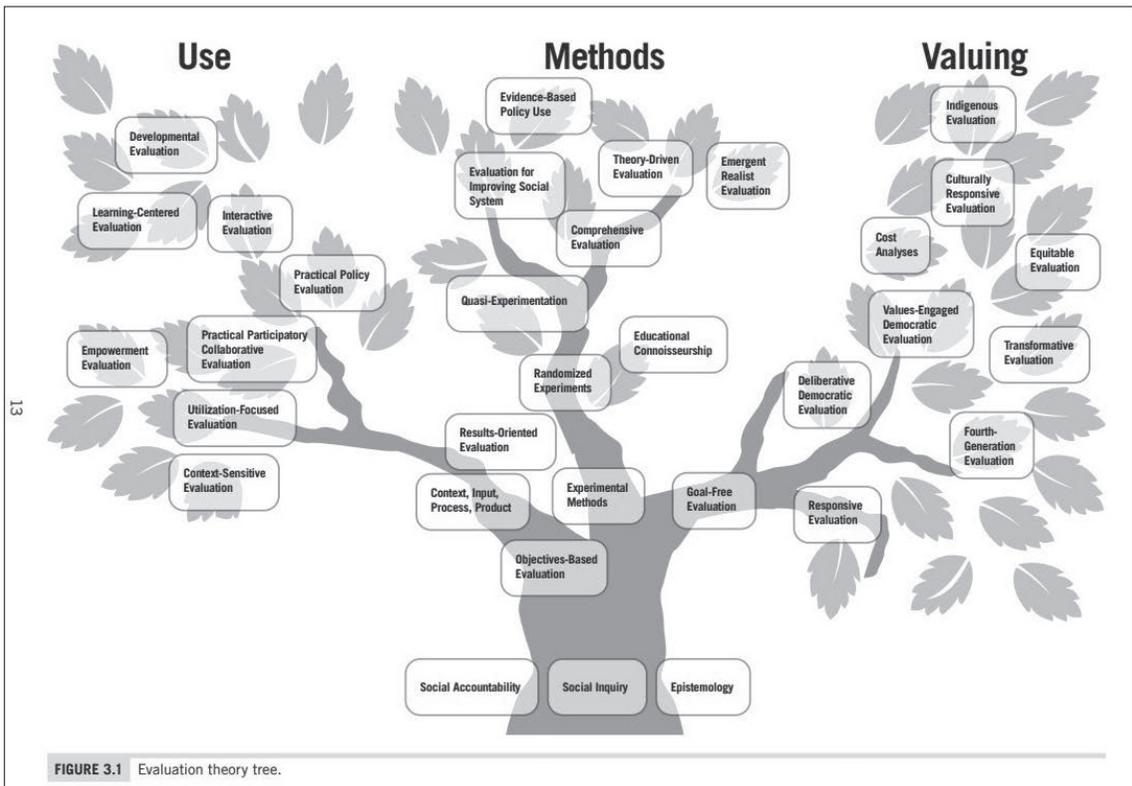


FIGURE 3.1 Evaluation theory tree.

(出所) Christie and Alkin (2023, p.13)

言える「評価＝事実特定＋価値判断」は、まさに事実と価値の分離を基礎としていると言える。

ハウスとハウ（House and Howe 1999）は、この二分法を否定し、むしろ事実と価値に関する言明は、線分の左端（事実）と右端（価値）を行ったり来たりする連続性のあるものだと解説する。「ダイヤモンドは岩より硬い」が左端の言明、「カベルネ・ソーヴィニオンはシャルドネより美味しい」が右端の言明だとすれば、「A君の方がB君より頭が良い」、「このテストは妥当だ」等の言明は、線分のどこかに位置するものと言う（House and Howe 1999, pp.5-7）。

これを出発点として、価値論の幹においては、評価作業における事実特定において、評価者を含めた特定の利害関係者の価値が優先され、その優先価値のメガネで事象が眺められ、評価設問が設定され、調査対象が決められていると考える。また、ゴール・フリー評価を提唱したスクリヴェンは、特に評価や評価者の価値中立性に関しては手厳しい批判を行っている（Scriven 2012, p.19）。

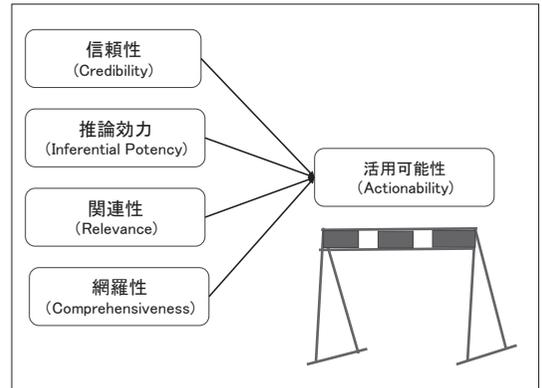
#### 4. 活用可能性の観点からのエビデンスの主要な問いかけ

##### (1) エビデンスの活用可能性の多面的要素

前節において、科学論における基本的対立とそれが評価研究に波及していることを概観した。この理解を得ると、昨今のEBPMが推進される国内外の潮流において、エビデンスの質とそれを産出する調査デザインと手法を選択する際に、エビデンスの信頼性のみを中核課題とするものの“物足りなさ”の感覚に至る。いくら信頼性の高いエビデンスが得られたとしても、それが新たな政策決定やプログラム・事業の設計に正しく活用されなければ、EBPMは成立せず、社会の改善活動につながらないからである。

このエビデンスの活用焦点をあてた視点が、エビデンスの活用可能性（actionability）である。評価結果の活用焦点をおくという点では、パットンにより長年推奨されてきた実用重視の評価（Patton 2008）や、「評価理論の木」における

図3 エビデンスの活用可能性の概念的枠組み



(出所) Mark (2015, p.279)

実用論の幹の主張との親和性が高いが、別の文脈において論じられ発展してきたエビデンスに特化した概念である。活用可能性は、「取りうるアクションの指針としてのエビデンスの適切性の度合」（Mark 2015, p.278）と定義付けられる。また活用可能（actionable）なエビデンスであるためには、そもそもエビデンスの信頼性が高いことが必須であるが、シュワント（2015）はそれ以外の側面が満たされる必要があると唱えている。以下シュワントの提示した活用可能性の概念的枠組みを元にマーク（Mark 2015, p.279）が整理したものを図示する。

エビデンスの活用可能性は、その度合が徐々に高まるものというのではなく、実践現場においてエビデンスが活用されるためには「閾値」があり、その境界線を超えているかという発想で捉えられる。そのためには、以下四つの側面の観点から、エビデンスの質が担保されている必要がある。

- ・エビデンスの信頼性（Credibility）とは、従来からのエビデンスを巡る議論の中心にある、エビデンスが真理の論拠として、信頼に値するかという視点のことを指す。
- ・エビデンスの推論効力（Inferential Potency）とは、特定のエビデンスが主張や仮説をどれだけ強く根拠付けているかということであり、エビデンスの効力の強さや決定力を示す。信頼性の高いエビデンスであっても、結果を推

論するための信頼区間が広がったり結果がいまいだったりしてエビデンスの推論効力が低いと、活用可能性が下がってしまう。

- ・エビデンスの関連性 (Relevance) とは、エビデンスを活用する側にとって、抽出されたエビデンスが個別具体の意思決定に対してどの程度の関連性が確保され得るかに着目するものである。
- ・エビデンスの網羅性 (Comprehensiveness) とは、回答が必要とされている設問に対して、エビデンスがどの程度網羅的に応えられているかである。

以上の四つの要素が組み合わさってエビデンスの活用可能性の判断につながると考えるのがこの枠組みだが、実際には各要素間の相互関連性の問題等もあり、活用可能性を満たす要件についての研究はまだ十分行われているとはいえない。しかし、エビデンスの質を信頼性からその一歩先の活用可能性の視点まで広げて捉えることによって、より活用されやすいエビデンスの構築が構想できるようになる—あるいは活用が予測できない場合はその理由が推測できるようになる—ことはEBPM推進にあたっても有用だと考える。

## (2) 信頼性が高くても活用可能性の低いエビデンス：What Works Clearinghouseの事例

What Works Clearinghouse (以下、WWC情報センター) は米国教育省の傘下の研究機関として2003年に設立された。WWC情報センターの設立背景には、特に教育分野において少なかった「厳密なエビデンス」に裏付けられた科学的に根拠のある方法論の推進と、教育現場における介入の有効性について信頼のおけるエビデンスに基づく判断が必要だという政策方針が存在する。なお、米国教育省は一貫して「厳密なエビデンス」とはRCTによるものだと定義し<sup>6</sup>、WWC情報センターの設定した信頼性の基準を満たした調査を集約すると同時に、そのような調査に対して積極的に予算をつけることを表明してきた。その結果、WWCのデータベースには2003年以降実施されてきた1,555件のRCTの調査結果が記

載されており<sup>7</sup>、一般利用者も参照可能である。

このWWC情報センターが認証したRCTの調査結果、即ち信頼性の側面においては疑義がないはずのエビデンスに対して、2016年にアメリカン・エンタープライズ研究所 (American Enterprise Institute、以下AEI) が「エビデンスの利活用」の観点からメタアナリシスを実施した結果、複数の問題点が指摘された。この報告 (Ginsburg and Smith 2016) を活用可能性の四つの観点から概観してみよう。

AEIの調査はWWC情報センターの数学のカリキュラム (小学校1年～高校最終学年) を対象とした27件のRCTを対象に分析を行っている。27件のRCT調査の中、19件 (70%) は教育現場では利用されなくなった古いカリキュラムを対象としたものであったという結果が出た。これは、エビデンスの関連性の観点から問題であるし、15件 (56%) は対照群が利用していたカリキュラムを明示していなかったため、エビデンスの網羅性の観点からの弱点が明らかにされた。特に後者の指摘は、教育現場において新しいカリキュラムを比較検討している関係者にとっては当然必須な情報であり、それがなくしてエビデンスの活用が損なわれてしまうことは容易に想像がつく。更に、カリキュラムの改訂による効果が発現するには一般的に最低3年間はかかると認識されている中、27件中23件 (85%) の調査は、カリキュラムが導入された初年度にRCTが実施されたことが調査から判明した。これは、推論効力の観点から問題があると考えられる。

このようにいくら信頼性が担保されているエビデンスが集約されていても、活用可能性の観点から分析し直せば、新たな政策・施策・介入に活用する上で障壁となるポイントが明らかになる。なお、本事例においては、適切な状況において実施されれば信頼性の高いエビデンスが期待されるRCTに焦点をおいた事例を元に考察したが、これが他の分析手法によって抽出されたエビデンスにも適用可能という点は言うまでもない。

## (3) 活用可能なエビデンス抽出と文脈の重要性

エビデンスの活用可能性を高めるためにもう

一つ鍵となる側面は、評価が実施される文脈を、エビデンスを抽出する前から十分把握することである。ジュールスとログは、ある評価デザインと手法が、他に比べてより活用可能なエビデンスを産出できるかどうかは、評価が実施される文脈の影響が大きいと主張している (Jules and Rog 2015)。即ち、活用可能なエビデンスを産出できる評価デザインと手法を的確に選択するには、文脈を読み解く力が重要となるのである。

ジュールスとログは、評価が実施される文脈に関して以下五つの側面を挙げ、それぞれどのようなことに留意すべきか整理している (Jules and Rog 2015, pp.222-240)。

- ・意思決定が行われる文脈 (Decision Context) とは、エビデンスが活用される場面の意思決定の種類や関係者の情報ニーズ、また評価が取り組むべきと合意された評価設問<sup>8</sup>の内容や評価タスクのことを指す。そのエビデンスを必要とする意思決定者が誰なのか、またその意思決定者が求めている結論のレベル感 (「観察できる変化は介入によって引き起こされているのか」、「介入の効果は受益者の属性によって違いはあるか」等) や望む結論の形式 (包括的な報告、数値的な結果、有効か否かの結果) 等も重要な文脈とされている。関係者が期待するエビデンスの厳密さを知ること、活用可能な手法の選択に必要不可欠である。
- ・評価対象としている現象にかかる文脈 (Problem/Phenomena Context) とは、評価調査が対象とする社会現象についての文脈を指す。例えば、その社会現象自体 (例：子どもの不登校、空き家の増加、公立校での教員不足等) が簡単に特定されるものか、また現象を引き起こしているメカニズムと因果関係の推定が容易なのかどうか、等についての検討が重要とされている。さらに、アウトカム等がどのようなプロセスを経て発現するのか、介入とアウトカムの道筋が線形なのか、規則性が予測されているのか、介入とアウトカムの間に時間的乖離があるか等の点も把握すべき重要な文脈としている。
- ・介入に関する文脈 (Program/Policy Intervention

Context) は、介入自体が置かれている状況のことで、特に複雑なシステムを対象としている評価の場合は、それに見合った評価手法の選択が必要となる。どの程度プログラム理論が明示されているかや、プログラムとしての成熟度合等も、評価デザインを検討する際に十分検討すべきところとなる。

- ・評価の実施にかかる文脈 (Evaluation Context) は、プログラム実施側の評価レディネスや関係者の評価能力に関する文脈で、評価デザインを決定する上で重要な情報となる。また評価実施の前提となる文脈、例えば評価資源や実施期間等に関する情報も把握すべき内容である。
- ・外部環境の文脈 (Environment Context) 評価が直面する倫理的な文脈、特に調査対象者に対しての配慮や、誰が評価に関わりを持っているか等関係者間の力学、既得権益や既存の格差を推進するような状況になりうるかの確認等、評価をめぐる外部環境の文脈がこれに含まれる。

このような文脈を評価の実践において十分把握する必要があるという指摘自体は、評価研究および実践において決して新しい議論ではない。パットン (Patton 2008) は、評価が置かれている文脈と評価デザインや評価結果の活用への影響について数多く発信してきた。また、西村と呉の論文においては、実在論的評価の認識論において介入の効果を引き起こしている「生成的メカニズムが作動する場合としない場合を分ける文脈的条件を確認することが重視されている」と述べている (西村・呉 2023)。

ここでのジュールスとログの貢献は、文脈的条件を体系的に整理し、かつ活用可能なエビデンスを得るために文脈に基づいて具体的な評価デザインや手法を事前に決定すべきということを示した点であり、活用可能なエビデンスを抽出するための重要な理路が提起されたと考えられる。

#### (4) エビデンスの活用可能性に関する実践例

前出の五つの文脈を読み解きながら適切に四

つの側面を満たす活用可能なエビデンスを産出するという事は、エビデンスを抽出する側と活用する側が、空間的にも時間的にも近い距離感でなければ実現しにくい。エビデンスをつくる側がエビデンスをつかう側から独立して調査を実施し、結果を蓄積して、活用されるのを待つという今日のEBPMの中心にある考え方や方法論は、活用可能性の観点からは実行が難しいことを意味している。では、具体的にどのような実践が可能なのか、以下で二つの事例を紹介する。

### ① Project Evident<sup>9</sup> (PE)

PEは米国にて2017年に活動を開始したネットワーク組織で、長年続いたソーシャル・教育セクターにおける研究者や資金提供者の関心に先導されたエビデンスの収集・集約に異を唱え、現場を抱える実務者を中心に据えた取り組みを、全国35か所を拠点に教育機関や草の根団体と協働で進めている。PEは、それまでの調査や評価が比較的調査リソースが潤沢な事業における説明責任の確保および資金提供の意思決定に焦点が置かれ過ぎているという問題意識を持ち、より社会的脆弱な立場の受益者を支える実務者のニーズの中心にある実施プロセスや、事業改善やスケーリング等についての戦略的な決断を行うために活用できるエビデンスの蓄積が進んでいないと考えた。そして、そのこと自体が社会的格差を助長していると主張する。

この状況を変えることを目的に、PEは一貫して実務者にとって活用可能なエビデンスを生成することを事業として進めている (Ginsburg and Smith 2016, p.2)。具体的には、実務者団体とパートナーシップを構築しながら、実務者が事業に活かせるエビデンスを「Actionable Evidence Framework」というツールを使いながら一緒に特定し、エビデンスの収集や活用までの一連のプロセスにおける能力開発と伴走支援を提供している。

### ② J-PAL<sup>10</sup>

J-PALは、米国の調査機関で国際開発分野においてRCTの普及を推進した中心的組織であり、貧

困削減のために科学的かつ厳密な調査の実施と分析、政策策定におけるエビデンスの活用、そしてそのためのパートナー団体の能力開発に力をいれてきた。J-PALもエビデンスの活用の推進を重視しているが、EBPMではなく「エビデンスに裏付けられた政策策定 (evidence-informed policy making)」という用語で事業の柱を立てており、エビデンス活用者側の主体性を尊重した表現を採用している。具体的な取り組みとしては、貧困削減に優先的に取り組んでいる政府機関やNGO等と直接パートナーシップを構築した上で、既存のエビデンスに裏付けられた政策立案プロセスをサポートしている。また、これと並行して、パートナー機関の中にエビデンスを生成するチームを養成し、有望な介入策を特定し、RCTを実行する専門部署やプロセスを組織内に設けるモデル作りを推進している。

PEにおいてもJ-PALにおいても、エビデンスを活用する主体者に積極的にアプローチし、そこでのエビデンスの生成と活用を能力開発や組織強化を通じて行っていることが最大の特徴であり、エビデンスの活用可能性に重点をおいた取り組みと言えるだろう。

## 5. 新たな問題提起～結論に代えて

ここまでの議論において、エビデンス議論の間口を信頼性から活用可能性に広げた場合に考察すべきことに関し、一定の見解を示した。最後に、この議論をさらに進めた場合、対峙しなければならない二つの論点について、「予告」的に紹介しておきたい。

### (1) 評価の妥当性考察

米国評価学会 (AEA) の2010年の年次大会において、アーネスト・ハウスの「評価の妥当性」に関する著作 (House 1980) を再考するという提案がなされたことを紹介し、エビデンスの信頼性や活用可能性の議論との関わりについて若干の検討を加えておきたい。この年次大会において、「評価の質」が大会テーマとして取り上げられ、同時にハウスの議論が参照点となった。学

会内の「評価理論」分科会 (Topical Interest Group) がパネルディスカッションを企画し、「評価の質に関する対立軸をいかに調停するか？」がテーマとなった (Griffith and Montrosse-Moorhead 2014)。この対立軸とは、評価研究において長く戦われてきた定量対定性論争にほかならず、論争は評価全般の質や妥当性の問題にも波及し、大きな議論となっていた。

評価の妥当性 (validity) は、一般的に因果推論の精緻さの程度で判断される<sup>11</sup>。しかし、ハウス (House 1980) は、評価の妥当性の三要素として「真であること (truth)」、「美しいこと (beauty)」、「正しいこと (justice)」を提示した。これについての再考を試みるのが2010年の大会テーマとなり、しかもこれは、21世紀における評価の妥当性をあらためて検討するという野心的な試みとなった。大会での議論はその後、2014年にAEAの学会誌における特集号にまとめられた (Griffith and Montrosse-Moorhead 2014)。

ハウスの三要素のうちの「真であること」において、エビデンスの役割は大きい。帰納推論において仮説を裏付ける論拠となるのがエビデンスだからである。しかし、社会的事象においては、仮説を完全に立証することは不可能であり、できることは推論 (argument) の精度を上げていくことであるとハウスは言う (House 1980, pp.70-73)。つまるところ、「評価は説得の行為に過ぎない」 (House 1980, p.72) のであり、だからと言ってそれは評価の妥当性を貶めるものではない。評価の妥当性は複合的な要素によって確保されるものだからである。ハウスは、特集号において、評価における妥当性はキャンベル流の内的妥当性と外的妥当性よりも広いものだという論を展開した自身の1980年の著作について振り返った上で、この三つの要素が三十年後においても有効だと述べている (House 2014, pp.10-12)。確かに、ハウスのこの見解は、少なくとも米国の評価者コミュニティの大きな部分で一般的に受け入れられており、本論のエビデンス議論に対しても有用な示唆を与えている。

では、ハウスによる評価の妥当性の他の要素である「美しいこと」、「正しいこと」とはどういうものであるだろうか。まず「美しいこと」

に関して言えば、これは「一貫性 (coherence)」であると紹介されている (House 1980, pp.97-117)。特集号の中で、デビッドソン (Davidson 2014, p.31) は、評価が説得行為であるというハウスの論をもとに、その説得性を増強するものは伝える相手によって言葉遣いや表現方法を変え、首尾一貫した主張を展開することだと確認する。しかし「美しいこと」はそれだけに留まらず、「美しさ」を支えるのは評価的推論 (evaluative reasoning) の力であると言う。「定量的エビデンスは“骨”、定性的エビデンスは“肉”、そして評価的推論は両者に生命を吹き込む“重要な内臓”である」という表現を使っている (Davidson 2014, p.34)<sup>12</sup>。つまり、評価におけるエビデンスは、評価的推論抜きにはその役割を発揮できず、内臓が骨や肉に働きを与えることによって、「美しいこと」が成立し、それが評価の妥当性の一要素を形成するという論になっている。

ハウスによる評価の妥当性の三つ目の要素である「正しいこと」については、以下の評価者と価値との関係も含め、別稿で詳しく整理する必要がある<sup>13</sup>。そもそも、評価によって事業や取り組み、ひいては社会の状態が改善されることが志向されるのであれば、それは誰にとってのどのような改善なのだろうか。そしてその善悪の価値判断は、誰が行うべきなのだろうか。ハウスは、このような考察は、評価活動の外にあるものではなく、それに内在し、評価の妥当性そのものを構成する重要な要素であると指摘する。

以上の評価の妥当性に関する議論は、本論で展開しているエビデンスの考察にも有用な示唆を与えている。エビデンスの考察が、信頼性の議論から出発して活用可能性の議論へと広がっていくと、評価理論や評価の妥当性全般の中でエビデンスの考察がいかなる位置付けを持つものとなるのか、一層注意深い検討が必要となっていく。

## (2) 評価者の役割、評価者と価値との関係

ここまでの議論を踏まえると、評価の妥当性を高める上での評価者の役割についての再考が

必要になってくる。

前項のハウスの言葉を引用すれば、評価者の第一義的な役割は「真であること」を追求し、信頼性の高いエビデンスをもって真理を描いていくことである。厳密な評価デザインを設計し、バイアスを最小限に抑えたデータ収集と分析をし、そしてそれに基づいた結果のとりまとめを執り行うことは、評価者としての基本的な役割と言える。この役割は一般的に期待される「外部評価者」や「第三者評価者」の姿と合致し、社会の要請に最も応えるものでもあるといえよう。

一方で、評価の妥当性の第二の要素である「美しいこと」、即ちデビッドソンが言う評価的推論を行う評価者の役割は、評価の専門性において見逃されがちな点である。評価の説得行為、即ち評価に係る利害関係者の関心を元に問いをたて、それに対する確にエビデンスをもって納得のいく結論を説いていく過程においては、事業・取り組みの目的を基本に、事業や評価に関わる関係者の優先課題や力関係を的確に調合し、異なる関係者にとって納得感の高い価値判断を下していかなければならない。評価推論に関する議論が繰り返しエビデンスを意識したものになるよう促し、最終的には関係者が納得のいく結論まで導き出すためのファシリテーターのような役割を果たすことも、評価者の極めて重要な役割と言える。

そして最後に、評価の妥当性の第三の要素である「正しいこと」は、評価者の役割についてのどのような示唆を与えているのであろうか。ポデムス (Podems 2019, p.228) は、評価者の社会的役割について、1) 評価者は社会の変化には関与しない、2) 評価者は社会の変化に通じるような事業や政策の意思決定のために情報や知識を提供する人々である、3) 評価者は社会の変化を起こす人々 (チェンジ・エージェント) であるという三つの見解があると紹介している。評価が評価対象の改善を通じての社会全体の改善に寄与するものだとしても、第一の立場に立てば、評価者は直接それには関与しないということになる。しかし、価値論の幹に属する評価理論家・実践者に限らず、北米等の評価者コミュニティ

を見ると、上記の第二または第三の立場をとる評価者一すなわち、観察者の価値中立性を首肯しない論者—が多いと見受けられる。特に、評価活動そのものを社会的格差の是正等社会正義の実現を目指す営みと捉える考え方は、ここ数年で勢いを増している<sup>14</sup>。

この文脈を受け、エビデンスの議論においても、「誰のためのエビデンスか」という問いも重要になってきており、評価者の役割としてもエビデンスを介入の関係者だけで活用することに留まらず、より社会的脆弱な立場にいる受益者、あるいはそのようなコミュニティを支える支援者まで視野を広げ、より活用可能なもののできるために尽力することが求められるといえよう。

## 注記

- 1 内閣府におけるEBPMへの取組。https://www.cao.go.jp/others/kichou/ebpm/ebpm.html (アクセス：9/27/2023)
- 2 Levels of evidence in research. https://scientific-publishing.webshop.elsevier.com/research-process/levels-of-evidence-in-research/ (アクセス：11/26/2023) .
- 3 GRADE: 医療や保健に関するエビデンスの質を評価するためのシステム。CASP: システムティックレビューやランダム化比較試験等の研究論文の質を評価するためのチェックリスト。CRAAP: ウェブサイトやニュース等の情報源の信用性を評価するための基準として用いられる。
- 4 例 えば、Ebrahim (2019) や、Gugerty and Karlan (2018)。
- 5 後者は、科学論 (科学史・科学哲学) においては「構成主義 (constructivism)」または「社会構成主義 (social constructivism)」という呼び名、社会学においては「構築主義 (constructionism)」という呼び名で言及されることが多い (松村 2021, p.32)。
- 6 この米国教育省の政策に対して、スクリヴェンをはじめとする、米国評価学会による反論声明と、その反論声明に異を唱える評価学会員有志による声明が発表され、米国評価学会内での「厳密なエビデンス」論争に至る (Donaldson 2015, pp.12-16)。
- 7 WWC情報センター。https://ies.ed.gov/ncee/wwc/

- FWW (アクセス: 9/26/2023)
- 8 ジュールズとログは、シュワント同様に、「エビデンス」は有効性に対する根拠に限らず評価において答えようとする設問に対しての根拠、と主張している (Schwandt 2015a)。
- 9 Project Evident <https://projectevident.org/> (アクセス: 9/27/2023)
- 10 J-Pal <https://www.povertyactionlab.org/evidence-to-policy/institutionalizing-evidence-use> (アクセス: 9/27/2023)
- 11 訳語の問題として、*validity*と*relevance*が「妥当性」という同じ用語に訳されていることは注意が必要である。DAC評価6項目の一つでもある*relevance* / 妥当性とは、事業実施がニーズや課題に合致しているかを問うもので、ここで議論されている評価の*validity* / 妥当性とは別の概念である。
- 12 この評価的推論 (evaluative reasoning) と評価的思考 (evaluative thinking) の違いについて、筆者の一人 (今田) がDavidson本人に尋ねたところ、「評価的推論とは、エビデンスと価値を足し合わせて評価の結論に至る一連のプロセス。評価的思考はより広い概念で、批判的思考の一種を評価の文脈に当てはめたもの」という回答だった (Davidson から今田への私信より、5/21/2019)。
- 13 三つ目のこの要素は、評価理論の三つの幹の中の価値論の幹の発展、特にその到達点であるTransformative EvaluationやFourth Generation Evaluationと関係する (Christie and Alkin 2023, pp.34-37)。特にここ数年の米国評価学会においてCulturally Responsive Evaluationへの注目度が増しており、この三つ目の要素が評価の妥当性の要であるという主張が強くなっている。
- 14 とはいえ、評価者はそのプロフェッショナルな立場においては「社会活動家」ではない。その線引きがどうあるべきかの議論もかなり丁寧かつ綿密に行われている。
- 参考文献**
- 浅岡浩章 (2016) 「国際開発分野におけるエビデンス活用の現状と課題」、『日本評価研究』、17 (1) :19-32
- 一般社団法人日本情報経済社会推進協会 (2015) 「エビデンス管理の課題と対策—コンプライアンス強化と記録・情報の利活用に向け」<https://www.jipdec.or.jp/archives/publications/J0005042>
- 岩崎久美子 (2017) 「エビデンスに基づく教育：研究の政策活用を考える」、『情報管理』、60 (1) :20-27
- 小林庸平 (2020) 「日本におけるエビデンスに基づく政策形成 (EBPM) の現状と課題」、『日本評価研究』、20 (2) :33-48
- 佐々木亮・正木朋也 (2020) 「「エビデンスに基づく政策立案 (EBPM)」の現状と課題」、『日本評価研究』、20 (2) :1-2
- 西村君平・呉書雅 (2023) 「実在論的評価の理論と日本のEBPMへの示唆」、『日本評価研究』24 (1) :29-44
- 日本評価学会 (2020) 「特集：エビデンスに基づく政策立案 (EBPM) の現状と課題」、『日本評価研究』、20 (2)
- 松村一志 (2021) 『エビデンスの社会学：証言の消滅と真理の現在』、青土社
- Alkin, M. C., and Christie, C. A. (Eds.) (2023). *Evaluation Roots: Theory Influencing Practice (3rd edition)*. New York, New York: The Guilford Press.
- Bernstein, R. J. (1983). *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis*. Philadelphia, University of Pennsylvania Press.
- Christie, C. A., and Fleisher, D. (2015). Social Inquiry Paradigms as a Frame for the Debate on Credible Evidence. In Donaldson, Christie, and Mark (Eds.), *Credible and Actionable Evidence: The Foundation for Rigorous and Influential Evaluations (2nd edition)*. Thousand Oaks, California: Sage Publications, Inc., 27-38.
- Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (4th edition)*. Thousand Oaks, California: Sage Publications, Inc.
- Davidson, E. J. (2014). How “Beauty” Can Bring Truth and Justice to Life. In Griffith and Montrosse-Moorhead (Eds.), *Revisiting Truth, Beauty, and Justice: Evaluating with Validity in the 21st century. New Directions for Evaluation*, 142, 31-43.
- Donaldson, S. I., Christie, C. A., and Mark, M. M. (Eds.) (2008). *What Counts as Credible Evidence in Applied*

- Research and Evaluation Practice?* Thousand Oaks, California: Sage Publications, Inc.
- Donaldson, S. I. (2015). Examining the Backbone of Contemporary Evaluation Practice: Credible and Actionable Evidence. In Donaldson, Christie, and Mark (Eds.), *Credible and Actionable Evidence: The Foundation for Rigorous and Influential Evaluations (2nd edition)*. Thousand Oaks, California: Sage Publications, Inc., 205-220.
- Ebrahim, A. (2019). *Measuring Social Change: Performance and Accountability in a Complex World*. Stanford, California: Stanford University Press.
- Feyerabend, P. (1975). *Against Method: Outline of an Anarchist Theory of Knowledge*. New York and London: New Left Books.
- Greene, J. C. (2015). How Evidence Earns Credibility in Evaluation. In Donaldson, Christie, and Mark (Eds.), *Credible and Actionable Evidence: The Foundation for Rigorous and Influential Evaluations (2nd edition)*. Thousand Oaks, California: Sage Publications, Inc., 205-220.
- Ginsburg, A., and Smith, M. S. (2016). *Do Randomized Controlled Trials Meet the "Gold Standard" ? A Study of the Usefulness of RCTs in the What Works Clearinghouse*. American Enterprise Institute.
- Griffith, J. C., and Montrosse-Moorhead, B. (2014). The Value in Validity. In Griffith and Montrosse-Moorhead (Eds.), *Revisiting Truth, Beauty, and Justice: Evaluating with Validity in the 21st century*. *New Directions for Evaluation*, 142, 17-30.
- Gugerty, M. K. and Karlan, D. (2018). *The Goldilocks Challenge: Right-Fit Evidence for the Social Sector*. Oxford: Oxford University Press.
- House, E. R. (1980). *Evaluating with Validity*. Thousand Oaks, California: Sage Publications, Inc.
- House, E.R., and Howe, K. (1999). *Values in Evaluation and Social Research*. Thousand Oaks, California: Sage Publications, Inc.
- House, E. R. (2014). Origins of the ideas in Evaluating with Validity. In Griffith and Montrosse-Moorhead (Eds.), *Revisiting Truth, Beauty, and Justice: Evaluating with Validity in the 21st century*. *New Directions for Evaluation*, 142, 9-15.
- Jules, G., and Rog, D. (2015). Actionable Evidence in Context. In Donaldson, Christie, and Mark (Eds.), *Credible and Actionable Evidence: The Foundation for Rigorous and Influential Evaluations (2nd edition)*. Thousand Oaks, California: Sage Publications, Inc., 221-258.
- Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.
- Marchal, B (2012). Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. *Evaluation*. 18 (2), 192-212.
- Mark, M. M. (2015). Credible and Actionable Evidence: A Framework, Overview, and Suggestions for Future Practice and Research. In Donaldson, Christie, and Mark (Eds.), *Credible and Actionable Evidence: The Foundation for Rigorous and Influential Evaluations (2nd edition)*. Thousand Oaks, California: Sage Publications, Inc., 275-309.
- Papineau, D. (Ed.) (1996). *The Philosophy of Science*. Oxford and New York: Oxford University Press.
- Patton, M. Q. (2008). *Utilization-Focused Evaluation: The New Century Text (4th edition)*. Thousand Oaks, California: Sage Publications, Inc.
- Pawson and Tilley (1997). *Realistic Evaluation*. London: Sage Publications.
- Podems, D. R. (2019). *Being an Evaluator: Your Practical Guide to Evaluation*. New York, NY: The Guilford Press.
- Schwandt, T. A. (2015a). Credible Evidence of Effectiveness: Necessary but Not Sufficient. In Donaldson, Christie, and Mark (Eds.), *Credible and Actionable Evidence: The Foundation for Rigorous and Influential Evaluations (2nd edition)*. Thousand Oaks, California: Sage Publications, Inc., 259-274.
- Schwandt, T. A. (2015b). *Evaluation Foundations Revisited*. Stanford, California: Stanford University Press.
- Schwandt, T. A., and Gates, E. F. (2021). *Evaluating and Valuing in Social Research*. New York, NY: The Guilford Press.
- Scriven, M. (2012). The Logic of Valuing. In G. Jules (Ed.), *Promoting Valuation in the Public Interest: Informing Policies for Judging Value in Evaluation*. *New Directions for Evaluation*, 133, 17-28.
- Scriven, M. (2015). Demythologizing Causation and

Evidence. In Donaldson, Christie, and Mark (Eds.), *Credible and Actionable Evidence: The Foundation for Rigorous and Influential Evaluations(2nd edition)*. Thousand Oaks, California: Sage Publications, Inc., 115-132.

Upshur R.E., VanDenKerkhof E.G. and Goel V. (2001). Meaning and Measurement: An Inclusive Model of

Evidence in Health Care. *Journal of Evaluation in Clinical Practice*, 7(2), 91-96.

Wanzer, D. L. (2021). What Is Evaluation?: Perspectives of How Evaluation Differs (or Not) From Research. *American Journal of Evaluation*, 42(1), 28-46.

(2023.12.20 受理)

## **Reconsideration of Contemporary Debates Around Evidence: Including Credibility and Actionability As Key Concepts**

Katsuji Imata

Blue Marble Japan Inc.

katsuji.imata@blue-marble.co.jp

Takara Tsuzaki

Western Michigan University,  
Interdisciplinary Ph.D. in Evaluation

takaratsuzaki@gmail.com

Minako Nakatani

Evaluation Consultant

minakonakatani@gmail.com

### **Abstract**

In recent years, discussions on evidence have been taking place in various fields such as healthcare, education, and international development. These discussions have also generated a multi-layered discourse about evidence in evaluation research and practice. However, it is difficult to consider that there are common understandings or recognitions of shared concepts. In this paper, we review the polysemy of the term “evidence” and discuss its usefulness, particularly its credibility and actionability. First, we introduce the fundamental conflicts in scientific discourse, the way evidence is treated in evaluation studies, and researchers’ and evaluators’ worldviews as they relate to the treatment of evidence. Second, we introduce four aspects of multifaceted elements that determine the actionability of evidence and present five contextual factors that are key to extracting actionable evidence. Finally, we provide two practical examples that ensure actionability and reconsider the role of evaluators in enhancing the validity of evaluations.

### **Keywords**

Evidence, Actionability of Evidence, Credibility of Evidence, Validity of Evaluation

## 【コメント】

## EBPM 批判の現在

西村 君平

東北大学

kunpei.nishimura.a6@tohoku.ac.jp

呉 書雅

岩手県立大学

wu\_s@iwate-pu.ac.jp

## 1. 論旨の確認

本論文は、近年のアメリカの評価研究を紐解き、エビデンスの信頼性に加えて、エビデンスの活用可能性がEBPMの推進にとって重要となることを提示するものである（本文ではエビデンスの活用可能性とは、取りうるアクションの指針としてのエビデンスの適切性の度合いと定義されている）。筆者たちは、信頼性が高くても活用可能性が低いエビデンスがあり得ること、そしてそういった場合にどのように対処していくことができるのかについて多角的に検討している。

本論文の主要な特徴の1つは、科学論を参照することで「信頼性だけではなく活用可能性も大事だ」という主張の背後にある認識論を明らかにしようとしている点である。この点において、本論文はエビデンスを「つかう」局面に関する議論（田辺 2020）をさらに発展させていくための有益な知見を提供している。

本論文では様々な認識論的立場が取り上げられているが、私たちコメンテーターの印象では筆者たちは特に社会的構築主義に共感的なようだ。この解釈をもとに論文の論旨を私たちに整理すると次のようになる。

どれほど信頼性が高いエビデンスであったとしても、そもそもどのような処置のどのような効果を取り扱うか（≒問い直すか）には個人の主観や集団の規範、ひいては個人や集団を取り巻く状況が入り込む。この意味で、エビデンスは、少なくとも部分的には社会的に

構築されている。よって、信頼性の高いエビデンスだとしても、エビデンスを構築する側と活用する側の諸条件（主観、規範、状況）がずれている場合には、エビデンスが当てはまらなかったり使えなかったりすることがある。だからこそEBPMの推進のためにはエビデンスの文脈に潜伏した諸条件を絶えず省察し、検証するべきである。その知見をもとにエビデンスを構築する側と活用する側を架橋していく必要があるし、そのための評価者の役割やスキルも整備していく必要がある。

## 2. 地球平面論者あるいはRCT懐疑論者は今どこで何をしているのか

私たちは本論文の要旨に納得しつつも、どことなく肩透かしを食ったような感覚を覚えた。その理由はかつてRossiが「地球平面論者」と揶揄したようなRCTおよびRCTの方法論を主軸にしたEBPMに対する懐疑論の顛末が少し不透明なことによる。

2001年のNo Child Left Behind Actをきっかけとして、当時のアメリカでは科学的根拠（エビデンス）に基づく教育政策に優先的に予算を与える階層付き補助金（Tiered Grant）が施行された。これに対してアメリカ評価学会はScrivenのリーダーシップのもとで「科学的根拠に基づくというのは支持できるが、実験デザインを最上位とすることには反対だ。因果関係を評価するためには、実験デザインが適用できない場合もあるし、実験デザインを用いなくても特定できる。」

と反対声明を表明した。さらにこの反対声明に対してRossiが反発し「なぜ地球平面協会のメンバーでなければならないのだ」との言葉を残し、学会を脱退する騒動となった（佐々木2007）。

もちろん本論文でもこうした顛末については触れられている。ただし、本論文を通してみると、アメリカの評価研究はEBPMのあり方に対して非常に融和的な態度を採っているようで、RCTに対する懐疑も鳴りを潜めているように見受けられる。EBPMを推進する立場からRCTを適切に運用しようという発想が色濃くなっている。「信頼性だけでなく活用可能性も大事だ」という考え方はまさにそうである。こうした考え方は非常に建設的なものであり、何も悪くない。ただ、かつての騒動を思い起こすと、あの頃のアメリカの評価研究で燃え上がったEBPM批判は現在どうなってしまったのか気になってしまう。本論文の論調含め、アメリカの評価研究の展開

は穏当ではあるものの、RCTの限界や暗黙の前提などを抉り出すようなラディカルな議論があってもいいのではないかなどと、無責任に期待してしまう。

そこでこの場を借りて筆者たちに質問を投げ掛けたい。今日のアメリカの評価研究において、EBPM批判、なかんずくRCTに対する懐疑論はどうなっているのか。

#### 参考文献

- 佐々木亮（2007）「評価研究の巨人ロッシ教授が残した影響といくつかの主要な論争」、立教大学大学院21世紀社会デザイン研究科・日本評価学会社会実験分科会共催セミナー（2007年12月3日）発表資料
- 田辺智子（2020）「エビデンスに基づく政策立案（EBPM）の推進に向けて—医療の経験からの示唆」、『日本評価研究』、20(2): 19-31

## 【リプライ】

## なぜミックスト・メソッドが普及したか

今田 克司

株式会社ブルー・マーブル・ジャパン

katsuji.imata@blue-marble.co.jp

本論において筆者らは、評価研究における「パラダイム・ウォー」に言及した。評価手法における定量対定性手法の選択の対立化、先鋭化を表現したのだが、これは単なる手法の論争ではない。本論では、その背景にあるものとして、科学論における「～イズム」の系譜について概説している。西村・呉コメントで指摘されたように、筆者らが「社会的構築主義に共感的」なのは間違いない。しかし同時に、筆者らは、何らかの「～イズム」に依拠し、それによって特定の方法論を選択することをよしとするものではない。

方法論のレベルにおいては、筆者らは、いわゆる不可知主義 (agnosticism) の立場を採る。不可知論とは、「神は存在するか」という命題に対して「我はそれを知り得ない」という言明を積極的に取りに行くもので、その言外には「いつまでも神学論争をしていても仕方がない」という主張がある。

西村・呉コメントの最後の投げかけである「今日のアメリカの評価研究において、EBPM批判、なにかなくRCTに対する懐疑論はどうなっているのか」という質問を見て、筆者らが抱いたのはこの言外の感想に近い。もちろん、RCTをゴールドスタンダードとする手法論争がなくなったわけではない。例えば、2019年に開発経済学のバナジー、デュフロ、クレマーの三氏がノーベル経済学賞を受賞し、RCTに代表される実験的アプローチが脚光を浴びたのを受けて、RCTを批判的に検証する冊子が編纂されたりもしている (Bédécarrats, Guérin and Roubaud 2020)。しかし、「今日のアメリカの評価研究」を見る限り、この争い

には停戦合意がなされていると言って良い。

停戦に至るいきさつは、本論で触れたドナルドソンらの著作 (Donaldson, Christie and Mark 2015) に記されている。西村・呉コメントにもある、米国の教育政策補助金制度に対する2003年のアメリカ評価学会の反対声明に端を発する論争は、2006年に開催された「信頼性のあるエビデンスとは何か？」シンポジウム<sup>1</sup>、これをまとめた2009年のドナルドソンらの著作、そして本論で取り上げたその第二版へと議論が進展していくなかで、エビデンスに関する多様・多層の理解が認められたことによって収束へ向かっていった。そこでは、「信頼性」に加えて「活用可能性」がエビデンスの考察の大きな要件として加えられたのは本論で論じたとおりである。

とはいえ、「～イズム」の対立が終焉したわけではなく、あくまでも停戦合意だというのが筆者らの見立てである。そして、筆者らが不可知主義を採用するのも、米国の評価研究者・実践者の多くがこの立場を採用しているからと言ってよい。評価が「社会改善の活動」ならば、「いつまでも神学論争をしていても仕方がない」のであり、評価の目的や置かれている状況、制約条件を勘案し、最適な方法を選択するのが評価者としての役割だろう。ホール (Hall 2013) は、この姿勢を、デューイ流のアメリカン・プラグマティズムに裏打ちされたものであると分析している。

そして、このプラグマティズムが方法論のレベルで具現化されたものが、「ミックスト・メソッド (mixed-methods)」と称される、定量、定性手法を混合して活用するアプローチである<sup>2</sup>。ミ

ックスト・メソッドについて、アメリカ評価学会の学会誌である *New Directions for Evaluation* では、これまで二回、1997年と2013年に特集号を組んでいる。1997年の特集号において編者のグリーンとカラセリ (Greene and Caracelli 1997) は、評価手法の「混ぜ方」には基本的に (1) 混ぜることをよしとしない「純粋派 (purist stance)」、(2) 評価対象が置かれている文脈等により手法を選択する「実践派 (pragmatic stance)」、(3) 評価手法が依拠する世界観の違いから安易な「混合」は避けるべきとしつつ、「混合」により評価の質が向上すると考える「弁証派 (dialectic stance)」の三つの派が存在すると整理している (p.1)。

それから16年後の2013年度の特集号において、グリーン (Greene 2013) は、1997年時点の整理を進化させ、ミックスト・メソッドに四つの立場があるとしている。「弁証派」はそのままとし、その上で「実践派」または「弁証派」の進化形として、多様な手法の強みを最終的につなぎ合わせる「相互補強派 (complementary-strengths stance)」と「～イズム論争」に陥ることを避ける「反パラダイム派 (a-paradigmatic stance)」を置き、さらに社会変革を評価の目的とした場合の手法の混成を検討する「代替パラダイム派 (alternative paradigm stance)」を置いている (p.111)。注目すべきは、1997年時点の整理で存在していた「純粋派」が消滅していることである。評価手法のあり方としてミックスト・メソッドが普及したことを物語っているのではないだろうか。

最後に、2013年の特集号において、ホワイト (White 2013) が、RCTの有用性を認めつつ、RCTは因果推論におけるアトリビューション (帰属性) を立証する反事実分析で強みを発揮するものの、それと定量定性のデータ分析を組み合わせることで因果推論を立体的に組み立てる評価デザインが必要だと論じている (p.62) ことに注目しておきたい。この論にもミックスト・メソッドの有効性が見られる。

このように、今日のアメリカの評価研究において、ミックスト・メソッドの普及によってRCTは取りうる手法の一つとなっているというのが現状であると言ってよいだろう。一方、EBPM推

進は進んでいる。しかし、EBPMを支えるエビデンスやその質に関する考え方が進化しているものであり、EBPM推進イコールRCT活用ではない。本論の「結論に代えて」において、評価の妥当性 (validity) に関するさらなる考察が必要だと予告的に紹介したが、EBPMについての的確に検討するための評価の妥当性に関する論考は、別稿で試みることにしたい<sup>3</sup>。

## 注記

- 1 <https://research.cgu.edu/dbos-events-and-conferences/home/what-constitutes-credible-evidence-in-evaluation-and-applied-research> (12/27/2023 アクセス)
- 2 ミックスト・メソッドの日本語訳としては、混合メソッド (手法)、混成メソッド (手法) などが考えられるが、ここでは、本文の「弁証派」のように「混ぜる」ことの危うさを指摘する向きもあり (であれば併存メソッド (手法) などの用語を充てるべきか)、特定の日本語のニュアンスに引っ張られることを避けるため、カタカナの用語を活用した。なお、MMIRA (Mixed Methods International Research Association) の日本支部的な位置付けのJSMR (Japan Society for Mixed Methods Research) は、日本混合研究法学会と命名されている。(http://www.jsmmr.org/home.html 12/27/2023 アクセス)
- 3 エビデンス、EBPM、評価の妥当性を結びつける論考として、本論で参考にした文献以外にガルガーニとドナルドソンの論文 (Gargani and Donaldson 2011) などがある。

## 参考文献

- Bédécarrats, Florent, Isabelle Guérin and François Roubaud (Eds.)(2020). *Randomized Control Trials in the Field of Development: A Critical Perspective*. Oxford: Oxford University Press.
- Donaldson, Stewart I., Christina A. Christie, and Melvin M. Mark (Eds.)(2015). *Credible and Actionable Evidence: The Foundation for Rigorous and Influential Evaluations (2nd edition)*. Thousand Oaks, California: Sage Publications, Inc., 205-220.
- Gargani, Jonn and Stewart I. Donaldson (2011). What Works

- for Whom, Where, Why, for What, and When? Using Evaluation Evidence to Take Action in Local Contexts. In H. T. Chen, S. I. Donaldson, & M. M. Mark (Eds.), *Advancing Validity in Outcome Evaluation: Theory and Practice*. *New Directions for Evaluation*, 130, 17-30.
- Greene, Jennifer C. (2013). Reflections and Ruminations. In D. M. Mertens & S. Hesse-Biber (Eds.), *Mixed Methods and Credibility of Evidence in Evaluation*. *New Directions for Evaluation*, 138, 109-119.
- Greene, Jennifer C. and Valerie J. Caracelli (1997). Advances in Mixed-Method Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms. *New Directions for Evaluation*, 74.
- Hall, Jori N. (2013). Pragmatism, Evidence, and Mixed Methods Evaluation. In D. M. Mertens & S. Hesse-Biber (Eds.), *Mixed Methods and Credibility of Evidence in Evaluation*. *New Directions for Evaluation*, 138, 15-26.
- Mertens, Donna M. and Sharlene Hesse-Biber (Eds.) (2013). *Mixed Methods and Credibility of Evidence in Evaluation*. *New Directions for Evaluation*, 138.
- White, Howard (2013). The Use of Mixed Methods in Randomized Control Trials. In D. M. Mertens & S. Hesse-Biber (Eds.), *Mixed Methods and Credibility of Evidence in Evaluation*. *New Directions for Evaluation*, 138, 61-73.



## 【研究論文】

## 実在論的評価の理論と日本の EBPM への示唆

西村 君平

東北大学

kunpei.nishimura.a6@tohoku.ac.jp

呉 書雅

岩手県立大学

wu\_s@iwate-pu.ac.jp

## 要 約

日本のEBPMの課題の一つとして、科学的根拠の外的妥当性の担保が挙げられる。本稿ではこの課題の解決に向けて実在論的評価の理論に着目する。実在論的評価の焦点の一つはランダム化比較実験を応用した実験的評価（experimental evaluation）の外的妥当性の批判的検討にある。評価方法論上の特徴は科学的根拠の構築や活用の際に、データ分析以上に理論構築を重視する点にある。実在論的評価は国内外の評価研究で注目を集めつつあるが、認識論的前提やその前提が評価理論にどのように反映されているかは不透明であり、このことが理論の理解や発展、実践への適用を妨げている。そこで本稿では実在論的評価の認識論的次元に焦点をあて、その理論の要諦を明らかにした。この知見を元に、実験的評価を主軸としたEBPMの限界とその克服の方策を理論的に明らかにした。その方策とは外的妥当性の高い評価結果の希求から応用可能性の高い中範囲理論の構築へEBPMの焦点を移すことである。

## キーワード

EBPM、実在論的評価（realist evaluation/realistic evaluation）、認識論

## 1. 背景と目的

本稿の目的は二つある。一つめは実在論的評価（realist evaluation/realistic evaluation）<sup>1</sup>の特徴を認識論的次元<sup>2</sup>に着目して明らかにすることである。二つめは、この知見に基づき、科学的根拠（evidence）の外的妥当性の担保の理路を明らかにすることである。

EBPM（Evidence-based Policy Making）は日本の政策的・社会的課題の一つである。EBPMとは政策<sup>3</sup>の立案や評価では科学的根拠が参照されるべきだという考え方である。科学的根拠の内容や形式は多様だが、その中心は政策のアウトカム（因果効果）<sup>4</sup>を明らかにした知見にある。そしてアウトカムを明らかにするための理想的な方

法はランダム化比較実験（Randomized Controlled Trial:以下適宜 RCTと略）とメタアナリシスだと考えられている。しかし、近年RCTの外的妥当性に疑義が生じている。

RCTは、対象者を実験群と対照群にランダムに割りつけることで処置以外の要素を実験群と対照群の間で均一化する。これにより処置のアウトカムに分析の焦点を当てる（Shadish et al. 2001, p.13）。こうした特徴ゆえに、RCTは内的妥当性（因果関係に関する推論の確かさ：Shadish et al. 2001, p.38）が高い。ただし、外的妥当性（因果的推論が別の対象や状況、処置、手法のもとで行われる研究にも適用できる程度：Shadish et al. 2001, p.38）については留保が付される。RCTをプログラム評価に応用した実験的評

価 (experimental evaluation: 呼称はPawson and Tilley (1997) による) がまさにこれである。

プログラム評価の場合、完全なランダム割り当てはほとんど不可能であり、攪乱要因の統制には限界がある。その結果、実験対象と科学的根拠を活用する対象に差異が生じる。この点に無配慮に特定の実験の結果を別の対象に適用すると外的妥当性の問題が生じる。つまり「ここ(実験)」と「そこ(活用)」の間にギャップが存在する場合、RCTの外的妥当性に疑念が生じるのである (Cartwright and Hardie 2012)。

このようにRCTの外的妥当性の問題は「RCTでは知り得ない何かがあるのではないか」という認識論上の問いを生じさせる。こうした問題意識は、認識論という概念こそ使われていないものの、日本の評価研究でも共有されている。小林 (2020) はRCTに固執せずに外的妥当性の高い評価を実現するための理路を解明していくことをEBPM推進の課題として指摘している。

この課題の解決に資する評価理論として、本稿は实在論的評価に着目する。实在論的評価は、理論駆動型評価 (theory-driven evaluation) の一種である。その主眼は「何が機能するのか (what works)」ではなく「何が、誰に、どのような条件のもとで、どのような点で、どのように機能するのか」にある。この目的の達成に向けて、社会プログラムのアウトカムに加えて、アウトカムを生成するメカニズム、メカニズムを作動させる文脈的条件を明らかにしたCMO型 (context-mechanism-outcome configuration) の理論を構築する点に方法論上の特徴がある (Pawson and Tilley 2005, pp. 364-365)。

本稿が实在論的評価に着目する理由は二つある。一つめの理由は实在論的評価の主要な問題意識の一つが評価、特に実験的評価の外的妥当性の批判に向けられており (Pawson and Tilley 1997, p. 28)、本稿と基本的な問題意識を共有しているからである。二つめの理由は、实在論的評価が科学的实在論 (scientific realism: 文脈に応じて「科学的」を略す) に依拠して、EBPMの基本概念である科学的根拠や因果関係を問い直しているからである (Pawson and Tilley 1997, pp. 55-56)。实在論的評価に内包された認識論は、外

的妥当性について再考するための一つの基礎を提供すると期待される。

一方、实在論的評価にも課題がある。实在論的評価が依拠する科学的实在論は決して系統的な理論や思想ではなく、様々なバリエーションがある (戸田山 2015, p.6)。そのため实在論的評価が科学的实在論のどの部分を参照し、そこからどのような示唆を導出しているのか、实在論的評価の評価理論としての特徴はどこにあるのかは自明ではない。

そこで本稿では、評価研究から一旦離れて、科学哲学における实在論の位置づけや含意について検討するというアプローチを採る。このアプローチによって、实在論的評価の認識論的次元を考察するための立脚点を構築し、その見地から实在論的評価の特徴と外的妥当性に対する示唆を明らかにする。

## 2. 先行研究の検討と本稿の課題

### (1) 实在論的評価に関する先行研究の概観

2005年、全米評価学会が発行した百科事典ではPawsonとTilleyによって实在論的評価の概要が紹介されている。先述の实在論的評価の特徴は、Pawson and Tilley (2005) に依拠している。

国際的な評価研究では、2000年頃から实在論的評価は注目を集めている。New Directions for Evaluation (Henry et al. 1998) では实在論的評価の特集が組まれている。この特集では、实在論的評価の拡張が積極的に検討され、創発实在論的評価 (emerging realism evaluation) という新しい方法論が提唱されている。Schmitt (2020, p.15) では实在論的評価は評価に関する因果的推論の主要な立場の一つとして紹介されている。

国内の評価研究でも实在論的評価に関する注目は高まっている。春田他 (2019) は实在論的評価に関する様々な適用例を検討した優れた総説である。他にも桐村 (2013) や西村・吳 (2022) では、関連領域の動向の整理の中で实在論的評価の考え方を紹介している。

## (2) 実在論を巡る混乱

先行研究の限界は、実在論的評価の認識論に関する精査が不十分な点である。このことが実在論的評価の認識論の理解を混乱させている。

実在論的評価にはEBPMの黎明期に見られた実験・準実験的方法論に対する過信に対するアンチテーゼとして注目されてきた経緯がある。その際に、社会科学で支配的な実証主義対社会的構築主義の図式を念頭に実在論的評価が受容された。結果として、実証主義のイメージ（量的調査、変数重視、法則定立：人間や社会に普遍的に妥当する法則や原理を確立すること）と社会的構築主義のイメージ（質的調査、事例重視、個性記述：個人や集団の実態をありのまま描写していくこと）を前提に、実在論的評価は後者に近い認識論的な立場と位置づけられる傾向にある。こうした傾向は桐村（2013）、春田他（2019）、Schmitt（2020、p.15）に共通して見られる。しかし、後述する通り、科学哲学において、実在論は実証主義対社会的構築主義の二項対立図式とは異なる位置づけを持っている。ここに齟齬が生じている。

Henry (et al. 1998) は創発実在論的評価という新しい理論の提案に主眼があり、この混乱の解消には有益ではない。西村・呉（2022）では高等教育研究への適用に向けて実在論を検討しているが、実在論的評価の取り扱い是非常に薄い。

## (3) 本稿の課題

先行研究を踏まえて①実証主義対社会的構築主義の二項対立図式とは異なる科学的実在論の認識論上の特徴とその評価研究への含意を明確化すること、②科学的実在論に関する理解を土台として、実在論的評価の実験的評価に対する問題意識を検討すること、③同様の見地から、実在論的評価の方法論を明確化すること、④実在論的評価の理論から日本のEBPMへの示唆を導出することを本稿の課題とする。

## 3. 実在論的評価の認識論

### (1) 科学的実在論

Papineau（1996、pp.1-20）は、科学哲学における実在論を実証主義と相対主義と対比する。以下ではPapineau（1996）をもとに、評価研究と関連づけながら三者の特徴を検討する。

実証主義、相対主義、実在論の関係は、認識の独立性「世界と人間の認識は独立していること」と認識の可能性「人間が世界を認識できること」の二つの論点で整理できる。

実証主義は、認識の独立性を信じる。一方、認識の可能性については「観察可能な範囲に限って、人間は世界を認識できる」と主張する。この二つの主張は、観察や実験により得られたデータによって科学を基礎づけようとする実証主義の態度に結実する。実験的評価は実証主義の系譜に位置づけられる方法である。ランダム化比較実験や実験的評価を主軸とした現行のEBPMは実証主義的である。

相対主義は認識の独立性にも可能性にも懐疑的である。現実是个々人の認識や解釈の相互作用を通して構成されるものだと位置づけられる。この説明が物語る通り、社会的構築主義は社会科学における相対主義の代表例である。また、教育学に顕著に見られる科学的根拠を単なる政治的なレトリックとして退ける見解（今井 2015）は、相対主義の認識論に位置づけられる。相対主義は科学それ自体に疑義を申し立てる認識論であり、自然科学の理解を重視する現代の科学哲学では下火である（戸田山 2005）。しかし、人文学や社会科学においては、相対主義は自然科学とは異なる知のあり方を模索する時の有力な手がかりである。評価研究では自然主義的探求（naturalistic inquiry）が典型例である（Lincoln and Guba 1985）。

実在論は認識の独立性と可能性の双方を信じる立場である。実在論は科学者には受け入れられやすい立場だが、科学哲学では科学への信頼が議論の格好的となっている（須藤・伊勢田 2013）。科学哲学において実在論に批判的な立場は実証主義であり、この点を強調する際には実証主義は反実在論と呼ばれる。両者の論争は実在論論争として知られている。実証主義と比較したとき、実在論の特徴は「観察不可能な範囲も含めて、人間は世界を認識できる」と主張す

る点に求められる。

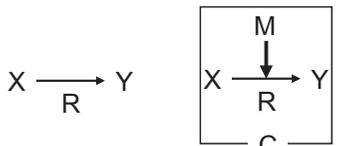
実在論論争では、世界は人間が観察可能な領域（例：水）と観察は不可能だが理論（概念や推論の体系）によって把握可能な領域（例：水分子）に区別される。実証主義にとって、後者は観察可能な領域を説明するための方便に過ぎない。よって実証主義は理論の構築や活用に対して消極的となる。一方、実在論は観察し得ない存在についても理論によってその構造や性質が把握されたのであれば（理論的対象とも呼ぶ）、それが実在していると信じる（戸田山 2015、pp.4-7）。

このように実在論論争は、一義的には観察不可能な対象の実在をめぐる論争である。しかし、その論点は科学における理論の役割や位置づけと密接に結びついている。

(2) 継起的因果論から生成的因果論へ

科学的実在論の観点からは、実証主義の因果モデルが批判の対象となる。この批判はR. Harré (1972、pp. 115-133)において、継起的 (successionist) 因果論と生成的 (generative) 因果論の対比を通して描かれる (Pawson and Tilley 1997、pp. 31-33 ; pp. 55-56)。

図1 継起的因果論と生成的因果論



継起的因果論 生成的因果論

(出所) Pawson and Tilley (1997、p.58、68) に基づき筆者作成

継起的因果論は「連続して生じること (succession)」を重視した因果モデルである。ある事象Xに続いて別の事象Yが観察されたとき、そこに因果関係を見出す。ただし、現象Xと現象Yのつながりはあくまでも人が見出すもの、人の心や認知の中にしか存在しないものとする。そして現象Xと現象Yの間の規則性R (regularity) に分析の焦点を当てる。その他の影響を除去す

ることができるRCTとは非常に親和的な考え方である。

一方、生成的因果論では、現象Xと現象Yに規則性があるのなら、その背後には現象Xと現象Yの規則性を生み出す生成的メカニズムM (mechanism) が存在するはずだと考える。生成的因果論では、こうした隠伏的な生成的メカニズムを解明してこそ、因果関係を明らかにできると考える。

メカニズムMを特定するためには生成メカニズムのデモンストレーション<sup>5</sup>が重要である (Pawson and Tilley 1997、pp. 57-63)。デモンストレーションは、メカニズムを実際に作動させて理論が予測する結果が得られることを確認することである。また、デモンストレーションでは、メカニズムが作動する場合と作動しない場合を分ける文脈的条件C (context) を確認することが重視される。メカニズムの作動と制御の成功が、メカニズムの実在の根拠となる。

一例を挙げて整理してみよう。工場で自然発生した火花が (X)、火薬の大規模爆発を起こしたとする (Y)。継起的因果論では火花に原因として注目する。この点を確認するためには、無作為に抽出した火薬に火花を点火し、実際に爆発が起きる確率やその平均的な効果を確認することになる。一方、生成的因果論では火薬の化学組成 (M) に着目する。火薬の化学的組成 (M) や火薬の状態 (C)、湿度 (C) を切り分けた爆発実験を繰り返し、微弱な火花が大規模な爆発を引き起こすプロセスを描き出す。そして静電気は単なるきっかけであり、火薬への無理解と杜撰な管理が問題の本質だと主張されるだろう。

このように継起的因果論と生成的因果論では因果関係の意味が異なる。問題はどちらの因果モデルが社会プログラムの評価に適しているかである。

(3) 社会プログラムの評価の主要課題

実在論的评价では、社会プログラムの評価の主要な課題として、行為主体と社会構造の相互作用の解明を掲げる。

行為主体と社会構造の相互作用は、A. Giddens の構造化理論 (structuration theory : Giddens

1984)、M. Archerの形態発生論 (morphogenetic approach : Archer 1995)、R. Bhaskarの批判的自然主義 (critical naturalism : Bhaskar 1979) といった科学的実在論に親和的な社会学が描き出してきた現象である。その要点は以下のように整理できる (図2)。

人は自らの行為を自ら選択する行為主体である。行為主体は社会に埋め込まれている。行為は、たとえ意図せざる形であれ、社会に影響を与える (再生産/社会変革)。それと同時に、社会の規範や習慣、歴史、文化は行為主体に影響を与える (社会化)。このような個人や集団における行為選択の過程や行為選択への社会構造の影響は直接観察することはできない。しかし、こうした行為主体と社会構造の相互作用は現に存在し、その帰結として個人や集団の行動が多様に生成され、変化し続けている (開放系)。

図2 行為主体と社会構造



(出所) Bhaskar (1979, p. 36) に基づき筆者作成

このような観点から社会科学における実在論的研究は、以下の論理をその中核に据える。

実在論的説明の論理：社会調査の基本的課題は、興味深く、不可解で、社会的に重要な規則性を説明することである (R)。説明は規則性を生成する基底メカニズムを措定する形を採り、社会構造と行為主体の相互作用がどのように規則性を構成したかについての命題からなる (M)。実在論的な研究では、メカニズムの作動が偶発的なのかあるいは条件的なのか、すなわち特定の地域的・歴史的・制度的文脈においてのみ発生するのかについての検討も行われる (C) (Pawson and Tilley 1997, p.71)。

自殺の社会学的な説明を取り上げてみよう

(Pawson and Tilley 1997, p.65)。自殺という結果 (Y) は、失職、病気、いじめなど様々なきっかけで生じる (X)。しかし、カトリックのような集団主義の地域よりもプロテスタントのような個人主義の地域の方が自殺率は高い (C)。その理由は、社会的結合が人の不安を大きく左右するからである (M)。このように自殺という一見個人的な選択も行為主体と社会構造の複雑な相互作用とは切り離すことはできない。

そして社会プログラムとは、この行為主体と社会構造の相互作用に介入し、特定の行動変容や社会変革を意図的に引き起こす施策である (Pawson and Tilley 1997, p. 74)。だからこそ、社会調査を通して、社会プログラムの成否を判断したり、その改善や開発に資する知見を得るためには、プログラムの背後に隠れた行為主体と社会構造の相互作用に接近する科学的実在論および生成的因果論が要請される。

ここでは実在論的評価が焦点を当てる生成的メカニズムは、プログラム理論のようなアウトカムを生成するための計画とは存在の次元が異なるという点に注意してほしい。生成的メカニズムは、プログラム理論の背後からその成否を左右するものであり、観察不可能な次元に存在している。

このような実在論的説明の論理および科学的実在論に依拠した社会調査論は実験的評価の限界を検討したり、実在論的評価を設計・運用するときの立脚点となる。

#### 4. 実験的評価の批判的検討—実在論的評価の問題意識—

##### (1) 実験的評価の盲点-処方と提案-

実験的評価の問題点は行為主体と社会構造の相互作用を見逃ごさざるを得ない点にある。

実験的評価では、原因は社会プログラムと見なされる。一方、行為主体と社会構造の相互作用は攪乱要因と見なされ、ランダム割り当てによって均一化される。しかし、社会プログラムのアウトカムは個人や集団の行動の選択やその背後で展開する様々な推論の複合体に深く根ざ

している。行為主体の「選択は、変化の重要な条件であり、変化を理解するための障害ではないのだ」(Pawson and Tilley 1997, p. 36)。

このような実験的評価の問題は「処方」と「提案」の混同と約言できる(Pawson and Tilley 1997, pp. 50-51)。薬は「処方」されれば化学や生理学のメカニズムを通して機械的に作用する。一方、社会プログラムは資源(もの、ひと、かね、知識など)の利用を対象者に「提案」するものであり、その作用は、意識的であれ無意識的であれ、対象者の選択と結びついている。「提案」のメカニズムには、「処方」とは比べ物にならないほどの多様性と流動性がついて回る。

## (2) 文脈に対する配慮の欠如

実験的評価は、評価から社会プログラムの文脈依存性に対する配慮を失わせる。Pawson (2006, p.59) はP. Rossiの評価事例をもとにこの問題を描き出している。

かつてRossiは「社会プログラムの効果量の総和の期待値はゼロだ。」(Rossi 1987, p.5) と述べた。この背後にはRossiが行った受刑者に対する金銭援助の再犯防止効果についての実験的評価の経験がある。Rossiは再犯防止の平均効果量がほぼゼロだったにも関わらず効果の分散に注目し、金銭援助は25%から50%の再犯減少をもたらすと報告した。このことは社会学において厳しい批判を招いた(Zeisel 1982)。

受刑者の金銭の利用はまさに選択そのものであり、分散は高く平均効果量はゼロに近づく。当時、Rossiは平均効果量を軽視して実験的評価の方法論からの逸脱を問題視された。しかし、むしろ問題なのは分散をもたらした文脈的要因に対する考察を欠いた点にある(Pawson 2006, pp.59-60)。

## (3) EBPMにもたらす問題

文脈に対する配慮の欠如は、EBPMの進展を妨げる。その典型例が、イギリスで実施された囚人を対象にした再犯防止プログラムのEBPMである。

Martinson (1974) は1945-67年の再犯防止プログラムの評価結果をレビューした。ここでは実

験群と対照群の比較可能性のある評価のみが取り上げられ、分析された研究は200以上に及んだ。レビューの中で、Martinsonは成功例や部分的な成功例の存在を認識していた。しかし、それでもなお「これらの例は孤立しており、特定の処置の有効性を示す明確なパターンは存在しなかった」(1974, p.49) と主張した。Martinson (1974) が呼び水となり、刑事政策の舵は再犯防止から厳罰化へと切られた。

实在論的評価の見地から見れば、Martinson (1974) が掲げる成功の規準は高すぎる。これでは有効に機能したプログラムの存在を軽視する結果を招く(Pawson and Tilley 1997, p.9)。

つまるところ、実証主義および継起的因果論に立脚する限り、別の対象や状況、処置、手法のもとで行われる研究にも適用できるような外的妥当性の担保は、社会科学ではほぼ不可能なのである。このことは科学的实在論および生成的因果論に立脚して外的妥当性を再考する必要があることを示唆している。

Pawson and Tilley (1997) は、外的妥当性の相対化や再定義を明確に主張しているわけではない。しかし、外的妥当性の相対化と再定義は日本のEBPMの課題に直結している。本稿では外的妥当性の相対化と再定義を念頭に、实在論的評価の検討を進めていく。

## 5. 实在論的評価の設計と運用

### (1) 实在論的評価のサイクル

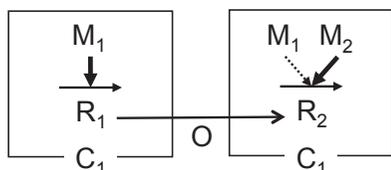
どのようにして科学的根拠を構築するのか。实在論的評価の基本的なデザインを検討しよう。

实在論的評価は以下のサイクルに即して展開する(Pawson and Tilley 1997, pp.84-86)。その過程は大きく「理論」「仮説」「観察」「特定」に即して整理される。なおこの区分は有名な「科学の車輪」(Wallace 1971) をアレンジしたものである。

「理論」は、文脈の中でメカニズムがどのように作動し、プログラムのアウトカムを生じさせるかについての命題で構成される。一般的な实在論における理論は規則性(R)、メカニズム

(M)、文脈(C)の三つの要素で構成される。実在論的評価の場合、規則性(R)に該当する要素は政策のアウトカム(Outcome)である。アウトカムは介入前に見られた $R_1$ と介入後に見られた $R_2$ の差や相違として位置づけられる。そして介入前に見られたメカニズム $M_1$ やプログラムが成功しているのであれば存在するはずの新しいメカニズム( $M_2$ )に焦点を当てた分析を行っていく。これによりプログラムとアウトカムをつなぐプロセスの複雑性や不確実性を紐解いていく(図3)。

図3 成功した社会プログラムの構成要素



(出所) Pawson and Tilley (1997, p.74) に基づき筆者作成

サイクルのスタート時点では、理論は評価対象となる社会プログラムに関連する先行研究のレビューや関係者に対する探索的調査に基づき指定される。

「仮説」は、理論を具体化したものである。プログラムの構成要素を整理し、処置から恩恵を受けやすい／受けにくい個人、集団、場所を検討し、アウトカムを生じさせ維持させる環境的な要因を浮かび上がらせていく。

「観察」は、仮説を検証するための調査である。プログラム理論の構築や改善に必要なデータを収集することが目的であり、特定の метод論が重要になるわけではない。量的方法・質的方法の区分に拘る必要はない。

「特定」は、ある状況において、誰にとって、どのようなプログラムが有効なのかを示す理論を構築する段階である。科学の車輪ではこの部分は「一般化」である。両者の違いは主に基礎研究と応用研究の違いによる。

## (2) 実在論的評価の運用例

実在論的評価の事例は、実在論を標榜しているかを別とすれば、無数に存在する。ここでは

二つ事例を検討する。最初の事例は生成的メカニズムを実際に作動させ、その作動に影響する文脈的条件を解明したものであり、デモンストレーションを体現した事例である(Pawson and Tilley 1997, pp. 89-94)。次の事例は実在論的評価の理論に即して実施されている(Pawson and Tilley 1997, pp.102-114)。

### ①プロパティマーキングの事例

プロパティマーキングは、家具に自分のマークを付けることで、家具の空き巣被害を抑制する窃盗対策である。【理論】プロパティマーキングの防犯効果は広く認知されていた(O)。その効果はマークにより家具を転売処分した際に逮捕されるリスクが増加するからだと説明されていた(M)。ただし、プロパティマーキングの効果を左右する条件は特定されていなかった(C)。また、プロパティマーキングは被害の対象者を変えるだけで(被害移行)、窃盗そのものを抑制しないとも報告されていた(C)。【仮説】こうした状況のもとで、Laycock (1985) は、プロパティマーキングのメカニズムが作動するための理想的な文脈を見つけるためのデモンストレーションを行った。南ウェールズの孤立した三つの自治体の2,234戸を対象として、プロパティマーキングの普及を徹底し参加率は70%を越えた。これにより被害移行の検証が可能となった。また、単に家具にマークを付けるのみならず、窓に防犯プログラムに参加していることを表すステッカーを付けることにした。プロパティマーキングの効果は、該当地域の潜在的な加害者の空き巣に関するリスクの認知に変化を生じさせることによるものなので、ステッカーを貼るによりその効果を簡単に高めることができると考えられたからである。【観察】対象地域におけるプログラム実施前の一年間の空き巣被害は115戸・128件であったが、実施後一年間では68戸・74件に低下した。ステッカーを貼った地域では、91件から35件に統計的に有意な形で減少した。一方、ステッカーを利用していない都市では37件から39件と微増したが、統計的有意差はなかった。また、メディアがプロパティマーキングの効果を大々的に報じたため、非対象地域でも空き巣被害が減少した。【特定】潜在的な加

害者に犯罪リスクを意識させることで (M)、プロパティマーキングは窃盗犯罪抑止の効果を上げる (O)。この効果が生じるためには、地域の大多数がプロパティマーキングに取り組む必要がある (C)。また、防犯ステッカーやメディア広報の併用も有効である (O)。

②刑務所内の高等教育の事例

サイモンフレーザー大学は、囚人に刑務所内で高等教育を受ける機会を提供しており、このプログラムには再犯防止の効果が期待された。20年に及ぶプログラムの実施実績があり、受講者は1,000人を越えていた。【理論】再犯防止の効果は漠然としており、その理論は未成熟であった。ただし、成人教育の研究には参考になる知見が存在した。受刑者に対する教育は自己実現欲求の向上、経済的事情の改善の機会、社会的な受容、道徳的・市民的な責任感の向上、認知的能力の向上をもたらし、これが再犯防止につながると説明されていた (M)。また受刑者の社会経済的状況、犯罪のタイプ、刑務所に対する態度、刑務所のタイプによって再犯防止効果が左右されることが明らかになっていた (C)。【仮説】実務家を対象にした探索的なインタビュー調査が行われた。この段階でメカニズムや文脈に関係する変数が50ほど浮上した。【観察】1,000人規模のデータセットを活用して、受刑者の一般的な非再犯率の期待値と受講者の実際の非再犯率の比較が行われた。一般的な非再犯率は58%と推定されたが、プログラム対象者は75%だった (O)。更に、特に再犯率の差に基づくサブグループ分析が行われ、先述の探索的調査の知見を手がかりとして、生成メカニズムと文脈的条件が検討された。【特定】受刑者は内省や合理的な推論の能力が向上し、再犯防止につながるという知見が得られた (M)。刑務所内の高等教育プログラムへの参画度が高いほど、教育水準が低く経済的に恵まれていない受刑者ほど、再犯防止の効果が発揮されることが確認された (C)。

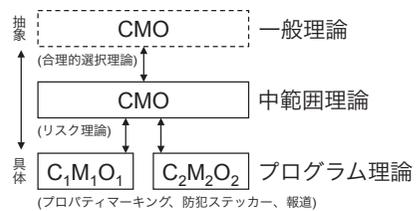
③実在論的評価の評価結果

実在論的評価は、特定の社会プログラムの総括的評価や形成的評価から始まる。それ以上に重要な成果として、実在論的評価が中範囲理論の構築 (図4) に寄与する点が挙げられる (Pawson

and Tilley 1997, p.116 ; pp.123-124)。

中範囲理論は社会学者R. マートンによる概念であり、抽象度の高い理論と具体的に操作化されたプログラム理論との間に位置する理論である。中範囲理論はプログラム理論よりも抽象度が高いため、直接的に特定の社会プログラムのアウトカムを明らかにしたり、予測したりするものではない。しかし、評価者や政策立案者が中範囲理論の内容や含意を理解すれば、そこから既存のプログラムの改善や新しいプログラムの開発の手がかりを導出できる。

図4 中範囲理論の位置づけと具体例



(出所) Pawson and Tilley (1997, p.122-127) に基づき筆者作成

Laycock (1985) がその典型例だろう。ここでは潜在的な加害者に犯罪のリスクを意識させることで事件の発生を未然に防ぐことができるといふ中範囲理論からプロパティマーキングのみならず、防犯ステッカーやメディア報道といったアイデアが導出されている。またLaycockの理論を、例えば合理的選択理論へと抽象化することもできる。これにより理論の内容的な意味は希薄化されるが、さらに広範囲の別の社会プログラムとの比較検討を行う可能性も生じる。

このように実在論的評価の評価結果は、特定の実証分析の結果というよりも、多種多様な実験や実証分析に支えられた抽象度の異なる理論の系 (一般理論／中範囲理論／プログラム理論) である。中範囲理論がその中核に置かれている。理論の系を科学的根拠の本質に据えるという発想は、我が国のEBPMにとっても示唆的である。

(3) 実在論的統合

科学的根拠の構築には、個々の評価結果を統合していく必要がある。従来、この役割はメタアナリシスが主に担ってきた。他方、実在論的

統合という考え方で、科学的根拠を構築していく。両者の考え方の違いは、前者がMartinson, R. (1974) のように「何が機能するか」を析出しようとするのに対して、後者が「何が、誰に、どのような条件のもとで、どのような点で、どのように機能するのか」を総合的に把握するための理論の構築を志向する点にある (表1)。

①理論構築としての実在論的統合

実在論的統合の目的は理論構築である。メタアナリシスが複数の一次評価の知見を要約することと対比的である (Pawson 2006, p.73)。

理論構築の過程には二つのポイントがある。

推論による理論の検討と抽象と具体の往復である (Pawson 2006, pp. 76-78)。

CMO型の実在論的な理論の構築のためには、平均効果量の算定そのものが重要なわけではない。重要なことは、一次評価およびそれが明らかにしてきた科学的根拠の整合性を検証し、必要であれば矛盾や葛藤の原因を探ることである。こうした過程は標準的な手続きに則って系統的に進むことはなく、反復的に進展する。

また、レビュー対象の共通概念 (例：対象者、介入、成果) の具体的な指示対象を明確化したり、その意味を抽象化したりすることが重要で

表1 実在論的統合

	1. 問いの明確化	2. 一次評価の探索	3. 質の精査	4. データの導出	5. データの統合	6. 知見の伝達
	プログラム理論の位置づけの検討	予備的探索				
	プログラム理論の優先順位づけ	プログラム理論に関する情報の探索		統合の候補選定に向けた注釈やノートの作成	統合の主な論点の選択と形式化	分析的・政策的焦点に関する意思決定者との交渉
	レビューのためのモデルの設定	プログラム理論の根拠の探索	一次評価とモデルとの関連性の検討	一次評価と理論との相互参照		
			一次評価の方法論的厳密性の検討	一次評価から科学的根拠の詳細の抜粋	主要な資料の統合	
		モデル改定のための追加的な探索	一次評価に関する理論的検討	科学的根拠の整理	科学的根拠に関する追加の理論的検討	今後の研究展望に関する協議
時間 ↓					プログラム理論に関する複合的・相互連関的な要素で構成されたモデルの改定	理論の要約と将来の政策の実装について「徹底的に考える」プロセスの開始

(出所) Pawson (2006, p.103) より筆者作成

ある。これにより社会プログラムの共通点や相違点を明確化することが可能となり、複数のプログラムに共通する理論の構築への手がかりが得られるからである。

## ②メタ回帰分析との差異

実在論的統合はメタ回帰分析と似た部分がある (Pawson 2006、p.56)。メタ回帰分析は処置とその因果関係を結ぶ調節変数をモデルに導入し、既存の研究の結果の差異を説明する多変量解析の手法で、メタアナリシスに援用される。両者を比較し、実在論的統合に特別な新規性はないと考える向きもあるかもしれない。

しかし、メタ回帰分析は複数のデータに現れたパターンを詳述したものであり、基本的には継起的因果論に基づいている。モデルを精緻にすれば目の前のデータを詳細に説明することはできるが、その結果としてモデルは目の前のデータだけに適用可能なものになっていく。さらなるデータとの統合が進めばより広範囲に適用可能なモデルが得られるかにも思えるが、その段階ではモデルは硬直的で複雑なものになる (Pawson 2006、pp.68-69)。

Cartwrightらの言葉を借りて補足すれば、メタ回帰分析でも、その知見を活用する段階では「ここ」と「そこ」の類似性を検討するための立脚点が必要となるということである。この立脚点を築く段階ではやはり理論が重要となる<sup>6</sup>。

## ③実在論的統合の科学的性質

実証主義に馴染んできた者は実在論的統合には違和感を覚えるかもしれない。確かにメタアナリシスに比べれば、実在論的統合は主観的である。アウトカムを確認することもできず、手続きは煩雑である。科学の名に値しないのではないかと疑問も生じかねない。

この疑念に対してPawson (2006、pp.181-182)では、Polanyi (1966) とCampbell (1984) を引き、科学における暗黙知の次元と相互検証の重要性を次のように指摘する。

科学にはひらめきや想像力のような暗黙知の次元が必ず存在し、その手続きは標準化し得ない。それでもなお科学が客観性や信頼性の高い知識を構築できるのは、科学者個人の能力や誠実さ故ではなく、科学という社会システムの内部で

は相互検証の過程が絶えず働いているからにはほかならない。

実在論的統合の過程は、まさに暗黙知的な相互検証の過程を、理論を架け橋として実現していく行為である。この意味において、実在論的統合により得られた知見は科学的根拠の名に値する。

## 6. 結論

### (1) 考察

#### 課題①実在論とその評価研究への示唆

##### i. 実在論的評価の認識論的次元

科学哲学において、実在論は観察し得ない対象の実在を信じる考え方と位置づけられる。この考え方は、科学における理論 (概念や推論の体系) の役割を重視する発想を伴う。この発想は、実在論的評価に顕著に反映されている。

また、実在論的評価は、特にHarréの継起的因果論と生成的因果論の対比を基礎として、社会学において展開してきた行為主体と社会構造の相互作用を重視する考え方から大きな影響を受けている。この考え方は実在論的説明の論理に結実している。実在論的説明の論理が社会調査やプログラム評価の理念型として、実在論的評価の根幹をなしている。

##### ii. 実在論的評価の認識論を巡る混乱の解消

先行研究に見られた実在論的評価の認識論的次元に関する混乱を解消しよう。

まず重要なことは、実在論的評価の理論的特徴とその含意は、実証主義と社会的構築主義 (認識論的な相対主義) の二項対立では理解できないという点である (表2)。実在論的評価と実験的評価ではそもそも想定している因果モデルが異なる。いわば「知りたいこと」がそもそも違うのだ。この相違はEBPMの課題にも現れる。実験的評価の発想では (内的妥当性に加えて) 外的妥当性の高い実証分析の産出が課題となるが、実在論的評価は中範囲理論を核とした理論の系の構築が課題となる。特に応用可能性の高い中範囲理論の構築と活用が重要課題である。

実在論的評価は、桐村 (2013) では質的方法

表2 実在論とその評価研究への示唆

	実証主義	相対主義	実在論
認識の独立性	成立	不成立	成立
認識の可能性	部分的に成立	不成立	成立
因果モデル	継起的因果論	不成立	生成的因果論
代表的手法	RCT 実験的評価	自然主義的探求	デモンストレーション 実在論的評価
EBPMの課題	外的妥当性の高い 実証分析の結果	政治的レトリックの批判	中範囲理論を核とした 理論の系の構築

(出所) 筆者作成

を代表するものとして、Schmitt (2020, p.15) では個性記述的な事例研究を方法論的な特徴とするものとして取り扱われている。確かに実在論的評価では、行為主体と社会構造の相互作用の分析に際して質的方法や事例研究を積極的に利用するため、この理解は決して間違いというわけではない。しかし、量的調査と質的調査、変数と事例、法則定立と個性記述を統合しながら理論構築へと至る、デモンストレーション型の評価デザインにこそ実在論的評価の方法論的な特徴があることをより強調するべきである。春田他 (2019) は優れた総説だが、認識論的な相対主義と親和性の強い部分を強調しすぎている。特に評価の頑健さを評価結果の利用者の納得性(読者にとって一貫性のある記述として理解可能か)により判断できるという説を採用している点には再考の余地がある。実在論的評価を相対主義に同化させるリスクがあるからである。実在論は認識の独立性と可能性を否定する相対主義の立場とは相容れないことを明記しておく(Pawson and Tilley 2005, p.365)。

春田他 (2019) とは逆に、西村・呉 (2022) には科学哲学における実在論の原理を強調しすぎているきらいがある。西村・呉 (2022) は科学哲学の知見に基づき、実在論的研究の成立の試金石を「その理論が未知の現象の存在や性質を予測できるか」という新奇的予言 (novel prediction) に求めている。新奇的予言は優れた理論の条件ではあるが、全ての理論にその規準を求めては社会科学どころか自然科学でも多くの理論は成り立たない。実在論的評価のように、

デモンストレーションによって理論の作動と制御を確認する方が評価デザインには適用しやすい。

Henry et al. (1998) では、創発実在論的評価の特徴は実在論を価値判断(規準の設定や運用)に応用する点に求められている。しかし、実在論の主眼は理論による認識の可能性に置かれており、それを価値判断に応用できるかどうかは別途検討が必要である。

課題②実験的評価の問題

RCTに対する疑念を呈してきた先行研究では(Cartwright and Hardie 2012; 小林 2020; 西村・呉 2022)、ランダム割り当ての意義を認めた上でその実現可能性を問題視してきた。実在論的評価はランダム割り当ての実行によって、行為主体と社会構造の相互作用に関する情報が消失することを問題視している。この指摘は、RCTは社会科学には馴染まないのではないかという懸念を認識論の次元から解き明かした重要な問題提起である。

この問題の裾野は広い。実験群と対照群の処置以外の条件を均一化するという発想はRCTのみならずそれを範として発展してきた準実験的手法に継承されている。また、近年、改めて注目を集めている因果推論<sup>7</sup>も基本的な考え方は共通している(大久保 2019)。この点については「課題④日本のEBPMへの示唆」で詳述する。

課題③実在論的評価の方法論

実在論的評価は理論駆動型評価の一形態である。よって既存の理論駆動型評価との対比を通して、その評価方法論としての特徴や評価研究

における位置づけを明らかにできる。

日本のEBPMで広く利用されている理論駆動型評価はロジックモデルである（小林2020）。そこで本稿では日本のEBPMへの示唆を念頭に、理論駆動型評価の中でも特にロジックモデルとの対比で実在論的評価の特徴を検討する。

#### i. 抽象度の高い理論の活用

ロジックモデルは二つの過程で構成された評価理論である。一つめは概念的なものであり、評価対象となる社会プログラムが意図するアウトカムとそれを実現するための処置の因果関係を示すプログラム理論を明らかにする過程である。二つめは実証的なものであり、プログラム理論が実際に機能しているかどうかを検証する過程である（Rogers et al. 2000、pp.5-6）。

実在論的評価と比較すると、ロジックモデルはプログラム理論の次元に特化しており、継起的因果論に近い発想を見て取れる。一方、実在論的評価はプログラム理論の背後にある中範囲理論に着目し、生成的因果を解明する点に特徴がある。

#### ii. 実在論的評価のコスト

プログラム理論およびその背後にある中範囲理論はM. Scrivenが評価の「贅沢品だ」（Scriven 1991、p.360）と指摘した部分である。確かに、個別のプログラムのアウトカムを明らかにするだけでもその時間や労力は甚大である。その生成メカニズムや文脈的条件を解明するとなれば、そのコストはロジックモデルの比ではない。

また、Scriven（1991、p.286）は、理論の構築は研究者の仕事であり、評価者が研究を行うことで、評価の本義である総括的評価が疎かになるとも指摘する。この指摘は評価の本質をめぐる問いかけである。

こうした議論を踏まえるなら、まず実在論的評価のコストの高さには留意が必要である。また、社会プログラムの理論構築がEBPMの推進に必要なとしても、そのコストを誰が負担すべきかについても検討の余地が大きい（実務家／研究者）。実在論的評価が有効に機能する局面を選んで、戦略的に評価に取り組む必要がある。

#### iii. 実在論的評価の主な適用範囲

実在論的評価がその効果を発揮するのは、生

成的メカニズムの解明が重要な局面である。社会プログラムの評価ではこうした局面は少なくないが、特に以下の二つの局面において実在論的評価が有益だと思われる。

一つめはプログラムのアウトカムを生成するメカニズムに関する知識基盤が不明確なときである。医療行為のように膨大な基礎研究や動物実験によって処置の生理学的なメカニズムが洗い出されている場合、実在論的評価の需要は低い。逆に、実務家の知恵や現場のアイデアに端を発するようなプログラムの効果を見定めるときには実在論的評価が重要である（Pawson 2006、p.46）。

二つめは行為主体と社会構造の相互作用が重要な領域や対象の評価を行う場合である。社会科学では、例えば教育や看護のように、状況の複雑性が問題となるような領域の存在はよく知られている（Schön1984=2007）。状況の複雑性は行為主体と社会構造の相互作用の帰結である。よって、こうした状況の複雑性が問題となる領域では実在論的評価が有益である。

#### 課題④日本のEBPMへの示唆

##### i. 理論の応用可能性

小林（2020）は実験的評価の方法論に固執すべきではないことを指摘しているが、この指摘は実在論的評価の観点からしても正鵠を射ている。また、個々別々の文脈の中で自前でエビデンスを構築していくような分散型のEBPMを提案している点も卓見である。

ただし、実験の外的妥当性の担保が難しいからといって幅広い文脈に通用する科学的根拠の構築が不可能になるわけではない。外的妥当性という概念の根底には、幅広い文脈に適用可能な知識へのニーズがある。実在論的評価に基づけば、このニーズを満たすものは特定の実証分析の結果ではなく、中範囲理論により架橋された理論の系である。もちろんこの理論の系は多彩な実証分析を内包していなければならない。

外的妥当性の高い実証分析の希求から応用可能性の高い中範囲理論の構築へと日本のEBPMの焦点を移す。おそらくこれが実在論的評価から得られる最も建設的な展望であり、本稿が最も主張したい点である。

## ii. 実験的評価の実在論的適用

実在論的評価では実験的評価とその前提にある実証主義が緊密に結びつけられ、強く批判されている（特にPawson and Tilley 2005, p. 365）。しかし、実証主義の問題と実験的評価の問題は切り分けて考えることができる。

本稿の3 (1) では科学哲学の観点から実在論、実証主義、相対主義とその評価理論への含意を検討した。ここで明らかになったことは、実証主義は理論の構築や活用に対して消極的にならざるを得ない認識論だという点である。確かに、実証主義に立脚する限り、生成的メカニズムを積極的に取り扱う動機や方法は持ち得ない。

しかし、実験的評価でも生成的メカニズムにアプローチすることは十分に可能である。例えば実験的評価を実施する場合にサブグループ分析を用いて積極的に対象者の社会経済的状況が処置に及ぼす影響を確認し、社会経済的状況に応じた最適な処置のあり方を模索していくといった改善案があり得る（青柳2018）。同様のアイデアは統計学の観点では効果の異質性への対応策として有望視されている（Stuart et al. 2010, p.372）。サブグループ分析は分析結果の恣意性が問題視されていた時期もあったが（青柳2010）、こうした評価実践は科学的実在論の観点から認識論的に擁護できる。

科学的実在論の観点からサブグループ分析を設計・運用すれば、実験的評価の強みである内的妥当性を活かして、対象者の社会経済的状況が処置に及ぼす影響に関する有益な情報を得ることができる。こうした情報は特にアウトカムの背後にあるメカニズムが作動する条件や範囲を見定める際に重要な手がかりとなる。

つまり「RCTでは知り得ない何かがある」というより「実証主義では知り得ない何かがある」という方が正確である。科学的実在論に基づく設計や運用の可能性は、RCTと基本的な論理を共有する準実験や因果推論にも当てはまる。

## (2) 本稿の意義、限界、課題

本稿のオリジナリティは科学哲学と評価研究の接続を改めて図るというアプローチにある。これまでも科学哲学は評価研究の重要な近接領

域であったが、科学的実在論の評価研究への示唆や含意の導出は本稿によって一歩進められたはずだ。これにより生じた理論的貢献は二つに整理できる。

一つめは科学的実在論に関する科学哲学の知見を紐解くことで、実在論的評価の本質的な特徴を明らかにした点である（評価理論の構築）。実在論的評価の本質的な特徴は生成的因果論の観点から、中範囲理論のレベルで行為主体と社会構造の相互作用を明らかにしていく点に求められる。実在論的評価の特徴が明らかになったことで、実験的評価とその類例に共通する限界、理論駆動型評価としての実在論的評価の特徴、その運用上の留意点が明らかになった。

二つめは実在論的評価の日本のEBPMへの示唆を明らかにした点である（評価慣行の理解）。実在論的評価に基づけば、外的妥当性の高い評価結果の希求から応用可能性の高い中範囲理論の構築へと舵を切ることがEBPM推進のカギとなる。この方向性の中では、実験的評価は科学的根拠の構築や活用に寄与しうる。

本稿の限界は実在論的評価の実践に至っていない点である。今後の課題は実在論的評価の観点からプログラム評価を実施し、実在論的統合を行うことで、プログラム開発や政策立案に資する科学的根拠を構築していくことである。また、実在論的評価の理論を踏まえた形で、EBPMに関する制度設計について多角的に検討していくことも実践上の重要な課題である。

## 謝辞

本論文の執筆にあたって有益な助言を下された島一則先生（東北大学）、原田健太郎先生（島根大学）に心よりお礼申し上げます。また、特集への参加をお声がけいただいた米原あき先生（東洋大学）に深く感謝の意を表します。（本研究はJSPS科研費22K02676の助成を受けたものです。）

## 注記

1 先行研究の春田他（2019）では、realist evaluation/

realistic evaluationは「リアリストアプローチ」と呼称されている。本稿では二つの理由でrealist evaluation/realistic evaluationを「リアリスト評価」といった訳語を用いず「実在論的評価」と訳出する。一つめの理由は、PawsonやTilleyの論考では、realist/realisticという概念が科学的実在論に依拠した概念として位置づけられているからである。他方、「リアリスト」の意味は多義的であり、現実主義や写実主義の訳語としても用いられる。realist/realisticに実在論的という訳語を当てること、PawsonやTilleyの議論の特徴を明確化できると判断した。二つめの理由は、PawsonやTilleyの論考ではrealist evaluationとrealistic evaluationは互換的に用いられているからである。稀にrealisticを評価の実用性を強調する文脈で用いられることもあるが（例えばPawson and Tilley 1997, p. XIII）、この用例の明確な該当例は少ない。

- 2 本稿では認識論を「知ること (episteme) に関する説明 (logos)」と定義する (Matthias and Neta 2020)。
- 3 本稿では政策を社会プログラムの一種と位置づける。社会プログラムは社会課題の解決に向けた処置の総称である。文脈に即してプログラムと呼称する場合がある。
- 4 本稿では因果効果を、社会プログラムの評価を念頭に、処置や介入によりもたらされたoutcomeを指すものと位置づけ、アウトカムと表記する。
- 5 実在論的な観点から行われる実験の名称は明確に定まっていない。ただし、メカニズムの作動と制御の意味での理論の検証には、demonstrationが当てられる傾向にある。
- 6 2023年時点の見地に立てば、実在論的統合は、理論に基づくシステマティックレビュー (theory-based systematic review: White 2018) と親和的と言える。
- 7 大久保 (2019) は、今日の因果推論の基本的な考え方を、反事実的条件を導入して、原因が存在した場合と存在しなかった場合の比較を通して、原因の影響を明らかにしていく発想に求めている。因果推論の手法の多くは、この考えかたを統計学によって精緻化したものであり、本質的な論理を実験の評価と共有している。

## 参考文献

- 青柳恵太郎(2010)「開発援助分野におけるRCT導入に見られる旧くて新しい課題」、『日本評価研究』、10(1)：53-61
- 青柳恵太郎(2018)「経済学および国際開発の視点から」第2回エビデンスに基づく実践と政策セミナー (2018年6月30日)発表資料
- 今井康雄(2015)「教育にとってエビデンスとは何か—エビデンス批判をこえて—」、『日本教育学会』、82(2)：88-210
- 大久保将貴(2019)「因果推論の道具箱」、『理論と方法』、34(1)：20-34
- 桐村豪文(2013)「因果関係の発見をめぐる認識論的・方法論的立場の多様性—米国における「科学的な教育研究」の内容をめぐる議論を参考に—」、『教育業財政論叢』、12：69-86
- 小林庸平(2020)「日本におけるエビデンスに基づく政策形成(EBPM)の現状と課題—Evidence-Basedが先行する分野から何を学び何を乗り越える必要があるのか—」、『日本評価研究』、20(2)：33-48
- 須藤靖・伊勢田哲治(2013)『科学を語るとはどういうことか—科学者、哲学者にモノ申す—』、河出ブックス
- 戸田山和久(2005)『科学哲学の冒険—サイエンスの目的と方法をさぐる—』、日本放送出版協会
- 戸田山和久(2015)『科学実在論を擁護する』、名古屋大学出版会
- 西村君平・吳書雅(2022)「EBPMに関する科学的認識論と高等教育研究への含意」、『日本高等教育研究』、25：237-256
- 春田淳志・小曾根早知子・後藤亮平(2019)「リアリストアプローチ：科学的方法論に基づいた複雑な介入や教育プログラムの評価」、『日本プライマリ・ケア連合学会誌』、42(3)：167-173
- Archer, M. (1995). *Realist social theory: the morphogenetic approach*. Cambridge: Cambridge University Press.
- Bhaskar, R. (1979). *The possibility of naturalism*. Brighton: Harvester.
- Campbell, D. (1984). Can we be scientific in applied science? In Connor, R. F. Altman, D.G. Jackson, C. (eds.), *Evaluation Studies Review Annual*, Newbury Park, CA: Sage, 24-48.

- Cartwright, N. and Hardie J. (2012). *Evidence-Based Policy: A Practical Guide to Doing It Better*. New York: Oxford University Press.
- Giddens, A. (1984). *The constitution of society : outline of the theory of structuration*. Cambridge: Polity.
- Harré, R. (1972). *The philosophies of science : an introductory survey*. Oxford: Oxford University Press.
- Henry, G. T., Julnes, G., and Mark, M. M.(Eds.) (1998). Realist Evaluation: An Emerging Theory in Support of Practice. *New Directions for Evaluation*, 78, 1-109.
- Laycock, G. (1985). *Property Marking: a deterrent to domestic burglary? Crime Prevention Unit Paper 3*. London: Home Office.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.
- Martinson, R. (1974). What works? Questions and answers about prison reform. *Public Interest*, 35, 22-45.
- Papineau, D. (1996). Introduction: In Papineau, David (Ed.), *The Philosophy of Science*. New York: Oxford University Press, 1-20.
- Pawson, R. (2006). *Evidence-based policy : a realist perspective*. London: SAGE.
- Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. London: Sage Publication Ltd.
- Pawson, R. and Tilley, N. (2005). Realistic Evaluation. In Mathison, S.(Ed.), *Encyclopedia of Evaluation*. California: Sage, 362-367.
- Polanyi, M. (1966). *The Tacit Dimension*. New York: Doubleday. (= 高橋勇夫訳 (2003) 『暗黙知の次元』、筑摩書房)
- Rogers, P. J. (2000). Program theory evaluation: Not whether programs work but how they work. In Stufflebeam, D. L. Madaus, G. F. and Kellaghan, T. (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation*. Boston, MA: Kluwer, 209-232.
- Rossi, P. (1987). The iron law of evaluation and other metallic roles. In Miller, J. H. and Lewis, M. (Eds.), *Research in Social Problems and Public Policy*. Greenwich, CT: JAI Press, Vol. 4, 3-20.
- Schmitt, J. (2020). The causal mechanism claim in evaluation: Does the prophecy fulfill?. J. Schmitt (Ed.), *Causal Mechanisms in Program Evaluation*. *New Directions for Evaluation*, 167, 11-26.
- Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think In Action*. London: Basic Books. (= 柳沢昌一・三輪健二訳 (2007) 『省察的実践とは何か—プロフェッショナルの行為と思考—』、鳳書房)
- Scriven, M. (1991). *Evaluation Thesaurus 4<sup>th</sup> edition*. Newbury Park, CA: Sage.
- Shadish, W. R., Cook, T. D. and Campbell, D. T. (2001). *Experimental and Quasi-Experimental Design for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Stuart, E. A., Cole, S. R., Bradshaw, C. P. and Leaf, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society*, 174, 369-386.
- Wallace, W. (1977). *The Logic of Science in Sociology*. New York: Routledge.
- White, H. (2018). Theory-based systematic reviews. *Journal of Development Effectiveness*, 10(1), 17-38.
- Zeisel, H. (1982). Disagreement over the Evaluation of a Controlled Experiment. *American Journal of Sociology*, 88(2), 378-389.

(2023.8.31 受理)

## **Theory of Realistic Evaluation: Implications for Evidence-Based Policy Making in Japan**

Kunpei Nishimura

Tohoku University  
kunpei.nishimura.a6@tohoku.ac.jp

Shuya Wu

Iwate Prefectural University  
wu\_s@iwate-pu.ac.jp

### **Abstract**

One of the challenges of Evidence-Based Policy Making (EBPM) in Japan is to ensure the external validity of evidence. We focus on the theory of realistic evaluation to analyze this aspect in our study. The theory's main feature is its emphasis on theory building rather than on data analysis while constructing and utilizing evidence. Although realistic evaluation methodology has attracted considerable attention in domestic and international evaluation research, how the epistemological premises are reflected in the evaluation theory remains ambiguous. This has hindered the understanding and development of the theory and its practical application. Therefore, the focus of this study is the epistemological dimension of realistic evaluation to clarify the core idea of the theory. We analyzed the significance and the limitations of the theory of realistic evaluation as a methodology for program evaluation. Furthermore, from perspective of realistic evaluation, we criticize the methodology of experimental and quasi-experimental methods. In conclusion, the promotion of EBPM requires a greater emphasis on constructing a middle-range theory with high applicability than attaining evaluation results with high external validity.

### **Keywords**

EBPM, Realistic Evaluation, Epistemology

【コメント】

# 伝統的評価及び犯罪学からの応答

津富 宏

静岡県立大学

tsutomi@u-shizuoka-ken.ac.jp

Campbell、Cook、Rossiの伝統を引き継ぐ評価（以下、伝統的評価と呼ぶ）及び、犯罪学の立場から応答する。

Campbell以下の伝統を引くTrochim（2007）は、

理論的と実証的という最初の2つの用語は、しばしば対比されることがある。（中略）社会調査とは理論的なものである。つまり、社会調査の多くは、世界がどのように作動しているかについて社会研究者が持っている理論や考えを発展させたり、探求したり、検証したりすることに関心がある。しかし、社会調査は、実証的なものでもある。つまり、現実の観察や測定、つまり私たちを取り巻く世界についての私たちの認識に基づいている。（中略）すなわち、（社会調査は）この世界の作動に関する観察を用いて、この世界がどのように作

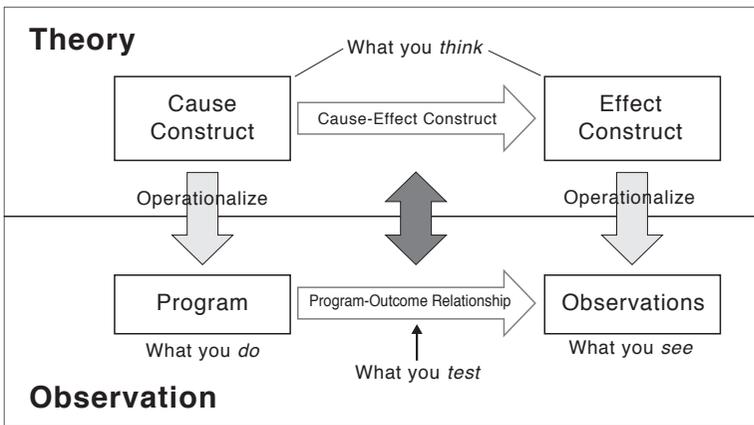
動しているかに関する理論を比較する営みなのである。

と述べる。すなわち、伝統的評価は、理論を実証（観察）によって比較する試みであり、その意味で実在論的評価である。

犯罪学では、理論的根拠のない介入が appreciateされることはない。犯罪学の分野における介入は、統制理論、分化的接触理論、合理的選択理論といった、なんらかの理論によって支えられていなければ思いつきとみなされる。理論的根拠を欠いた介入が有効であっても、有効である理由がわからず、その介入は応用できないからである。このように、犯罪学における介入効果研究は、実在論的評価である。

Trochimは、理論と実証の関係を下図に示す。上段は、理論レベルの因果関係、そして、下段

図1 研究に関わる2つの世界



(出所) Trochim (2007)

は、その理論を観察可能／測定可能にする「操作化」を経た実証レベルの因果関係を示す。なお、上段の理論は、プログラム理論ではなく、アカデミックな理論である。

では、西村・呉の实在論的評価と、伝統的評価（や犯罪学における介入効果研究）はどこが異なるのか。

第一に、伝統的評価では、研究のよさは、外的妥当性、内的妥当性、構成概念妥当性、統計的結論妥当性で表現される。外的妥当性とは、標本抽出に基づく「一般化可能性」を意味し、構成概念妥当性とは、理論を「操作化」して実証に移す際のよさ、すなわち、具体化（測定）のよさ、裏返せば、実証研究の「抽象化可能性」を意味する。すなわち、構成概念妥当性は、理論（上段）と実証（下段）の関係を担保する。しかしながら、西村・呉の枠組みは、構成概念妥当性という概念を欠いており、抽象（理論）と具体（実証）の関係を捉えていない。

第二に、西村・呉は、RCT（やメタ分析）を単なるデータ解析と理解しているが、伝統的評価においては、RCT（やメタ分析）は、データ解析ではなく理論検証を行う道具である。たとえば、メタ分析は、同一の理論的構成概念を（さまざまに）操作概念化した一次研究を束ねることによって、個々の一次研究を超えた構成概念妥当性をもつからこそ価値がある。

第三に、西村・呉は、理論を、中範囲の理論と措定するが、伝統的評価は、理論及びその実

践としての介入はその適用範囲を「中範囲」に限定することはない。その理論が普遍的であるかどうか、裏返せば、文脈依存であるかどうかは、実証的そして理論的課題である。

第四に、西村・呉は、平均効果を求めることを、効果が文脈依存であることを軽視している（さらに、外的妥当性を平均効果の外延として理解して、外的妥当性を文脈効果を見捨てた概念である）と批判的に捉えているが、伝統的評価は、普遍的な平均効果を措定しつつ、同時に、その効果の大きさに影響を与える条件付け変数の探求であり、平均効果を求めること自体を否定的に捉えることはない。

第五に、西村・呉は、文脈効果を、行為主体と社会構造の相互作用として捉えるが、伝統的評価では、あらゆる条件付け変数を「文脈」と捉える。たとえば、医療行為のような生理学的介入であっても、遺伝形質など因果関係にもたらず「個体」レベルの文脈効果はありうる。

以上、評者は、西村・呉と同様、評価研究における理論の復権を求めつつ、それは、伝統的評価の枠組みで十分に取り扱おうと主張する。

## 参考文献

- Trochim, W. 2007. *Research Methods Knowledge Base*. (<https://conjointly.com/kb/five-big-words-in-research/>から2023年12月28日にダウンロード)

【リプライ】

## 伝統的評価と実在論的評価の比較に基づく回答

西村 君平

東北大学

kunpei.nishimura.a6@tohoku.ac.jp

呉 書雅

岩手県立大学

wu\_s@iwate-pu.ac.jp

### 1. 5つの批判に対する回答

【①構成概念妥当性概念の欠如と抽象と具体の関係性を捉えていないという批判】本論文は構成概念妥当性について言及していないので前半部分の指摘は確かにそうである。ただし、科学的根拠の妥当性を取り上げる時には、常に内的妥当性、外的妥当性、構成概念妥当性、統計的結論妥当性をセットで論じなければならないわけではない。妥当性という概念は科学の発展の中で拡張されたり（村山2012）、それぞれの関係が絶えず見直されたりしているので（Shadish et al. 2001, pp. 462-483）、妥当性を取り扱う時には本論文のようにどのような妥当性に注目するのか、それはなぜなのかを説明するしかない。つまり指摘の前半部分は前提に問題がある。批判の後半部分については、本論文では継起因果と生成因果の区別や中範囲の理論を核とした理論の系という伝統的評価の枠組みとは違う切り口から抽象（理論）と具体（実証）の関係を論じているので妥当な批判ではない。

【②RCTやメタアナリシスへの誤解】本論文では、EBPMの評価慣行において、RCTやメタアナリシスが単なるデータ解析として理解され運用されてきた経緯を問題として取り上げている。私たちはRCTやメタアナリシスの評価理論を単なるデータ解析として位置付けているわけではないのだが、この点を本論文で適切に表現できていない点は筆者らの責任である。RCTやメタアナリシスの背後にあるCampbell流の因果推論を掘り下げること、この個人的な反省を有意義な評価研究に昇華できるかもしれない。なお理論に

対するコミットメントの欠如およびそれに伴うRCTやメタアナリシスを単なるデータ解析と理解してしまうことの弊害は再現性の危機の原因の一つとして社会科学に散見される現象である（Lavelle 2022）。つまりこの評価慣行上の問題は、政策コミュニティや評価コミュニティに限った話ではなく、研究コミュニティでも生じる問題である。犯罪学がこの問題を回避できたのであれば、示唆的な事例として注目に値する。

【③理論を中範囲の理論に限定しているとの批判】実在論的評価は中範囲の理論を重視するが、それはあくまでも抽象度の異なる理論の系を活用していくための核としてであり、理論を中範囲理論に限定しているわけではない。実在論的評価では理論の抽象度の階層性については自覚的であり、本論文でも強調している。

【④平均効果への誤解】論点②と同じく、やはり評価理論と評価慣行の区別を明確化しなかったため誤解を招いてしまったと反省している。実在論的評価は平均効果を求めること自体を否定しているのではなく、平均効果の算定やその結果への素朴な着目によって、行為主体と社会構造の相互作用に由来する文脈効果に対する配慮が失われてきた経緯を指摘し、この問題を掘り下げている。この評価慣行上の問題が依然として重要であることも、論点②と同じである。

【⑤行為主体と社会構造の相互作用とは異なる文脈効果の軽視】遺伝的形質であれなんであれ、それが重要だと判断されるのであれば文脈効果として取り上げれば良いのは確かにそうである。単に実在論的評価の焦点がそこにはなく、行為主体と社会構造の相互作用に由来する文脈効果

に置かれているだけである。これは実在論的評価の問題というよりも特徴である。私たちにはむしろ伝統的評価の文脈効果に対する議論は、伝統的評価の射程を広くしているというより、伝統的評価から焦点を失わせているように見える。(自己批判も込めて言うと) 文脈を語り尽くすことはできないのだから、どのような文脈に注目するかを問うことが重要である。

## 2. 伝統的評価と実在論的評価の比較： 実在論的評価に対する理解の深化に向けて

津富氏のコメントは「以上、評者は、西村・吳と同様、評価研究における理論の復権を求めつつ、それは、伝統的評価の枠組みで十分に切り抜くという主張」と結ばれている。しかし、ここまでの議論が示唆している通り、実在論的評価も私たちが単に理論の復権を求めているわけではない。

一見すると実在論的評価は「理論と経験の往還」という伝統的な社会科学の考え方をプログラム評価に忠実に落とし込んだ評価理論であり、特に理論の重要性を声高に叫んでいるように思える。この解釈は正しいが表面的である。実在論的評価には続きがある。社会科学で理論と呼ばれる観念の論理的な体系には様々なタイプがある。だからこそ評価の設計や運用において、どのタイプの理論がどのような役割を果たすのかを問い直すことが評価研究上の重要な課題となる。実在論的評価は継起因果と生成的因果の区別や行為主体と社会構造の相互作用に着目した評価の認識論を展開している。

一方の伝統的評価では理論は「世界がどのように作動しているか」を明らかにするものである。理論は複数の実証分析を紐づけて知見を集積したり、その知見の適用可能な範囲を見定めたりする役割を担うことも示唆されている(Trochim 2007)。このような伝統的評価の理論観に異論はない。ただし、ここで語られていることは科学に関する一般論であり、先に挙げた評価研究上の課題に対する積極的な貢献は相対的に希薄である。

改めて比較してみると、実在論的評価の特に認識論的次元における評価理論上の貢献は伝統的評価の枠組みから導出されたものではないし、導出できるものでもないことは明らかである。

最後に、津富氏のコメントによって実在論的評価の特徴がより明確化されたことに心よりお礼申し上げたい。

## 参考文献

- 村山航(2012)「妥当性-概念の歴史の変遷と心理測定学的観点からの考察」、『教育心理学年報』、51:118-130
- Lavele, J. S. (2022). When Crisis Becomes an Opportunity: The Role of Replications in Making Better Theories. *The British Journal for the Philosophy of Science*, 73(4), 965-986.
- Shadish, W. R., Cook, T. D. and Campbell, D. T. (2001). *Experimental and Quasi-Experimental Design for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Trochim, W.(2007). *Research Methods Knowledge Base*. (<https://conjointly.com/kb/five-big-words-in-research> 2024/01/05最終閲覧)

## 【総説】

## EBP プログラムの構築を目指す形成的評価手法におけるエビデンスの捉え方

新藤 健太

日本社会事業大学

shindo@jcsu.ac.jp

## 要 約

社会福祉をはじめとする対人援助サービス分野においてはEBP（Evidence-Based Practice）プログラムを構築し、実施・普及していくことへの関心が高まっている。近年、EBPM（Evidence-Based Policy Making）の取組みもあり、プログラムの提供者や利用者のみならず国や地方自治体等、政策立案者にとっても「エビデンスに基づく」は大きな関心事となっている。こうしたEBPプログラムをはじめとする効果的な実践プログラムを生み出すためにはプログラムの有効性をより高めるための情報を生み出す形成的評価手法が欠かせない。しかし、形成的評価手法におけるエビデンスの捉え方については、必ずしも共通の理解があるわけではない。そこで本稿では、エビデンスに基づく実践プログラム、あるいは政策を生み出す際に用いるべき評価情報とは何か、そもそもEBPやEBPM等、エビデンスに基づくと表現される場合のエビデンスとはどういったものか等の観点から、改めて形成的評価手法におけるエビデンスの捉え方を考察する。

## キーワード

エビデンス、EBPプログラム、形成的評価、質的データ／数量的データ、CD-TEP法

## 1. はじめに

近年、社会福祉をはじめとする対人援助サービス分野においてはEBP（Evidence-Based Practice）プログラムを構築し、実施・普及していくことへの関心が高まっている。

例えば、社会福祉の担い手である社会福祉士の倫理綱領には「（最良の実践を行う責務）社会福祉士は、自らが属する組織・職場の基本的な使命や理念を認識し、最良の業務を遂行する」ことが倫理基準の1つとして位置付けられている（公益社団法人日本社会福祉士会 2020：2）。このことにも関連して、2019年に改訂された社会福

祉士養成課程のカリキュラムでも教育に含むべき事項として「ソーシャルワークにおける評価」が位置付けられており、ここで想定される教育内容の例として「根拠に基づく実践（EBP）」が挙げられている（厚生労働省 2019：13）。

また、2017年頃からは、EBPM（Evidence-Based Policy Making）の動きも加速しており、社会政策の様々な領域でエビデンスに基づく政策立案の取組みが進められている（古矢 2017；水口ら 2022；大島ら 2022：3）。こうした動きは国や地方自治体等、政策立案者のEBP等に対する関心をより一層高めることに貢献しており、社会福祉領域についても、厚生労働省がEBPM推進に係る

有識者検証会を設け、EBPMの対象にすべき事業は何かについて議論する等、まだ十分ではないものの取組みが進んでいる。

このように、実践的にも政策的にもEBPプログラム等エビデンスに基づく効果的な実践プログラムが必要とされている状況にあるが、現状、社会福祉領域では、RCT（ランダム化比較試験）等、厳格な評価手法によって検証された有効性に関する知見やこれをもとにして生み出されるEBPプログラム等は不足しており、様々な社会福祉課題に対して効果的な実践が設けられていない状況にある（佐藤 2002；Holosko 2010；山口ら 2013：70；大島 2016；大島ら 2019：7）。

こうした状況に対して、EBPプログラム等の構築に有効なプログラム評価の理論と方法（Rossi et al. =2005；源ら 2020）、とりわけ形成的評価の取組みが必要とされている（大島ら 2019；大島ら 2022；新藤ら 2022；新藤 2023）。ただし、後述するように、EBPにおけるエビデンスの捉え方や形成的評価手法におけるエビデンスの捉え方については、プログラムの実施に関与する実践家、資金提供者、政策立案者、そして、形成的評価の実務を担う評価専門家等、様々な関係者間で必ずしも共通の理解がもてているわけではない。

そこで本稿では、特に、「EBPやEBPM等、Evidence-Based（エビデンスに基づく）と表現される場合のエビデンスとはどういったものか」の観点から、改めて形成的評価手法におけるエビデンスの捉え方を概観し、一定の捉え方を提示することを目的とする。

## 2. EBP（Evidence-Based Practice）プログラムにおけるエビデンスとは何か

### （1）EBP（Evidence-Based Practice）におけるエビデンスとは何か

まず、EBPにおける「エビデンス」とは何かについて概説する。EBPとは科学的根拠に基づく実践を意味しており、実践家の経験や勘に頼るのではなく、信頼できる根拠をもって介入方法を選択し、実践することをいう。

しかし、そもそも、「エビデンス」とは「根拠・証拠・裏付け」を意味する言葉であり、ここでいう根拠や証拠には様々なものが含まれる。例えば、社会福祉領域の実践現場で広く導入されている福祉サービス第三者評価の実践マニュアルには「自らの思い込みで評価せず、記録等のエビデンスに基づいて評価するよう注意を促す」という記載があり（社会福祉法人全国社会福祉協議会 2014：49）、大学基準協会が発行している大学認証評価の実施ガイドには「自己点検・評価を進めるうえで重要となるのが、適切な根拠（エビデンス）を示すこと」という文章とともに適切な根拠に該当するものとして大学の基礎データ等を挙げている（公益財団法人大学基準協会 2016）。このように、ここで用いられる「エビデンス」とは単に根拠・証拠・裏付けのことであり、記録や大学が保有している基礎情報等を意味している。

一方でEBPはEBM（Evidence-Based Medicine）がその源流になっていることからわかるように、ある社会課題の解決を目指す介入・実践の決定に際して、客観的で信頼性のある科学的根拠を用いることを意味しており（Soydan et al. 2014：36-37）、そのプロセスは、①解決すべき問題をサービス利用者とともに明確化する、②明確化された問題へのアプローチ法とその根拠を探る（この時、様々な科学的根拠の妥当性やサービス利用者への役立ち感等を批判的に検討する）、③検討したアプローチ法を適用してより良い介入を実践する（検討した科学的根拠を必要に応じて応用し、実践する）、④行った介入の結果を評価する、の4ステージから進められることが知られている（Davies et al. 2000；米倉 2009：147）。

そして、志村（2021：206）は②の過程で活用できるインターネット上のデータベースとして、教育・刑事司法・社会福祉・国際開発・障害者支援等に関するSR（系統的レビュー）やRCT等の知見を公開する「キャンベル共同計画（Campbell Collaboration）」を紹介している。以上のようにEBPにおける「エビデンス」とは「何が効果があるのか（What works?）」という問いに答えるものであり、前述した一般的なエビデンスの定義（記

図1 EBPMに適した政策領域・適さない政策領域

実施プロセスへの関与可能性

高	大企業支援施策	中小企業振興策、輸出促進策、イノベーション支援、農業振興施策、犯罪抑止、政府開発援助	教育、 <u>児童福祉・子育て支援</u> <u>施策</u> 、 <u>医療・介護・健康</u> 、 <u>就労支援</u> 、税等の滞納予防、3R(リデュース・リユース・リサイクル)
中	基礎研究の大規模支援	インフラ整備、公害	<u>福祉(生活保護等)</u> 、減税措置、CO2排出抑制・大気汚染抑制
低	安全保障、外交	法人向け税制、貿易協定	個人向け税制、財政政策・金融政策
	少ない	中程度	多い

潜在的な政策対象者数

(注) 社会福祉に関連する政策領域を太字・下線にて強調した。

(出所) 大竹ら (2022 : 24) を一部加筆

録や組織保有の基礎情報等、根拠・証拠・裏付けになる情報全般)を意味するものではないことに注意が必要である。

なお、時折、社会福祉領域の実践現場では「実践と研究は違う(研究によって得られる知見は実践現場では役に立たない)」や「EBPは科学的根拠が絶対であり、常に実践がこれに従わなければならないとすれば、それは非現実的である」等の意見が聞かれるが、こうした批判の多くはEBPに対する誤解に基づくものである。実際には、科学的根拠に基づく介入・実践はサービス利用者のアウトカムがより効果的に達成される可能性があり、近年の社会福祉実践では推奨されるものであることが知られている。さらに、EBPにおいて選択され、提供される実践・介入は、サービス利用者とともに決定するものであり、決して研究の知見のみが唯一の決定要因ではないことは前述のとおりである。むしろ、社会福祉実践における専門職は自身の行う実践の効果を知る必要があり、サービス利用者や行政、その他多くの関係者に対して当該実践の意義についての説明責任を果たさなくてはならない。EBPはこれを可能にし得る社会福祉の実践原理なのである(山口ら 2013 : 68)。

また、社会福祉の実践領域でRCT等の厳格な評

価研究を行うことは不可能であり、科学的根拠を生み出すことすらできないという批判もある。これについても、社会福祉に関連する政策領域は潜在的な政策対象者数が多く、例えば、ランダム割付によって政策対象者を介入群・統制群に割り付けることが可能等政策実施プロセスへの関与可能性も高いこと、アウトカムの測定可能性、アウトカムの分散、データの取得コスト等様々な側面から効果検証の可能性も高く、またその必要性も高いことから、むしろEBPMに適した政策領域であることが指摘されている(大竹ら 2022 : 23-24) (図1)。

そもそも、EBPやEBPMにおいて活用可能なエビデンスはRCTによるもののみではなく、複数のRCT研究をとりまとめたSRを頂点にして、次にRCT、次にRDD(回帰不連続デザイン)やマッチング・デザイン、DID(差の差分分析)等の準実験的デザイン、次に統計手法による統制やコホート研究、次に単純な事前・事後比較デザインやベンチマーキング、ケーススタディや専門家の意見等へと続く「エビデンスレベル」の捉え方がある(津谷 1999 ; Soydan et al. 2014 : 36-37 ; 家子ら 2016 : 4 ; 大島ら 2019 : 36)。EBPやEBPMはこうした介入・実践の有効性に関わる様々な研究から得られる最善の知見を選択する

ものであり、必ずしもSRやRCTによる知見の活用に限定されるものではない。こうしたことを踏まえれば、社会福祉領域においてEBPやEBPMを実践することは十分に可能であり、むしろやりやすい領域である可能性も示唆しているといえる。

## (2) EBP (Evidence-Based Practice) プログラムにおけるエビデンス

こうしたEBPの実践には活用可能なEBPプログラム等効果的な実践プログラムの存在が不可欠である。そこで、次に、EBPプログラムについて、また、EBPプログラムにおいてエビデンスをどのように捉えているのかについて概説する。

大島ら (2019: 35-36) はEBPプログラムについて、「社会プログラムが目指すべき問題解決や支援ゴールの達成に一貫した有効性のあることが、科学的な評価方法を用いて体系的に立証されたプログラム」、「RCT等、内的妥当性の高い評価方法によって、効果性に関するエビデンスを蓄積し、蓄積したエビデンスのレベルに基づいて国際的あるいは国内的レベル等、社会の認証・認定を得たものがEBPプログラムと呼ばれる」と説明している。

また、大島ら (2019: 45) は、エビデンスレベルに基づく「効果的な実践プログラム」を①EBPプログラム (十分に蓄積されたエビデンスがある)、②ベストプラクティスプログラム (EBPほどのエビデンスはないが、それが蓄積され、かつ十分な実践的裏付けがある)、③エキスパート・コンセンサスプログラム (専門領域のエキスパートの多くが推奨する)、④実践の中で有効性の裏付けが徐々に得られているプログラム、⑤エビデンスが明確でないプログラム、の5つに分類して説明している。なお、代表的なEBPプログラムには、地域包括ケアマネジメントACT (西尾 2004; 大島 2016) やIPS援助付き雇用プログラム (Becker et al. =2004; 大島 2016)、家族心理教育 (Dixon et al. 2001; 大島ら 2016) があるが、ここではIPS援助付き雇用プログラムを取り上げて、EBPプログラムにおけるエビデンスの捉え方について述べる。

IPS援助付き雇用プログラムは、働きたいとい

う希望を持つ精神障害のある人を対象にし、彼ら・彼女らの就労実現を強力に支援する実践プログラムである。このプログラムは有効なアウトカムを得るために必要な8つのプリンシパル (①競争的雇用) に焦点が当てられている、②仕事探しをいつ始めるのかはクライアントの選択に基づいている。③リハビリテーションと精神保健サービスの統合、④クライアントの好みを尊重する、⑤個別の経済的カウンセリング、⑥迅速な職探し、⑦系統的な職場開拓、⑧無期限の個別支援) を備えている (JIPSA n.d.a)。このプログラムの特徴として、施設内で長い訓練を行い、就労準備性を高めてから就職を目指すtrain-then-placeの支援ではなく、まずはサービス利用者が望む場で就職をし、そこで働きながら必要な訓練を受けるplace-then-trainの支援であることがよく知られている。そして、こうした8つのプリンシパルが守られたIPS援助付き雇用プログラムによる実践は、競争的雇用の獲得というアウトカムの達成に効果的であり、逆に施設内訓練中心の支援はこうしたアウトカムに対して然程効果的でないことが明らかにされている (Becker et al. =2004; JIPSA n.d.a)。さらにこのプログラムが開発された米国のみならず、日本においても同様に十分な有効性をもつことが明らかになっている (大島ら 2000)。

ここで重要なのは、8つのプリンシパルが遵守されていること、すなわちIPS援助付き雇用プログラムが正しく、適切に行われる必要があるという点である。なぜなら、EBPプログラムをはじめとする実践プログラムは意図したように、正しく適切に実施することが簡単ではないからである。place-then-train型の介入・実践を行うはずのIPS援助付き雇用プログラムでサービス利用者本人が望まない長い施設内訓練を提供していたら、最早この実践はIPS援助付き雇用プログラムとはいえなくなる。EBMにおける投薬という介入は、基本的には「薬を処方し、提供する」という比較的単純な方法で行われるため、実施方法に様々な留意点が存在し、これを遵守しなければならぬEBPプログラム等に比べれば容易に正しく、適切に実施することができる (それでも誤薬等のトラブルが起こることもある

が)。

すなわちEBPプログラムにおける「エビデンス」とは、正しく、適切な実施が可能な実践プログラムを評価対象とし、RCT等、科学的な評価方法を用いてその有効性が立証された知見といえる。言い換えれば、複数人の実践家によってでも、正しく適切に再現することが可能な実践プログラムの有効性に関する知見であるといえる。そしてさらには、こうした検証が複数の国、地域でも行われることで、内的妥当性のみではなく外的妥当性も一定程度担保され、プログラムの有用性が社会的に認められた知見であるともいえる。前述したEBPプログラムはいずれも日本で生まれたものではないが、日本でもその有効性が立証され、現在もいくつかの実践現場で用いられている実践プログラムである。

### (3) フィデリティ尺度というアイデア

IPS援助付き雇用プログラムをはじめとするEBPプログラムは、正しく、適切に実施するためにフィデリティ尺度と呼ばれる評価ツールを備えている (Bond et al. 2000; 大島 2016; 大島ら 2019)。フィデリティ (Fidelity) という言葉は忠実性・忠実度という意味であり、この言葉を用いたフィデリティ尺度とは、あるプログラムが意図したとおりに忠実に実行されているかを確認するための、プロセス・モニタリング、プロセス評価に用いるための尺度である。

例えば、IPS援助付き雇用プログラムであれば、①スタッフ配置 (3項目)、②組織 (8項目)、③サービス (14項目) からなるフィデリティ尺度を備えているが、③サービスのなかにある「③-4.一般雇用のための迅速な求職活動」という項目では、雇用担当窓口との最初の接触が平均して「9か月以上後 (1点)」、「5か月以上9か月未満 (2点)」、「2か月以上5か月 (3点)」、「1か月以上2か月未満 (4点)」、「1か月未満 (5点)」という指標・測定基準を用いており、これを測定することでIPS援助付き雇用プログラムに取り組む実践家は自身の実践が「迅速な職探し」というこのプログラムのプリンシパルを遵守できているかを把握することができる (JIPSA n.d.b)。

このように、IPS援助付き雇用プログラムのフ

ィデリティ尺度項目はこのプログラムの8つのプリンシパルと密接に関連して作成されているため、この尺度を用いて自身の実践を振り返ることで、自身がIPS援助付き雇用プログラムを正しく、適切に行っているかを評価することができる。

## 3. EBP (Evidence-Based Practice) プログラムの構築を目指す形成的評価におけるエビデンスとは何か

### (1) 形成的評価 (Formative Evaluation) とは何か

次に、EBPプログラムの構築を目指す形成的評価における「エビデンス」とは何かについて考察する。

形成的評価について、Rossi et al. (=2005: 35) は「形成的評価は、プログラムの改良を導くための情報提供を意図した評価であり、その目的はプログラムをよりよく遂行できるように形成、あるいは形作ることを援助すること」と説明し、Weiss (=2014: 39) は教育カリキュラムの評価を例に挙げ、「形成的評価は、カリキュラム開発中に、カリキュラム改善支援のためにフィードバックされるべき情報を生成するもの」と説明している。また、形成的評価・総括的評価 (Summative Evaluation) という区別を導入した Scriven (1991: 168-169) は、「形成的評価は通常、プログラム等の開発または改善中に、しばしば複数回実施される」と説明している。

Rossi et al. (=2005: 35) は、形成的評価にかかわる関係者は、通常は、プログラム効果を最適化することに関心のある者であるとし、求められる情報に「プログラムに対するニーズ」、「プログラムのデザイン」、「その実施やインパクト (影響)」、「効率性」に関連することを挙げている。これはプログラム評価における評価階層 (Rossi et al. =2005: 77; 源 2014: 6; 源ら 2020: 32) に関連した情報である (表1)。このことから、形成的評価とはプログラムの改善に関心をもつ様々な関係者とともにもプログラム評価を行い、評価階層に関する情報を得て、これらを用いながらプログラムの効果性を高めていくこと

表1 評価階層と評価の視点

評価階層	評価の視点
①ニーズ評価 Assessment of Need for the Program	・プログラムの活動や戦略は社会のニーズと合致しているかどうか
②セオリー評価 Assessment of Program Design and Theory	・プログラムがどのように組み立てられているか、その設計は目的を達成するために効果的であるか
③プロセス評価 Assessment of Program Process and Implementation	・プログラムは意図したとおりに実施されているのか、プログラムの実施過程で何が起きているのか、アウトカムに繋がっているか
④アウトカム／インパクト評価 Assessment of Program Outcome/Impact	・プログラムの効果があがっているか、それはプログラム実施の結果であるか（帰属性）
⑤効率性評価 Assessment of Program Cost and Efficiency	・プログラムが効率的に実施されているか、あるいは実施されるか

(出所) 源 (2014 : 6)

であると理解できる。こうした形成的評価の全てがEBPプログラムの構築を目指して行われるわけではない。しかし、EBPプログラム等効果的な実践プログラムを構築するためには、こうした形成的評価の手法が欠かせない。

## (2) EBP (Evidence-Based Practice) プログラムの構築を目指す形成的評価におけるエビデンス

山口ら (2013) は、Medical Research Council (2000) を参考に、社会福祉実践のような複雑な実践に関するエビデンス構築の過程を図2のように整理している。それぞれの介入研究が行われる背景や根拠は、理論研究や歴史研究という土台があって成り立っている (Pre-clinical)。また、介入の内容は横断研究、ケース・コントロール研究、コホート研究あるいは質的研究等で発見された知見に基づいて構成される (Phase I)。ここまでの知見は、効果的な実践プログラムを構成するための情報 (根拠・証拠・裏付け) として重要なものであり、欠かせないものである。プログラム評価における評価階層と対照させればニーズ評価やセオリー評価の問いに答える情報に該当する。その意味では本稿冒頭に述べた一般的な「エビデンス」の定義には当てはまる。ただし、Evidence-Based (エビデンスに基づく) と表現される場合の「エビデンス」はここから先の情報を意味する。

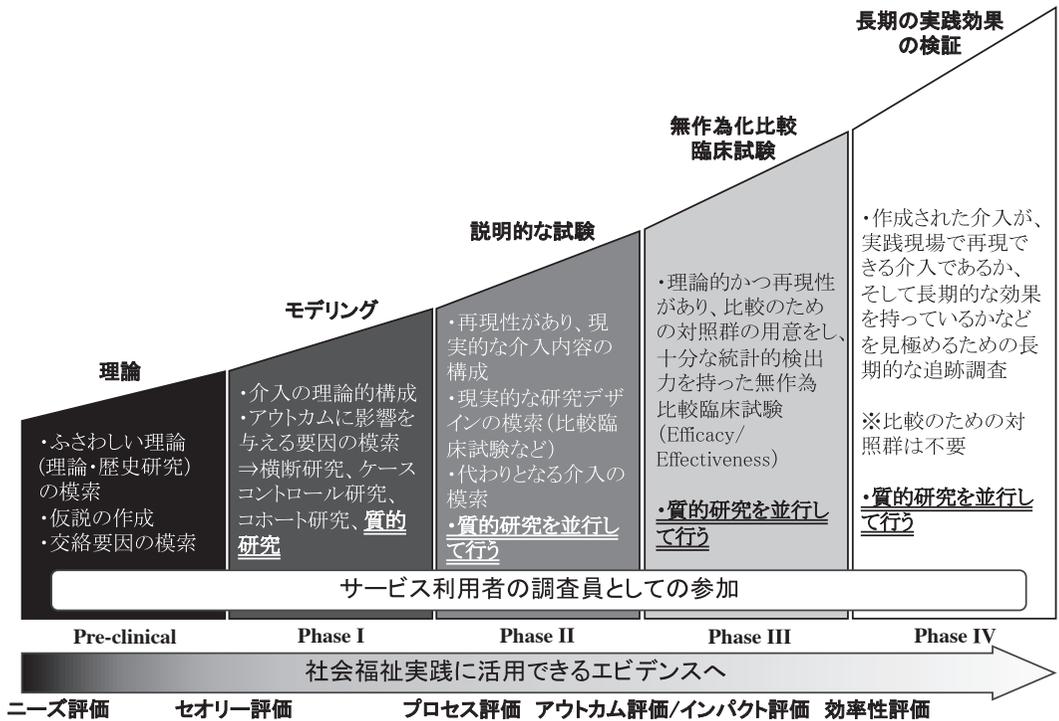
介入の理論的枠組みが構成された後は、再現

性のある介入を構成し、その実践が本当に効果的であるのかを測定する (Phase II)。特に、RCTによる実践評価と明確な実践効果に関するエビデンスの提示は社会福祉実践においても説明責任の遂行と効果的な実践の普及に大きく貢献する (Phase III)。これらはプログラム評価における評価階層と対照させればプロセス評価・アウトカム評価、そしてインパクト評価の問いに答える情報に該当する。なお、図2において効率性評価の位置付けは必ずしも明確ではないが、行うとすればPhase IIIであると考えられる。

そして最終段階にあるのが介入の長期的効果である。ある社会福祉実践が一時的に良好な成果を上げたとしても、それが持続せず、一定期間後にまた問題が繰り返されるようではサービス利用者の本質的な生活改善にはたどり着かない (Phase IV)。しかし、こうした長期的な効果の検証は、必ずしもプログラム評価には明確な位置付けがない。近年、社会的インパクト評価という言葉が流行し、この文脈においては長期的な社会経済的変化を「インパクト」と称することもあるが、津富 (2020 : 111) や佐々木 (2023) はプログラム評価における「インパクト (介入における純粋な変化量)」とこうした長期的な効果は別ものであることを説明している。

ちなみに、山口ら (2013 : 70) は、RCTには2つのタイプがあることを指摘している。1つは、研究対象者の厳しい選定基準や介入内容の限定等、極めて実験的な空間で仮説となる介入の効

図2 複雑な実践に関するエビデンス構築の過程におけるフレームワークと評価階層の対照



(注) 質的研究部分を太字・下線で強調し、評価階層は筆者が加筆した。  
(出所) 山口ら (2013 : 69)

果を測るもの (こうして得られる結果をEfficacy (効能) と呼ぶ)、もう1つは厳しい選定基準を設けずに研究対象者を比較的広く集め、そして地域の中で多様な実践が複雑に作用する中で介入の効果を測るものである (こうして得られる結果をEffectiveness (効果) と呼ぶ)。これらのうち、様々な地域実践の中で展開される社会福祉実践においては、「Effectiveness (効果)」に関するエビデンスがより重要なものとなる。

なお、こうした介入研究によるプログラムの有効性評価 (すなわちインパクト評価) を行う際、重要なことは、本稿でも前述したとおり、プログラムを正しく、適切に実施することである。Rossi et al. (=2005 : 180-182) は、プログラムが正しく、適切に実施されない状態を「実施上の失敗 (implementation failure)」と表現し、具体的な失敗として、①「無プログラム」と不完

全介入 (プログラムが実際には全く行われていない、あるいは行われているが十分な質・量の水準で実施されていないことで起こるプログラムの失敗)、②誤った介入 (プログラムが当初想定していたものと違う方法 (内容) で実施されたことによるプログラムの失敗)、③標準化されていない介入 (プログラムが場所、人によってばらばらに実施されたことによるプログラムの失敗) の3つを挙げている。こうしたことが頻発するプログラムは「正しく、適切な実施が可能なプログラム」とは呼べないため、EBPプログラム等効果的な実践プログラムにはなり得ない。前述したフィデリティ尺度を備えるというアイデアは上記に示した実施上の失敗①~③のうち、少なくとも②及び③の解決に貢献することができる。そして、これらは、プログラムは正しく、適切に実施されたのか、そのうえで十分に有効

なものであったかという問いに答えるものであり、Evidence-Based（エビデンスに基づく）と表現される場合の「エビデンス」に該当する。

### (3) 質的（定性的）データの活用とサービス利用者の評価活動への参加

こうしたRCT等、プログラムの有効性を明らかにする厳格な評価手法によって得られるエビデンスは重要なものであり、EBPやEBPM等、いわゆるEvidence-Based（エビデンスに基づく）と表現される様々な取組みの基盤を支えている。しかし、社会福祉実践あるいはその他の対人援助サービス分野の実践においては、数量的（定量的）には測りにくいアウトカムも多く存在する。

例えば、前述のIPS援助付き雇用プログラムは、働きたいという希望を持つ精神障害のある人を対象にし、彼ら・彼女らの就労の実現やその後の就労定着を達成すべきアウトカムとして設定している。しかし、就労を実現したことで、あるいは、その後しばらくの時間が経過した後で得られるアウトカムには個別的なものも多い。例えば、あるサービス利用者は「仕事を得たことで、自分に自信がついた」と語るかもしれないし、また別のある人は「毎日に張り合いが来た」と語るかもしれない。逆に、「仕事は得たが毎日辞めたくて苦しい想いをしている」と語る人がいるかもしれない。こうした変化は個別的なものであり、全てのサービス利用者にも共通の指標を設定して測定することがすぐわない場合もある。「仕事は得たが苦しい」と語るサービス利用者については、就労もしており、（本心では辞めたいが）仕事も続けているということから、数量的（定量的）分析ではプログラムの提供によるポジティブな結果（事例）の1つとしてカウントされてしまうだろう。

山口ら（2013：71-72）は、こうした類のアウトカム評価に関しては質的研究法が向いていると指摘し、社会福祉実践においては、質的研究からのエビデンスが不可欠であり、EBPと社会福祉実践の発展のために重要なエビデンスとして扱われるべきであると述べている。また、Soydan（2008）はこうしたことから、社会福祉実践において、RCTが実践評価のゴールド・スタンダー

ド、すなわち唯一・最良の実践にはなり得ないと指摘している。そのため、社会福祉実践のような複雑な実践に関するエビデンス構築の過程では質的研究を並行して行うべきであり（Phase I～IV）、質的研究から得られるエビデンスはRCT等から得られる知見と同等に扱われるべきであるといえる。さらに山口ら（2013：72）は、RCT等有効性に関する厳格な評価研究と質的研究が同時に行われ、両方の研究から実践プログラムにおける長期的な効果についてのエビデンスが得られたなら、そのエビデンスは実践の普及に大きく貢献することが予想されると述べている。

社会福祉実践のような複雑な実践に関するエビデンス構築の過程でもう1つ重要なことはサービス利用者の評価研究への参加である。山口ら（2013：73）は、サービス利用者がプログラムの評価研究に参加することの利点として、プログラムの理論的枠組み及び実際の介入の構築とその評価の質を高められることを指摘している。具体的な利点として、サービス利用者が調査員としてあるプログラムの評価に関与した場合、当該プログラムが提供するサービスについてより否定的な意見、つまりより正直な意見を引き出せる可能性があること、また、データの分析等に当たって、サービス利用者の経験や感情に基づいた分析ができることを挙げている。

## 4. プログラム理論・エビデンス・実践間の円環的対話による効果的プログラムモデル形成のためのアプローチ法（CD-TEP法）

次に、EBPプログラム等効果的な実践プログラムを形成・構築するための形成的評価手法について概説する。近年、社会福祉をはじめとする日本の対人サービス領域では、既存サービスでは対応できない社会課題が拡大しており、こうした課題に対してEBPプログラム等効果的な実践プログラムが生み出され、それが様々な地域・実践現場で用いられること、すなわち効果的な実践プログラムを用いたEBPの実践を行うことが

表2 CD-TEP改善ステップ (12ステップ)

<p><b>【第Ⅰステージ：設計・開発評価ステージ】</b></p> <p>(Step1) 現状分析・ニーズ評価</p> <p>(Step2) 評価可能性アセスメントの実施と予備的効果モデル（暫定版）の作成</p> <p>(Step3) GP(Good Practice)事例調査の実施</p> <p>(Step4) 質的データ分析と実践家参画型ワークショップの準備</p> <p>(Step5) 実践家参画型ワークショップ：第1次効果モデル（試行版）の構築</p> <p>(Step6) 第1次効果モデル（試行版）の形成・構築・改善：効果モデル5アイテムの作成</p> <p><b>【第Ⅱ-1ステージ：形成・改善評価ステージ（導入期）】</b></p> <p>(Step7) 広域的事業所調査：第1次効果モデル（試行版）の広域的な検証</p> <p>(Step8) 広域的試行評価調査①：単一グループで行う多施設協働調査</p> <p>(Step9) 質的・量的データ分析と実践家参画型ワークショップの準備：効果モデル改定案作成</p> <p>(Step10) 実践家参画型ワークショップ：第2次効果モデル（提示版）への形成・改善</p> <p><b>【第Ⅱ-2ステージ：形成・改善評価ステージ（成熟期）】</b></p> <p>(Step11) 広域的試行評価調査②：比較による有効性研究（CER）で行う多施設協働調査</p> <p>(Step12) 第3次効果モデル（エビデンス版）への形成・改善：質的・量的データ分析と実践家参画型ワークショップによる効果モデル（エビデンス版）への形成・改善</p>
---

(出所) 大島ら (2019: 129) を参考に筆者作成

期待されている。そのためには、EBPプログラム等効果的な実践プログラムの構築やその実施・普及を支援する形成的評価手法が必要である。それも、手続きが可視化されており、誰もが取り組みやすい具体的な方法論が必要である。

大島ら (2019; 2022) はこうした課題に対して「プログラム理論・エビデンス・実践間の円環的対話による効果的プログラムモデル形成のためのアプローチ法 (CD-TEP法)」を開発し、その実践と普及に取り組んでいる。CD-TEP法は、評価対象となる実践プログラムに関する「プログラム理論 (Theory)」と「エビデンス (Evidence)」、「実践 (Practice)」のそれぞれを、相互に円環的に関連させながら、より良い効果的な実践プログラムの発展に反映させるプログラム開発・評価の方法論である (大島ら 2019: 77)。ここで構築される効果的な実践プログラムは、①プログラムゴールとインパクト理論 (プログラムのゴール、アウトカムに関する設計図)、②プロセス理論 (プログラムのプロセスに関する設計図であり、サービス利用計画・組織計画からなる)、③効果的援助要素リスト (プログラム・プロセスのより詳細な内容)、④評価ツール (フィデリティ尺度とアウトカム指標からなる)、⑤実施マニュアル (これら①～④を格納した実

施マニュアル) という5つのアイテムから構成されている (大島ら 2019: 94-95)。

また、こうした効果的な実践プログラムの発展を「(第Ⅰステージ) 設計・開発評価ステージ」、「(第Ⅱ-1ステージ) 形成・改善評価ステージ (導入期)」、「(第Ⅱ-2ステージ) 形成・改善評価ステージ (成熟期)」、「(第Ⅲステージ) 実施・普及評価ステージ」から整理し、こうした発展段階の取組みを具体的に進めていくために表2に示した「CD-TEP改善ステップ (12ステップ)」を設けている (大島ら 2019: 88-92)。

CD-TEP法はEBP等効果的な実践プログラムの形成・構築のためにプログラム評価 (形成的評価) を用いる方法論である。そのため、Step1～4辺りはニーズ評価に、Step5～7辺りはセオリー評価に、そして、Step8以降はプロセス評価・アウトカム評価/インパクト評価に該当している。

つまり、CD-TEP法が形成・構築を目指す効果的な実践プログラムはStep11やStep12に位置付けられる有効性に関する厳格な評価研究を用いて確かめられる。なお、その前提として、CD-TEP法を用いて形成・構築が目指される効果的な実践プログラムにはフィデリティ尺度が備えられているため、複数のプログラム提供者、複数の地域においても正しく、適切な実施が可能とな

っている。また、途中途中で質的データ分析が位置づいており、効果的な実践プログラムの形成・構築に数量的（定量的）データと質的（定性的）データの両方を用いることが可能となっている。以上のことから、EBPプログラム等効果的な実践プログラムの形成・構築において、CD-TEP法が有効な手法であることがわかる。

EBPプログラム等効果的な実践プログラムの形成・構築を目指す人たちは、CD-TEP法におけるこれらの手順を進めていくことでこれを実現することができる。実際、これまでもいくつか効果的な実践プログラムの開発に用いられ、主には社会福祉・精神保健の対人サービス分野で成果を上げている（山野 2015；新藤ら 2017；古屋ら 2021）。しかし、こうした分野以外での適用事例は筆者が知る限りでは存在しない。社会福祉・精神保健の対人サービス外で適用可能かどうかは、今後の試行とその検証が求められる。

## 5. おわりに

本稿では、特に、「Evidence-Based（エビデンスに基づく）」と表現される場合のエビデンスとはどういったものか」の観点から、形成的評価手法におけるエビデンスの捉え方を概観した。最後に、ここではこれまでに述べた主要な論点を振り返りながら、本稿のまとめとして、EBPプログラム等効果的な実践プログラムの構築を目指す形成的評価手法における「エビデンス」について一定の捉え方を提示したい。

まず、一般的な「エビデンス」という言葉には様々なものが含まれる。記録や組織の基本情報も、それが何かを説明するための根拠・証拠・裏付けになっているのであれば「エビデンス」といえる。したがって、これらはどのような文脈においても「エビデンス」にはなり得ないというのは誤った理解である。むしろ、EBPプログラムの構築を目指す形成的評価において、プログラムを開発する最初の段階である、ニーズ評価やセオリー評価の裏付けにはこうしたデータや理論研究、歴史研究による知見を「エビデンス」として用いることが推奨される。しかし、

EBPやEBPM等、Evidence-Based（エビデンスに基づく）」と表現される場合の「エビデンス」は、実践・介入の有効性に関するものに限られる。記録や組織の基本情報がこうした実践・介入の有効性を裏付けるものであれば該当するが、そうでなければEvidence-Based（エビデンスに基づく）」における「エビデンス」にはなり得ない。

ただし、こうした実践・介入の有効性に関する「エビデンス」がSRやRCTによって得られた知見に限定されるかということ必ずしもそうではない。エビデンスレベル（津谷 1999；Soydan et al. 2014：36-37；家子ら 2016：4；大島ら 2019：36）のなかから当該実践・介入が置かれた文脈で採用すべきものを判断することになる。なお、低いエビデンスレベルの知見しか得られていない実践・介入についてはより高いレベルのエビデンスが得られるよう、評価研究を継続していくことが求められる。そして、この時生み出すべき有効性に関するエビデンスは効能（Efficacy）ではなく効果（Effectiveness）に関するものである。だからこそ、EBPプログラムの構築を目指す形成的評価の取組みは日々の実践のなかで行われることが望ましい。図2やCD-TEP法が志向しているのはそうした取組みである。

次に、Evidence-Based（エビデンスに基づく）」と表現される場合の「エビデンス」には、サービス利用者の語り等、質的（定性的）データによるものも含まれる。むしろ、質的（定性的）データによるエビデンスを数量的（定量的）データによるエビデンスと同等に扱うことが重要である。なぜなら、社会福祉実践をはじめとした対人援助サービス分野の実践・介入では、共通指標で大人数の傾向を把握するのでは捉えられない極めて個別なアウトカムが多く存在し、しかも、サービス利用者一人ひとりにとってはこうした個別なアウトカムの達成が極めて重要な意味をもつからである。加えて、質的（定性的）な研究を用いて、介入・実践による長期的な実践効果の検証を行うことも必要である。

ただし、この時、十分に気を付けておかなければならないことは質的（定性的）データの特性、すなわち、数量的（定量的）データと質的（定性的）データの違いである。杉澤（2022：

100) は数量的 (定量的) データの特徴について、「すでに符号化されたり、数字で表現されたりしているデータを扱うことで、分析には統計的な手法を用いることができる」と説明している。つまり、調査対象となった人たちの大まかな傾向を把握したり、変数同士の関係を明らかにしたりすることに向いているといえる。一方で、安田 (2022: 138) は質的 (定性的) データの特徴について、「文字が中心の資料、及び非言語が中心の資料であり、これらを扱うことで数量的 (定量的) データでは扱えないことに迫ることができる」と説明し、具体的には「人々の日常生活や活動、これまでの経験、その変化の過程、他者との相互行為やそのときの思い、感じ、考えていたこと」等を挙げている。これらは数量的 (定量的) データ分析の弱点を十分に補完する。なお、形成的評価を行い、良質な質的 (定性的) データを得るためには、サービス利用者の参加も重要な要因となる。

最後に、Evidence-Based (エビデンスに基づく) プログラムが備えなければならない前提は、複数のプログラム提供者によってでも、あるいは複数の地域においてでも正しく、適切に実施することができるという点である。本稿では、これがそれほど容易なことではない点を指摘し、この部分への十分な備えがなければRossi et al. (=2005: 180-182) が危惧する実施上の失敗が引き起こされてしまうことについて述べた。こうした問題を解決する1つの方法が形成・構築を目指すEBPプログラム等効果的な実践プログラムに「フィデリティ尺度」を備えることである。これを備えることで、プログラムの提供者は自身の実践が効果的な実践プログラムに準拠したのかを評価し、問題があれば見直し、改善することができるようになる。

以上、本稿の目的であるEBPプログラムの構築を目指す形成的評価における「エビデンス」の捉え方を提示した。こうした捉え方は、プログラム提供者・実践家や資金提供者、政策立案者、そして形成的評価の実務を担う評価専門家等の間で一定の共通理解がもてるものと考えられる。一方で、本稿で取り上げた文献の多くが社会福祉領域、あるいはこれに近いものであったため、

異なる学問的背景をもつ者等からすれば、受け入れられない部分があったかもしれない。そのため、本稿がテーマにしたEBPプログラムの構築を目指す形成的評価手法における「エビデンス」の捉え方については、引き続きの継続した議論が必要であると考えられる。

## 参考文献

- 家子直幸・小林庸平・松岡夏子・他 (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」、『政策研究レポート』、三菱UFJリサーチ&コンサルティング
- 大島巖・梅原芳江・久米和代・他 (2000) 「広域地域活動支援センターにおけるIPS援助付き雇用 (個別職業紹介とサポートプログラム) 導入とその評価 (2)」、『平成19年度厚生労働省科学研究補助金：精神障害者の一般就労と職場適応を支援するためのモデルプログラム開発に関する研究 (研究代表者：西尾雅明) 分担研究報告書』
- 大島巖 (2016) 『マクロ実践ソーシャルワークの新パラダイム エビデンスに基づく支援環境開発アプローチ：精神保健福祉への適用例から』、有斐閣
- 大島巖・源由理子・山野則子・他 (2019) 『実践家参画型エンパワメント評価の理論と方法 CD-TEP法：協働によるEBP効果モデルの構築』、日本評論社
- 大島巖・新藤健太・源由理子 (2022) 「対人サービスのイノベーションをEBPプログラムに発展させる形成的評価法の開発：CD-TEP法を用いた実践家・当事者参画型エンパワメント評価の可能性」、『日本評価研究』、22 (2) : 3-14
- 大竹文雄・内山融・小林庸平 (2022) 「第1章 EBPMとは何か」、大竹文雄・内山融・小林庸平編『EBPM：エビデンスに基づく政策形成の導入と実践』、日本経済新聞出版
- 公益社団法人日本社会福祉士会 (2020) 『日本社会福祉士会の倫理綱領・行動規範』、(<https://www.jacsw.or.jp/citizens/rinrikoryo/>、2023年9月23日閲覧)
- 公益財団法人大学基準協会 (2016) 『第3期認証評価における大学評価の実践ガイド』、公益財団法人大学基準協会
- 厚生労働省 (2019) 『社会福祉士養成課程のカリキュ

- ラム (案)』、(<https://www.mhlw.go.jp/content/000525183.pdf>、2023年9月23日閲覧)
- 佐々木亮 (2023) 『インパクト評価事例集 実験デザイン (RCT) から専門家判断まで18の評価事例：よりよい「社会的インパクト評価」の実施のために』、([https://www.idcj.jp/9evaluation/sub5\\_files/impact\\_eval\\_jirei\\_28july2011.pdf](https://www.idcj.jp/9evaluation/sub5_files/impact_eval_jirei_28july2011.pdf)、2023年11月24日閲覧)
- 佐藤豊道 (2002) 「第4章 社会福祉実践研究方法試論」、仲村優一編『講座 戦後社会福祉の総括と二十一世紀への展望Ⅳ：実践方法と援助技術』、ドメス出版
- JIPSA (日本IPSアソシエーション) (n.d.a) 『IPS援助付き雇用の理念とエビデンス』、(<https://jipsa.jp/ips/about-ips-3>、2023年9月24日閲覧)
- JIPSA (日本IPSアソシエーション) (n.d.b) 『日本版個別援助付き雇用フィデリティ尺度およびGeneral Organization Index』、(<https://jipsa.jp/tools/jpnipsmanual>、2023年9月24日閲覧)
- 志村健一 (2021) 「第6章第2節 ソーシャルワークにおける評価の意義」、一般社団法人日本ソーシャルワーク教育学校連盟編『社会福祉調査の基礎』、中央法規出版
- 社会福祉法人全国社会福祉協議会 (2014) 『福祉サービス第三者評価実践マニュアル (Version2)』、社会福祉法人全国社会福祉協議会政策企画部
- 新藤健太・大島巖・浦野由佳・他 (2017) 「障害者就労移行支援プログラムにおける効果モデルの実践への適用可能性と効果的援助要素の検討：全国22事業所における1年間の試行的介入研究の結果から」、『社会福祉学』、58 (1)：57-70
- 新藤健太・大島巖 (2022) 「実践家参画型エンパワメント評価の基盤を支える「EBP効果モデル」技術支援センターの意義と役割：評価キャパシティ形成に向けた役割を中心に」、『日本評価研究』、22 (2)：69-82
- 新藤健太 (2023) 「EBPM (Evidence-Based Policy Making) に資する評価方法についての一考察：プログラム理論・エビデンス・実践間の円環的対話による効果的プログラムモデル形成のためのアプローチ法 (CD-TEP法) に焦点を当てて」、『評価クォータリー』、66：2-19
- 杉澤秀博 (2022) 「第5章 量的データの整理と分析」、潮谷有二・杉澤秀博・武田丈編『社会福祉調査の基礎』、ミネルヴァ書房
- 津谷喜一郎 (1999) 「EBMとコクラン共同計画」、矢野栄二編『医療と保健における評価』、南江堂
- 津富宏 (2020) 「第7章 プログラムの介入後の評価」、源由理子・大島巖編『プログラム評価ハンドブック：社会課題解決に向けた評価手法の基礎・応用』、晃洋書房
- 西尾雅明 (2004) 『ACT入門：精神障害者のための包括型地域生活支援プログラム』、金剛出版
- 古矢一郎 (2017) 「政府における「証拠に基づく政策立案 (EBPM)」への取組みについて」、『季刊行政管理研究』、160：76-85
- 古屋龍太・大島巖 (2021) 『精神科病院と地域支援者をつなぐ みんなの退院促進プログラム：実施マニュアル&戦略ガイドライン』、ミネルヴァ書房
- 水口綾子・鈴木智大 (2022) 「政策の効果検証 (EBPM) の推進による各府省の政策立案等の支援：「政策効果の把握・分析手法の実証的共同研究」と行政評価局内の取組」、『評価クォータリー』、10：41-47
- 源由理子 (2014) 「地域ガバナンスにおける協働型プログラム評価の試み」、『評価クォータリー』、30：2-17
- 源由理子・大島巖・他 (2020) 『プログラム評価ハンドブック：社会課題解決に向けた評価方法の基礎・応用』、晃洋書房
- 安田美予子 (2022) 「第6章 質的データの整理と分析」、潮谷有二・杉澤秀博・武田丈編『社会福祉調査の基礎』、ミネルヴァ書房
- 山口創生・米倉裕希子・岩本華子・他 (2013) 「社会福祉実践におけるエビデンスとエビデンス構築の過程：理論から実践へ」、『社会問題研究』、62：67-79
- 山野則子 (2015) 『エビデンスに基づく効果的なスクールソーシャルワーク：現場で使える教育行政との協働プログラム』、明石書店
- 米倉裕希子 (2009) 「ソーシャルワークにおける根拠に基づく実践：Evidence-based Practiceの現状と課題」、『社会問題研究』、53 (1)：145-163
- Becker, D. R. and Drake, R. E. (2003). *A Working Life for People with Severe Mental Illness*. Oxford University Press. (=2004、大島巖・松為信雄・伊藤順一郎監訳『精神障害のある人たちのワーキングライフ IPS：チームアプローチに基づく援助付き雇用ガイド』、金剛出版)
- Bond, G. R., Evans, L., Salyers, M. P. et al. (2000).

- Measurement of Fidelity in Psychiatric Rehabilitation, *Mental Health Services Research*, 2, 75-87.
- Davies, H., Nutley, S. and Smith, P. (2000). Introducing evidence-based policy and practice in public services, Amann, R. ed. *What works? Evidence-based practice in public services.*, The Policy Press.
- Dixon, L., McFarlane, W., Lefley, H. et al. (2001). Evidence-based Practices for services to families of people with psychiatric disabilities. *Psychiatry Services*, 52(7), 903-910.
- Holosko, M. J. (2010). What types of designs are we using in social work research and evaluation?, research on *Social Work Practice*, 20(6), 665-673.
- Medical Research Council. (2000). *A framework for development and evaluation of RCTs for complex intervention to improve health*, Medical Research Council.
- Rossi, P. H., Lipsey, M. W. and Fetterman, H. E. (2004). *Evaluation: A Systematic Approach*, 7th Ed., Sage Publications ( =2005、大島巖・平岡公一・森俊夫・他訳『プログラム評価の理論と方法：システムティックな対人サービス・政策評価の実践ガイド』、日本評論社)
- Scriven, M. (1991). *Evaluation Thesaurus Fourth Edition*. Sage Publications.
- Soydan, H. (2008). Applying randomized controlled trials and systematic reviews in social work research, *Research on Social Work Practice*, 18(4), 311-321.
- Soydan, H. and Palinkas, L. A. (2014). *Evidence-based Practice in Social Work: Development of a new professional culture*. Routledge.
- Weiss, H. C. (1998). *Evaluation: Methods for studying programs and policies*. 2<sup>nd</sup> edition, Prentice Hall (=2014、佐々木亮監修・前川美湖・池田満監訳『入門評価学：政策・プログラム研究の方法』、日本評論社)

(2023.12.18 受理)

## **How to grasp the evidence in the formative evaluation methodology aimed at constructing an EBP program**

Kenta Shindo

Japan College of Social Work  
shindo@jcs.w.ac.jp

### **Abstract**

In fields such as social welfare and interpersonal assistance services, there is a growing interest in constructing, implementing, and disseminating Evidence-Based Practice (EBP) programs. In recent years, there also have been efforts in Evidence-Based Policy Making (EBPM), making “evidence-based” a matter of significant concern, not only for program providers, but also for governments and local authorities, among other policy makers.

To create effective practice programs like EBP programs, it is essential to employ formative evaluation methodologies that generate information to enhance program effectiveness. However, there is not always a common understanding of how to capture evidence in formative evaluation methodologies.

In this paper, we reconsider the approach to capturing evidence in formative evaluation methodologies focusing on what evaluation information should be used when creating evidence-based practice programs or policies, and what constitutes evidence in cases where terms like EBP or EBPM are used.

### **Keywords**

Evidence, EBP(Evidence-Based Practice)Program, Formative Evaluation,  
Qualitative Data / Quantitative data, The CD-TEP approach

## 【コメント】

## 問題解決に資するプログラムの開発過程

青柳 恵太郎

株式会社メトリクスワークコンサルタンツ

aoyagi@metricswork.co.jp

## 1. はじめに

2010年代後半から“Evidence-Based”という表現が国内、とりわけ行政現場を席巻しているが、社会福祉領域では遥か以前からEvidence-Based Practice (EBP：根拠に基づく実践)の重要性が謳われてきていた。すなわち、問題解決に資するプログラム開発への需要である。

そのようなプログラムの開発過程においては、様々な事項を段階的に検討する必要がある。その際、種々の情報を参照・提示することで各検討結果が適切であることを主張していくことが不可欠となるが、これらの情報をエビデンス<sup>1</sup>という用語から捉えた時に関係者間で共通認識が醸成されていない状況が継続してしまっている。新藤論文は問題解決に資するプログラムを開発する過程に注目し、それを評価論における形成評価と位置づけ、参照・提示される情報とevidenceの包含関係やevidenceの射程を整理している。端的にまとめるならば、①“Evidence-Based”という文脈におけるevidenceは、有効性を検証した結果を意味するものである（ニーズを示す情報や介入内容を下支えする理論等はevidenceではない）、②検証結果の確証度合い、すなわちevidenceにはレベルがある、③定性的に把握をしたアウトカムに関する質的データもevidenceになりうるということになる<sup>2</sup>。新藤論文では、プログラムの開発過程でこのようなevidenceが積み上げられ、有効性が確立したものを「EBPプログラム」と呼んでいる<sup>3</sup>。

新藤論文が示した問題解決に資するプログラムの開発過程は社会福祉領域固有のものではな

い。本稿では、評者が専門とする国際開発領域における類似の議論を紹介することで、evidenceに基づく解決策の構築過程として新藤論文の整理が一般にも通用するものであるとの援護射撃を行いたい。

## 2. 国際開発領域におけるEvidence-Based Project Cycle

図1は国際開発領域でキャリアを重ねキャンベル共同計画のCEOを務めた経験もあるHoward White氏がキャンベル共同計画の公式Blogに提示したEvidence-Based Project Cycleの概念図である(White 2019)。

このプロジェクトサイクルの最初の工程は、データに基づく現状把握と問題設定、問題を生じさせている原因の特定、同種の問題に関する既存evidenceの参照によってプログラムを考案するところから始まる。既存evidenceから有効性の確認が取れたプログラムであったとしても、文脈が異なれば効果は発現しないかもしれない。そこで考案したプログラムは3段階の試行的実施(検証)に付されていく。まずは形成評価<sup>4</sup>として、プログラムを小規模に実施することで実行可能性や受益者の受容性に問題がないかを確認する。導入上の問題がないことが確認できたら、RCTをはじめとする様々なリサーチデザインを駆使したプログラムの有効性検証に焦点が移る。この段階で行う最初の検証は効能評価である。やはり小規模に可能な限り理想的な状況下でプログラムを実施し、想定していた効果が生じるかを

確認する。検証の最終段階は有効性評価である。ここでは規模を拡大し、より現実の状況に即した形で介入効果を検証する。躓いた場合は、原因を踏まえてプログラムを改訂し、検証を継続することとなる。こうした反復のプロセスを経て最終的に効果を確認することができれば、介入を行った背後に存在する母集団に対する有効なプログラムを開発できたことになる。もし別の母集団にプログラムをスケールアウトする場合はこうした一連の検証を積み重ねていく。そして複数のevidenceが蓄積されてきたら、これらの統合を図る。こうして作られたevidenceは次のプログラムを考案する際に活用されていくことになる。

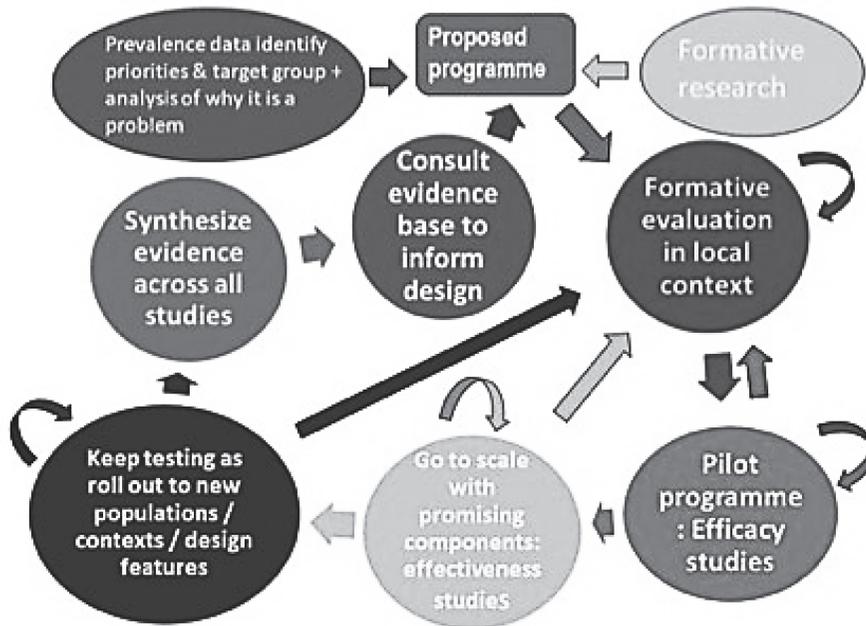
以上がWhite (2019) が提案するEvidence-Based Project Cycleの理想像である。有効性の確認を中心に据えた作業を通じて問題解決に資するプログラムを開発していく過程が描かれており、社会福祉領域で論じてきたEBPプログラムの開発工程との類似性が見てとれる。

### 3. evidenceに裏付けられたプログラム

新藤論文では形成評価を念頭にevidenceの捉え方(≒定義)を示している。形成評価におけるevidenceを論じた3(2)では山口他(2013)の図を引用し、プログラム開発過程で積み上げられていく情報からevidenceとなるものを切り出している。しかし、山口他(2013)は同図に示されている情報(理論や歴史研究から導かれる知見等)は全てevidenceとしてみるのが実践の発展には不可欠という捉え方をしている。またWhite(2019)はEvidence-Based Project Cycleにおいてプログラム考案時に参照すべき情報は全てevidenceであるという記載をしている<sup>5</sup>。

いずれにせよ“Evidence-Based”な取組の本質は、問題解決における最善の意思決定を行うという点にあり。その意味では、問題解決に資するプログラムを開発するという点に重要性があり、有効性を論じない“Evidence-Based”な取組は存在し得ないという点については共通認識があるだろう。新藤論文は“Evidence-Based”

図1 Evidence-Based Project Cycle



(出所) White (2019)

なるプログラムとそれ以外との境界について明確な整理をしたといえる。

## 注記

- 1 新藤論文では、エビデンスという用語の意味合いは“Evidence-Based”という文脈で用いられる場合と日常用語では異なることを強調している。この点を踏まえ、以下では前者の用法に焦点を当てる場合は、evidenceと英語表記することとする。
- 2 三つ目の点については、新藤論文の主張には不明瞭な点が残る。Evidenceは有効性を検証した結果である（新藤論文では「『何が効果があるのか（What works?）』という問いに答えるもの」という記述がなされている）という捉え方は、介入とアウトカム間の因果推論の結果であると言い換えることができるだろう。つまり、定性的に把握したアウトカムであっても、そのアウトカムが介入によってもたらされたことが論じられていればevidenceとなる。しかし、新藤論文では定性的アプローチで得られた質的アウトカムデータそれ自体がevidenceであり、その情報が定量的なevidenceを補完すると位置付けている。

- 3 Evidenceにはレベルがあることから、正確には高いレベルのevidenceに裏打ちされたプログラムのみをEBPプログラムと位置付けている。
- 4 この用語の使い方は新藤論文とは異なる。新藤論文では、White（2019）のEvidence-Based Project Cycle全体を形成評価として捉えていると考えられる。なお、形成評価を新藤論文のように捉えると、形成評価固有のエビデンスの捉え方というものはなく、この問題設定自体が成立しないように思われる。
- 5 新藤論文においても有効性に関する検証を経なければEBPプログラムは確立し得ないということを強調しているのであり、他の情報の必要性を否定しているわけではない。

## 参考文献

- 山口創生・米倉裕希子・岩本華子・他(2013)「社会福祉実践におけるエビデンスとエビデンス構築の過程:理論から実践へ」、『社会問題研究』、62:67-79
- White, H. (2019). *Putting evidence to work, to do what works: The evidence-based project cycle* (<https://www.campbellcollaboration.org/blog/the-evidence-based-project-cycle.html> 2024年1月8日最終閲覧)

## 【リプライ】

## 定性データの扱い及びEBPプログラムの視点に基づく回答

新藤 健太

日本社会事業大学

shindo@jcsu.ac.jp

## 1. 定性データをどう扱うべきか

まずは、エビデンスと定性データの扱いについて改めて検討したい。青柳コメントには「定性的に把握したアウトカムであっても、そのアウトカムが介入によってもたらされたことが論じられていればevidenceになる。しかし新藤論文では、質的アウトカムデータそれ自体をevidenceと捉えている」という指摘がなされている。

確かに、本論において筆者が主張したエビデンス<sup>1</sup>は「介入とアウトカムの間の因果推論の結果」を意味している。そうであれば青柳コメントが指摘している通り、定性データの分析に際してもプログラム効果 (Program effect)、すなわち当該プログラムが特有に寄与し得るアウトカム変化の部分 (Rossi et al. =2005) を取り出し、これを検証しなければならない。例えば定性データを収集する際にも、介入群・比較群それぞれからデータを収集して比較するデザインを検討すべきかもしれない。あるいは最低限「もしこの介入が存在しなかったら現在どのような状況だったと考えるか」と尋ねなければいけないかもしれない<sup>2</sup>。

この点については、本論においても、本論が引用したいくつかの文献においても明確な言及はなかったが (こうした視点を明確には持ち得ていなかったが)、青柳コメントの主張はもっともであり、定性データを分析することによってエビデンスを明らかにしようとするのであれば意識しなければならない視点であるといえる。

## 2. EBPプログラムにおけるエビデンスをどう捉えるか

次に、EBPプログラムにおけるエビデンスについて補足したい。青柳コメントにあるように本論ではエビデンスを①有効性を検証した結果、②検証結果の確証度合い (evidenceのレベル)、③定性データもevidenceになり得ることを述べている<sup>3</sup>。本論で言及したことでこれらに加えるならば、④ (例えばフィデリティ尺度を備えることで) 複数の提供者によってでも正しく当該介入を実施できるようにすることがあった。

EBPプログラムにおけるエビデンスは、特定の提供者のみではなく、他の提供者によっても正しく実施することができ、同じようにその効果を発揮できるからこそ意味がある。そのためEBPプログラムとして位置づけられるIPS援助付き雇用やACTプログラム、家族心理教育にはこのフィデリティ尺度とこれを基にプログラムの概要・実施手順等を説明したプログラム実施マニュアルが備えられている。

近年、こうした発想はソーシャルワーク以外の分野でも見受けられる。例えば、英国政府が1998年に外郭団体として設立したNestaが公表しているエビデンスの基準 (Standards of Evidence) では、RCT等による因果関係の検証 (レベル3) 及びスケールアウト事例に関する因果関係の検証 (レベル4) の先にマニュアル・システム・手順書による一貫したスケールアウト方法の確保 (レベル5) が位置付けられている (Nesta 2013; 家子 2019) (図1)。また、青柳コメントでも紹介のあったWhite (2019)の別の記事では、2010年以

降、ナレッジ・ブローカーの台頭があり、エビデンスの活用を制度化するための構造（図2）が構築されつつあることを説明するとともに、図2に示された上位3つのレベル（エビデンスのポータルサイト、ガイドライン、チェックリスト）が備わっていることで、意思決定者（エビデンスの主要な活用者）は個別の研究論文を確認することなく、エビデンスに基づいた意思決定が可能になることを述べている。こうしたマニュアルや手順書、チェックリスト等を備える、そのために評価情報を生み出し、活用するというのであれば、それは形成的評価に位置づく取組みといえるのではないかと考える。本論でもこうした点を意識しながらEBPプログラムの構築を目指す形成的評価におけるエビデンスについて再考した<sup>4</sup>。

一方で、White (2019) は、(図2に示したような) ピラミッドの階層は新たな基準であり、「従来のようにエビデンスの基準（検証結果の確証度合）を表したのではない」と明確に述べている<sup>5</sup>。そうであれば、本論で言及した「Evidence-Basedと表現する場合のエビデンスとは『介入の有効性に関するものである』』という主張と合致しにくい部分もあり、本論においては、これらの整理が不十分であったり、いくらかの矛盾ははらんでしまっていたりしたのかもしれない。

## 注記

- 1 ここでのエビデンスとはEvidence-Basedと言われる場合のエビデンスであり、青柳コメントにおけるevidenceと同義である。
- 2 すなわち、定量データの分析と同じく事実と反事実

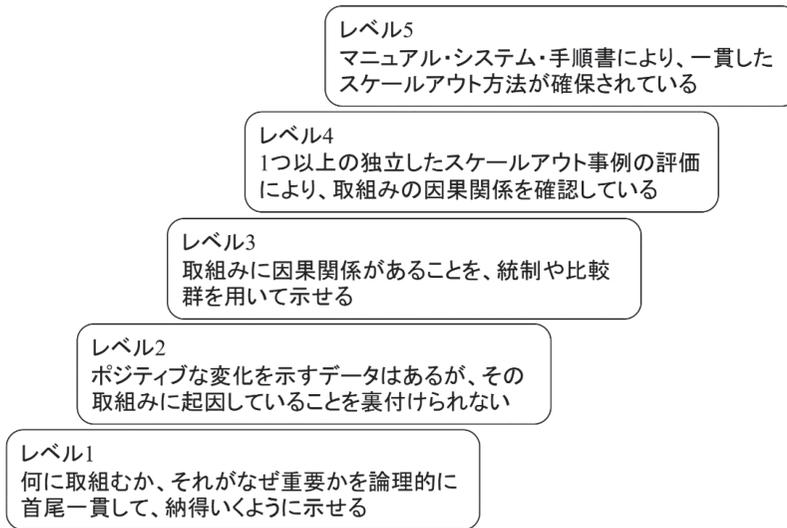
の比較を行う必要がある。

- 3 エビデンス（介入とアウトカム間の因果推論の結果）とするならば、プログラム効果を意識すべき点は前述のとおりである。
- 4 ただし、White (2019) の主張については青柳コメントを受けて、本稿を執筆するにあたって出会った主張であり、本論執筆時から意識していたものではない。
- 5 Nestaにおけるエビデンスの基準 (Standards of Evidence) に規定されるレベル1も介入の有効性とは別次元のものを位置付けている。

## 参考文献

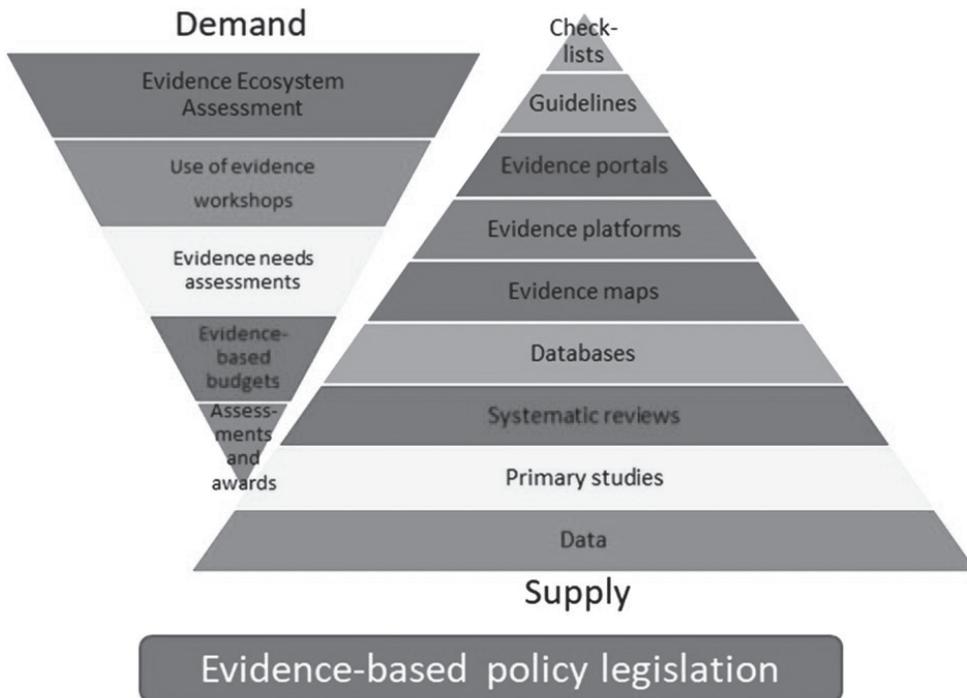
- 家子直幸 (2019) 「対人サービスでのイノベーションをEBPに発展させるための英国での取組み：What Works Centerを中心に」、『日本評価学会第20回全国大会抄録集』、25-28
- Rossi, P. H., Lipsey, M. W. and Fetterman, H. E. (2004). *Evaluation: A Systematic Approach, 7th Ed.*, Sage Publications (=2005、大島巖・平岡公一・森俊夫・他訳『プログラム評価の理論と方法：システマティックな対人サービス・政策評価の実践ガイド』、日本評論社)
- Puttick, R. and Ludlow, J. (2013). *Nesta Standards of Evidence: An approach that balances the need for evidence with innovation* ([https://media.nesta.org.uk/documents/standards\\_of\\_evidence.pdf](https://media.nesta.org.uk/documents/standards_of_evidence.pdf) 2024年1月8日閲覧)
- White, H. (2019). The twenty-first century experimenting society: the four waves of the evidence revolution, *Palgrave Communications*, 5 (<https://www.nature.com/articles/s41599-019-0253-6> 2024年1月8日閲覧)

図1 Nestaによるエビデンスの基準



(出所) Rush et al. (2013)；家子 (2019)

図2 エビデンスの構造



(出所) White (2019)

【実践報告・調査報告】

## 民間非営利セクターにおけるエビデンス活用の現状と課題 — 児童家庭福祉分野団体へのアンケート調査から —

清水 潤子

武蔵野大学

jshimizu@musashino-u.ac.jp

### 要 約

社会課題とその解決主体が多様化する中で、エビデンスに基づく政策立案や実践に関心が高まっている。一方、日本の民間非営利セクターの実践現場において、どの程度エビデンスが活用されているかについては明らかになっていない。本稿は民間非営利セクターの児童家庭福祉分野に従事する団体へのアンケート調査を通じて、エビデンス活用の実態について探索的な調査・分析を行った。調査結果からは、エビデンスの定義に関する認識が多様であり、その多様な認識に基づく活動が展開されていることが分かった。また、エビデンスを介入の効果を実証的に明らかにした知見だと認識する回答は全体的に低く、対人支援サービスの提供の有無をはじめとした団体や活動分野の特徴とエビデンスの定義認識の関係を見ても、これらの項目との関係に統計的に有意な差までは見られなかった。これらの結果を踏まえ、民間非営利セクターの昨今の特性を考慮した上で、エビデンスを適切なカタチで認識し、活用を推進していく際の評価者の役割を考察した。

### キーワード

エビデンスに基づく実践 (EBP)、エビデンス活用、民間非営利セクター、児童家庭福祉、評価

## 1. 問題の所在

### (1) EBPMを取りまく国家の概況

世界恐慌における「市場の失敗」の学びから、包括的な社会保障制度の整備によって誕生した福祉国家は、1970年代以降の官僚主義的な政策への批判や、新自由主義的な理念に基づく社会保障改革により、再び市場経済を強化する方向へより戻しが起きた（金子 2016）。

しかし、その後持続可能な成長に代表される世界経済の動向における変化や、グローバル化、脱工業化、労働市場の変化、ジェンダーや家族

役割の変化、移住や文化的変化、人口動態の変化などは新たな社会的リスクを生み、ガーランド（2016）が福祉国家3.0<sup>1</sup>を提唱していることから明らかなように、福祉国家は過渡期を迎えている。福祉政策や社会保障に代表される国家による福祉の充実は依然追求されるべきものであるが、こと日本においては、経済状況の悪化に加え、当面の人口減少という免れない未来が見えている中で、国家主導の福祉政策の限界も目に見えている。

このような状況下で、より効率的・効果的な政策立案・政策運営を進めていくという点で、

近年のエビデンスに基づく政策立案 (Evidence-Based Policy Making: EBPM) 推進の動きは見逃せない。日本では、2016年12月の「統計改革の基本方針」により政府全体におけるEBPMの定着を目的に、統計改革推進会議が設置された(総務省 2017)。その後、2017年5月の統計改革推進会議の「最終取りまとめ」を踏まえ、EBPMの推進が図られ、各府省のEBPM統括責任者らによるEBPM推進委員会が開催されてきた。同委員会は2023年4月1日付けで内閣総理大臣が議長を務める行政改革推進会議の下に位置づけられている。2023年の政府の骨太方針においても、「効果的・効率的な支出の推進とEBPMの徹底強化」という題目のもと、「事後的な検証が可能な形で事前にKPIの設定と政策効果を検証するためのエビデンス・成果の提出を求め、政策の優先順位の見える化を進める。」(内閣府 2023: 36)とあり、このような動きからも、政府のEBPMへの関心の高まりが見て取れる。

## (2) 社会課題と解決主体の多様化

一方、複雑化する人々の福祉的ニーズの充足を行政とは異なる立場から行う民間非営利セクターに寄せられる期待も増大している。これは単に課題解決の主体の多様化だけの問題ではなく、これまでの制度・施策による行政事業という基盤の上に成り立っていた専門職中心のサービス提供から、必ずしも福祉や教育、心理といった専門職教育を受けた実践家ばかりではなく、多様な経験や背景をもつ個人・団体が現場の支援プログラムに従事したり、開発を進める主体となって活動したりすることが包含されている。しかし、日本の民間非営利セクターにおいて、活動団体によって提供されるサービスや社会的プログラムの質や有効性、社会的意義については、体系的に明らかになっているとは言い難い。

津富(2018)は、エビデンスを活用する最大の目的は「害を与えないため」だとしている。社会課題の解決や、福祉の充実を目指す際、支援対象を取り巻く状況が少しでも良くなることを目的として介入し、害のない効果的な支援を行っていく責任は、課題の解決の主体が行政であれ、民間非営利活動の各主体であれ、本質的

には変わらないものである。専門職集団においては、その倫理的責任が行動指針等で定められ、エビデンスをはじめとする科学的知見の活用に基づく実践 (Evidence-Based Practice: EBPや Evidence-Informed Practice: EIP) が求められるが、民間非営利セクターにおける多様な解決主体による実践において、エビデンスがどのように理解され、活用され、実践されているのか。本稿では、民間非営利セクターの中でも、児童家庭福祉に関連する取り組みを行う団体へのエビデンス活用に関するアンケート調査の分析を通して、民間非営利セクターにおけるエビデンス活用の実態について、探索的に明らかにする。そして、今後の民間非営利セクターにおけるエビデンス活用に向けた評価者の役割を考察する。

## 2. 日本の民間非営利セクターにおけるエビデンスの活用

### (1) 日本の民間非営利セクターの概要

民間非営利セクターといってもその定義は多様である。日本NPOセンター (n.d.) の定義では、「Nonprofit Organization (NPO) は直訳すると非営利組織であるが、より正確な意味では「民間非営利組織」とするのがよく、民間とは「政府の支配に属さないこと」とし、「非営利」は「利益があがっても構成員に分配しないで、団体の活動の目的を達成するための費用に充てること」としている。また「組織」については、「社会に対して責任ある体制で継続的に存在する人の集まり」であるとし、センターが支援する対象は「医療・福祉・環境・文化・芸術・スポーツ・まちづくり・国際協力・交流・人権・平和など、あらゆる分野の市民活動団体等の民間非営利組織で、民間の立場で活動するものであれば、法人格の有無や種類を問わない」(n.d.)としている。

このようなNPOの定義の中に含まれる団体は多く、多くのNPOは法人格をもたない任意団体であり、公益法人や社会福祉法人、一般法人(非営利型)なども民間団体として意識をもって活動していればNPOであり(日本NPOセンター (n.d.))、単に特定非営利活動法人ないし認定特

定非営利活動法人を指すのではないという点で、全体像の把握が難しいとされる<sup>2</sup>。

## (2) 昨今の日本の民間非営利セクターの特徴

民間であることの定義として「政府の支配に属さない」とあるように、民間非営利セクターの特徴としては、一定程度の行政との牽制関係が見て取れる。なかでも、特定非営利活動法人や認定特定非営利活動法人などは広く社会公益的な活動を推進するにあたり、団体のミッションに忠実であることに力点がおかれ、行政事業のように、制度・施策によって事業活動の細目がしばられず、民間性を活用した比較的自由な取り組みや、イノベーション創出に期待が寄せられている。特に制度・施策は対象が厳密に限定される中で、制度の狭間に落ち込んでしまう人々の生活・福祉ニーズに対してサービス提供や互助の推進を通じて貢献したり、まだ行政が認知していない社会的なニーズを拾い上げ、運動や組織化を通じてアドボカシーにつなげていたりする役割を担っている。

一方、2000年に行われた社会福祉基礎構造改革以降、行政が措置的に行っていた事業が、民間に移行されたり、公共の担い手の拡充が進んだりすることで、民間非営利セクターの活動団体の中にも、行政事業の委託を受け行政に代わってサービス提供をする主体も多数存在している。本稿の冒頭に述べた福祉国家におけるサービス提供主体と関連付けて考えると、日本の社会福祉の目的や理念、対象別の福祉サービスに共通する基本的事項を規定している社会福祉法がある。利用者保護の必要性が高い第一種社会福祉事業の経営主体は、原則として国や地方公共団体と社会福祉法人しか担うことができないとされている。しかし、同法で経営主体の定めがなく、利用者への影響が比較的少ないとされ、届出により事業をすることが可能な第二種社会福祉事業においては、特定非営利活動法人や株式会社が参入でき、サービス提供主体として事業を行っている。しかし、本稿が対象分野とする児童家庭福祉に関連する領域でいうと、例えば保育所、助産施設といった児童福祉法に規定される施設や、里親への委託措置や児童養

護施設等への施設入所を解除された児童に対して、自立に向けた支援を行う児童自立支援生活援助事業、家庭において保育を受けることが一時的に困難となった乳幼児に対して保護を行う一時預かり事業、障害児の発達を支援する障害児通所支援事業などが該当する。ここに紹介した施設や事業は、第二種社会福祉事業のほんの一部であるが、果たして本質的に「利用者への影響が比較的に少ない」事業なのだろうか。市民の福祉的ニーズを満たすためのサービス提供事業は、市民活動の一部に過ぎないが、善意に基づく取り組みを超えて、サービス提供主体としての責任も求められている。実践の質の担保や向上を考えるにあたっては、エビデンスの検討は、もはや不可避といっても良いだろう。

## (3) 民間非営利活動団体におけるエビデンス活用に関する状況

「NPO エビデンス」でCiNiiで先行研究を参照し、特定非営利活動法人法が施行された1998年以降に絞り検索をしたところ、44件が論文として該当した。NPOと関係性が不明瞭な論文も含まれたが、論文として目立ったものは、NPO団体の利用者を対象として医学・心理・教育的な介入結果の報告であった。対象としては、発達障害児、がん患者、きょうだい児、自殺対策などであり、教育・心理・医学的な介入研究やその視点からの検討であったが、エビデンスの報告として挙がる件数は圧倒的に少ないといわざるを得ない。また「民間非営利活動 エビデンス」での検索結果は0件であり、同領域におけるエビデンス活用のメゾ的な研究については、研究報告がないことが明らかとなった。

また、「エビデンス」という言葉が多義的に使われている現状は、民間非営利活動領域でも散見される。例えば、環境保全・改善に関する研究・活動に対して助成する東急財団(2023)は、「民間非営利団体(NPO/NGO等)が一定の科学的根拠(エビデンス)に基づいて行う活動等」を助成対象としている。

民間非営利セクターにおいて、昨今の大きな動きとしては2019年度から助成がスタートした休眠預金等活用制度がある。休眠預金等活用に

あたっては、法律・制度文の中に「エビデンス」という言葉は見当たらないが、休眠預金等を活用して民間公益活動推進業務を行う指定活用団体の機能・業務として「知の構造化」が位置づけられている。「知の構造化」とは、「自律分散的に創造された知識と知識の関係性を明らかにすること。これにより、知的価値、経済的価値、社会的価値、文化的価値に結びつけるための方法論を構築し、成果の実装に結び付ける」(内閣府 2017: 2)としている。休眠預金等活用制度が対象とする助成事業は、地域や分野などの多様性を踏まえた社会の諸課題への対応や、イノベーション創出・革新的手法の開発・実装が含まれ、これまででない社会実験的な試みも応援する側面がある。このような介入が想定した成果を生み出すことができたのかを、社会的インパクト評価を用いて検証することを求め、その評価情報を「知」と称しているように読み取ることができる。

一方、助成・委託事業を受けることがあるNPO団体では、助成財団や委託元から請求される領収書や契約書、従事記録といった書類や証憑といった証拠書類を「エビデンス」と呼ぶような文脈も存在している。このように、「エビデンス」を求める側も差し出す側も、多様な「エビデンス」の定義の中におかれており、現場における混乱も想像に難くない。

### 3. 諸外国の福祉的課題解決分野におけるエビデンス活用

諸外国におけるエビデンスに基づく実践や政策に関する先行研究は多数存在するが、福祉的課題の解決分野においては代表的なものとして、Boazら(2019)が、イギリス、アメリカ、カナダ、北欧諸国、オーストラリア・ニュージーランドのエビデンス活用の実態についてまとめている。本研究は国際比較を主眼としているものではないため詳細は割愛するが、先進諸国におけるエビデンス活用を後押しする動きとして2000年に始まったキャンベル共同計画はあまりに有名である。EBPM発祥の地であるイギリスで

は、質の高いエビデンスをつくり・伝え・使いながら政府の意思決定に活かしていくためのイニシアチブとして、What Works Centre (WWC)が進められている。WWCは財政的には政府から独立し、明確な政策領域が設定されていることが基本とされ、社会的に不利な状況におかれている子どもたちの学力向上に関する政策領域ではEducation Endowment Foundation、虐待予防や青少年非行・暴力予防などの早期介入に関する政策領域ではEarly Intervention Foundation、福祉や多面的な豊かさ、ウェルビーイング領域ではWhat Works Centre for Wellbeing、高齢社会の課題においてはCentre for Aging Betterといったセンターがネットワークメンバーとして機能している(家子 2018)。これらのセンターは、日々の実践開発の一貫としても、評価活動の拡大を支持しており、成果(アウトカム)の重要性を強調することを通じて、専門的なサービスの中心にあるべき受益者にとっての価値の重要性を再認識することに貢献している(Ghate and Hood 2019)。

しかし、実践における課題も複数示されている。例えば質の高いエビデンスが見つかり、それを導入するとしても、基礎自治体からの予算や運営方針に左右され、実際に掲載されているエビデンスを活用できる団体は限られるともいわれている(Ghate and Hood 2019)。

### 4. 方法

本研究では「児童福祉領域における非営利組織を対象としたエビデンスの活用に関するアンケート」のデータを活用した。同調査は2019年度トヨタ財団助成「児童福祉領域における知識仲介の研究—機能のモデル化と試行的実装」(研究代表: 家子直幸)<sup>3</sup>の一環として実施し、筆者を含む同助成事業の研究班によって設計したものである。「児童福祉領域における知識仲介の研究—機能のモデル化と試行的実装」では、児童家庭福祉領域において、効果的な予防的プログラムが実践で導入されておらず、エビデンス・プラクティス・ギャップが存在していることから、このギャップを埋めるために、効果的なエビデ

ンスに関する情報を現場実践へとつなぐ Knowledge Broker（知識仲介者）の役割に焦点を当てた。助成プログラムの目的は、日本の児童家庭福祉領域における知識仲介が機能することによって、児童家庭福祉領域における予防的プログラムの実装が試行的に実現することを志向したものである。知識仲介という介入行為が効果的に行われるために、介入対象となる現場のエビデンスに関する認識や、「エビデンス活用」に関するニーズ、エビデンス活用の現状を把握するために、同アンケート調査を実施した。また「エビデンス」という言葉は社会的にいろいろなニュアンスで使われているが、本調査におけるエビデンスの定義は、介入の効果を実証的に裏付けるための情報として、「介入の効果についての実証的な知見」を第一義とした。

調査の実施時期は、2020年11月1日から11月30日であり、調査対象は日本財団のCANPANのデータベース上で、2020年10月末日時点で、主たる活動分野として①こども、②青少年、または③教育・学習支援を行っている登録している団体と、主たる活動分野ではないが、活動分野として①こども、②青少年、または③教育・学習支援を実施していると登録のある団体の内、メールアドレスが公開されていた4,164の非営利組織である。調査の実施方法は、ウェブのアンケート様式「Questant」を用いた。

調査項目は、大きく8つのセクションから構成されており、「①団体の基礎情報（法人格、有給職員の人数、回答者の役職）」、「②法人が行っている児童福祉関連事業の活動分野」、「③対人サービス提供の有無」、「④回答者のエビデンスの定義」、「⑤④で定義したエビデンスの活用度合い」、「⑥「介入の効果についての実証的研究による知見」としてのエビデンスに関する関与・活用度合い」、「⑦「介入の効果についての実証的研究による知見」としてのエビデンスに関する活用希望や活用支援」、「⑧回答者の団体名等属性情報（任意）」である。

「④回答者のエビデンスの定義」に関する指標としては、田辺（2018）のエビデンスに関する概念区分を参照し、加えてNPO支援に携わる実務家を含む研究班での検討の上、指標を追加した。

分析方法は、単純集計によってデータの概観をつかんだ上で、属性や回答結果のクロス集計や検定を行った。統計分析に当たっては、SPSSを活用した。データクリーニングで明らかに入力ミスと考えられる回答や論理的に整合しない回答については、集計から除外し欠損値として処理した。

分析にあたっては、多様な経営主体を内包する民間非営利セクターにおいては、その事業内容や法人種別などによって、エビデンスの認識や活用状況によって違いがあるのではないかとこのことを問いとして設定し、直接的な対人サービスの有無や、法人種別、事業活動分野における違いに着目していく。その背景としては、福祉サービスの担い手が拡張している状況がある。専門職によってサービス提供が推進されることが多い経営主体や（例：社会福祉法人）、前述した先行研究論文が対象とするようなEBPの起源となる医療領域に近接する領域（例：保健医療）において、エビデンスを本調査が第一義とする定義である「介入の効果についての実証的な知見」として認識し、活用している実態があるのではないかと想定した。

また、民間非営利セクターにおける行政からの委託事業の広がりや考慮し、当初は事業単位に焦点を当てた調査を検討したが、民間非営利セクターで活動する主体の多様性や、組織の大きさなどの関連から、調査項目が複雑化するため、事業単位でのエビデンス活用については今回の調査では問わず、法人格や活動分野単位での調査を実施した。

本調査データ分析にあたっては、武蔵野大学人間科学部研究倫理委員会の承認を得ている（受付番号2023-16-01）。

## 5. 結果

### (1) 分析対象団体の基本的な属性

有効回答数438（回答率10.5%）であり、回答団体の法人格では、特定非営利活動法人が212で約半数（48.4%）を占めた。次いで、任意団体69（15.8%）、一般社団法人52（11.9%）、社会福祉法

人41 (9.4%)、認定特定非営利活動法人36 (8.2%)、公益財団法人8 (1.8%)、公益社団法人6 (1.1%)、一般財団法人5 (1.1%)、その他5 (1.1%)、株式会社3 (0.7%)、協同組合1 (0.2%)であった。

団体の有給職員数 (N=438) の最大値は4,965人、最小値は0人 (SD=263.92) であり、平均値は43.93人、中央値は4人、最頻値は0人であった。

回答者の組織・団体における役職 (N=438) については、代表が230 (52.5%)、理事・評議員・役員が89 (20.3%)、マネージャー (中間管理職) が64 (14.6%)、スタッフが48 (11.0%)、ボランティアが6 (1.4%)、その他が1 (0.2%) であった。

団体が主に活動している児童家庭福祉領域における活動分野 (N=435) に関しては、教育が166 (38.2%)、児童健全育成が154 (35.4%)、障害児支援が151 (34.7%) と多数を占めたが、それ以外では家庭支援が111 (25.5%)、その他が78 (17.9%)、保育が55 (12.6%)、保健医療が48 (11.0%)、要保護児童支援が40 (9.2%) であり、また不明が3件あった。

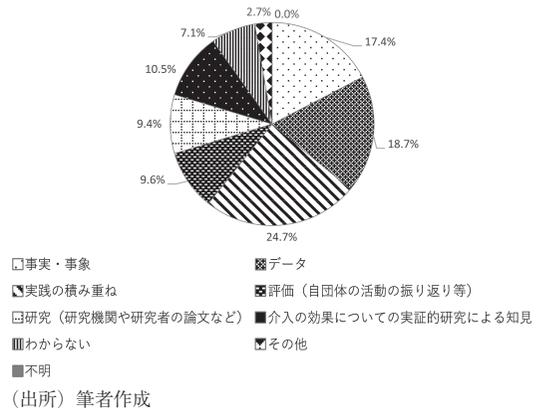
また主に活動している児童家庭福祉に関する事業において、団体のスタッフ自身が対面・オンラインを問わず直接的な支援 (対人援助) を行っているかという設問 (N=438) に対し、行っているが364 (83.1%)、行っていないが74 (16.9%) であった。

## (2) エビデンスに関する考え方

「エビデンスを活用する」と聞いたときの「エビデンス」の定義として、団体の考えに最も近いものを聞いた設問 (N=438) では、実践の積み重ねが108 (24.7%)、データが82 (18.7%)、事実・事象が76 (17.4%)、介入の効果についての実証的研究による知見が46 (10.5%)、評価 (自団体の活動の振り返り等) が42 (9.6%)、研究 (研究機関や研究者の論文など) が41 (9.4%)、わからないが31 (7.1%)、その他が12 (2.7%) という結果であった (図1)。

また法人種別ごとのデータを表1に示す。法人種別ごとの選択項目順位で並べると、「実践の積み重ね」を多く選択した法人として、任意団体、特定非営利活動法人、公益社団法人、社会福祉

図1 エビデンスの定義認識 (SA) (パーセント)



法人があった。次点の「データ」を多く選択した法人として認定特定非営利活動法人、一般社団法人があった。「介入の効果についての実証的研究による知見」を一番多く選択したカテゴリはその他であり、属性としては、地方公共団体、医療法人社団、大学の研究室などであった。その他を除いて、「介入の効果についての実証的研究による知見」の選択の順位が高かった法人は社会福祉法人であった。これらの法人種別であることが、「介入の効果についての実証的研究による知見」をエビデンスとして認識することと関連するかを確認するために、「介入の効果についての実証的研究による知見」の回答を1、それ以外の回答を0とするダミー変数を作成し、統計分析を行ったが、社会福祉法人 ( $p=.359$ )、その他 ( $p=.592$ ) とともに統計的に有意ではなかった。

また、活動領域ごとのデータを表2に示す。一法人が複数の領域にまたがって事業・活動を行っているため、複数回答となっている。「実践の積み重ね」は児童健全育成、家庭支援、保育、障害児支援、教育、その他の分野において一番選択された項目となった。要保護児童支援は「事実・事象」が、保健・医療分野では、「研究」が一番選択された。「介入の効果についての実証的研究による知見」については、障害者支援と保健・医療領域において、第3位となった。活動領域においても、法人種別と同様、特定の活動分野についても、「介入の効果についての実証的研究による知見」のダミー変数を用いて分析を行

表1 法人種別ごとのエビデンスの定義認識の順位 (SA) (回答数)

法人種別	回答数	1位	2位	3位	4位	5位	6位	7位	8位
任意団体	69	実践の積み重ね (24)	事実・事象 (12)	データ (12)	評価 (自団体の活動の振り返り等) (7)	研究 (研究機関や研究者の論文など) (7)	わからない (5)	介入の効果についての実証的研究による知見 (2)	その他 (0)
特定非営利活動法人	212	実践の積み重ね (49)	事実・事象 (41)	データ (34)	研究 (研究機関や研究者の論文など) (25)	介入の効果についての実証的研究による知見 (23)	評価 (自団体の活動の振り返り等) (22)	わからない (13)	その他 (5)
認定特定非営利活動法人	36	データ (10)	実践の積み重ね (7)	評価 (自団体の活動の振り返り等) (6)	事実・事象 (4)	わからない (4)	介入の効果についての実証的研究による知見 (3)	研究 (研究機関や研究者の論文など) (1)	その他 (1)
一般社団法人	52	データ (14)	実践の積み重ね (12)	事実・事象 (11)	介入の効果についての実証的研究による知見 (8)	評価 (自団体の活動の振り返り等) (3)	研究 (研究機関や研究者の論文など) (2)	その他 (2)	わからない (0)
公益社団法人	6	実践の積み重ね (2)	事実・事象 (1)	研究 (研究機関や研究者の論文など) (1)	わからない (1)	その他 (1)	データ (0)	評価 (自団体の活動の振り返り等) (0)	介入の効果についての実証的研究による知見 (0)
一般財団法人	5	データ (1)	実践の積み重ね (1)	研究 (研究機関や研究者の論文など) (1)	介入の効果についての実証的研究による知見 (1)	わからない (1)	事実・事象 (0)	評価 (自団体の活動の振り返り等) (0)	その他 (0)
公益財団法人	8	データ (2)	評価 (自団体の活動の振り返り等) (2)	研究 (研究機関や研究者の論文など) (2)	事実・事象 (1)	その他 (1)	実践の積み重ね (0)	介入の効果についての実証的研究による知見 (0)	わからない (0)
社会福祉法人	41	実践の積み重ね (12)	データ (8)	介入の効果についての実証的研究による知見 (7)	わからない (6)	事実・事象 (5)	その他 (2)	評価 (自団体の活動の振り返り等) (1)	研究 (研究機関や研究者の論文など) (0)
協同組合	1	データ (1)	事実・事象 (0)	実践の積み重ね (0)	評価 (自団体の活動の振り返り等) (0)	研究 (研究機関や研究者の論文など) (0)	介入の効果についての実証的研究による知見 (0)	わからない (0)	その他 (0)
株式会社	3	実践の積み重ね (1)	研究 (研究機関や研究者の論文など) (1)	わからない (1)	事実・事象 (0)	データ (0)	評価 (自団体の活動の振り返り等) (0)	介入の効果についての実証的研究による知見 (0)	その他 (0)
その他	5	介入の効果についての実証的研究による知見 (2)	事実・事象 (1)	評価 (自団体の活動の振り返り等) (1)	研究 (研究機関や研究者の論文など) (1)	データ (0)	実践の積み重ね (0)	わからない (0)	その他 (0)

(出所) 筆者作成

ったが、いずれの活動領域においても統計的に有意な関連は認められなかった (要保護児童支援:  $p=1.00$ 、児童健全育成:  $p=.193$ 、家庭支援:  $p=1.00$ 、保育:  $p=.468$ 、障害児支援:  $p=.07$ 、教育:  $p=.336$ 、保健・医療:  $p=.142$ 、その他:  $p=.308$ )。ただし、有意水準を下回りはしなかったが、他の活動領域と比較して、「介入の効果についての実証的研究による知見」をエビデンス定義として認識した人の割合が多かった「障害児支援」と「保健・医療」においては、 $p$ 値も他

と比較して低くなった。

相談支援など直接的な対人支援の提供をしている団体と、そうでない団体の間にエビデンスの認識に違いがあるかについても、エビデンスの定義のダミー変数を用いて同じようにクロス集計 (表3) を用いて、カイ二乗検定を行い分析した。対人援助サービスを実施している方が、していない方に比べて、エビデンスを「介入の効果についての実証的研究による知見」として認識する度合いは多かったが、こちらについて

表2 活動分野ごとのエビデンス定義認識の順位 (MA) (回答数)

活動分野	回答数	1位	2位	3位	4位	5位	6位	7位	8位
要保護児童支援	40	事実・事象 (11)	データ (9)	実践の積み重ね (7)	介入の効果についての実証的研究による知見 (4)	評価 (自団体の活動の振り返り等) (3)	研究 (研究機関や研究者の論文など) (3)	その他 (2)	わからない (1)
児童健全育成	154	実践の積み重ね (36)	事実・事象 (29)	データ (28)	研究 (研究機関や研究者の論文など) (20)	評価 (自団体の活動の振り返り等) (15)	介入の効果についての実証的研究による知見 (12)	わからない (9)	その他 (5)
家庭支援	111	実践の積み重ね (27)	データ (23)	事実・事象 (18)	研究 (研究機関や研究者の論文など) (14)	介入の効果についての実証的研究による知見 (12)	評価 (自団体の活動の振り返り等) (10)	わからない (5)	その他 (2)
保育	55	実践の積み重ね (18)	データ (9)	事実・事象 (7)	研究 (研究機関や研究者の論文など) (7)	評価 (自団体の活動の振り返り等) (5)	介入の効果についての実証的研究による知見 (4)	わからない (3)	その他 (2)
障害児支援	151	実践の積み重ね (44)	事実・事象 (29)	介入の効果についての実証的研究による知見 (22)	データ (18)	評価 (自団体の活動の振り返り等) (10)	研究 (研究機関や研究者の論文など) (10)	わからない (9)	その他 (5)
教育	166	実践の積み重ね (48)	データ (35)	事実・事象 (22)	研究 (研究機関や研究者の論文など) (18)	評価 (自団体の活動の振り返り等) (16)	介入の効果についての実証的研究による知見 (14)	わからない (9)	その他 (4)
保健・医療	48	研究 (研究機関や研究者の論文など) (11)	事実・事象 (10)	介入の効果についての実証的研究による知見 (8)	実践の積み重ね (6)	データ (5)	評価 (自団体の活動の振り返り等) (4)	わからない (4)	その他 (0)
その他	78	実践の積み重ね (18)	事実・事象 (14)	データ (14)	介入の効果についての実証的研究による知見 (11)	評価 (自団体の活動の振り返り等) (8)	わからない (6)	研究 (研究機関や研究者の論文など) (4)	その他 (3)

(出所) 筆者作成

も統計的に有意な関連は認められず (p =.677)、直接的な対人支援の有無とエビデンスを「介入の効果についての実証的研究による知見」として認識するかには関連しないことが分かった。

表3 対人支援サービスの提供と介入の効果選択ダミーのクロス表

		対人援助サービスの提供		合計	
		している	していない		
ダミー介入の効果選択	0	度数	327	65	392
		対人援助サービスの提供の%	89.8%	87.8%	89.5%
	1	度数	37	9	46
		対人援助サービスの提供の%	10.2%	12.2%	10.5%
合計		度数	364	74	438
		対人援助サービスの提供の%	100.0%	100.0%	100.0%

(出所) 筆者作成

(3) エビデンスの活用に関する実態

各団体のイメージに近い定義に基づく「エビデンス」の活用を行っているかという問い (N=437) に対して、「行っている」が272 (62.2%)、「行っていない」が79 (18.1%)、「わからない」が86 (19.7%) という回答であった。各団体のイメージに近い定義に基づく「エビデンス」の活用率については表4にまとめた通りである。「わからない」という回答を除き、どのエビデンスの定義の項目においても6割近く、もしくはそれ以上の活用の実態があることが分かった。定義ごとの活用率で一番高い項目は「評価 (自団体の活動の振り返り等)」(81.0%) であり、次に「実践の積み重ね」(72.2%)、そして「介入の効果についての実証的研究による知見」(67.4%) となり、「実践の積み重ね」を除き、必ずしもより広く認識されている定義 (例: 「事実・事象」や「データ」) が活用されているのではないことが分かった。

表4 各団体の定義に基づく活用状況 (SA)

	定義 (A)	定義に基づく活用 (B)	B/A
事実・事象	76 17.4%	48 17.6%	63.2%
データ	82 18.7%	48 17.6%	58.5%
実践の積み重ね	108 24.7%	78 28.7%	72.2%
評価 (自団体の活動の振り返り等)	42 9.6%	34 12.5%	81.0%
研究 (研究機関や研究者の論文など)	41 9.4%	26 9.6%	63.4%
介入の効果についての実証的研究による知見	46 10.5%	31 11.4%	67.4%
わからない	31 7.1%	0 0.0%	0.0%
その他	12 2.7%	7 2.6%	58.3%
合計	438 100.0%	272 100.0%	

(出所) 筆者作成

各団体に、「介入の効果についての実証的研究による知見」を活用しているかについて聞いた設問 (N=438) では、「はい」が111 (25.3%)、「いいえ」が169 (38.6%)、「わからない」が158 (36.1%) と回答した。同質問に「はい」と回答した111団体の法人種別は任意団体 (10.8%)、特定非営利活動法人 (46.8%)、認定特定非営利活動法人 (9.9%)、一般社団法人 (15.3%)、一般財団法人 (0.9%)、公益財団法人 (4.5%)、社会福祉法人 (7.2%)、協同組合 (0.9%)、その他 (3.6%) であった。

「介入の効果についての実証的研究による知見」を活用している111の団体と、各団体のエビデンスの定義に基づく活用状況を比較したデータを表5に示す。このデータから読み取れることとして、「介入の効果についての実証的研究による知見」を活用していると回答した団体は、かなり高い割合で各団体の定義認識に基づくエビデンスの活用を行っていることが分かった。もともと「介入の効果についての実証的研究による知見」を自団体のイメージに近いエビデンス定義として認識している団体に加え、「評価 (自団体の活動の振り返り等)」を自団体のイメージに近いエビデンス定義として認識している団体における「介入の効果についての実証的研究による知見」の活用状況は100%であった。

「介入の効果についての実証的研究による知見」の活用目的を聞いた問い (複数回答) の結果において一番多く選択されたものは、「事業を実施する過程で有効性を確認するため」(74) であり、次いで「事業改善のため」(62)、「スタッ

表5 各団体の定義に基づく活用状況 (SA) と実証的知見の活用状況 (SA) の比較

	介入の効果についての実証的な知見を活用している団体のエビデンス定義別内訳 (A)		介入の効果についての実証的な知見の活用も各団体の定義に基づく活用もしている団体数 (B)		B/A
事実・事象	11	9.9%	10	90.9%	
実践の積み重ね	25	22.5%	22	88.0%	
評価 (自団体の活動の振り返り等)	10	9.0%	10	100.0%	
研究 (研究機関や研究者の論文など)	15	13.5%	14	93.3%	
介入の効果についての実証的研究による知見	30	27.0%	30	100.0%	
わからない	0	0.0%	0	0.0%	
その他	1	0.9%	0	0.0%	
合計	111	100.0%	104		

(出所) 筆者作成

フのモチベーション向上のため」(33)、「広報やアドボカシーを実施するため」(33)、「事業立ち上げのため」(26)、「ファンドレイジングのため」(14)、「その他」(9)と続いた。「その他」の中には、会員サービスの向上のためや、社会への啓発活動、政策提言が含まれた。全体として法人種別によって、エビデンス活用の目的に大きな特徴の違いは見られなかったが、「介入の効果についての実証的研究による知見」の活用の目的として、任意団体、一般財団法人、社会福祉法人、協同組合、その他の団体はファンドレイジングでの活用は行っていなかった。

また同じく、「介入の効果についての実証的研究による知見」を活用していると回答した団体に対して、「今までどのようにその知見を入手してきたか」(複数回答)について尋ねた問いの結果は「本・報告書などの文献」(67)、「学会やセミナー」(54)、「大学、個人研究者、研究機関」(48)が上位3つのカテゴリとなった。次いで、「専門職の研修、能力開発講座など」(43)、「インターネット情報(オンライン情報)」(39)、「政府機関」(24)、「その他」(21)、「助成財団、中間支援団体」(17)、「新聞、雑誌、大衆紙、ラジオ、テレビ」(11)となった。「その他」の項目の中には、自団体の知見や、協働している団体からの提供、自団体で行った調査研究や実地調査(アンケート調査、受益者・当事者からの聞き取り内容)、などが含まれた。

「介入の効果についての実証的研究による知見」を今後活用したいと思うかという問いに対して、「とてもそう思う」が103(23.5%)、「そう思う」が170(38.5%)、「どちらでもない」が66(15.1%)、「そう思わない」が8(1.8%)、「全くそう思わない」が0(0%)、「その他」が91(20.8%)であった。

今後「介入の効果についての実証的研究による知見」を活用しようと考えた際に、どのようなハードルがあるか(複数回答)については、「現場に落とし込むための財政的リソースが足りない」(160)、「どのように探せばいいかわからない」(125)、「探したり読む時間がない」(119)、「現場に落とし込む裁量がない」(112)、「活用の仕方がわからない」(110)、「わからない」(60)

「日本語の論文・文献がない」(35)、「特にない」(33)、「その他」(41)の順で回答が得られた。

また、「介入の効果についての実証的研究による知見」を活用する際に、どのようなサポートがあると良いかという問い(複数回答)については、「利用しやすい形でまとめられていること」(247)、「活動支援の一環として、活用に関する活動のための財源があること」(218)、「活用ができる人材を紹介してくれること」(157)、「研究者とつなげてくれること」(138)、「探し方を教えてくれること」(102)、「読み方・見分け方を教えてくれること」(93)、「わからない」(50)、「特にない」(14)、「その他」(10)という結果であった。

## 6. 考察

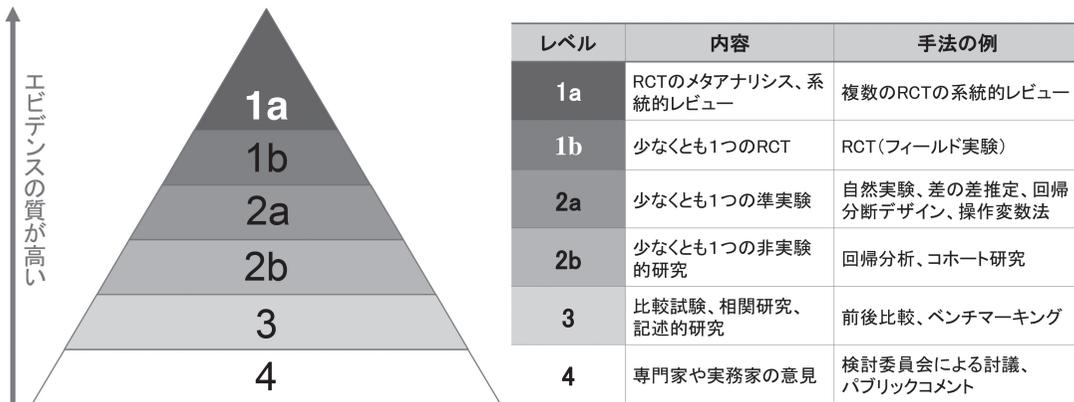
### (1) 調査結果から見えてきたこと

本稿では、民間非営利セクターの中でも、児童家庭福祉に関連する取り組みを行う団体へのエビデンス活用に関するアンケート調査を通して、民間非営利セクターにおけるエビデンス活用の実態について探索した。

アンケート調査では、エビデンスを活用する際の定義に近いものを候補から選択してもらい、その定義に基づく「エビデンス」を活用しているかについて尋ねた。その結果、約6割が何らかの定義に基づくエビデンスを実践で活用していることが分かったが、エビデンスと定義されているものが多様な状況の上で成り立っている活用の実態が浮き彫りになった。加えて、エビデンスの定義について民間非営利セクターにおける多様な児童家庭福祉の提供主体において、特定の活動分野や対人支援活動事業の有無から特徴があるかを捉えようと試みたが、エビデンスを「介入の効果による実証的な知見」と認識しているかどうかと、これらの要素の間に統計的に有意な関係性までは認められなかった。

これらの結果が示唆するものとして、今回サンプル数は限られてはいるものの、児童家庭福祉の支援に取り組む民間非営利セクターの団体においては、まず何をエビデンスとして定義す

図2 エビデンスレベルと手法の具体例



(出所) 惣脇 (2010) 等を参考に三菱UFJリサーチ&コンサルティング作成

るかという点で、「介入の効果による実証的な知見」をエビデンスだと認識している団体は、全体的に少ないことと、多様なエビデンス認識に基づく活動がある程度行われているが、すべての実践において、エビデンスが重視されているとは言い難い状況がある。

これは、例えば事実がエビデンスと認識されるのであれば、「クライアントが〇〇と言った。ゆえに、介入をした。」もエビデンスの活用になりうる状況があり、データがそのままエビデンスと解釈されるのであれば、「市民にとつたアンケートの結果が〇〇だった。ゆえに、事業を開発した」もエビデンスの活用である。実践現場からの応答で一番回答数の多かった「実践の積み重ね」に関しては、「実践=介入」だとするのであれば、それを行った結果がどうであったかという振り返りに基づく実践(シングル・システム・デザイン)や、実践による知見が形式知化され、その知見に基づいて、次の介入が検討されているのであれば、それは図2のピラミッドが示すように、エビデンスの質のレベルは高くはないが、本稿が定義する「エビデンス」の活用は行っていると言える可能性がある。ただ、本調査においては、「実践の積み重ね」という言葉以上の具体的な積み重ね方や、そこからの知見の抽出をすることはできなかった。

「介入の効果についての実証的研究による知見」のイメージとして近いものは、例えば子育て

に悩む親の支援を行う必要があるときに、どのような介入を行うことがその時点においてベストか(過去に効果が証明された介入手法がないか)を参照する際に活かす知見がある。また、過去に参照できるエビデンス情報がない場合は、自らが介入のデザインを検討し、その介入が効果があったのかどうかを確かめたことで得られる知見がある。後者は、本調査で示した定義例にも示したような評価(自団体の活動の振り返り等)における中核的なテーマであると言える。現場実践における評価に関連する活動や省察が、エビデンスの概念や考え方と的確に結びついていないまま、現場実践が行われている様相が見て取れ、改めてこの点について整理することが、現場における実践や最終受益者の便益を高める可能性が見えてきた。

## (2) 民間非営利セクターの特徴を踏まえたエビデンス活用の検討

本結果が民間非営利セクターの団体のエビデンス活用の実態を捉えたものとするとき、統計的に有意とまではいかなかったが、法人種別や分野によってエビデンスの認識や活動度合いに特徴が見られたことは興味深いことである。例えば、「介入の効果についての実証的研究による知見」をエビデンスとして認識をするかという点においては、法人では社会福祉法人や、医療法人社団や地方公共団体といった科学的な知見

を日常から活用する専門家を多く含む法人や行政において、比較的多くそれがエビデンスだと認識されていたり、活動分野においても、障害児支援や保健・医療といった項目でより多く認識されていたりしていた。

また、「介入の効果についての実証的研究による知見」を活用していると回答した団体の多くが、自団体の定義のイメージに近いエビデンスの定義に基づくエビデンスの活用を行っていたことも興味深い結果となった。これはある意味で、エビデンスの認識の違いはあれ、何かしらの情報を参照したり、活用したりしながら、事業を開発したり改善したりしていく土壌が民間非営利セクターにおける実践現場にあるとも考えられる。

しかし、本調査の半数以上の回答者となる特定非営利活動法人や認定特定非営利活動法人といういわゆるNPO法人や、その他の法人については、「介入の効果についての実証的研究による知見」をエビデンスとして認識する状況は限定的だと言える。この現状については、いくつかの課題点がある。ひとつは、前述したように、第二種社会福祉事業のような形で事業を受託するようなことや、同事業に限らず子どもやその家族に対人支援を提供する際に、サービスの質や内容を考慮していくためにも、実証的な知見からの学びを参照することが「害を少なくする」ことにつながることを、考える必要がある。

一方、前述のNPOの定義に内包される市民活動の中には、社会プログラム化されていない運動や互酬性による活動も含まれる。隣のおばさんのおせっかいにエビデンスを求める人はいないだろう。しかし、近年ではおせっかいが事業化して、居場所事業となり、そこに孤独や貧困の解消という目的が加わると、エビデンスを求められるようなことが起きている。特に、子どもや家庭といった対象に直接・間接的問わず介入をすることで「害」があってはならない領域であれば、なおさらである。エビデンスを探しているよりも、行動力や、「誰かのために何かをしたい」という想いが先にくることがある活動において、エビデンスとどのような距離感を保つかは、その事業や団体の運営主体によっては

厄介な問題でもある。おそらく本調査の結果についても、より科学的な知見を背景にもつ者が見れば、この活用状況を課題だと捉える可能性が高い。一方で、エビデンスを活用するというプロセスには相応の手続きが必要になる。加えて、アンケート結果にも出ていたように、活用したいと思っても財政面や時間、リテラシーの不足といった課題点も挙げられる。内発的な取り組みや地域の自主的な取り組みや運動においては、社会的なプログラムの意図のように明確でないものや、互酬のように働きかけや介入の起点が不明瞭なものもあり、その価値や価値創出のプロセスをやみくもなエビデンス重視論で減じる可能性があることにも自覚的である必要がある。

### (3) 評価者の役割

今回の調査から、「介入の効果についての実証的研究による知見」の活用を今後してみたいかという問いについては、約6割がその意向を示していた。一定、民間非営利セクターの児童家庭福祉領域において関心も寄せられている中で、民間非営利セクターの現場にエビデンスや知を適切なかたちで認識し、活用する動きを推進するために、評価者ができることにどのようなことがあるだろうか。

まずエビデンスの認識については、本調査でも浮き彫りになった多様な認識の状況を踏まえて、評価者は事業者やステークホルダーとの間での言葉の定義や意味の目線合わせをガイドできるだろう。プログラム評価を実施する際において、エビデンスの定義が関係者間でずれていると、ニーズ評価やセオリー評価の内容や質にも影響する。ただし、民間非営利セクターの活動主体は、誰しもが実証的な知見へのアクセスを持っていたり、慣れている環境におかれていたりするわけではないことから、エビデンス自体に不慣れである可能性もある。評価活動の各ステップにおいて、厄介な言葉であるエビデンスを丁寧に扱う必要があるだろう。

活用については、本研究で明らかとなったように、エビデンスを活用する際の課題として、財政的な指摘もあった一方、「どのように探せば

いいかわからない」という回答が多く見られた。また、「利用しやすい形でまとめられていること」が活用に向けて助けになるという声や、専門家の紹介、リサーチスキルを高めるようなテクニカルな介入を求める声もあった。

社会的なプログラムを対象とするプログラム評価にそって事業評価を進めていく際には、たとえば事業をデザインするときに、自分たちが検討している介入プログラムは、最善なものなのか、すでに同じような対象に対する同様なプログラムはないか、といったことを問いかけ、そのような評価設問に答えていくための支援が必要だと言える。形成的な評価の段階でできることである。

または実際の介入プログラムが、本当に有益かどうか（エビデンスがあるのか）を明らかにするために、実験デザインやその他の方法を用いて調査を進め、そこから分かったデータをもとに、活動を改善し、アドボカシーにつなげていくことなども可能である。

ただ、エビデンスに照らした実践を進めていくにあたり、前述したイギリスのようにエビデンスが参照できるシステムの構築も必要であるほか、エビデンスについても、介入効果に関する知見にのっとった上で、頑健なものから、弱いものまで幅広く射程に収める必要もあるだろう。また、いくら頑健なエビデンスでも、それが使える環境でなければ意味を持たない。良いエビデンスが、よく実装されるかは別の話である。事業デザインが思い付きではなく、過去から学びながら、それがうまく実装できていない状況などをクリティカルに考えるプロセスに評価者が伴走することで、実践家のエビデンス活用の視点を含む評価キャパシティの形成に貢献できる余地がある。そしてそれは当然各主体の評価可能性や、存在目的などに寄り添いながら考慮されるべきであろう。

また、エビデンスをある介入の効果を表付けるための情報と考え、それを確かめていくには当然実証的な方法論が必要となる。しかし、冒頭にも述べたように複雑化している社会において、人々の福祉的なニーズを満たしていく中で attribution を特定していくことは困難を伴うこと

も多い。現場では官民協働による包括的な支援体制を組むことや、政策課題的には異なっても、同じアジェンダのもとで集合的なインパクトを個々が生み出し、その成果を統合的に見ていくような動きが見え始めており、実証主義・ポスト実証主義的な評価哲学の守備範囲も越えつつある。このようなパラダイムにおけるエビデンスのあり方について、多様な評価手法や哲学的基盤などを参照しながら考え続ける必要があるだろう。

## 7. 本研究の限界と今後の課題

本研究では「児童福祉領域における非営利組織を対象としたエビデンスの活用に関するアンケート」をもとに、法人格や活動分野といった属性をもとに分析を行った。しかし、民間非営利セクターにおける市民活動の内容は、これらの変数にとどまることなく多様であり、活動の性質などによっても違いがある可能性がある。エビデンス活用の必要性が相対的に低いような取り組みも含まれるが、今回はこの点に関しては直接的な対人支援サービス提供の有無程度でしか把握することができなかった。より深く活動の性質をとらえ、適切な活用支援を考慮していくことが、やみくもにエビデンスを推奨することの防止にもつながりうるため、今後の課題としたい。

また、活動の性質は当然組織や事業、財政状況にも影響するものだが、エビデンスの活用に向けたハードルとして財源不足が多く指摘され、エビデンス活用に向けた財源支援ニーズも顕在化されたが、それを今回深掘りすることはできなかった。この点についても団体単独の努力ではなく、諸外国で見られるようなエビデンス活用に向けたシステムや支援構築を含む体制整備と不可分な課題であるため、継続的な分析課題としたい。

## 謝辞

本稿の執筆にあたり、「児童福祉領域における

非営利組織を対象としたエビデンスの活用に関するアンケート」に回答して下さった民間非営利活動団体の皆様と、「児童福祉領域における知識仲介の研究－機能のモデル化と試行的実装」の研究班の家子直幸氏、石川陽介氏、加藤剛氏、西郷民紗氏、越智真奈美氏（アンケート実施当時、順不同）に書面を借りて御礼申し上げます。また、本稿の修正に際し、匿名の査読者より、大変貴重なご指摘をいただきましたことに感謝申し上げます。

## 注記

- 1 ガーランド（2016）は第二次世界大戦後、1950～60年代に新たな統治様式として生まれた福祉国家を福祉国家1.0、1970年代の経済的混乱の中で、「ニューライト」の改革者たちによる、より限定的かつ緊縮的なバージョンの福祉国家を福祉国家2.0と定義している。
- 2 2023年4月末時点で、特定非営利活動法人は50,310、認定NPO法人は1,268ある。また、2020年12月末時点で、公益社団法人は4,175、公益財団法人は5,439あり、2014年以降漸増が続いている（助成財団センター 2023）。また社会福祉法人は2021年時点で、21,003法人ある。
- 3 <https://toyotafound.my.salesforce-sites.com/psearch/JoseiDetail?name=D19-R-0153>（2023年9月30日閲覧）

## 参考文献

- 家子直幸（2018）『イギリスWhat Works Movementsの5年史－研究と政策・実践をつなぐWhat Works Network－』、第2回エビデンスに基づく実践と政策セミナー資料（2018年6月30日開催）
- 金子充（2016）「第7章 市場経済への不信」、坪洋一・金子充・室田信一著『問いからはじめる社会福祉学－不安・不利・不信に挑む』、有斐閣、151-174
- ガーランド、D.(2016)『福祉国家－救貧法の時代からポスト工業社会へ』、白水社
- 総務省（2017）『参考資料2 統計改革の経緯等について』 [https://www.soumu.go.jp/main\\_content/000522778.pdf](https://www.soumu.go.jp/main_content/000522778.pdf)（2023年9月30日閲覧）
- 惣脇宏（2010）「英国におけるエビデンスに基づく教

- 育政策の展開」、『国立教育政策研究所紀要』、第139集
- 田辺智子（2018）『エビデンスと政策の間－評価利用の研究が示すもの－』、第2回エビデンスに基づく実践と政策セミナー資料（2018年6月30日開催）
- 津富宏（2018）『なぜ、エビデンスを「つかう」のか？「つかう」とは何か？』、エビデンスを「つかう」を学ぶ入門ワークショップ資料（2018年10月24日開催）
- 東急財団（2023）『東急財団2024年度多摩川の美しい未来づくり助成』 <https://foundation.tokyu.co.jp/kankyo-josei/>（2023年9月30日閲覧）
- 内閣府（2017）『休眠預金等活用制度における各主体の役割等』 [https://www5.cao.go.jp/kyumin\\_yokin/shingikai/20171129/shiryou\\_1-1.pdf](https://www5.cao.go.jp/kyumin_yokin/shingikai/20171129/shiryou_1-1.pdf)（2023年9月30日閲覧）
- 内閣府（2023）『経済財政運営と改革の方針2023』 [https://www5.cao.go.jp/keizai-shimon/kaigi/cabinet/honebuto/2023/2023\\_basicpolicies\\_ja.pdf](https://www5.cao.go.jp/keizai-shimon/kaigi/cabinet/honebuto/2023/2023_basicpolicies_ja.pdf)（2023年9月30日閲覧）
- 日本NPOセンター（n.d.）『NPOに関するQ&A：NPOの基礎知識』 [https://www.jnpoc.ne.jp/?page\\_id=134#a01](https://www.jnpoc.ne.jp/?page_id=134#a01)（2023年9月30日閲覧）
- 三菱UFJリサーチ&コンサルティング（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～』 [https://www.murc.jp/wp-content/uploads/2022/10/seiken\\_160212.pdf](https://www.murc.jp/wp-content/uploads/2022/10/seiken_160212.pdf)（2023年9月30日閲覧）
- Boaz, A., Davis, H., Fraser, A., & Nutley, S. (2019). *What Works Now?: Evidence-informed Policy and Practice*, Policy Press.
- Ghate, D., & Hood, R. (2019). Using Evidence in Social Care, in Boaz, A (Eds.), *What Works Now?: Evidence-informed Policy and Practice*, Policy Press, 89-109. (2023.12.31 受理)

## **Evidence Utilization in the Japanese Nonprofit Sector -Analysis on a Survey of Child and Family Welfare Organizations-**

Junko Shimizu

Musashino University  
jshimizu@musashino-u.ac.jp

### **Abstract**

Since social issues and the entities that solve them have become increasingly diverse, evidence-based policymaking and practice have gotten much attention. Nevertheless, the extent to which evidence is utilized in the Japanese nonprofit sector must be clarified. This research study conducted an exploratory survey and analysis of the nature of evidence utilization among the organizations engaged in child and family welfare in the nonprofit sector. The survey results indicate that diverse perceptions of the definition of evidence exist. Therefore, their activities are being developed based on these insights. Overall, responses recognizing evidence as findings that empirically reveal the effectiveness of interventions were low, and no statistically significant differences existed in perceptions of the definition of evidence when examining its relationship to organizational characteristics, including whether interpersonal support services were implemented. Based on these results, this paper discussed the role of evaluators in supporting nonprofit players in understanding evidence definitions and utilizing them by considering the characteristics of the Japanese nonprofit sector.

### **Keywords**

Evidence-Based Practice, Evidence Utilization, Nonprofit Sector,  
Child and Family Welfare, Evaluation

## 【コメント】

## エビデンスの活用を促進するために

新藤 健太

日本社会事業大学

shindo@jcsu.ac.jp

清水論文では、「介入の効果についての実証的な知見」<sup>1</sup>をエビデンスと認識している団体が少ないこと、そしてサービスの質を向上させるためにエビデンスのより積極的な活用が必要であることが指摘されている。こうした指摘は児童福祉領域のみではなく社会福祉領域全般に当てはまる。そもそも、エビデンスを活用するためにはこうした知見に関する一次研究が不可欠であるが、それが豊富に存在しているわけではない。EBM（科学的根拠に基づく医療）やEBP（科学的根拠に基づく実践）には「つくる」「つたえる」「つかう」のプロセスがあることが知られているが、我が国の社会福祉領域においては「つくる」の段階で十分な取組みが行われていない現状がある<sup>2</sup>。

なお、仮にこうした一次研究、あるいはシステマティック・レビューを経た頑強なエビデンスがあったとしても、それがなかなか活用されない現状があることは清水論文が指摘する通りである。例えば、Lehman et al. (1998) はPORT (Patients Outcomes Research Team) 勧告という国家レベルの治療ガイドラインに必要な治療・支援と判断された介入のうち、薬物療法は90%以上の患者へ提供がなされているのに対して、家族心理教育やACTプログラム、IPS援助付き雇用といった社会プログラムについてはニーズを持つ対象者の10～30%程度にしか提供されていない状況があることを明らかにしている。大島(2010)は、その対策としてエビデンスに基づく実践の実施システム（例えば、社会福祉領域の実践現場における組織体制等）及び社会システム（例えば、それぞれの実践現場における取組

みをバックアップするための法制度等）の整備が必要であることを指摘している。これは主に「つたえる」や「つかう」に関連した課題であるといえるが、このことは必ずしも個々の実践現場にのみ課題があるのではなく、社会システム（制度等の仕組み）にも大きな課題があることを示唆している。

こうした状況に対して清水論文では「評価者ができること」として①関係者間でのエビデンスの認識合わせ、②エビデンスを利用しやすい形で整理し提供すること、③評価専門家の紹介や実践家のリサーチスキルを向上させるためのサポート等を挙げている。さて、こうした取組みがエビデンスの活用を促進し得るだろうか<sup>3</sup>。

前述した、実施システムを整備するという点に関してはProctor et al. (2009) が「実装研究の概念モデル」を作成し、実装研究（すなわちエビデンスの実装）が達成すべきアウトカムを①実装アウトカム、②サービスアウトカム、③対象者アウトカムの3レベルに区分した上で、これを達成するための実装戦略を説明している（図1）。また、アメリカ連邦保健省薬物依存精神保健サービス部（2009）はこうした実施システムを整備し、エビデンスに基づく実践の活用を促進するためにEBPツールキットプロジェクトを開始し、エビデンスに基づく実践の実施・普及に貢献している。

一方、社会システムを整備するという点に関しては政策立案者に向けてPolicy Briefを作成する取組みなどが行われている。Policy Briefには、エビデンスを紹介するのみではなく、エビデンスを実践現場に実装するとしたら、どのような

障壁が生じ得るか、こうした障壁にどのように対処すべきかを記載する。さらに、障壁やその対処に対して不確実性が存在するのであれば、それが何なのかをも記載し、これを受けて、何をモニタリングし、どう評価すべきなのかについての提案も行う（松本 2020）。WHOは2012年にEVIPNet (Evidence-informed Policy Network) という情報ネットワークを構築し、SURE (Supporting the Use of Research Evidence) というフレームワークを用いて前述のPolicy Briefの作成に取り組んでいる。

このように、エビデンスの活用を促進するためには、情報提供を超えた積極的な実施システム・社会システムへの介入が必要になる。エビデンスを蓄積し、情報提供さえ行えば、自然とその活用が進むだろうという考えはやや楽観的であり、エビデンスとはこれを扱う土壌を耕して、初めて活用され、社会の改善に貢献できる代物といえるのではないだろうか。

注記

- 1 清水論文ではエビデンスのことをこのように表現しているため、このコメントでもそれに倣う。
- 2 ここでの「つくる」には海外でつくられたエビデンスの日本での適応可能性を検証すること（外的妥当性の検証）も含む。

- 3 エビデンスを活用する動きを推進するためということだが、ここでの①や③は「つくる」に関連する事項かもしれない。

参考文献

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

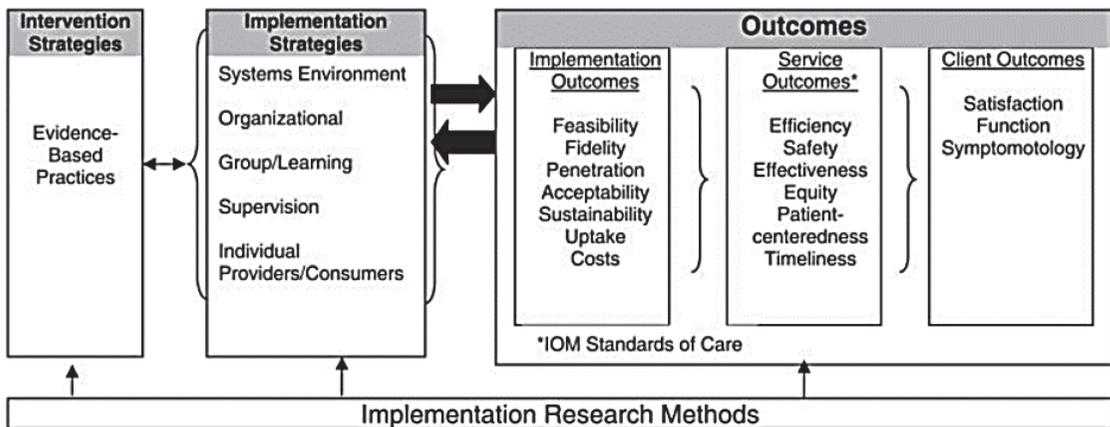
大島巖 (2010)「精神保健福祉領域における科学的根拠にもとづく実践 (EBP) の発展からみたプログラム評価方法論への貢献：プログラムモデル構築とフィデリティ評価を中心に」、『日本評価研究』、10(1):31-41

松本晴樹 (2020)「第3章コメント EBMから学ぶEBPMに必要なステップとEBPMならではの注意点」、大橋弘編『EBPMの経済学：エビデンスを重視した政策立案』、東京大学出版会

Lehman AF and Steinwachs DM (1998). Patterns of usual care for schizophrenia: Initial results from the Schizophrenia Patient Outcomes Research Team(PORT) Client Survey, *Schizophrenia Bulletin*. 24(1), 11-20.

Proctor EK, Landsverk I, Aarons G and et al. (2009). Implementation Research in Mental Health Service. *Adm Policy Ment Health*, 36(1), 24-34.

図1 実装研究の概念モデル



(出所) Proctor et al. (2009)

【リプライ】

## エビデンス活用の促進に関する指摘への回答

清水 潤子

武蔵野大学

jshimizu@musashino-u.ac.jp

筆者は本論において、民間非営利セクターにおける児童家庭福祉分野団体へのアンケート調査から見えたエビデンス活用の現状と課題を踏まえ、エビデンスを適切な形で認識し、活用を促す際の、評価者の役割について言及した。新藤コメントのポイントは、エビデンス活用の促進にあたって、エビデンスを蓄積し、情報提供を行えばエビデンスの活用が進むという観点では不十分であり、実施システムや社会システムの観点からエビデンス活用に向けた検討が不可欠であるという指摘だと捉えている。

本論の「諸外国の福祉的課題解決分野におけるエビデンス活用」でもエビデンス活用を促進するイニシアチブの存在にふれた通り、グローバルな潮流から遅れをとる日本において、エビデンスに基づく実践の実施システムや社会システムといったマクロな体制整備を進める必要性があるという指摘には大筋合意する。特に児童家庭福祉分野においては、英国のWhat Works Centre (WWC) の取り組みは注目に値する。WWCは公共政策に携わる地方自治体職員や対人援助職などの現場の実践者の意思決定において、何が有効か (what works) に関する最良のエビデンスを提供する機関であり、エビデンスの創出、活用に向けた翻訳、プログラムの普及や実装という役割をもつ (家子 2022)。特に本論で焦点化したエビデンスの活用とも重なるエビデンスの「翻訳」においては、システムティック・レビュー等の二次研究によって創出したエビデンスを、その有効性や頑健性といった軸で数値化したり、意思決定・実装支援を目途にエビデンスを比較可能な形で見える化したりするといった取り組

みが行われている (家子 2022)。本論で活用したアンケート調査もエビデンスに関する情報を現場実践へとつなぐKnowledge Broker (知識仲介者) の介入を効果的に行う前段階として実施したものであることは記述の通りであり、日本におけるエビデンス仲介機関の在り方や実装を意識していないものではない。

では、なぜ本論の帰結がシステム構築や体制整備に向けた提唱に及ばなかったのか。

第一に、本調査でも明らかになったように、実践者によってエビデンスの定義認識も異なる状況では、仮にシステムが構築されたとしても、活用される状況には程遠いことが予想される。「エビデンス」というようなカタカナ言葉が現場に混乱を生じさせる事象は散見されるが、その間に生まれる多様な解釈の差を埋め、その活用のレベルや求めるエビデンスの質にグラデーションがあったとしても、実践者がエビデンスを一定程度「自分事」だと認識するためには、マクロ的な体制整備だけではなく、よりミクロな介入が必要になると考えられたからである。

第二に、諸外国でも推進されているエビデンス活用のプラットフォームの運営やエビデンス実装がすべてうまくいっているかという点、課題もある。詳細は本論で言及している通りであるが、諸外国とはサービスの提供体制や、政策を支える国家の体制も違う中で、システムを構築していくための検討は、一評価者の役割を超えており、単なる必要性の主張は、現場にとって空虚なものになりかねないと考えたからである。

第三に、本論はエビデンス活用を推進する対

象者を、制度福祉や公共政策の中だけでなく、広く民間非営利セクターという中で捉えているところに特徴がある。これは、特に福祉的なサービス提供の責任が、国から民間へと移行されてきている中、対人支援、ないしはそのプログラム提供を行う主体が多様化し、それによって実践者のエビデンスリテラシーも、求められるエビデンスの種類や質も、その実践に関わるステークホルダーも多様化したことを意味する。

一方、行政の役割を補完する民間と、民間の創発性や自由さを活かす取り組みとの間で、社会的に求められるものも異なり、私たちの暮らしの延長線上にある市民活動にはプログラム化されないものや、介入が意図的でないことも多い。このような状況において、より政策的観点からの影響を色濃く残すエビデンス活用やシステム構築のやみくもな推奨は、場合によっては市民性や民間性を脅かすことにもつながりかねない。エビデンスが提供するものは、特定の状況下において何が有効かという情報であったとき、それが最大の利害関係者である受益者の意思決定にどのように反映するかは、受益者の知識や環境、価値観と統合して行われるものである（西郷 2018）。そして、それが実践家の価値観や、彼らが属する団体のミッション等にも影響するものであるとき、何が価値で、何が利害関係者にとって有益で、現実的に活用可能なのかについての視点を持つ者こそが評価者であり、調査者による知識情報の紹介とは一線を画すものだと考える<sup>1</sup>。

この点で、本論では、より現場に近いところで実践に寄り添う評価者を想定し、現場におけるエビデンスのリテラシーや評価キャパシティを高める役割を意識的に検討した。新藤コメントでは「情報提供を超えたシステムへの介入」という文言があり、何を「システム」として捉えるかにもよるが、現場に伴走することを、ある意味積極的な介入とするのであれば、大きくスタンスは変わらないのかもしれない。

一方、評価者といわれる者が、筆者のようにより現場に近いところで関わるような場合と、例えば政策評価等を行うような立場における評価者とは、おそらく観点も変わってくるだろ

う。また、諸外国で先行するような知識仲介やエビデンス活用推奨機関において、「評価者」が実際どのような役回りを果たしているのかについては、不明瞭な部分も多い。本論やリプライでも述べたようなエビデンス活用のミクロな支援を提供していくには、人材育成が必要であり、それこそシステムとして考えていく必要がある。新藤コメントの指摘からは、エビデンス活用に向けた社会システム構築に、知識仲介者側の人材育成が不可欠であるということに改めて気づかされた。そしてそこにどう多様な哲学やディシプリンを背景にもつ「評価者」が関わりうるのか、関わるべきなのかといった議論を、日本の現場の課題に照らしてさらに深めていく必要があるという示唆を頂いた。

## 注記

- 1 評価者 (Evaluator) と調査者 (Researcher) の違いについては、Podems (2019) やWanzer (2021)を参照のこと。

## 参考文献

- 家子直幸 (2022) 「対人サービスにおけるイノベーションをエビデンスに基づく実践プログラムに発展させるための英国の取組み－児童福祉分野の What Works Centre を中心に－」、『日本評価研究』、22(2):39-50.
- 西郷民紗 (2018) 『エビデンスのつかいかた～現場の取組みと政策形成への示唆～』、エビデンスを「つかう」を学ぶ入門ワークショップ資料 (2018年10月14日)
- Podems, D., R. (2019). *Being an Evaluator: Your Practical Guide to Evaluation*. Guilford Press.
- Wanzer, D., L. (2021). What is Evaluation? Perspectives of How Evaluation Differs (or Not) from Research. *American Journal of Evaluation*. 42(1), 25-46.

## 【研究論文】

不完全一対比較表を活用した参加型評価手法の提案  
— 評価結果に差がつく相対評価を目指して —

飯田 洋市

公立諏訪東京理科大学

youichi.iida@gmail.com

## 要 約

多くの地方自治体では、緊縮財政を迫られる一方、社会保障に係る費用の増加や住民からの要望の多様化などにより財源への要求が減らないという課題がある。このような背景から、地方自治体では、予算面を念頭に、どの施策あるいは事業を重点化するか決定するための評価結果の高低に差がつく相対評価が求められている。一方で、地方自治の観点から、このような評価には住民の声が反映されることが望まれる。本論文の目的は、階層分析法における一対比較を用いることで、評価結果の高低に差がつく相対評価手法を提案することである。階層分析法の中でも、不完全一対比較表に関するハーカー法による相対評価値を用いる。また、この評価手法が評価結果の高低に差がつく相対評価であること、また、この評価手法が実用的であることを、2019年から2023年に実施してきた長野県岡谷市の商業活性化計画での活用事例により示す。

## キーワード

参加型評価、階層分析法、相対評価、事前評価、合意形成

## はじめに

近年、人口減少問題などを抱える多くの地方自治体では、多様化する住民からのニーズの全てには対応できないという課題を抱えている。そこで、地方自治体として取り扱う施策あるいは事業（以下、これらを区別せず施策と呼ぶ）をビルド&スクラップ<sup>1</sup>しながら施策を推進していくために評価結果の高低に差がつく相対評価が求められている。本論文は、この相対評価に関するものである。

さて、国において、エビデンスに基づく政策形成、いわゆるEBPM（Evidence-Based Policy

Making）が推進される中、自治体においてもEBPMに関する取組みが進められている。事業をビルド&スクラップしていくためにも、エビデンスに基づく議論が期待されている。ただ、地方自治体の場合、各施策の予算規模がそれほど大きくないことや、住民に近い施策を扱っていることを考慮する必要がある。また、費用対効果などの数値だけではなく、住民の価値観や意見も取り入れていくことが望まれる。この視点から、本論文が提案する評価手法は、住民の参加しやすさを意識している。EBPMには達しないが、データドリブンな政策立案・決定を目指すものである。

ところで、佐藤（2021）は、自治体におけるエビデンスとはなにかに関する文脈の中で、「住民意識調査の結果に基づいて施策を立案するといっても、住民が重視する施策における目的と行政の活動との間には距離があります」と指摘している。実際、本論文が事例として取り上げる商業活性化計画でも、計画の策定段階では、商業者と行政で意識の乖離が見られた。「行政がやることだから、意見は求めるが、既定路線に落ち着くに違いない」「商業者にとっては、市が企画する会議は、市が決めた事業を追認するための場でしかない」という商業活性化会議の委員からの声もあった。筆者はこの計画の策定段階から現在まで知識経験者として参加しているが、評価の伴走者の立場から、住民でもある委員と行政職員の距離を、評価を通じて縮めたいという問題意識があった。

さて、本研究の目的は、施策（あるいは計画）に関係する住民の意見を集約して評価することで、複数の施策を対象に評価結果の高低に差がつく相対評価を得るための手法を提案することである。ここで提案する手法は、市民の意見を集約する視点から、経験や実感に基づく主観的判断を定量化できる意思決定手法である階層分析法（the Analytic Hierarchy Process。以下、AHP）を応用したものになる。本論文では、その簡略化の仕方から「roughly AHP」と呼ぶことにする。具体的には5節で説明する。そして、6節で、この簡略化した手法の実用性と有用性を示すために、2019年から2023年までに継続的に実施した長野県岡谷市の商業活性化計画の重点施策の重要度と優先度への適用事例を示す。

本研究での適用事例は、規模も小さく、評価者も少数（10人強）である。また、商業という枠組みの中での施策の相対評価ということから、評価者間で強い対立を引き起こすようなものでもない。理想的な環境での意思決定である包括合理性モデル<sup>2</sup>での適用になる。

## 2. 先行研究

行政評価の分野におけるAHPの活用事例など

を概観する。

佐々木（2002）は、美術館評価の一部として実施した「美術館のバリアフリーに関する実態調査」においてAHPを活用し、美術館の職員に対して、6つの評価基準の全ての組み合わせについて、3つの選択肢「同じくらい」「やや重要である」「重要である」から一つ回答を求めるとともに、調査紙において評価基準を比較する前に、それらを重要と判断する順番に並べることを勧めている。

小川（2004）は、自治体の歳出配分評価で川崎市の住民を対象にAHPを活用し、そのためのアンケート調査（インターネット調査、サンプル数500人）を実施し、集団としての評価値の集約に幾何平均を使っている。松田（2004）は、政策間予算配分でAHPを活用し、アンケートの結果などは架空としながらも、2,000人を対象に1,832人から有効回答があったとし、評価基準については15組、代替案については全部で90組について2要素間の比較を行っている。

佐藤（2009）は、AHPにより政策の優先順位付け評価モデルを構築し、大阪府豊中市などで実証実験と被験者を対象としたアンケート調査によりその有効性を検証し、基本事業の優先順位付けに「経済性」「効率性」「有効性」を評価基準としてあげ、また、自治体レベルでの活用として、川西市、瀬戸市、横須賀市、三重県、宮城県などでAHPの適用が試行あるいは導入されていると述べている。高村（2003）は「国および地方自治体またはこれに準ずる機関が行う、国民・市民・住民等に関連する意思決定」を公共意思決定と定義し、住民の価値観が多様化する時代にあって、公共意思決定の新しい枠組みが必要であると主張、それを支援する数理的手法の一つとしてAHPを取り上げている。

島田（1999）は、住民参加型評価について、住民意識調査は広く行われているが、行政評価にその結果が活用されている例はきわめて少ないことを指摘し、AHPにも触れている。

以上より、AHPは行政評価において有効な手法であり、また、多数の住民を対象に活用できるといえる。本研究で提案するroughly AHPは、行政評価の文脈で一般市民を対象にAHPを適用

しようとするものであるが、従来のAHPと異なり、主観的判断を定量化するために必要な設問の回数を減らすアプローチを採用している。先行研究では見つけられなかった新しい試みといえる。

### 3. 階層分析法の概略

階層分析法（AHP）は、1970年代にT.L.サーティにより創始された、不確実な状況や多様な評価基準における意思決定手法で、主観的判断とシステムアプローチを組合わせた問題解決型意思決定手法として知られている。主観的判断を数値化するところに特徴があり、相対評価したい2つの要素をサーティ尺度<sup>3</sup>（表1）により比較する。これをAHPにおける一対比較と呼ぶ。AHPの目的は、この一対比較により、各代替案の総合評価値を算出することである。

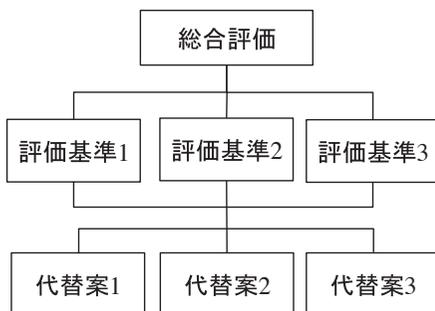
#### (1) AHPの手順

AHPは次の3つの手順からなる。

##### 手順1：階層図を描く

まず、相対評価する「代替案」（施策や事業など）を決め、次に、それらを相対評価する「評価基準」を決める。そして、これらを階層図としてまとめる（図1）。

図1 階層図のイメージ



（出所）小野・田淵（2001）

##### 手順2：各階層にある各要素の重みづけ

「評価基準」については総合評価に関する相対評価を、「代替案」については各評価基準に関する

相対評価を行う。この相対評価では、AHPにおける一対比較を利用する。

ここでの一対比較は、同じ階層にある2つの要素（たとえば、2つの評価基準）について、一つ上の階層にある要素（たとえば、総合評価）に関して「どちらがより重要か」「それはどれくらい重要か」という質問に回答することをいう。この質問は「どちらがどれくらい重要か」とまとめることもできる。なお、比較内容により、「どちらがどれくらい優先的か」などとする。

そして、この質問に対する回答は、サーティ尺度（表1）を参考に形容詞で回答することになる。そして、その回答は表1により一対比較値に変換される。なお、サーティ尺度は英語による表現であるため、AHP活用者が母語にあわせた表現に変更することになるが、これらには定められた規則はなく、活用する状況に応じて工夫することになる。AHPが日本に紹介された経緯から、日本語表現では、刀根（1986）や木下（2000）を参考にした表現がよく使われている<sup>4</sup>。

実際の回答として、たとえば、総合評価に関して「評価基準1が評価基準2よりかなり重要」と判断されれば、「評価基準1は評価基準2より7倍重要」と数値化される。これは、同時に「評価基準2は評価基準1より1/7倍重要」も意味する。

これらの結果を表にまとめたものを一対比較表と呼ぶ。また、これを数学で扱うために、行列  $A = (a_{ij})$  で表現し、これを一対比較行列と呼ぶ。今の場合、 $a_{12}=7$ 、 $a_{21}=1/7$ となる。以下、一対比較表と一対比較行列を適宜読み替えるものとする。一対比較行列の全ての対角成分は $a_{ii}=1$ と決められている。この一対比較行列は $a_{ji}=1/a_{ij}$ という性質を持つが、AHPでは一対比較の結果を反映する重要な性質としている。

何らかの理由で一対比較の結果が得られない場合は、空欄を持つ一対比較表が得られるが、これを不完全一対比較表（あるいは、不完全情報一対比較表）と呼ぶ。これに対して、全ての一対比較結果が得られる場合は完全一対比較表（あるいは完全情報一対比較表）と呼ぶ。

以下、完全一対比較表から各要素の重みを算出する方法を説明する。さて、完全一対比較行列  $A = (a_{ij})$  は、実数倍を除いて、ただ一つの全

表1 サーティ尺度

刀根 (1986) の定義	サーティ (1977) の定義	重要度の強度 (一対比較値)
同じくらい重要	Equal importance	1
若干重要	Weak importance of one over another	3
重要	Essential or Strong importance	5
かなり重要	Demonstrated importance	7
絶対的に重要	Absolute importance	9

(※) 2,4,6,8は補間的に用いる。上の数値の逆数は、後の項目から前の項目をみた場合に用いる。

(出所) 刀根 (1986) のp.14

での成分が正の固有ベクトルを持つこと、そして、この固有ベクトルは絶対値が最大となる正の固有値  $\lambda_{max}$  に属することが知られている (ペロンの定理)。そこで、得られた一対比較行列の固有値  $\lambda_{max}$  を求め、それに属する固有ベクトルを計算する。さらに、このベクトルの成分の和が1となるように実数倍して求められるベクトルの成分が各要素の重みとなる。成分の和を1とする計算をAHPでは正規化と呼び、最終的に得られる正規化された固有ベクトルを重みベクトルと呼ぶ。

### 手順3：総合評価

最後に、手順2で求めた各代替案の重みについて、同じく手順2で求めた評価基準の重みによる加重総和を計算することで、各代替案の総合評価値が求められる。

上記の手順にはないが、総合評価値による各要素の順位を説明するために、手順2で求めた一対比較表を吟味し、最終的な結果への影響などを確かめることが推奨されている。また、主観的判断により得られる一対比較表は矛盾を含むことが少なくない。一対比較行列  $A = (a_{ij})$  は、任意の3つの成分について、 $a_{ij} \times a_{jk} = a_{ik}$  の関係を満たすとき、完全整合であると呼ぶ。これは、一対比較の結果による任意の3つの要素間に、一対比較値の関係を含めた推移性に矛盾が全くないことを意味する。たとえば、推移性の関係から「評価基準1は評価基準2より3倍重要 ( $a_{12}=3$ )」と「評価基準2は評価基準3より3倍重要 ( $a_{23}=3$ )」から、「評価基準1は評価基準3より9倍重要 ( $a_{13}=9$ )」が期待されるが、回答者の主観的判断から「評価基準1は評価基準3より7倍重要 ( $a_{13}=7$ )」

や「評価基準1は評価基準3より1/3倍重要 ( $a_{13}=1/3$ )」という回答が得られている場合に矛盾を含むとされる。

そもそもサーティ尺度に9を超える数値は設定されていないため、整合性に注意する必要がある。そこで、サーティは整合性の目安となる整合度指標C.I.を提案している ( $n$ は比較対象とする要素の個数)：C.I. =  $(\lambda_{max} - n)/(n - 1)$ 。一般に、 $\lambda_{max} \geq n$  が成り立ち、一対比較行列が完全整合の場合はC.I.=0で、整合性が悪くなると数値が大きくなることが知られている。サーティは、C.I.が0.1未満のとき (状況によっては0.15未満のとき) に、一対比較行列は整合的であるとしてよいとしている<sup>5</sup>。

### (2) 不完全一対比較表とハーカー法

不完全一対比較表から各要素の重みを推定する方法にハーカー法がある ((木下2000, pp.64-66) や (Kułakowski 2020, pp. 68-71) を参照のこと)。

まず、推定したい重みベクトルを  $W = (W_j)$  ( $n$ 個の成分を持つ列ベクトル) とし、空欄のある一対比較行列  $A = (a_{ij})$  から、空欄の無い一対比較行列  $B = (b_{ij})$  を次のように作る。一対比較値  $a_{ij}$  が値を持つ場合は  $b_{ij} = a_{ij}$ 、一対比較値が得られていない場合は  $b_{ij} = w_i/w_j$  とする。簡単な計算により、この行列  $B$  は、全ての成分が非負である行列  $\tilde{B}$  に変形できることがわかる。そして、行列  $B$  が既約と呼ばれる性質を持つ時、完全一対比較行列と同様の性質を持つことが知られている (ペロン・フロベニウスの定理)。この場合、行列  $\tilde{B}$  の固有値  $\tilde{\lambda}_{max}$  に対する正規化された固有ベクトル  $\tilde{W}$  を求めることができ、このベクトル  $\tilde{W}$

が不完全一対比較表による重みベクトルになる。

この場合の整合度指標C.I.は、 $C.I. = (\bar{\lambda}_{max} - n) / (n - 1)$  で与えられる (Kulakowski 2020, p.135)。ハーカー法により得られた重みベクトル  $\bar{W}$  の成分を代入することにより得られる不完全一対比較行列  $B$  ( $\bar{B}$ でも同じ) が完全整合であれば  $\bar{\lambda}_{max} = n$  であり、 $C.I.=0$ である。

### (3) 集団による意思決定

$n$ 人からなる集団の意見をAHPで集約する場合、サーティは、各メンバーから得られる一対比較表  $A^{(k)} = (a_{ij}^{(k)})$  の成分ごとの幾何平均をとることで、集団としての一対比較表  $A = (\sqrt[n]{\prod_{k=1}^n a_{ij}^{(k)}})$  を作成するとしている。これにより、AHPの一対比較表  $A = (a_{ij})$  が持つ  $a_{ji} = 1/a_{ij}$  という性質が満たされる ((刀根 1986, pp. 40-41) や (木下 2000, pp. 67-68) を参照のこと)。

## 4. 本研究の目的

本研究の目的は、住民の意見を集約し、施策に優先順位をつけるための評価結果の高低に差がつく相対評価手法を提案することである。具体的には、AHPの手順2を大幅に簡略化する手法の提案である。そして、事例を通して、この手法の実用性と有用性を示すことである。

ところで、意思を反映させる相対評価では、誰が評価するかが大きな課題になる。本論文では、限定的にはなるものの、1節の最後で触れた包括合理性モデルにおいて意思表示が可能な人達、すなわち、宮川 (1994) がいう「各代替的行動案に対してあらゆる面にわたって結果を予測し、その結果の望ましさを目的に照らして判断して、代替案を順位づけることができる」人達を想定している。

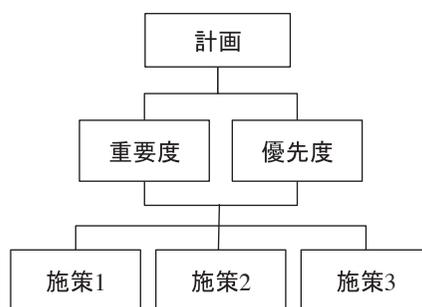
本論文の適用事例 (商業活性化計画) では、評価対象となる事業に直接関係する住民であり、市内商業関連の代表者でもある商業活性化会議の委員に、今回提案する手法で評価してもらうことを通して実用性などを検証した。この評価は、次年度予算申請で予算化の根拠のための評価と位置付けられている。これにより、評価

が形骸化することなく、また、市にとっても価値ある評価といえる。

## 5. 本論文で提案する手法

本論文で提案する手法を説明するために、ここでは、ある計画に紐づく三つの施策1、2、3を相対評価することを考える。評価基準は「重要度」と「優先度」にする。手順1として、これらを階層図にまとめる (図2)。

図2 施策評価の階層図



(出所) 筆者作成

さて、本論文で提案する手法は、手順2の設問数を大幅に減らすものである。従来のAHPでは、たとえば、評価基準の「重要度」に関して、3つの施策に関する全ての組み合わせについて一対比較し完全一対比較表を得ることで、「重要度」に関する各施策の重みを算出することになる。一方、本論文で提案する手法では、①評価基準の「重要度」に関して3つの施策を重要性の大きい順に並べ、②隣り合う施策に関してのみAHPの一対比較を行うことで不完全一対比較表を作り、③ハーカー法で各要素の重みを計算するものである。つまり、評価基準に関する代替案 (この場合は施策) の重みを、不完全一対比較表から算出するところに特徴がある。これにより、たとえば、比較する対象となる要素が6つであれば、AHPでは15回の比較が、5回の比較で済むことになる。ただし、従来のAHPと比較してかなり粗い計算となることから、本論文では、roughly AHPと呼ぶことにした。

一方、これには次のような重要な長所がある。まず、意識的に作り出す不完全一対比較表は、完全一対比較表から重みベクトルを算出する方法に比べて、重みの意味づけが理解しやすい。また、従来のAHPと同様に固定した一つの要素を基準としないながらも、最も少ない比較回数で全ての代替案の相対評価値を算出できる。さらに、この不完全一対比較表は必ず完全整合となるため、C.I.の値が大きいという理由により、回答者による一対比較表を、集団としての重みの計算から排除したり、一対比較値を恣意的に変更する必要が全く無いなどが挙げられる。

なお、適用事例では、サーティ尺度も簡略化し「少し重要」「重要」「かなり重要」から回答を選択する方法を採用した<sup>6</sup>。これにより、サーティ尺度のうち3つの一対比較値(3、5、7)以外は使わないことになる<sup>7</sup>。これは、本論文の目的が、評価者の潜在的な意思に関心があるわけではなく、かつ、細かな重みの差を数値化することより、評価結果の高低に差がつく相対評価手法を提案することにあるためである。

以下、次節の適用事例を想定して、roughly AHPの手順2を説明する。まず、重要と思われる順に右から並べる(表2の2行目)。これはどちらの施策が重要かのみに関する2要素間の一対比較により簡単に行うことができる。表2では、施策2が最も重要、続いて、施策1、施策3である。次に、一番左に配置された施策3を基準にすぐ右隣の施策1がどれくらい重要か判断する。続いて、施策1を基準にすぐ右隣の施策2がどれくらい重要か判断する。選択肢は3つあるが、表2は「少し重要」と「かなり重要」が選択されている。

表2 一対比較値を得るための回答欄

重要度	3位	2位	最重要
施策番号	3	1	2
該当するものに○をつける		<div style="border: 1px solid black; padding: 2px; display: inline-block;">少し重要</div> 重要 かなり重要	少し重要 重要 <div style="border: 1px solid black; padding: 2px; display: inline-block;">かなり重要</div>

(出所) 筆者作成

次に、「少し重要」は3倍、「重要」は5倍、「かなり重要」は7倍とし、表2と組み合わせること

により、以下を得る:

施策3の相対評価値:施策1の相対評価値:施策2の相対評価値=1:1×3:1×3×7.

同様にして、評価者2の各施策に関する相対評価値を得る(表3の3行目)。最終的に集団としての相対評価値および重みを求めるため、ここで正規化する必要はないが、この相対評価値を正規化すればハーカー法による重みと同じになる(完全整合していることからほぼ明らか)。

最後に、この2人の集団としての重みを計算する: 施策1の相対評価値:施策2の相対評価値:施策3の相対評価値= $\sqrt{45}:\sqrt{105}:\sqrt{1} \approx 0.374:0.571:0.056$ .

以上より、2人の評価者による評価を集約した「集団としての重要度」は、施策1の重みが0.374、施策2の重みが0.571、施策3の重みが0.056となる(表3の5行目)。

表3 集団としての重要度(評価者2人)

評価者	施策1	施策2	施策3
1	3	21	1
2	15	5	1
幾何平均	$\sqrt{45}$	$\sqrt{105}$	$\sqrt{1}$
重み(重要度)	0.374	0.571	0.056

(出所) 筆者作成

同様に、優先度の重みを計算する。ところで、適用事例では、手順3を用いず、ここまで求めた重要度と優先度の重みを、そのまま最終的な重要度と優先度としている。「重要度」と「優先度」を一対比較して重み付けすることで、手順3により総合評価値を得ることもできる。

以上により、重要度に関していえば、集団として最重要とした施策は施策2(0.571)、続いて、施策1(0.374)、施策3(0.056)となる(表3の5行目)。従来のAHPでの重みの算出方法と比較して、固有ベクトルなどを使わない分、算出過程が分かりやすいといえる。

## 6. 事例

### (1) 岡谷市商業活性化計画の概要

岡谷市商業活性化計画（岡谷市、岡谷商工会議所）は、2018年3月に取りまとめられた「岡谷市の商業環境に関する調査報告書」にもとづいて、5年10年後の商業振興に向けたあるべき姿を示すために策定されたものである。

岡谷市商業活性化計画では、最終的に3つの基本戦略と11の重点施策が設定されている（図3）。商業活性化計画に当たっては、各重点施策について重要度と優先度を設定し、それぞれの施策の重要度と優先度をABCの3段階で評価して記載する<sup>8</sup>。Aとする事業は2つか3つとし、次年度の予算編成で重点化を目指すことになる。重要度と優先度のABCは以下のように決められている（岡谷市・岡谷商工会議所 2019）：

<重要度>

- A: 取組みを進める上で、集中的に資源投入を行い、課題解決に向けて取組むもの
- B: マンパワーでの取組みを含め、早急に課題解決に向けて取組みを講じるもの
- C: 各施策の進捗を踏まえて取組みを講じるもの

<優先度>

- A: 改善、解決に向けて取組むべき喫緊の課題として、最優先に取組むべきもの（今年度中に何らかの成果を求めるべきもの）
- B: 早急に改善、解決に向けて対処すべきもの
- C: 各施策の進捗を踏まえて検討を行うもの

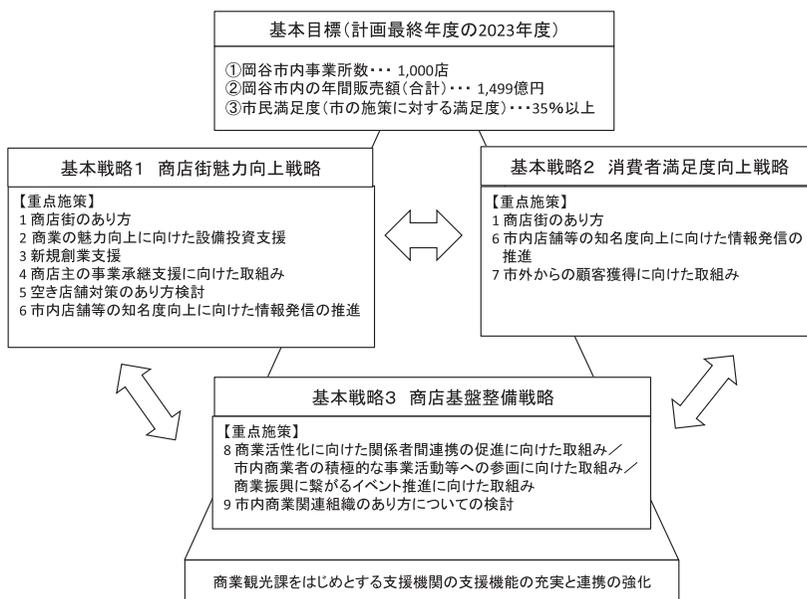
この重要度と優先度のABCは、行政だけで決定するのではなく、商業活性化計画を策定した全委員が関与するのが望ましいとなり、研究の一環と明示し、本論文で提案する手法を適用したものである（飯田 2019, 2021）。

### (2) 重要度調査によるABCの決定

商業活性化会議の各委員の意見を集約して各重点施策のABCを決定するために、roughly AHPを適用する。このために、「岡谷市商業活性化計画に関する重要度調査」をこれまでに4回実施してきた。本論文では、2020年度に実施した2回目の重要度調査を事例として紹介する<sup>9</sup>。重要度調査の実施概要は、以下の通りである：

調査目的：基本戦略に紐づく重点施策の重要度と優先度の相対評価値と評価の決定

図3 岡谷市商業活性化計画の全体像



（出所）長野県岡谷市の「岡谷市商業活性化計画」より抜粋

図4 基本戦略1に関する設問の例

【問】「基本戦略1商店街魅力向上戦略（市内店舗、商店街、商業会向けの取組み）」には6つの重点施策があります。

6つの重点施策：①、②、③、④、⑤、⑥

- ① 重点施策1 商店街のあり方（消費者の商店街（買物）に求める価値の多様化への対応）  
 キーワード 買物をきっかけに人が集う街、  
 街が一つのモール（おかやモール）と見立てた取組み  
 子育て世帯支援、高齢者支援、買物弱者支援
- ② 重点施策2 商業の魅力向上に向けた設備投資支援  
 キーワード 店舗の魅力向上、キャッシュレス決済対応
- ③ 重点施策3 新規創業支援  
 キーワード 創業前支援、創業後支援、  
 効果的な情報発信と創業希望者の市内への呼び込み
- ④ 重点施策4 商店主の事業承継支援に向けた取組み  
 キーワード 事業承継に向けた啓発、親族内承継、親族外承継
- ⑤ 重点施策5 空き店舗対策のあり方検討  
 キーワード 情報発信のあり方、空き店舗の価値向上に向けた取組み
- ⑥ 重点施策6 市内店舗等の知名度向上に向けた情報発信の推進  
 キーワード 継続した情報発信、ネット広告等SNSによる情報発信

上記の6つの重点施策について、重要と思える順に、2行目の右から施策番号（①、②、③、④、⑤、⑥）をご記入下さい。また、3行目で、すぐ左の項目（一つ下位の項目）と比べて、どれくらい重要か、「少し重要」「重要」「かなり重要」のいずれかに○をして下さい。

重要度	6位	5位	4位	3位	2位	最重要施策
施策番号 の記入欄						
該当するものに○を つけて下さい	(6位より) 少し重要 重要 かなり重要	(5位より) 少し重要 重要 かなり重要	(4位より) 少し重要 重要 かなり重要	(3位より) 少し重要 重要 かなり重要	(2位より) 少し重要 重要 かなり重要	

(出所) 筆者作成

実施期間：2021年3月8日（月）から3月15日（月）  
 調査方法：質問紙による郵送法（発送は岡谷市、  
 返送先は筆者が所属する大学の研究室）  
 調査対象：2020年度岡谷市商業活性化会議委員  
 14名中13名（筆者を除く<sup>10</sup>）  
 回答数：13名（回収率100%）

2019年度から2020年度になるにあたり、2019  
 年度の委員のうち4名が入れ替わり、1名が追加  
 となった。よって、評価者13名のうち5名（38.5

%）は、初年度の商業活性化計画の策定に関わ  
 っていない委員である。ただし、2020年10月に  
 2020年度の第1回商業活性化会議を開催し、重点  
 施策に関わる事業の進捗状況などを共有してい  
 る。

以下、重要度調査の詳細について、重点施策  
 数が一番多い基本戦略1の重点施策の重要度によ  
 り説明する。基本戦略1に関する設問は図4の通  
 りである。ここに記載した以外に、自由記述欄  
 を設けている。評価者である委員は、調査紙の

上部に書かれている施策を比較しながら、回答欄を埋めていくことになる。計画書にあるキーワードを併記し、重点施策の内容が分かるように配慮している。

5節で説明したように、まず、重要と思われる順に右欄から重点施策の番号を埋めていく。一対比較しながら並べてもいいし、全体を考えながら分るところから並べることもできる。一番右に記入されるものが、評価者が考える最重要施策となる。ここに記入された最重点施策の回数は、最終的な評価結果が実感と合うか確認するために使われる(表7、9、11)。

重要度に関する降順が決まったところで、次に、6位に記入した重点施策と、5位に記入した重点施策について一対比較を行う。すでに5位に記入された重点施策の方が6位に記入された重点施策よりも重要であることはわかっているのので、「6位より、どれくらい重要ですか」という意味の質問になる。回答は、調査紙にあるように、「少し重要」「重要」「かなり重要」の3つから総合的に考えて最も適しているものを選択し、それに○をつけることになる。事前に会議は行われているので、各重点施策の詳細が知りたければ、第1回商業活性化会議で配布された資料を参照することになる。なお、今回の2回目の重要度調査は、3回目以降と異なり、会議と調査のタイミングが少し離れている。次年度より改善している<sup>11)</sup>。

同様にして「5位より、どれくらい重要ですか」と続き、最後に、「2位より、どれくらい重要ですか」に回答する。これで目的の不完全一対比較表が求められたことになる。

この設問に関する重要度調査の結果をまとめた、実際に使用した表が表4である。ここには、順位と順位が隣り同士の重点施策との一対比較の回答が記入されている。たとえば、委員1の行を横にたどると「3」がある。凡例をみると、サーティ尺度の「かなり重要」を意味することがわかる。また、5位と書かれた列にあるので、6位とした重点施策(今の場合はP5)より「かなり重要」と判断したことがわかる。つまり、P3はP5より、かなり重要であると回答したことがわかる。

そして、表4を相対評価値に数値変換したものが表5(下の2行を除く)である。同じ委員1のところを横に見ていくと、P5の下に1とある。つまりP5の相対評価値は1であることがわかる(P5は6位)。次にP3の下を見ると7とある。すぐ前で説明したように、委員1は「P3はP5より、かなり重要」としていたので、P3の評価値は $1 \times 7$ から求められる数値である。同じ要領で、一対比較値から相対評価値を数値化できる。最後に、13人分の評価値を集約するために幾何平均を計算し、正規化することで重み、すなわち重要度が得られる(表5の下の2行)。

以上の結果をまとめたものが表6である。2行目に重要度が記載されている。これは表5の一番下の行を転記したものである。同様にして得られる、優先度も記入してある。最後に、この重要度調査の結果として、重みが大きい方から2つあるいは3つをAとする。

なお、表7は「該当する重点施策が最重要施策(1位)として選ばれた回数」を示した表である。たとえば、重点施策1(P1)は8人の評価者が最重要施策と判断していることを表す。この数値は、ここまでに得られた重要度(あるいは優先度)である表6の2行目(あるいは3行目)が、評価者達の主観的評価と整合しているか確認するための数値として利用するものである。最重要施策として多数回選択された重点施策は、重要度でも上位となることが期待される<sup>12)</sup>。

ちなみに、表6と表7のそれぞれの2行目を比較することで、重点施策2と6は、最重要施策として選ばれた回数がいずれも2回で一致しているが、重要度には差があること、また、重点施策1は、重要度が1位となった回数が8回と飛びぬけており、これに相応して、相対評価値も飛びぬけて大きくなっていることなどが確認できる。基本戦略2と3は、表8から11に結果だけ示す。

以上より、重要度調査による全ての重点施策についてABCが決定した<sup>13)</sup>。この結果は、2021年3月25日に開催された第2回商業活性化会議で説明し、協議のうえ承認を得ている。

表4 重要施策に関する各委員の重要度評価結果  
(生データ)

委員	重要度の順位						一対比較結果 (左との差)				
	6位	5位	4位	3位	2位	1位	5位	4位	3位	2位	1位
1	P5	P3	P4	P2	P6	P1	3	2	3	3	3
2	P4	P5	P3	P6	P2	P1	2	1	1	3	1
3	P4	P3	P5	P2	P1	P6	2	2	3	3	3
4	P4	P3	P6	P5	P1	P2	1	1	1	2	2
5	P5	P4	P3	P6	P2	P1	1	1	3	1	3
6	P5	P4	P3	P2	P6	P1	1	1	2	2	3
7	P4	P5	P3	P2	P1	P6	2	2	2	2	3
8	P5	P4	P3	P2	P6	P1	1	1	2	3	2
9	P2	P1	P5	P6	P4	P3	2	1	1	1	1
10	P5	P4	P3	P1	P6	P2	1	1	2	1	1
11	P6	P4	P5	P2	P3	P1	1	2	1	1	2
12	P4	P6	P3	P5	P2	P1	1	1	2	1	1
13	P3	P5	P6	P4	P2	P1	2	2	2	2	2

一対比較結果：1=少し重要、2=重要、3=かなり重要

(出所) 筆者作成

表6 基本戦略1の重点施策のABC

相対評価値	P1	P2	P3	P4	P5	P6
重要度	0.712	0.144	0.015	0.005	0.006	0.118
優先度	0.087	0.354	0.040	0.007	0.035	0.478
ABC	A/B	A/A	B/B	B/B	B/B	A/A

(出所) 筆者作成

表8 基本戦略2の重点施策のABC

相対評価値	P1	P6	P7
重要度	0.551	0.341	0.108
優先度	0.423	0.448	0.129
ABC	A/B	B/A	B/B

(出所) 筆者作成

表10 基本戦略3の重点施策のABC

相対評価値	P8	P9
重要度	0.740	0.260
優先度	0.626	0.374
ABC	B/B	C/C

(出所) 筆者作成

(3) 2019年から2023年の調査結果

本論文で提案する手法が継続的に活用でき、これにより実用性を示すために、これまで実施した重要度調査の結果を示す(表12、13)。

ところで、3回目となる2021年度以降は、例年

表5 重点施策の重要度

委員	サーティ尺度による数値化					
	P1	P2	P3	P4	P5	P6
1	12005	245	7	35	1	1715
2	945	315	15	1	5	45
3	1225	175	5	1	25	8575
4	135	675	3	1	27	9
5	1323	189	9	3	1	63
6	1575	45	9	3	1	225
7	625	125	25	1	5	4375
8	1575	45	9	3	1	315
9	5	1	405	135	15	45
10	45	405	9	3	1	135
11	675	45	135	3	15	1
12	405	135	9	1	45	3
13	3125	625	1	125	5	25
幾何平均	574.33	115.76	12.16	4.24	4.86	95.24
重要度	0.712	0.144	0.015	0.005	0.006	0.118

(出所) 筆者作成

表7 基本戦略1で最重要施策/  
最優先施策に選ばれた回数

回数	P1	P2	P3	P4	P5	P6
重要度	8	2	1	0	0	2
優先度	1	3	1	0	2	6

(出所) 筆者作成

表9 基本戦略2で最重要施策/  
最優先施策に選ばれた回数

回数	P1	P6	P7
重要度	9	4	0
優先度	7	6	0

(出所) 筆者作成

表11 基本戦略3で最重要施策/  
最優先施策に選ばれた回数

回数	P8	P9
重要度	11	2
優先度	9	4

(出所) 筆者作成

3月に開催される第2回商業活性化会議の開催前に、まず委員長、市の担当職員、筆者でどの重点施策をAとするかなどの素案を作成したうえで、それを基に当該年度2回目の商業活性化会議(通常3月開催)で協議して決定されることにな

った。このようなプロセスで最終承認されるため、重要度調査から得られるABCと、最終的なABCは必ずしも一致しなくなった<sup>14</sup>。しかしながら、最終的な評価が決定されるまでに、この委員による重要度調査の結果が重要視され、かつ、事業を見直すうえでの重要な役割を果たしていることは確かである。

そこで、ここでは重要度調査から得られた重要度と優先度の数値だけを取り上げる。表12が重要度の推移、表13が優先度の推移である。これをグラフ化したものが図5から10になる。なお、2020年度は2019年度の結果をそのまま利用している。これは商業活性化計画が2019年10月に発行されたもので、まだ十分な事業成果が得られていないため、そのまま継続となったことによる。このため、ここでは、2020年度から2023年度として表やグラフを作成した。重要度調査は事前評価の位置づけであり、該当する重点施策の実施年度は、重要度調査の実施年度の翌年となる。これにより、表には2023年度まで記載がある。AHPには時間軸が入らないが、この事例では、時間的な連続性を取り入れることができている。

## 7. 考察

本研究の目的は、住民の意見を集約し、施策に優先順位をつけるための評価結果の高低に差がつく相対評価手法を提案することであり、AHPの手順2を大幅に簡略化した手法（本論文ではroughly AHPと呼んだ）の提案であった。ここで、継続的に実施した重要度調査の結果の推移を確認することで、提案した手法の実用性や有用性を検討する。

まず、図5から10で注目したい部分は、グラフが交差するところである。特に、1位となる重点施策が動的に推移していることがわかる。たとえば、2020年にはコロナ感染症拡大の影響で、商業会には大きな影響を及ぼしたが、図5の2020年度から2021年度の重要度の推移にそれが表れている。また、グラフの交差はみられないが、図9からは商業関係者間連携の促進（P8）の重要性が高まっていることがわかる（P8の内容は図3を参照のこと）。

1位となる重点施策は確実にA評価になるが、Aは予算配分と関係することから、委員の意見の傾向が見える化できることの意義は大きい。重要度調査による重要度と優先度の結果は第2回商業活性化会議で説明しているが、毎回、委員からの納得を得ている。

表12 重要度の推移（2020年度から2023年度）

年度	基本戦略1						基本戦略2			基本戦略3	
	P1	P2	P3	P4	P5	P6	P1	P6	P7	P8	P9
2020年度	0.296	0.552	0.026	0.005	0.081	0.040	0.643	0.194	0.163	0.521	0.479
2021年度	0.712	0.144	0.015	0.005	0.006	0.118	0.551	0.341	0.108	0.740	0.260
2022年度	0.547	0.289	0.031	0.003	0.014	0.117	0.347	0.426	0.227	0.682	0.318
2023年度	0.443	0.356	0.068	0.011	0.041	0.080	0.252	0.556	0.192	0.742	0.258

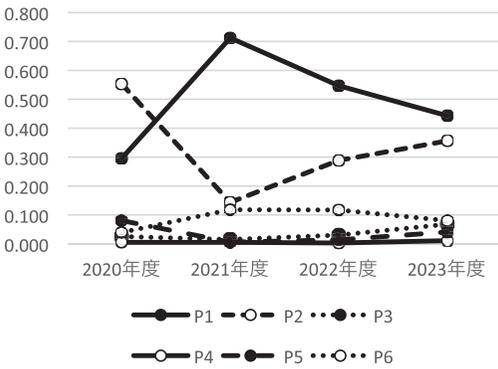
（出所）筆者作成

表13 優先度の推移（2020年度から2023年度）

年度	基本戦略1						基本戦略2			基本戦略3	
	P1	P2	P3	P4	P5	P6	P1	P6	P7	P8	P9
2020年度	0.097	0.784	0.028	0.004	0.051	0.036	0.542	0.284	0.174	0.619	0.381
2021年度	0.087	0.354	0.040	0.007	0.035	0.478	0.423	0.448	0.129	0.626	0.374
2022年度	0.288	0.447	0.021	0.013	0.059	0.173	0.289	0.499	0.212	0.770	0.230
2023年度	0.254	0.326	0.066	0.007	0.055	0.291	0.142	0.673	0.185	0.729	0.271

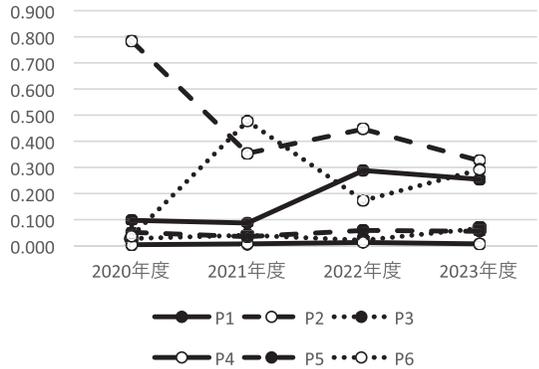
（出所）筆者作成

図5 基本戦略1の重要度の推移



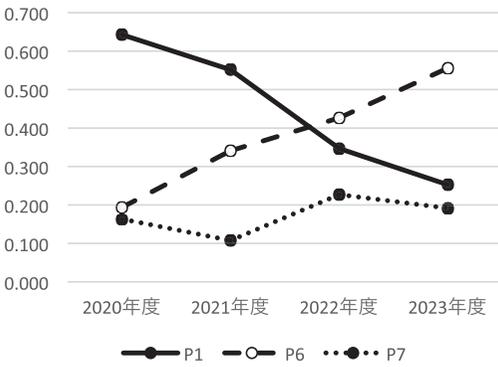
(出所) 筆者作成

図6 基本戦略1の優先度の推移



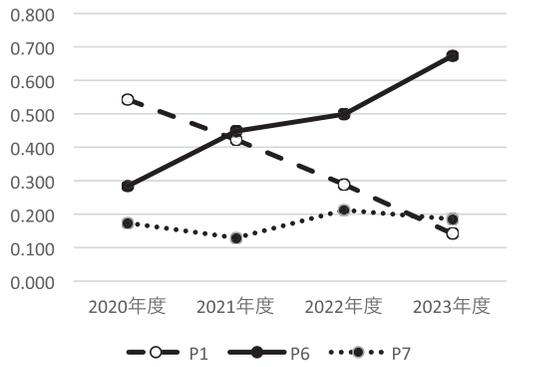
(出所) 筆者作成

図7 基本戦略2の重要度の推移



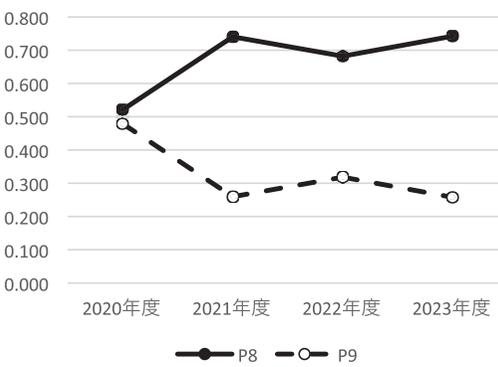
(出所) 筆者作成

図8 基本戦略2の優先度の推移



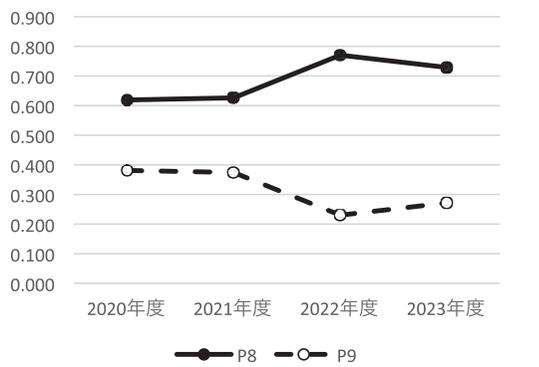
(出所) 筆者作成

図9 基本戦略3の重要度の推移



(出所) 筆者作成

図10 基本戦略3の優先度の推移



(出所) 筆者作成

また、毎年、商業活性化会議の委員の何人かが入れ替わるが、問題なく回答している。これ

には、調査前に施策に紐づく事業内容の確認を行う会議を開催していることが関係していると

考えられ、本論文で提案する手法による評価結果の信頼性を高めるためにも、事前の会議に参加してもらうか、欠席の場合は担当の行政職員より説明を受けてもらう必要がある。この意味では、回答者は限定的になるといえる。

同様に、全ての調査を通して、委員からは一度も回答方法への質問がなく、また自由記述欄への記入も少なくなかった。さらに、継続的に実施することに対する批判もなく、会議では好意的な意見を聞くことができた。これらのことから、大幅に簡略化した設問により、住民が集中力を欠くことなく、かつストレスを感じることもなく回答でき、しかもうまく意見を集約できているといえる。

以上より、施策にABCを決めるための基礎となる評価結果に差をつける相対評価手法として、roughly AHPとそれによる重要度と優先度の重みがうまく機能したと考える。一方、重点施策名の重複による評価の乖離問題や、重点施策の基本戦略への配置問題なども明らかになった。計画を立案する際に、一対比較の特性を考慮した階層図が求められる。

## 8. おわりに

本論文では、住民の意見を集約し、施策に優先順位をつけるための評価結果の高低に差がつく相対評価手法を提案し、事例により実用性と有用性を示した。特に、継続的にこの手法を活用することで、本手法に対する信頼性を高めてきたといえる。

一方で、回答とする3つの選択肢やその形容詞の問題、さらに、それらの回答に対する一対比較値の妥当性については、一つの事例での検証となっていることから、十分に検証できたとは言いがたい。また、他の計画などで同じように成果が得られるか検証する必要もある。これらが今後の課題である。

## 謝辞

本論文の初稿に対して、論旨を明確にするう

えで大変有益なご指摘を多数いただいた匿名の査読者に深く感謝したい。また、アンケート調査にご協力いただいた岡谷市商業活性化会議委員の皆様、評価手法の研究にこの商業活性化計画の場を提供していただくことにご理解とご協力をいただいた関係者の皆様に御礼申し上げる。本研究は、JSPS科研費基盤研究(C)(一般)(課題番号20K01480)の助成を受けたものである。

## 注記

- 1 今村(2018)が提唱する財政健全化手法の一つである。北川正恭元三重県知事の言葉との記載がある(今村2018, p.111)。
- 2 宮川(1994)は、政策選択へのアプローチとしての理想的な意思決定手法は包括合理性モデルであるとし、「まず、意思決定者は明確に定義された目的を持ち、異なる価値観の対立に悩まされることはない。価値の対立があるにしても、それはなんらかの手続きによってある一般の嗜好指標の中に解消できる。第二に、選択の対象となる代替的行動案の集合は、意思決定者に完全に既知のものである。第三に、意思決定者は各代替的行動案に対してあらゆる面にわたって結果を予測し、その結果の望ましさを目的に照らして判断して、代替案を順位づけることができる。そして最も望ましい結果をもたらす代替案(これを「最適な代替案」という)を選択する」(宮川1994, p.263)、「しかし意思決定においてこのようなアプローチをきっちりとしたかたちでとることのできるのは、考慮すべき要因の数が少なく、価値の対立の問題が大きくないような比較的小規模かつ単純な構造の問題解決においてである」(宮川1994, pp.166-167)、「複雑な問題に対して人間がとるアプローチは、それとはかなり異なり、目的や代替案の考慮や分析が逐次的であり、また限定的である」(宮川1994, p.267)と述べている。
- 3 サーティ自身は、The Scale and Its Description (Saaty 1977, p.246) や Fundamental Scale of absolute numbers (Saaty 2010, p.105) と呼んでいる。また、後者では「Equal importance」「Moderate importance」「Strong importance」「Very strong or demonstrated importance」「Extreme importance」としている。これらの表現は、日本語の文献でも相違がみられる。

- 4 木下（1993）は「同じくらい重要」「やや重要」「かなり重要」「非常に重要」「極めて重要」としている（木下 1993、p.11）。また、木下（2000）では「やや重要」の代わりに「少し重要」としている（木下 2000、p. 6）。
- 5 サーティ尺度（表1）から、本質的に完全整合（C.I.=0）となることはほとんどないことがわかる。
- 6 回答の選択肢は、形容詞により評価者による感覚の差が表れないようにすることと、roughly AHPでは中間の表現を使用しないことなどを考慮し、刀根（1986）の日本語表現を参考に、「重要」を中心に「少し重要」と「かなり重要」とした。
- 7 回答の選択肢を大幅に簡略化した理由として、AHPを利用するからには、大まかな評価値が得られれば十分という考えがある。ただ、現実の評価場面において、やり直しができないという事情も影響している。評価者が出席する会議で結果を細かく報告する責任は重い。
- 8 策定段階では「重要度」「緊急度」とし、前者を商業全体への影響の大きさに関する評価、後者を着手時期に関する評価として提案した。しかし、緊急性が高い重点施策とするなら、なぜAとしてすぐに実施しないかという問題が起こることが指摘され「重要度」「優先度」となった。
- 9 初回はまだ計画の策定段階であったこと、また、このため全8回の会議後に実施されるなど特別な状況であったことによる。なお、重点施策の重要度と優先度を定めるための調査であるが、初回に重要度調査と名付け、その後もこの用語を使っているため、本論文でも「重要度調査」と呼ぶことにした。
- 10 筆者は知識経験者として岡谷市商業活性化計画に参加しており、中立な立場を取るために重要度調査の調査対象者から外している。
- 11 2020年度は、第1回商業活性化会議が2020年10月13日に開催され、重要度と優先度のABCを決めるための重要度調査は2021年3月8日から15日の期間で実施している。これは3月25日に開催された第2回商業活性化会議の直前にあたる。重点施策に関わる事業内容などを共有したあと期間を置かず重要度調査を実施した方が良かったため、2021年度からは第1回商業活性化会議のすぐあとに実施時期を変更している。実際、2021年10月20日に第1回商業活性化会議が開催され、11月8日から26日の期間で重要度調査を

施した。

- 12 重みの降順と、最重要施策とされた回数の降順が一致しないことがある。その際は、選択回数に関して2位以下も検討する。実際、2021年度に実施した3回目の重要度調査ではそのようになった。そこでは、重点施策番号が同じものが有るため混乱したと結論付けた。重点施策6は基本戦略1と2にある（図3）。
- 13 AHPあるいはroughly AHPでは、手順3として、「総合評価」に関して「重要度」と「優先度」を一对比較し、総合評価値としての「優先度」を求めることになるが、商業活性化計画では、そこまで算出していない。2つの評価基準が「重要度」と「緊急度」であれば、これらを一对比較することで「総合評価」に関して「優先度」を求めることも考えられた。
- 14 計画の方向性、経年での事業の推進状況、外部環境としての社会状況なども考慮しなければならないため、重要度調査の結果だけで最終的なABCが決まるわけではない。

## 参考文献

- 飯田洋市（2019）「地方自治体におけるメリハリある行政評価を目指した相対評価手法の研究」、日本評価学会第20回全国大会予稿集：249-256
- 飯田洋市（2021）「予算編成に向けた市民参画型事前評価の手法の研究～長野県岡谷市商業活性化計画を事例にして～」、日本評価学会第22回全国大会予稿集：1-8
- 今村寛（2018）『“財政が厳しい”ってどういうこと？、ぎょうせい
- 小川顕正（2016）「階層化分析法（AHP）を用いた自治体の歳出配分評価」、日本財政学会『アベノミクスと税財政改革』、有斐閣、187-204
- 岡谷市・岡谷商工会議所（2019）『岡谷市商業活性化計画』（2023年9月28日時点で岡谷市のホームページで参照可）
- 小野達也・田淵雪子（2001）『行政ハンドブック』、東洋経済新報社
- 木下栄蔵（1993）『問題解決型意思決定法 AHP手法と応用技術』、総合技術センター
- 木下栄蔵（2000）『入門AHP—決断と合意形成のテクニック』、日科技連出版社
- 佐々木秀彦（2002）「博物館をめぐる状況」、村井良子

- 編著『入門ミュージアムの評価と改善～行政評価や来館者調査を戦略に活かす～』、株式会社ミューゼ、7-34
- 佐藤徹 (2009) 『自治体行政と政策の優先順位づけ“あれもこれも”から“あれかこれか”への転換』、大阪大学出版会
- 佐藤徹編著 (2021) 『エビデンスに基づく自治体政策入門～ロジックモデルの作り方・活かし方』、公職研
- 島田晴雄・三菱総合研究所政策研究部 (1999) 『行政評価』、東洋経済新報社
- 高村善晴 (2003) 『地方自治体の公共意思決定—実践的プロセス志向』、日本経済評論社
- 刀根薫 (1986) 『ゲーム感覚意思決定法』、日科技連
- 中井達 (2005) 『政策評価 費用便益分析から包絡分析法まで』、ミネルヴァ書房
- 松田敏幸 (2004) 『政策評価と予算編成—新たな予算配分方法—』、晃洋書房
- 宮川公男 (1994) 『政策科学の基礎』、東洋経済新報社
- Iida, Y. (2020). Analytic Hierarchy Process for evidence-based policy making, *Proceedings of the 16th International Symposium on the Analytic Hierarchy Process 2020*, 1-5. DOI: 10.13033/isahp.y2020.014
- Iida, Y. (2022). Consolidation of citizens' opinions with the AHP for PDCA cycle in local governments, *Conference paper of the 17th International Symposium on the Analytic Hierarchy Process 2022*, 1-4. DOI: 10.13033/isahp.y2022.026
- Kułakowski, K. (2020). *Understanding the analytic hierarchy process*, Boca Raton, FL: CRC Press.
- Saaty, T. L. (1977). A Scaling Method for Priorities in Hierarchical Structures, *Journal of Mathematical Psychology*, 15, 234-281.
- Saaty, T. L. (2010). *Principia Mathematica Decernendi: Mathematical Principles of Decision Making*, Pittsburgh: RWS Publications.

(2023.12.15 受理)

## **A Participatory Evaluation Method using incomplete matrices of the AHP -Toward a Meaningful Relative Evaluation of Projects-**

Yoichi Iida

Suwa University of Science  
youichi.iida@gmail.com

### **Abstract**

While many local governments are under pressure for fiscal austerity, the demands on financial resources due to increasing social security costs and diversifying residents' needs and so on are not decreasing. So, local governments need to make a meaningful relative evaluation to determine which programs or projects to focus on. At this time, from the perspective of local autonomy, it is desirable that the voices of residents are reflected in such evaluations.

The purpose of this paper is to propose an evaluation method that facilitates residents' participation in the evaluation and that can utilize the evaluation results as an expression of residents' will, specifically, a method based on incomplete pairwise comparison matrices of the Analytic Hierarchy Process. We demonstrated that this evaluation method can be a meaningful and relative evaluation by using a case study of a relative evaluation of programs of the commercial revitalization plan of Okaya City, Nagano Prefecture, conducted from 2019 to 2023.

### **Keywords**

Participatory Evaluation, Analytic Hierarchy Process, Relative Evaluation, Preliminary Evaluation,  
Consensus Building

## 【実践報告・調査報告】

少人数学級の政策決定におけるエビデンス  
—エビデンスは決定に影響したのか—

荒木 進太郎

東京都教育庁

shintarou.araki004@gmail.com

## 要 約

本稿は、2011年度及び2021年度の公立小学校における少人数学級（35人学級）導入に関する予算編成過程を対象にし、財務省と文部科学省が使用したエビデンスを確認する。ここでは、エビデンスの多様性と多義性を前提としたうえで、どのようなエビデンスが用いられたのかに加え、それが両省でどのように扱われ、政策決定に影響を与えたのかについて確認する。

この事例においては、政策の採否に争いがあるところでは見解が相違し、エビデンスによる政策決定が困難となっていたこと、アジェンダ設定の際にエビデンスが政治に結びついていたことが確認された。

## キーワード

エビデンス、EBPM、少人数学級、35人学級

## 1. はじめに

## (1) 本稿の目的

評価は政策へ反映すること、すなわち政策決定に寄与することが眼目の1つである。これを政策の側からいけば、評価の結果はエビデンスの1つとしてとらえることができる。エビデンスを政策に結びつける手法として注目されているのがEBPM（Evidence-Based Policy Making）である。

これまで、エビデンスとEBPMの制度設計に関する議論は様々になされており、今後は、それが実装可能かどうかについての検証が必要な段階へと進むだろう。しかしながら、エビデンスが実際の政策決定にどのように影響しているのかという研究までは十分な展開が不足している

というのが筆者の立場である。

そこで、本稿では、エビデンスの多様性と多義性を前提とし、教育政策の重要課題の1つである公立小学校における少人数学級（具体には、後述する学級編制の標準を40人から35人へ引き下げた学級）の導入にかかる予算編成過程において、何がエビデンスとして用いられ、どのように扱われたのか、最終的な結論にエビデンスがどのように影響したのかを確認する。なお、本稿は予算編成過程を対象とするが、これは予算案への計上により政府内における政策決定に至ると考えるためである。

## (2) 本稿の構成

本稿の構成は以下のとおりである。

まず、本稿におけるエビデンスの定義、多様

性及び多義性についての定義を行う。

次に、小学校第1学年への少人数学級の導入にかかる2011年度の予算編成過程において、財務省と文部科学省がどのようなエビデンスを使用していたのかを整理する。そのうえでエビデンスがどのように扱われているのかについて分析を行い、どのような結論に至ったのか、結論に影響を与えた要因は何であったのかを確認する。

同様に、小学校第2学年以降への少人数学級の導入にかかる2021年度の予算編成過程について確認する。

最後に、本稿のまとめとして、エビデンスと政策決定について確認された点の整理と若干の考察を行う。

## 2. 本稿におけるエビデンスの定義

### (1) エビデンスの定義

エビデンスには確立した定義が存在しないが、例えば、小池・落(2020:10)は、行政の実務、政府の見解においてこれを広くとらえることから、エビデンスを「政策プロセス全般に関わるデータ、統計、分析結果など」として論じており、エビデンスを広義にとらえる立場といえる。これへの反対概念として、「エビデンスとはデータとイコールではなく政策効果を因果関係として示したデータ分析結果である」(伊藤 2016:5)というように、エビデンスを政策介入とそれに基づく効果の因果関係を示唆する根拠として狭義にとらえる立場がある(ここにはエビデンスの質についての議論もあるが本稿では扱わないこととする)。

また、エビデンスを求める主体に着目して論ずることもできる。岩崎(2022:8-10)は、「政策立案者の需要に応じて用いられる『政策決定支援型のエビデンス』」と「研究者主導による『知識支援型のエビデンス』」の2つがあるとしている。前者は、「政策立案者からの需要に応じるもので、政策決定時に有用となるデータや情報など、時には量的データのみならず質的データなど」までを含むものとしており、後者は、「政策や実践への活用のため、有効で質の高い知識(デ

ータ)を科学的に整理しようとしたもの」(岩崎 2022:9)であるとしている。

本稿においては、具体的政策決定を確認する趣旨から、これらの定義について、広義にとらえる立場及び政策決定支援型のエビデンスを「広義のエビデンス」、狭義にとらえる立場及び知識支援型のエビデンスを「狭義のエビデンス」と2つに整理し、そのうえでエビデンスについて両者を含むものと定義して考察する。

### (2) エビデンスの多様性と多義性

本稿においては、研究目的をふまえ、エビデンスの多様性、多義性を焦点とした論考は他稿に譲ることとしたい。そのうえで、本稿が前提とするエビデンスの多様性及び多義性の定義を示すとすれば、次のとおりとなる。

まず、エビデンスの多様性については、エビデンスを広義にとらえる立場をとっており、その点から様々なものがエビデンスとして採用されることになる。つまり、本稿における多様性は、政策課題にあわせて様々なエビデンスが用いられることを意味する。

次に、多義性についてである。先に示した岩崎(2022:8-10)は、政策立案者と研究者の間でエビデンスの意味合いが異なることをもって多義性と述べている。本稿においては上述のとおりエビデンスを広義と狭義の両者を含むものと定義していることから、本稿では、岩崎のいう多義性の定義を前提として論ずる。

## 3. 事例研究

### (1) 少人数学級とは

公立小学校については、基礎自治体である市区町村にその設置義務が課せられている(学校教育法第38条)。学校には校長をはじめとした教員や事務職員などを含めた教職員が配置されており、原則として学校の経費は設置する市町村が負担する(学校教育法第5条)。しかし、財政的に不安定な自治体においても教育の機会均等と教育水準の向上を図る目的から、政令指定都市を除く市町村立小・中学校の教職員の給与は、

財政的に安定している国と都道府県が負担するしくみとなっている。

国が教職員の給与を負担する際には「公立義務教育諸学校の学級編制及び教職員定数の標準に関する法律」（以下「標準法」）に基づいて教職員数が決定される。この標準法により、1学級当たりの児童生徒数の上限（以下「学級編制の標準」）が定められており、それに基づいて教職員数が決まる（以下「教職員定数」）。

「少人数学級」<sup>1</sup>とは、政府による公式な定義はないが、学級編制の標準を従前のそれよりも引き下げることにより、規模が縮小された学級を指す。これにより教職員定数、つまり学校に配置される教職員数が増加し、国の財政負担も増加することになる。教育環境の改善や教員の多忙化などが政策課題となり、少人数学級は自治体や学校からの要望も多いものであるが、予算額に与える影響は大きく<sup>2</sup>、1991年に学級編制が40人で行われるようになって<sup>3</sup>以降、学級編制の標準の引き下げは行われていなかった。この学級編制の標準を35人とした条件下で編制された学級を「35人学級」という。

## (2) 2011年度予算：小学校第1学年への少人数学級（35人学級）の導入

### ① 事例の概況

学級編制の標準について転機がおとずれたのは民主党政権下の2011年度予算編成時であった。この際、文部科学省は、少人数学級を押し進めるべく、小学校第1・2学年における学級編制の標準を40人から35人へ引き下げる予算要求を行い<sup>4</sup>、その結果、小学校第1学年について学級編制の標準が35人とされた<sup>5</sup>。

### ② エビデンスの整理

文部科学省は、2011年度予算要求に向けて、大学教授や教育委員会関係者などの有識者ヒアリングを実施した。2010年5月の有識者ヒアリングでは、中央教育審議会委員でもある小川正人から、米国テネシー州のSTAR計画（Student/Teacher Achievement Ratio:1985年）における少人数学級研究についての報告が行われた<sup>6</sup>。

一方、財務省は、「人材の育成・活用に関する研究会」を設立し、財務省としてのエビデンス

の整理を行った。この研究会では、一橋大学経済研究所教授の小塩隆士が、「資源要因のクラス規模については、小規模クラスであれば成績向上に有利であるとの見方もある中、日本ではそうした結果は得られなかった。一方、他国においては寧ろ大規模クラスで成績が良い傾向にある。」と主張し、1学級における児童生徒数（以下「学級規模」）と成績の関連性について疑問を示した<sup>7</sup>。

また、文部科学省を担当する神田真人主計官が、教職員定数と児童生徒数の推移の比較や学級規模の状況などについて財務省としての見解を述べていた<sup>8</sup>が、これは前年に財務省が公表した主張<sup>9</sup>を概ね踏襲するものであった。

文部科学省は、これへの反論<sup>10</sup>もふまえつつ要求<sup>11</sup>を行った。2011年度予算編成に関連する両省の主要な見解（2009年から2010年）を整理すると〈表1〉のようになる。

### ③分析

【1】教職員定数と児童生徒数<sup>12</sup>については、児童生徒数と教職員定数の比較という指標に関して財務省は、児童生徒数の減少程には減少していないことから、実質的に教職員定数は増加していると主張していた。他方で、文部科学省は、学級編制とは別に措置されてきた教職員定数を含めていること、また、通常学級以外の特別支援学級を含めていることを指摘し、その指標を作成する視点を批判していた。

特別支援学級とは、小学校または中学校において、障害による学習上又は生活上の困難を克服するための教育を行う目的で特別に編制された学級のことである。特別支援学級の学級編制の標準は、8人であり通常学級の標準よりも少ない。このため、特別支援学級を含めて算出するか否かにより、児童生徒40人当たりの教職員数は大きく変動する。

次に【2】学級規模の現状について財務省は、1学級当たりの児童生徒数は30人を下回っていると主張した。他方、文部科学省は、先にあげたように学級編制の標準が異なる特別支援学級を指標に算入していることを指摘し、指標に参入する要素が不適であると反論した。

【3】授業時間などの国際比較については次の

とおりとなる。まず、①の教員の授業時間数という指標に関して、財務省は、日本の授業時間数についてアメリカ、ドイツ、フランス及びOECD加盟国平均のそれに比して少ないとの認識を示していた。他方で、文部科学省は、比較対象との状況の違いをふまえていないとしてその指標を単純に比較することが適切ではないと反

論した。②については、小学校教員1人当たりの児童生徒数という指標に関して、財務省は主要先進国（アメリカ、イギリス、ドイツ、フランス）と比較しても遜色ないと主張したが、文部科学省はOECD加盟国平均という別の比較対象をとりあげて、それと比較すると低く、改善すべき格差であると反論した。

表1 2011年度予算編成をめぐる両省の主な見解

論点	財務省	文部科学省																								
【1】 教職員定数と児童生徒数	1989年以降、児童生徒数が3割減となる一方で、教職員数(公立小中学校)は▲8%にとどまっていることから、児童生徒40人当たり教職員数は34%増加している。	児童生徒数の減少割合に比べて教職員数が減少していないのは、1989年以降、40人学級の実施やティームティーチング <sup>13</sup> 、少人数指導 <sup>14</sup> の実施などを行ってきたことによる。 近年は、障害のある児童生徒数の急増により、特別支援学級などの在籍者数が増加しており、全体的に児童生徒数が減少しても、教職員定数の算定基礎となる学級数は微増している。財務省の見解にはこの視点が欠けている。																								
【2】 学級規模の現状	公立小学校においては、1学級当たりの児童生徒数は30人を切っている。 学級数で見ても、全国の公立小学校の約5割は30人学級を達成している。	学級編制の標準は、通常学級で40人、特別支援学級で8人であるが、財務省の示す1学級当たり児童生徒数には特別支援学級も含んでおり、30人を切っているとするのは不適切である。																								
【3】 授業時間などの国際比較	① 日本の教員の授業時間は短い。 小学校教員1人当たりの年間授業時数(2007) <table border="1" style="margin-left: 20px;"> <tr><td>日本</td><td>705</td></tr> <tr><td>アメリカ</td><td>1,080</td></tr> <tr><td>ドイツ</td><td>806</td></tr> <tr><td>フランス</td><td>914</td></tr> <tr><td>OECD加盟国平均</td><td>798</td></tr> </table> <small>※イギリスの「授業時間」のデータは無し。</small> ② 日本の教員数は、主要先進国と比較しても遜色ない水準。 小学校教員1人当たりの児童生徒数(2007) <table border="1" style="margin-left: 20px;"> <tr><td>日本</td><td>19.0</td></tr> <tr><td>アメリカ</td><td>14.6</td></tr> <tr><td>イギリス</td><td>19.4</td></tr> <tr><td>ドイツ</td><td>18.3</td></tr> <tr><td>フランス</td><td>19.7</td></tr> <tr><td>G5平均</td><td>18.2</td></tr> <tr><td>OECD加盟国平均</td><td>16.0</td></tr> </table>	日本	705	アメリカ	1,080	ドイツ	806	フランス	914	OECD加盟国平均	798	日本	19.0	アメリカ	14.6	イギリス	19.4	ドイツ	18.3	フランス	19.7	G5平均	18.2	OECD加盟国平均	16.0	① わが国の教員は授業以外に生徒指導や学校行事などを多く行っており、授業時間以外に子供を指導する時間を多く持っている。 ② 教員1人当たりの児童生徒数は、財務省が示すデータからも明らかなように、OECD加盟国平均と比較すると低い水準(小学校で3.0人の格差)である。
日本	705																									
アメリカ	1,080																									
ドイツ	806																									
フランス	914																									
OECD加盟国平均	798																									
日本	19.0																									
アメリカ	14.6																									
イギリス	19.4																									
ドイツ	18.3																									
フランス	19.7																									
G5平均	18.2																									
OECD加盟国平均	16.0																									
【4】 少人数学級の効果	学級規模については、小規模学級であれば成績向上に有利であるとの見方もある中、日本ではそうした結果は得られなかった。一方、他国においてはむしろ大規模学級で成績が良い傾向にある。	少人数学級・少人数指導の有効性は、STAR計画など様々な調査・研究で指摘されている。全国学力学習状況調査で4年連続上位の秋田県は、早くから県独自に少人数学級を実施してきた。																								

(出所) 人材の育成・活用に関する研究会資料<sup>15</sup>、財務省資料<sup>16</sup>、文部科学省資料<sup>17</sup>、<sup>18</sup>をもとに筆者作成。

【4】少人数学級の効果については、双方が実証研究に基づく見解を示していた。エビデンスの質についての疑義は残されるものの、財務省及び文部科学省の双方ともに学力向上を目標としたうえで少人数学級という政策についての実証研究に基づく見解を示している点は注目に値する。しかしながら、双方とも自省の主張に資する研究を持ち出しているため見解が相違している点にも注目しておきたい。

#### ④政策決定

それでは、当該予算編成の帰結はどのようになったのであろうか。さらにいえば、エビデンスの質の高さにより政策は決したということがいえるのであろうか。この点に言及してみよう。

2011年度予算要求において文部科学省が求めた小学校第1・2学年の学級編制の標準を40人から35人とする要求に対して、最終的には小学校第1学年についてのみ35人とする事となった。しかしながら、この決定は文部科学省の提供したエビデンスによりもたらされたものではなかった。報道<sup>19</sup>によれば、この決定に影響を与えた大きな要因は、民主党が先の参院選挙で少人数学級の推進を公約に掲げていたという点にあった。言い換えれば、予算編成における両省のエビデンスの妥当性よりも政治的な要因が最終的な決定に大きな影響を与えたということである。

### (3) 2021年度予算：小学校第2学年以降への少人数学級（35人学級）の導入

#### ①事例の概況

2011年度予算編成において小学校第1学年の学級編制の標準が35人とされた以降、少人数学級の他学年などへの拡大は行われていなかった。しかし、10年後の2021年度予算編成時において、小学校第2学年<sup>20</sup>以降及び中学校第1学年から第3学年の学級編制の標準の引き下げを財務省に要求する動きが生じた。その結果、標準法が改正され、小学校第2学年から第6学年までの学級編制の標準が35人へ引き下げられることとなった。この学級編制の標準を一律に引き下げるとは、1980年以来、約40年ぶりのものであり、近年の教育政策において特に重要なものの1つとなった。

#### ②エビデンスの整理

2021年度予算要求において文部科学省は、「GIGAスクール構想」<sup>21</sup>のもと、個別最適な学びを実現することができるよう効果的なICTの活用や身体的距離の確保など、新しい時代の学びを支える環境を整備するとして学級編制の標準の引き下げを要求した<sup>22</sup>。

これに対して、財務省は、2020年10月26日の財政制度等審議会財政制度分科会歳出改革部会などにおいて文部科学省の要求に反対した<sup>23, 24</sup>。これに対する文部科学省の反論<sup>25</sup>も含め両省の見解の主なものをまとめると〈表2〉のとおりとなる。

#### ③分析

【1】教職員定数と児童生徒数について財務省は、1989年以降、児童生徒当たりの比較において、実質的に教職員定数は増員となっていると従前と同様に主張した。これに対して文部科学省は、特別支援学級などを含めた指標で比較することを否定した。加えて、2006年度以降の通常学級の児童生徒40人当たりの教職員定数の増加は2%程度であると反論した。文部科学省の反論の意味するところは、40人学級編制が完成した以降から2005年度までの定数改善計画による増員は、少人数指導の実施や通常1名となる教頭・養護教諭を大規模学校などに複数配置するなど学級編制とは異なる目的で措置するものであるから、これを除いて指標を定めるべきと従前と同様の主張をしたのである。

【2】学級規模の現状の①では、35人以下となっている学級の割合について、財務省は、約9割が達成していることからこれ以上の措置が必要な状況ではないとの見解を示していた。これに対して、文部科学省は、都道府県・指定都市別の1学級当たりの児童生徒数についてばらつきが大きい（最大と最小でおよそ10.8人の差）ことを指摘し、全国の一定の教育水準の均衡を図る環境整備が必要との見解を示していた。さらに②では、財務省は、東京都を含む5都府県（東京、埼玉、愛知、神奈川、大阪）において35人以上の学級が多く存在し、これらでは加配<sup>26</sup>の多くを自治体独自の少人数学級ではなく、少人数指導などに活用している一方で、加配定数を利用し

て自治体独自の少人数学級を実施している県も存在する状況をとらえ、学級編制の標準の引き下げは不要であるとの見解を示していた。これに対して文部科学省は、都市部において

も自治体独自の少人数学級の取組は行われており、また、国に対して少人数学級を求める要望もあるとして、同じ状況について異なる見解を示していた。

表2 2021年度予算編成をめぐる両省の主な見解

論点	財務省	文部科学省														
【1】 教職員定数と児童生徒数	1989年以降、児童生徒数の減少ほど教職員定数は減少していない。 したがって、児童生徒数当たりの教職員数を1989年度と同水準とした場合の教職員数(約48万人)と比べれば、実質20万人の増加である。児童生徒数40人当たりの教職員定数は2.0人から2.9人へ増加。	教職員定数が児童生徒数ほど減少していないのは、特別支援学級などに通う児童生徒数の増加による影響が大きい。 定数の改善計画を実施した以降(2006年度以降)の小・中学校の通常学級に通う児童生徒40人当たりの教職員定数の増加は約2%程度である。														
【2】 学級規模の現状	① 小学校の学級の9割が35人以下の学級となっている。 ② 小学校36人以上学級の約6割は5都府県に集中している。これらの都府県では、加配された定数の多くを少人数指導などに活用。一方、既存の加配定数を自治体独自の少人数学級に充てている県も存在する。少人数学級とするか少人数指導とするかは各自治体が地域や学校の実情に応じて判断すべき。	① 35人以下の学級数が多く存在する要因として県の独自負担や国の加配の活用による自治体独自の少人数学級によるものがある。一方、都道府県における学級規模のばらつきがあり、全国の一定の教育水準の均衡をはかる環境整備が必要。 ② 都市部でも自治体独自の少人数学級の取組は行われており、また、国に対し少人数学級を求める要望などもある。														
【3】 授業時間などの国際比較	① 日本の教員の年間授業時間数は、主要先進国の中では低水準。教員と教員以外の人材との役割分担が不可欠。教員以外の人材についても予算人員は増加。  小学校教員1人当たりの年間授業時数(2019) <table border="1" style="margin-left: 40px;"> <tr><td>日本</td><td>747</td></tr> <tr><td>アメリカ</td><td>1004</td></tr> <tr><td>イギリス</td><td>855</td></tr> <tr><td>フランス</td><td>900</td></tr> <tr><td>ドイツ</td><td>698</td></tr> </table> ② 日本は諸外国に比べ学級規模が大きいとの指摘があるが、教員1人当たりの児童生徒数は主要先進国並み(日本は学級担任以外の教員が多い)。  小学校教員1人当たりの児童生徒数(2018) <table border="1" style="margin-left: 40px;"> <tr><td>日本</td><td>16.2</td></tr> <tr><td>G5平均</td><td>17.1</td></tr> </table>	日本	747	アメリカ	1004	イギリス	855	フランス	900	ドイツ	698	日本	16.2	G5平均	17.1	① 年間授業時数は算定方法が各国で異なり、一概に日本が低水準とは言えない。受け持つ授業時間数を一定に保ちつつ、授業準備の時間などを確保することにより、質の高い授業を提供し、世界トップレベルの学力水準を維持していくことが重要である。 ② 学級担任以外の教員数の影響により教員1人当たりの児童生徒数はOECD並み <sup>27</sup> となっているが、学級担任以外の教員は特別な支援が必要な児童生徒の対応などを行っている。
日本	747															
アメリカ	1004															
イギリス	855															
フランス	900															
ドイツ	698															
日本	16.2															
G5平均	17.1															
【4】 少人数学級の効果	① 学級規模が学力に与える影響について、最近の新しいデータを使った研究ほど、学級規模の縮小の効果はないか、あっても小さいことを示している研究が多い。他方、社会経済的背景 <sup>28</sup> が低い学校の子供には有意な学級規模効果が確認されたとする研究結果もある。	① 学級規模が学力に与える影響については、社会的経済的背景が低い子供が多い学校や非認知能力 <sup>29</sup> の観点からは効果があるなど様々な研究結果がある。 ② 個別最適な学びの実現や感染症対策などの観点からも学級編制の標準の引き下げを求める現場の声がある。教育再生実行会議などにおいても効果や必要性が述べられている。														

(出所) 財務省資料<sup>30</sup>、文部科学省資料<sup>31</sup>、OECD資料<sup>32</sup>をもとに筆者作成。

【3】授業時間などの国際比較についても2つの指標が取りあげられていた。まず、①教員の授業時間数という指標が取りあげられていた。これは近年、教員の長時間労働などが社会問題化している状況をふまえてのことである。これについて財務省は、主要先進国（アメリカ、イギリス、フランス、ドイツ）に比べて日本の教員は授業時数が少なく、また、代替可能な業務は教員以外の人材に任せることで多忙化を解消できるとしていた。補習や部活動指導のための人材をこれまで増加させてきたことから教員の増員ではなく、それらの活用をはかるべきと主張したのである。これに対して文部科学省は、年間授業時数の算定方法が各国で異なることを指摘し、また、PISA（Programme for International Student Assessment:OECD生徒の学習到達度調査）では比較対象国よりも高い学力水準を示していることを述べていた。そのうえで日本において教員を中心とした学校運営が行われてきたことを念頭に、長時間労働の解消や授業準備の確保のために教員の増員が必要であるとする見解を示していた。

②については、教員1人当たりの児童生徒数という指標に関して、財務省は主要先進国、文部科学省はOECD加盟国平均と、それぞれが異なる比較対象を採用しつつも両省とも概ね同指標については日本のそれと同程度であるとしていた。しかしながら、学級担任以外の教員に関して財務省はこれを含めて指標を算出するべきとの見解であるのに対して、文部科学省はこれを除いて算出するべきとの見解を示していた。

【4】少人数学級の効果については、①で、双方が社会経済的背景の低い学校の子供に対しては効果を認めていた。しかしながら、その先の政策についての主張は異なっていた。財務省は、地域の実情や効果を見極めた弾力的な教員配置を行っていくべきとして、学級編制の標準を引き下げることに反対していたが、文部科学省は政策効果が認められるのであるから学級編制の標準を引き下げるべきという見解であった。また、学級規模が学力に与える影響について、財務省は前回に引き続き否定的な見解を示していた。さらに、後述するが、②において「個別最

適な学びの実現や感染症対策等」という新たな政策アジェンダを提示している点には留意しておく必要がある。

#### ④政策決定

当初、財務省は、先に示したように文部科学省の要求に反対していた<sup>33</sup>。しかし、この対立状況下において、2020年12月17日に文部科学大臣と財務大臣による大臣折衝<sup>34</sup>が行われ、標準法を改正して小学校について、学級編制の標準を5年かけて35人に引き下げることに両省が合意した。では、この点についてエビデンスはどのように影響したのであろうか。

今般の決定においては、新型コロナウイルス感染症（以降「新型コロナ」）の感染拡大で教室の「密」を避ける要望を無視できなかったとされる<sup>35,36</sup>。2020年2月から、新型コロナの感染拡大防止を企図して休校が全国で相次いだ。その後、感染拡大を防ぎながら学校教育を継続させるために分散登校が行われた。これは、学級を2つに分け、例えば午前、午後に分けて登校させることで教室内の密集を避ける措置である。この分散登校は、実態として少人数学級と同じ状態となっていた。また、1人1人にタブレットなどのICT端末を配布してリモートで学習を行うことが進められ、学校における教育活動の形が大きく変化した。

このような背景から全国知事会、全国市長会、全国町村会の地方3団体が、新しい学びの環境整備を求めて7月3日、7月8日と立て続けに財務省や政府に対して要望を提出した<sup>37,38</sup>ように、少人数学級を含む学びの環境整備を求める機運は高まっていった（小林 2021:21）。こうした声に押されて（前川 2023:41）7月17日に閣議決定された「経済財政運営と改革の基本方針2020」<sup>39</sup>には次のように記載されていた。

「学校の臨時休業等の緊急時においても、安全・安心な教育環境を確保しつつ、全ての子供たちの学びを保障するため、少人数によるきめ細かな指導体制の計画的な整備やICTの活用など、新しい時代の学びの環境の整備について関係者間で丁寧に検討する。」

これ以降も9月8日には、教育再生実行会議<sup>40</sup>初等中等教育ワーキング・グループが少人数学級導入の検討を促す取りまとめを行っていた<sup>41; 42</sup>ほか、9月24日には自民党の教育再生実行本部が30人学級の実現、つまり学級編制の標準を30人とするための法改正を文部科学大臣に申し入れていた<sup>43</sup>。

ここで注目すべきことがある。先の閣議決定に先立ち7月8日に開催された2020年第10回経済財政諮問会議において「経済財政運営と改革の基本方針2020（仮称）原案」<sup>44</sup>が示されていたが、ここでは次のように記載されていた。

「学校の臨時休業等の緊急時においても、安全・安心な教育環境を確保しつつ、全ての子供たちの学びを保障するため、少人数指導（傍点は筆者による。）によるきめ細かな指導体制の計画的な整備やICTの活用など、新しい時代の学びの環境の整備について関係者間で丁寧<sup>45</sup>に検討する。」

ここに登場する「少人数指導によるきめ細かな指導体制」と「少人数によるきめ細かな指導体制」の表記は、一見すると小さな記載の変化のようであるが、教育政策（とりわけ少人数学級）の観点からはその意味するところが大きく異なる。前者は特定教科について児童生徒をいくつかのグループに分けて学習の習熟度にあわせて指導することなどを指す概念であり、学級編制の標準を引き下げることには含まない。つまり、少人数学級という政策は、政府の基本方針として予算編成に大きな影響を与える「経済財政運営と改革の基本方針2020」で否定されることになる。

他方で、後者は少人数指導に加えて学級編制の引き下げ、つまり少人数学級も含む概念となる。この区別は教育政策を論ずるうえでは常識であるので、単なる誤りとは考えられない。7月17日の2020年第11回経済財政諮問会議において原案からの変更点の説明が行われていたが、ここでは「少人数指導」から「少人数」へ変更した旨の説明はなされていなかった。

地方3団体の要望書には「少人数学級編制（傍

点は筆者による。）を可能とする教員の確保」とあった。これを考慮すると、原案に「少人数指導」と記載されていたものが、少人数学級を求める機運を無視できず、「少人数」というより大きな概念に引き戻されたと考えた方が妥当であろう。これらのことから、2021年度予算編成においてもエビデンスは大きな影響を果たさなかったと考えることができるだろう。

## 4. 考察

### (1) 広義のエビデンス

2011年度及び2021年度における【1】【2】【3】は、広義の意味でエビデンスがとらえられている。また、2011年から10年が経過した際にも、一部の指標が変更されたものの、基本的な構造は変わらなかったといえることができる。広義のエビデンスに関しては、財務省が提示した指標（児童生徒数の減少と教職員定数の比較など）に対し、文部科学省はその指標の前提となる視点を批判し、双方の立場は異なっていた。以上の事例から確認されたのは、どのようなエビデンスを採用するのか、また、エビデンスをどのように解釈するのかについて見解が相違していたことである。なお、これらの見解は各組織のもつ価値観に裏付けられている。このことをふまえていえば、広義のエビデンスを用いて何らかの政策的帰結を導くことが困難となる可能性が指摘できる。

### (2) 狭義のエビデンス

2011年度と2021年度において、少人数学級の効果に関する研究成果が提示されていた。2011年度の【4】及び2021年度【4】①については、両年度とも狭義のエビデンスによる議論であったといえる。しかし、ここでは財務省及び文部科学省の双方が異なる研究成果を示しており、政策効果の有無について見解が異なっていた。この構造は、両年度にわたり変わらず存在していた。また、2021年度【4】①において、政策効果を両省が認めていてもその先の政策の採否については方向性が異なっていた。これらのことから、1

つの政策について様々なエビデンスのなかで何を採用するのか、また、そのエビデンスをどう評価し政策に結びつけるのかなどについて見解の相違が生じ、ここでもエビデンスにより政策的帰結に至ることが困難となる事態が確認された。

### (3) エビデンスとアジェンダ設定

2011年度の政策決定は、マニフェストに示されていたように民主党政権の政治的スタンスによりもたらされた。しかし、2021年度【4】②では異なるアジェンダが設定され、文部科学省が少人数学級を「感染症対策」にも結びつけて提案し、これが国民の支持を受けて決定に至った。ここでは、エビデンスの質よりも政治的な支持が政策決定に影響を与えた可能性を指摘できる。

このアジェンダ設定には、地方3団体が提言した「新しい時代の学びの環境整備に向けた緊急提言」<sup>45</sup>が影響していた。同提言では、公立小・中学校の普通教室の平均面積は64㎡であって現在の40人学級では、感染症予防のために児童生徒間の十分な距離を確保することが困難であるとして、感染症の再拡大時にも子供たちの学びを保障するために少人数学級が必要であるという主張を展開していた。これに対して、財務省はワクチンの開発状況などを考慮し、感染症対策の必要性を疑問視した<sup>46</sup>。しかし、最終的には感染症対策というアジェンダが採用され、少人数学級（35人学級）が導入される結果となった。

キングダム（Kingdon 2011=2017）の政策の窓理論によれば、政策決定は「政策」「問題」「政治」の3つの流れが合流したときに実現する可能性が高まるという。従来、少人数学級の導入という「政策」は学力という「問題」に結びつけられていたが、これは国民の支持が得られず「政治」が合流しなかったために決定に至らなかった。しかしながら、感染症対策という新たな「問題」が少人数学級という「政策」に結びつき、国民の支持を受けて「政治」の流れが合流し、決定がもたらされたと考えられることができる。

他方で、2021年度予算案決定後の2021年3月18日に行われた財政制度等審議会財政制度分科会において、ある委員は「来年度の政策に小学校

の35人学級とありますが、これも少しコロナの風に乗って、エビデンスがないまま導入されたようなイメージもございます（後略）」と述べていた<sup>47</sup>。この発言からは、少人数学級の導入という政策に対するアジェンダが従来の学力から感染症対策へと変化した一方で、財務省がこれと異なる認識を持っていることが伺える。

ところで、この少人数学級に関する2021年3月30日の国会質疑において、萩生田光一文部科学大臣（当時）は、次のように述べていた<sup>48</sup>。

「私、気を付けなきゃいけないのは、今回少人数やるときに、エビデンスがないということ物を物すごく当局から言われました。じゃ、その小学校で少人数化することによるエビデンスって何を指標にするのかというのは、これはまさしく立法府の先生方がいろんな角度から提案していただくことがすごく必要で、不登校が一人でも減ったら私は大きな成果だと思し、（中略）一点刻みの点数で、全校一斉学力テストが平均点が（原文ママ）0.幾つ伸びたとか、そういうことだけに着目をしないで是非支えていただきたい」

萩生田の発言には、「何を指標にするのか」という点が触れられているが、これは結局のところ少人数学級という「政策」と国民が求める「問題」とを結びつける「政治」の役割を述べているといつてよい。

### (4) 考察のまとめ

本節では、次の2つの点が確認された。

第1に、政策の採否に争いがあるところでエビデンスによって政策決定を行うことが困難となっていた点である。広義のエビデンスにせよ狭義のエビデンスにせよ、見解が相違し、エビデンスによって決定することが困難となっていた。

第2に、アジェンダ設定の際に、国民の支持を与えるためにエビデンスが政治に結びついていたことである。EBPMについては、政治的なイデオロギーを排除することで良質な政策立案を志向する（杉谷 2022:196）という説明も見られるが、むしろアジェンダ設定において、エビデンスが

政治に結びついていた状況が確認された。

## 5. おわりに

本事例では、見解が相違しエビデンスによって決定を行うことが困難となっている事態が確認された。これについて、悪意をもって都合よくエビデンスを選択し解釈していると解することもできるかもしれない（いわゆる「チェリーピッキング」）。

他方で、都合のよいエビデンスを選択しているというよりも、多様な政策目的にあわせてエビデンスが用いられていると考えることもできる。また、エビデンスが政治に結びつき、様々な主張がなされることが民主的な政策決定に資するものととらえることもできるかもしれない。

前者の立場に立てば、多様性及び多義性を前提としたエビデンスは政策決定にとって警戒すべきものであり、後者の立場に立てば民主的な決定にあたって望ましい状況ととらえることもできるだろう。

しかしながら、本稿はエビデンスの作用を包括的、構造的に示すものではなく、1つの事例に過ぎない。また、この判断は、読者のよって立つスタンスにより異なることとなろう。よって、本稿はこれについてなんら方向性を示すことはできないが、今後、より多くの研究がなされるための一助となれば幸いである。

## 補遺

筆者は地方教育行政に関わる者であるが、本稿の全ては、公開されている資料にもとづき分析を行っていること、また、筆者個人の見解であり、所属する組織の見解とは一切関係が無いことを付言しておく。

## 謝辞

本稿執筆にあたり査読者及び編集委員、また、武藤博己法政大学名誉教授の主催する公共・行政研究会の方々から丁寧なご指摘を頂いた。こ

の場をおかりし感謝申し上げます。

## 注記

- この他、自治体が独自に、国の定める学級編制の標準を下まわって編制した学級を指す場合もある。この場合、本稿では、自治体独自の取組であることを表記する。特に断りのない限りは、本文記載のとおり学級編制の標準を従前のそれよりも引き下げることにより、規模が縮小された学級を指すものとして論ずる。
- 2010年度で約1兆5千9百億円であり、文部科学省の一般会計予算5兆5千9百億円のうち28.5%を占める最大の予算となっていた。その状況は2023年度においてもほぼ変わらない（2023年度予算において28.7%を占める）。
- 標準法が制定された1958年当時の各県の1学級当たりの児童生徒数は60人程度であったとされる。その後、学級編制の標準は、第1次定数改善計画（1959～1963年）による教職員定数の改善により50人、第2次定数改善計画（1964～1968年）により45人、第5次定数改善計画（1980～1991年）により40人となった。
- 文部科学省（2010a）「平成23年度 概算要求主要事項」、2、[https://warp.ndl.go.jp/info:ndljp/pid/11293659/www.mext.go.jp/component/b\\_menu/other/\\_icsFiles/afieldfile/2011/07/27/1297091\\_01\\_1.pdf](https://warp.ndl.go.jp/info:ndljp/pid/11293659/www.mext.go.jp/component/b_menu/other/_icsFiles/afieldfile/2011/07/27/1297091_01_1.pdf)（2023年12月13日閲覧）
- 文部科学省（2010b）「平成23年度文部科学省予算（案）のポイント」、1-2、[https://warp.da.ndl.go.jp/info:ndljp/pid/1621348/www.mext.go.jp/component/b\\_menu/other/\\_icsFiles/afieldfile/2011/01/31/1300823\\_12.pdf](https://warp.da.ndl.go.jp/info:ndljp/pid/1621348/www.mext.go.jp/component/b_menu/other/_icsFiles/afieldfile/2011/01/31/1300823_12.pdf)（2023年12月13日閲覧）
- 小川正人（2010）「今後の学級編制及び教職員定数の改善に関する有識者ヒアリング（小川正人氏提出資料）」、「1. 学級規模をめぐる論議と課題の整理—米国の少人数学級研究からの示唆」の項、[http://www.mext.go.jp/b\\_menu/shingi/chousa/shotou/072/shiryo/attach/1293877.htm](http://www.mext.go.jp/b_menu/shingi/chousa/shotou/072/shiryo/attach/1293877.htm)（2023年12月13日閲覧）
- 人材の育成・活用に関する研究会（2010a）「学力を決めるのは学校か家庭か—アジア主要国の比較分析—」、『「人材の育成・活用に関する研究会」【セッション2】理論的検討 第1回会合 2010年11月19日』、後

- 半部分の(2)発表、[http://warp.da.ndl.go.jp/info:ndljp/pid/10342933/www.mof.go.jp/pri/research/conference/zk090/zk090\\_03.htm](http://warp.da.ndl.go.jp/info:ndljp/pid/10342933/www.mof.go.jp/pri/research/conference/zk090/zk090_03.htm) (2023年12月13日閲覧)
- 8 人材の育成・活用に関する研究会 (2010b)「教育・科学技術施策について(データの視座からの一考察)」、「『人材の育成・活用に関する研究会』【セッション1】現状把握 第2回会合 2010年10月13日」、発表資料及び議事要旨(5)、[http://warp.da.ndl.go.jp/info:ndljp/pid/10342933/www.mof.go.jp/pri/research/conference/zk090/zk090\\_02.htm](http://warp.da.ndl.go.jp/info:ndljp/pid/10342933/www.mof.go.jp/pri/research/conference/zk090/zk090_02.htm) (2023年12月13日閲覧)
- 9 財務省 (2009)「文部科学省予算について」『平成22年度予算編成上の主な個別論点「文教関係」』、4-6、[https://warp.ndl.go.jp/info:ndljp/pid/11100613/www.mof.go.jp/budget/budger\\_workflow/budget/fy2010/bunkyo.pdf](https://warp.ndl.go.jp/info:ndljp/pid/11100613/www.mof.go.jp/budget/budger_workflow/budget/fy2010/bunkyo.pdf) (2023年12月13日閲覧)
- 10 文部科学省 (2009)「財務省公表資料『平成22年度予算編成上の主な個別論点(文部科学省予算について)』(平成21年12月3日)に対する文部科学省の見解」、7-11、[https://warp.da.ndl.go.jp/info:ndljp/pid/10201762/www.mext.go.jp/component/b\\_menu/other/\\_icsFiles/afieldfile/2009/12/15/1287932\\_2\\_1.pdf](https://warp.da.ndl.go.jp/info:ndljp/pid/10201762/www.mext.go.jp/component/b_menu/other/_icsFiles/afieldfile/2009/12/15/1287932_2_1.pdf) (2023年12月13日閲覧)
- 11 文部科学省 (2010c)「小学校1・2年生における35人学級の実現」、[https://warp.da.ndl.go.jp/info:ndljp/pid/10201762/www.mext.go.jp/component/b\\_menu/other/\\_icsFiles/afieldfile/2010/09/21/1297903\\_5.pdf](https://warp.da.ndl.go.jp/info:ndljp/pid/10201762/www.mext.go.jp/component/b_menu/other/_icsFiles/afieldfile/2010/09/21/1297903_5.pdf) (2023年12月13日閲覧)
- 12「児童」とは、概ね6歳から12歳までの小学校などに在籍する子供を指し、「生徒」とは13歳から18歳までの概ね中学校や高等学校などに在籍する子供を指す。本稿においては、小学校教員について論じているが、本事例は中学校も含めて議論されていた部分もあるため「児童生徒」に統一して記述している。
- 13 加配された教員を活用し、1学級を複数の教員が役割を分担して指導する形態のこと。
- 14 児童生徒を学習の習熟度や興味関心などに応じて少人数の学習集団を編制し、異なる教員が指導すること。学級編制の標準を引き下げることには含まない。
- 15 前掲注7及び注8
- 16 前掲注9
- 17 前掲注10
- 18 前掲注11
- 19 日本経済新聞 (2010)「来年度予算、関係折衝で合意、農家戸別補償8,000億円、求職者支援は来年10月導入 (2010年12月18日)」
- 20 2012年予算において、小学校第2学年を35人編制とするのに必要な教員定数が措置され、実態上としては小学校第2学年において35人編制が可能となっている。しかし、これは学級編制の標準を35人としたものではなく、相当する教員数を別途措置(「加配」という。注26を参照されたい)したものであり、標準法の改正も行われていない。
- 21 文部科学省が打ち出した政策構想。タブレットなどのICT端末を1人につき1台配備することと高速大容量の通信ネットワークを一体的に整備することで、多様な子供たちを誰一人取り残すことなく、公正に個別最適化され、資質・能力が一層確実に育成できる教育環境を実現するとしている。また、これまでの我が国の教育実践と最先端のベストミックスを図ることにより、教師児童生徒の力を最大限に引き出すとしている。
- 22 文部科学省 (2020a)「令和3年度概算要求のポイント」、5、[https://www.mext.go.jp/content/20200929-mxt\\_kouhou01-000010168\\_1.pdf](https://www.mext.go.jp/content/20200929-mxt_kouhou01-000010168_1.pdf) (2023年12月13日閲覧)
- なお、この際は、事項要求(項目だけを記載し、政策の予算要求額や規模を明示しない予算要求)となっており、学級編制の標準についても30人を目指すべきとの自民党からの要望もあった(注42を参照されたい)。
- 23 財務省 (2020a)「財政制度等審議会財政制度分科会歳出改革部会(令和2年10月26日開催)資料文教・科学技術」、4-8、[https://warp.da.ndl.go.jp/info:ndljp/pid/12901305/www.mof.go.jp/about\\_mof/councils/fiscal\\_system\\_council/sub-of\\_fiscal\\_system/proceedings\\_sk/material/zaiseier20201026/01.pdf](https://warp.da.ndl.go.jp/info:ndljp/pid/12901305/www.mof.go.jp/about_mof/councils/fiscal_system_council/sub-of_fiscal_system/proceedings_sk/material/zaiseier20201026/01.pdf) (2023年12月13日閲覧)
- 24 財務省 (2020b)「令和3年度予算の編成等に関する建議」、60-64、[https://www.mof.go.jp/about\\_mof/councils/fiscal\\_system\\_council/sub-of\\_fiscal\\_system/report/zaiseia20201125/01.pdf](https://www.mof.go.jp/about_mof/councils/fiscal_system_council/sub-of_fiscal_system/report/zaiseia20201125/01.pdf) (2023年12月13日閲覧)
- 25 文部科学省 (2020b)「財政制度等審議会財政制度分科会歳出改革部会(令和2年10月26日)資料についての文部科学省の見解(教職員定数関係)」、1-7、[https://www.mext.go.jp/content/20201027-mxt\\_kouhou02-000010677\\_1.pdf](https://www.mext.go.jp/content/20201027-mxt_kouhou02-000010677_1.pdf) (2023年12月13日閲覧)

- 26 標準法で学級数などに応じて算出されるものとは別に、少人数指導（注14を参照されたい）やいじめ、不登校対応などの特定の政策目的のため毎年の予算措置により配分される教職員のこと。
- 27 文部科学省は「OECD並み」と表現するのみであるが、OECD加盟国平均を指していると考えられる。その場合のOECD加盟国平均の教員1人当たりの児童生徒数は14.6人である。出典は注32を参照されたい。
- 28 社会経済的背景（SES:socio-economic status）とは、子供たちの間にある経済的格差や不平等のこと。金銭的・経済的な資源の他、保護者の教育への意識など社会的・文化的な資源も含む。
- 29 非認知能力は、意欲・意志、人と協力する力などを含み、数値などで測定が困難なものとされる。
- 30 前掲注23及び注24
- 31 前掲注25
- 32 OECD(2020). *Education at a Glance 2020. Ratio of students to teaching staff in educational institutions, by level of education (2018)*. Retrieved January 12, 2024, from [https://www.oecd-ilibrary.org/education/ratio-of-students-to-teaching-staff-in-educational-institutions-by-level-of-education-2018\\_dda3e8a8-en](https://www.oecd-ilibrary.org/education/ratio-of-students-to-teaching-staff-in-educational-institutions-by-level-of-education-2018_dda3e8a8-en)
- 33 前掲注23及び注24
- 34 文部科学省（2020c）「萩生田光一文部科学大臣記者会見録－予算－（令和2年12月17日）」、冒頭発言付近、[https://www.mext.go.jp/b\\_menu/daijin/detail/mext\\_00121.html](https://www.mext.go.jp/b_menu/daijin/detail/mext_00121.html)（2023年12月13日閲覧）
- 35 同様の指摘は、小林（2021:21）、前川（2023:40-43）も行っている。
- 36 日本経済新聞（2020）「『社説』少人数学級に自治体の裁量を（2020年12月18日）」
- 37 全国知事会（2020a）「令和2年7月3日『新しい時代の学びの環境整備に向けた緊急提言』の要請活動について」、資料、[https://www.nga.gr.jp/committee\\_pt/committee/shuryo/bunkyo/r02/1593757127030.html](https://www.nga.gr.jp/committee_pt/committee/shuryo/bunkyo/r02/1593757127030.html)（2023年12月13日閲覧）
- 38 全国知事会（2020b）「令和2年7月8日『新しい時代の学びの環境整備に向けた緊急提言』に係る要請活動について」、資料、[https://www.nga.gr.jp/committee\\_pt/committee/shuryo/bunkyo/r02/0708.html](https://www.nga.gr.jp/committee_pt/committee/shuryo/bunkyo/r02/0708.html)（2023年12月13日閲覧）
- 39 経済財政諮問会議（2020a）「経済財政運営と改革の基本方針 2020～危機の克服、そして新しい未来へ～」、27、[https://www5.cao.go.jp/keizai-shimon/kaigi/minutes/2020/0717/shiryu\\_02-1.pdf](https://www5.cao.go.jp/keizai-shimon/kaigi/minutes/2020/0717/shiryu_02-1.pdf)（2023年12月13日閲覧）
- 40 教育改革を推進するため官邸に設置された会議。内閣総理大臣、内閣官房長官、文部科学大臣（兼）教育再生担当大臣、有識者などにより構成された。
- 41 文部科学省（2020d）「教育再生実行会議 初等中等教育ワーキング・グループ第1回会議を開催」、[https://www.mext.go.jp/b\\_menu/activity/detail/2020/20200908.html](https://www.mext.go.jp/b_menu/activity/detail/2020/20200908.html)（2023年12月13日閲覧）
- 42 教育再生実行会議（2020）「少人数によるきめ細かな指導体制・環境整備について」、[https://warp.ndl.go.jp/info:ndljp/pid/12251721/www.kantei.go.jp/jp/singi/kyouikuseisei/jikkoukaigi\\_wg/syotyutou\\_wg/dai1/sidou\\_kankyouseibi.pdf](https://warp.ndl.go.jp/info:ndljp/pid/12251721/www.kantei.go.jp/jp/singi/kyouikuseisei/jikkoukaigi_wg/syotyutou_wg/dai1/sidou_kankyouseibi.pdf)（2023年12月13日閲覧）
- 43 文部科学省(2020e)「萩生田光一文部科学大臣記者会見録（令和2年9月25日）」、1人目の記者質問への大臣発言部分、[https://www.mext.go.jp/b\\_menu/daijin/detail/mext\\_00095.html](https://www.mext.go.jp/b_menu/daijin/detail/mext_00095.html)（2023年12月13日閲覧）
- 44 経済財政諮問会議（2020b）「経済財政運営と改革の基本方針 2020（仮称）（原案）」、26、[https://www5.cao.go.jp/keizai-shimon/kaigi/minutes/2020/0708/shiryu\\_02.pdf](https://www5.cao.go.jp/keizai-shimon/kaigi/minutes/2020/0708/shiryu_02.pdf)（2023年12月13日閲覧）
- 45 前掲注37及び注38
- 46 前掲注24
- 47 財務省（2021）「財政制度分科会（令和3年3月18日開催）議事録」、後半部分の赤井委員発言、[https://www.mof.go.jp/about\\_mof/councils/fiscal\\_system\\_council/sub-of\\_fiscal\\_system/proceedings/proceedings/20210318zaiseia.html](https://www.mof.go.jp/about_mof/councils/fiscal_system_council/sub-of_fiscal_system/proceedings/proceedings/20210318zaiseia.html)（2023年12月13日閲覧）
- 48 参議院（2021）「文教科学委員会 第7号 令和3年3月30日」、141萩生田大臣発言、<https://kokkai.ndl.go.jp/simple/txt/120415104X00720210330/141>（2023年12月13日閲覧）

## 参考文献

- 伊藤公一朗（2016）「政策の効果をどう測定するか？海外における『エビデンスに基づく政策』の最新向（筆者注：原文ママ）」、5、<https://www.rieti.go.jp/jp/>

- events/bbl/16102501\_ito.pdf (2023年12月13日閲覧)
- 岩崎久美子 (2022) 「教育における『エビデンス』の多義性」、『指導と評価』、68(2):8-10
- 小池拓自・落美都里 (2020) 「第1章 我が国におけるEBPMの取組」、『EBPM (証拠に基づく政策形成)の取組と課題 総合調査報告書』、国立国会図書館調査及び立法考査局、10
- 小林美津江 (2021) 「義務標準法改正案をめぐる国会論議－小学校35人学級の実現－」、『立法と調査』、435、21
- 杉谷和哉 (2022) 『政策にエビデンスは必要なのか－EBPMと政治のあいだ－』、ミネルヴァ書房、196
- 前川喜平 (2023) 『コロナ期の学校と教育政策 (電子書籍版)』、論創社、40-43
- John W.Kingdon(2011). *Agendas, Alternatives, and Public Policies, updated 2nd ed.* (=笠京子訳 (2017) 『アジェンダ・選択肢・公共政策 政策はどのように決まるのか』、勁草書房)

(2024.1.18 受理)

## **Evidence in policy making on small class sizes -Did the evidence influence the decision?**

Shintaro Araki

Tokyo Metropolitan Government Office of Education  
shintarou.araki004@gmail.com

### **Abstract**

This paper reviews the evidence used by the Ministry of Finance and the Ministry of Education, Culture, Sports, Science and Technology to study the policy decision of small class sizes in public primary schools in Japan in the 2011 and 2021 academic years. Given the diversity and polysemy of the evidence, this paper confirms what was used as evidence, as well as how it was perceived by the two ministries and how it influenced the policy decision.

In this case study, it was confirmed that where there were disputes over the adoption or rejection of policies, there were differing views on evidence, making it difficult to make policy decisions based on evidence, and that evidence was linked to politics when setting agendas.

### **Keywords**

Evidence, EBPM(Evidence-Based Policy Making),  
small-sized classes in public elementary schools, classes of 35 students or less

**日本評価学会第24回全国大会**  
**「評価は社会に何をもたらすのか～ What does the evaluation bring to society?～」**  
**開催の報告とお礼**

2023年12月16日（土）、17日（日）開催（於：大阪大学）の日本評価学会第24回全国大会には、120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もとに、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。

企画委員長 南島和久（龍谷大学）  
 第24回全国大会実行委員長 齊藤貴浩（大阪大学）

**第1日目 2023年12月16日（土）**

※カッコ内は座長名

	会場 A	会場 B	会場 C
	共通1	共通2	自由1
9:30～11:30	「政策学と評価学」  (湯浅孝康)	「RTD：実務者と研究者の交流に基づく学術論文執筆ワークショップ：『日本評価研究』誌の発展にむけて」  (米原あき)	「開発協力」  (石田洋子)
11:40～	理事会		
	共通3	共通4	自由2
13:00～15:00	「政策評価と研究評価の交差点：責任ある研究イノベーション（RRI）と根拠に基づく政策（EBPM）」  (白川展之)	「『学校評価マップ』の開発：学校評価学の構築のために」  (小澤伊久美)	「コロナとICT」  (佐藤由利子)
15:10～16:50	シンポジウム「大学の社会的インパクト評価を考える」(齊藤貴浩)		
17:00～17:40	総会		

## 第2日目 2023年12月17日(日)

※カッコ内は座長名

	会場 A	会場 B	会場 C
	共通5	共通6	自由3
9:30 ~ 11:30	「政策形成・評価に関する改革の 動向：総務省行政評価局におけ る最近の取組」  (渡邊洋平)	「What DE in Action Looks Like： 発展的評価の実践における鍵概 念と事例研究」  (今田克司)	「データとエビデンス」  (田辺智子)
11:40 ~	各委員会		
	共通7	共通8	自由4
13:00 ~ 15:00	「日本における業績測定の開拓と その意義の再評価：ハリー・ハ トリーの業績と関連づけて」  (田中啓)	「Overview of Evaluation Education in the Asia Pacific Region and Cases of Graduate Programs in Australia and Japanese Universities (アジア 太平洋地域における評価教育の 概要と豪州及び日本の大学院教 育の事例)」  (佐藤由利子)	「CBAと社会的インパクト」  (今田克司)
	共通9	共通10	自由5
15:10 ~ 17:10	「RTD：日本評価学会「評価倫理 ガイドライン」の改定案に関す る会員意見交換会」  (小林信行)	「国際協力機構（JICA）の事業評 価における様々な評価手法の活 用について」  (佐藤真司)	「国立大学と文化芸術」  (源由理子)

「評価は社会に何をもたらすのか～What does the evaluation bring to society?～」  
プログラム詳細

○印は発表者または主発表者

1日目 (2023年12月16日 (土))

09:30 ~ 11:30

会場A

共通論題1：政策学と評価学

座長・司会・討論者：湯浅孝康 (大阪国際大学)

- ・企画趣旨の紹介：湯浅孝康 (大阪国際大学)
- ・研究発表1：山谷清秀 (青森中央学院大学)「住民参加と評価」
- ・研究発表2：鏡圭佑 (朝日大学)「アジャイル型政策評価における政策学と評価学の定位」
- ・研究発表3：三上真嗣 (同志社大学)「グローバルな政策の架橋－評価と計画－」
- ・指定討論
- ・質疑応答・フロアとの討論

会場B

共通論題2：実務者と研究者の交流に基づく学術論文執筆ワークショップ：『日本評価研究』誌の発展にむけて (RTD)

座長・司会：米原あき (東洋大学)

- ・趣旨説明：米原あき (東洋大学)「実践と研究を架橋する「日本評価研究」の在り方とは？」
- ・話題提供1：村上裕一 (北海道大学)「『日本評価研究』のカイゼン：特に種別や査読などについて」
- ・話題提供2：橋本昭彦 (日本女子大学)「研究論文の社会的役割とその構成を考える」
- ・話題提供3：窪田好男 (京都府立大学)「実務経験を『日本評価研究』に寄稿する：研究論文になる場合、実践報告になる場合、そしてそれぞれの書き方」
- ・話題提供4：新藤健太 (日本社会事業大学)、大澤望 (インパクト・マネジメント・ラボ)、和田泰一 (日本民間公益活動連携機構)「実務家が研究することの意義とその際に配慮すべき倫理的事項」
- ・フロア討論

会場C

自由論題1：開発協力

座長・司会：石田洋子 (広島大学)

- ・○佐々木亮 (国際開発センター)、藪田みちる (国際開発センター)、Maree Soos (Ministry of Education, Palestine Authority)「Impact Evaluation-Palestine Education Program (PIQMAS)」
- ・○宗像朗 (JICA)、杉山卓 (コーエイリサーチ&コンサルティング)、志賀圭 (コーエイリサーチ&コンサルティング)「バングラデシュの郡自治体ガバナンスのパフォーマンス評価結果について」
- ・丸山隆央 (JICA)「Using data to improve and expand operations for educational development: An example of Indian NGO Pratham」
- ・牟田博光 (国際開発センター/大妻女子大学)「ミャンマー連邦共和国における大学合格基準変更が合格者の男女比に及ぼす影響に関する研究」

13:00 ~ 15:00

会場A

共通論題3：政策評価と研究評価の交差点：責任ある研究イノベーション (RRI) と根拠に基づく政策 (EBPM)

座長・司会：白川展之 (新潟大学)

討論者：南島和久 (龍谷大学)

- ・趣旨説明：白川展之 (新潟大学)
- ・発表1：標葉隆馬 (大阪大学)「責任ある研究イノベーション (RRI) と研究評価」
- ・発表2：高谷徹 (三菱総合研究所)「科学技術イノベーション政策における研究評価と政策評価：国プロジェクトと政策過程における知識生産の実務的視点から」
- ・発表3：野呂高樹 (政策研究大学院大学)「欧州連合 (EU) における Smart Specialisation の取組に関する一考察－地域政策と科学技術イノベーション政策の包摂－」
- ・発表4：白川展之 (新潟大学)「政策評価と研究評価の交錯：責任ある研究イノベーション (RRI) と根拠に基づく政策 (EBPM) をどう読み解くか？」
- ・討論

<p><b>会場B</b></p> <p>共通論題4：「学校評価マップ」の開発：学校評価学の構築のために          座長・司会：小澤伊久美（国際基督教大学）          討論者：中谷美南子（評価コンサルタント）</p> <ul style="list-style-type: none"> <li>企画趣旨・学校評価マップの概要紹介：小澤伊久美（国際基督教大学）</li> <li>研究発表1：橋本昭彦（日本女子大学）「学校評価制度の諸課題と向き合うための研究マッピング」</li> <li>研究発表2：石田健一（計画と評価専門家）「学校評価の変貌への希望-体験学習の評価実践を再訪して」</li> <li>研究発表3：佐々木保孝（天理大学）「『学校関係者評価』の課題と展望」</li> <li>研究発表4：小澤伊久美（国際基督教大学）「学校外の学びの場を支える評価」</li> <li>指定討論</li> <li>質疑応答・フロアとの討論</li> </ul>
<p><b>会場C</b></p> <p>自由論題2：コロナと ICT          座長・司会：佐藤由利子（日本学生支援機構）</p> <ul style="list-style-type: none"> <li>小西敦（静岡県立大学）「COVID-19 における都道府県間格差はなぜ生じたのか」</li> <li>○中谷大洋（滋賀大学）、杉本知之（滋賀大学）、大里隆也（滋賀大学）「コロナ禍における宿泊飲食業の正規・非正規雇用調整の効果検証」</li> <li>○周鵬（麗澤大学大学院）、秋山誠（NetExplorer）、ラウ・シンイー（麗澤大学大学院）「ICT 産業における受託開発企業の規模の経済の基礎的研究」</li> <li>○白井宏隆（麗澤大学大学院）、加藤郁夫（あしがら地域振興協会）、竹内正興（国際開発センター）「循環型発電事業による地域振興への影響評価に関する検討：北海道釧路市における企業廃業に伴う跡地再利用と循環型発電による産業創出と地域振興の事例」</li> </ul>
<p>シンポジウム：大学の社会的インパクト評価を考える          座長・討論者：齊藤貴浩（大阪大学）</p> <ul style="list-style-type: none"> <li>講演者1：今田克司（CSO ネットワーク／ブルー・マーブル・ジャパン）「社会的インパクト評価の系譜と展開—混乱を乗り越えるために—」</li> <li>講演者2：林隆之（政策研究大学院大学）「大学改革・研究評価改革における大学の社会的インパクト評価」</li> <li>講演者3：吉川徹（大阪大学）「人文社会科学の社会的インパクト評価をめぐる不確定性」</li> <li>質疑応答・フロアとの討論</li> </ul>
<p><b>2 日目（2023 年 12 月 17 日（日））</b></p>
<p><b>09:30 ～ 11:30</b></p>
<p><b>会場A</b></p> <p>共通論題5：政策形成・評価に関する改革の動向：総務省行政評価局における最近の取組          座長・司会：渡邊洋平（総務省）          討論者：小野達也（鳥取大学）</p> <ul style="list-style-type: none"> <li>趣旨説明</li> <li>報告1：風早勇作（総務省）「政策評価制度の運用見直しと最新の状況について」</li> <li>報告2：菊池明宏（総務省）「政府におけるEBPM 推進の取組～EBPMの実証的共同研究を中心に～」</li> <li>報告3：玉置賢（総務省）「『一人暮らしの高齢者に対する見守り活動』及び『遺留金等』に関する調査」</li> <li>討論者と各発表者のディスカッション</li> <li>質疑応答・意見交換</li> </ul>
<p><b>会場B</b></p> <p>共通論題6：What DE in Action Looks Like：発展的評価の実践における鍵概念と事例研究          座長・司会：今田克司（ブルー・マーブル・ジャパン）          討論者：米原あき（東洋大学）</p> <ul style="list-style-type: none"> <li>研究発表1：千葉直紀（ブルー・マーブル・ジャパン）「発展的評価（DE）の鍵概念の整理：実践にむけて」</li> <li>研究発表2：津崎たから（ウエスタン・ミンガン大学大学院）「ウィズコロナを機にDE的評価の導入に取り組んだ自治体と中間支援組織の挑戦事例」</li> <li>研究発表3：中谷美南子（チームやまびこ）、浅井美絵（チームやまびこ）「発展的評価（DE）の実践としてのアウトカム・ハーベスト手法の導入：子ども支援におけるコレクティブ・インパクトの評価事例より」</li> <li>指定討論者および座長からのコメント</li> <li>セッション参加者からの質疑応答</li> </ul>

○印は発表者または主発表者

会場C	<p>自由論題3：データとエビデンス          座長・司会：田辺智子（早稲田大学）</p> <ul style="list-style-type: none"> <li>・村上裕一（北海道大学）「[アジャイル・ガバナンス]と政策評価」</li> <li>・西村健（日本公共政策研究所）「行政経営・各種政策にかかわる指標・KPIの問題～EBPMの観点から」</li> <li>・本田正美（関東学院大学）「デジタル庁の提示する「オープンデータ取組の質評価指標」を拡張したオープンデータ施策の評価のあり方」</li> <li>・森俊郎（名古屋大学博士課程／養老町立東部中学校）「ニュージーランド the Iterative Best Evidence Synthesis programme の概要と動向」</li> </ul>
13:00 ～ 15:00	
会場A	<p>共通論題7：日本における業績測定の実態とその意義の再評価：ハリー・ハトリの業績と関連づけて          座長・司会・討論者：田中啓（静岡文化芸術大学）</p> <ul style="list-style-type: none"> <li>・研究発表1：田中啓（静岡文化芸術大学）「ハトリーによる業績測定の方法論—その功績を偲んで—」</li> <li>・研究発表2：左近靖博（浜銀総合研究所）「米国の業績測定から我が国取組が学ぶべきこと—日本の行政評価の取組を振り返る—」</li> <li>・研究発表3：小野達也（鳥取大学）「業績測定に何を求めるか—EBPM と評価疲れの間隙を巡って—」</li> <li>・研究発表4：上野宏（アジア都市コミュニティ研究センター）「政策評価における業績測定型評価の役割—プログラム評価型、RCT 型、費用便益型評価との比較—」</li> <li>・討論者によるコメント</li> <li>・質疑応答</li> </ul>
会場B	<p>共通論題8：Overview of Evaluation Education in the Asia Pacific Region and Cases of Graduate Programs in Australia and Japanese Universities（アジア太平洋地域における評価教育の概要と豪州及び日本の大学院教育の事例）          座長：佐藤由利子（日本学生支援機構）          司会：佐々木亮（国際開発センター）          討論者：石田洋子（広島大学）</p> <ul style="list-style-type: none"> <li>・趣旨説明と論点提起</li> <li>・発表1: Dr. Shyam Singh (Institute of Rural Management Anand, India) “Approach to M&amp;E Education: Ideas Based on a Review of M&amp;E Programmes Across the Globe”（モニタリング・評価教育へのアプローチ：世界のモニタリング・評価教育のレビューから）</li> <li>・発表2: Gullickson, Amy and Arbour, Ghislain (University of Melbourne) “Graduate Program in Evaluation Studies at the University of Melbourne”（メルボルン大学大学院における評価教育の概要）</li> <li>・発表3: 賛川信幸、新藤健太、ヴィラーグ・ヴィクトル、小原真知子（日本社会事業大学）、大島巖（東北福祉大学）“Welfare Program Evaluation Course for Social Work in the Graduate School of the Japan College of Social Work”（日本社会事業大学大学院における福祉プログラム評価履修コースの概要）</li> <li>・討論者によるコメントと発表者からの応答</li> <li>・質疑応答・フロアとの意見交換</li> </ul>
会場C	<p>自由論題4：CBAと社会的インパクト          座長・司会：今田克司（ブルー・マーブル・ジャパン）</p> <ul style="list-style-type: none"> <li>・石田健一（計画と評価専門家）「ソーシャル・イノベーション活動を評価する」</li> <li>・佐藤夢乃（関西学院大学大学院）「国際協力における社会的インパクト評価のあり方検討：財源基盤のない組織が評価を実装するために」</li> <li>・○桑原亜希子・柴田吉隆（日立製作所）、保井俊之（日立製作所/観啓大学）「社会イノベーション事業における社会的インパクト評価方法の検討」</li> <li>・宮本幸平（神戸学院大学）「CBAによる事務事業のインパクト測定の妥当性」</li> </ul>

15:10 ~ 17:10

## 会場A

共通論題9：日本評価学会「評価倫理ガイドライン」の改定案に関する会員意見交換会（RTD）

座長・発表者：小林信行（OPMAC/評価倫理・スタンダード分科会代表）

司会：中谷美南子（評価コンサルタント）

- ・ 討論者1：佐々木亮（国際開発センター）
- ・ 討論者2：西野桂子（関西学院大学）
- ・ 主旨説明：本ラウンドテーブルの目的について
- ・ 発表：小林信行（OPMAC/評価倫理・スタンダード分科会代表）「『評価倫理ガイドライン』改定案について」
- ・ 討論者による意見・質問、発表者の回答
- ・ ラウンドテーブル：「評価倫理ガイドライン」改定案に係るディスカッション
- ・ まとめ：議論のポイントと今後について

## 会場B

共通論題 10：国際協力機構（JICA）の事業評価における様々な評価手法の活用について

座長・司会：佐藤真司（JICA）

討論者：正木朋也（JICA）

- ・ 研究発表1：阿部俊哉（JICA）「JICA 事業評価の概要」
- ・ 研究発表2：森貴志（JICA）、田村愛弥（元JICA）、正木朋也（JICA／北里大学大学院）「世銀SWIFTを活用した高頻度家計調査による事業モニタリング・評価～マラウイ国における実践紹介（第2報）」
- ・ 研究発表3：石原和伊（JICA）、山口みちの（JICA）「質的比較分析（QCA）の手法を用いた事業間連携と効果の持続に関する考察」
- ・ 総合討論：質疑応答・フロアとの意見交換

## 会場C

自由論題5：国立大学と文化芸術

座長・司会：源由理子（明治大学）

討論者：米原あき（東洋大学）

- ・ 押海圭一（人間文化研究機構）「国立大学法人評価における研究の社会的インパクト評価はどうあるべきか」
- ・ ○洪井進（大学改革支援・学位授与機構）、金子明日香（大学改革支援・学位授与機構／日本女子大学大学院）「国立大学法人評価の影響に関するアンケート調査の縦断的分析」
- ・ 中村美亜（九州大学）「文化事業におけるプログラム評価の課題」
- ・ 金谷重朗（国立美術館）「アウトカム文化政策へのプログラム評価のアプローチ」

## 日本評価学会第24回全国大会セッション報告

### シンポジウム

座長・司会 齊藤 貴浩 (大阪大学)  
シンポジスト 今田 克司 (CSO ネットワーク/  
ブルー・マーブル・ジャパン)  
吉川 徹 (大阪大学)  
林 隆之 (政策研究大学院大学)

大学は優れた人材の育成のみならず、高度な研究を進め、得られた成果を広く社会に提供することなどによって社会に大きな成果をもたらしているはずだが、それを目に見える形で表現できていないのではないか。本シンポジウムでは、社会における大学の存在意義に照らして、どのように大学の社会的インパクト評価を行うべきかを考えることを目的に開催された。シンポジウムに先駆けて、大阪大学理事・副学長福田祐一氏より始めの言葉をいただき、座長の齊藤貴浩会員より大学の教育成果としての卒業生、研究成果としての論文等の先の話であるという説明を経て、3名の登壇者からの報告を得た。

最初に今田克司会員による「社会的インパクト評価の系譜と展開—混乱を乗り越えるために—」では、分科会の議論を踏まえて今までの社会的インパクト評価の流れについてご紹介いただき、大学に対しては諸活動の「意図」（大学ではミッションと呼ばれる要素）から始めることをご提案いただいた。

次に林隆之氏による「大学改革・研究評価改革における大学の社会的インパクト評価」では、英国のREFを中心に情報提供をいただき、インパクトを生む環境の重要性や、大学が社会と共にインパクトを生み出し、社会変革することをどのように評価するかという問題が提起された。

最後に吉川徹氏による「人文社会科学研究の社会的インパクト評価をめぐる不確定性」では、学内プロジェクトの紹介とともに、現在の研究評価が個人単位で学業成績重視であるため、社会的インパクトにアプローチできていないことと、評価方法が定まっていないために踏み出せないことが示された。新たな評価に対応すべきと思いつつも無駄なことはできないジレンマが示されたと言えよう。

これら発表を踏まえ、フロアの押海圭一会員からは政府の言う社会的インパクトと評価学のそれとの違いは何かという問いをいただいた。また、なぜ大学は社会的インパクト評価を行わなければならないのか等、短い時間ではあったが全体討議が行われ、有意義な時間となった。

## 共通論題セッション

### 共通論題1 「政策学と評価学」

座長・司会・討論者 湯浅 孝康（大阪国際大学）

共通論題セッションである本セッションでは、まず座長から企画の趣旨を説明し、3名の会員がそれぞれ報告した。その後、各報告についての討論が行われた。報告時間は各20分、討論時間は計30分であった。最後に、フロアからの質疑応答の時間を設けた。その際には複数の質問が寄せられ、関心の高さがうかがえた。

第1報告は山谷清秀会員（青森中央学院大学）からの「住民参加と評価」であった。本報告では、住民参加論において評価がどのように位置付けられてきたのかについて論じられた。参加の概念拡大と協働概念との合流を経て、内外問わず政策過程の各段階への参加が想定されるようになったものの、参加論特有の規範性の強さと個別の参加から実践的に得られた知見とのギャップは大きく、規範と実践の結節点を用意する必要があるとのことであった。

第2報告は鏡圭佑会員（朝日大学）からの「アジャイル型政策評価における政策学と評価学の定位」であった。本報告では、アジャイル型政策形成・評価の概念から見直された日本の政策評価制度における各府省の自律性について論じられた。各府省が裁量を適切に行使するにあたり、評価学は評価デザインに関する裁量の適切な行使やある資料を評価書として採用する際の適切な判断に関して、政策学は政策の特性把握や過去の教訓の提供に関して有用であるとのことであった。

第3報告は三上真嗣会員（同志社大学）からの「グローバルな政策の架橋—評価と計画—」であった。本報告では、グローバル化を背景として転換期を迎えつつある政策学と評価学の関係について論じられた。評価活動が政策形成のあらゆる段階に大きな影響力をもつようになるなか、政策学はそれに合わせて評価学の培ってきた知見に学びながら、政策や組織の管理やマネジメントについて再考する必要があるため、両学問領域に携わる実務家や研究者を増やしていく必要があるとのことであった。

### 共通論題2 「実務者と研究者の交流に基づく学術論文執筆ワークショップ： 『日本評価研究』誌の発展にむけて (RTD)」

座長・司会 米原 あき（東洋大学）

日本評価学会の魅力の一つは、学術研究と実践のコラボレーションにある。現在、研究者会員と実務者会員の比率は約半々となっている一方で、多くの実務者にとって学術論文の執筆に対する敷居は高く、実務者会員のあいだには豊富な実践経験や事例データが蓄積されているながら、それらが活字となって共有されるケースは限られている。本ラウンドテーブルは、『日本評価研究』誌が、学会の強みを活かした魅力

ある学術誌となるために、また、評価学と評価実践の架橋に貢献する学術誌となるために、オープンな議論を交わす場として編集委員会によって企画されたものである。

本ラウンドテーブルでは、以下の4つの話題提供のあと、フロア討論が展開された。村上裕一会員（北海道大学）の『「日本評価研究」のカイゼン』では、各投稿種別の存在意義と研究者・実務者双方の情報発信の場としての『日本評価研究』の可能性が提示された。橋本昭彦会員（日本女子大学）の『研究論文の社会的役割とその構成を考える』では、「研究論文」執筆の作法と同時に、「実践報告」「調査報告」がもつ役割との違いが分かりやすく説明された。窪田好男会員（京都府立大学）の『実務経験を「日本評価研究」に寄稿する』では、実務経験者が論文を書く時のコツが共有された。新藤健太会員（日本社会事業大学）の『実務家が研究することの意義とそれに配慮すべき倫理的事項』では、多忙な実務者が論文を執筆する動機付けについて、実務者会員である大澤望会員（インパクト・マネジメント・ラボ）、和田泰一氏（日本民間公益活動連携機構）の意見と共に、見解が述べられた。これらの話題提供に基づいて活発な討論が交わされ、編集委員会に持ち帰るべき論点も明らかになった。

### 共通論題3「政策評価と研究評価の交差点：責任ある研究イノベーション（RRI）と根拠に基づく政策（EBPM）」

座長・司会 白川 展之（新潟大学）  
指定討論者 南島 和久（龍谷大学）

科学技術評価分科会設立後3度目となる本セッションでは、異なる専門の研究者によりあまり交渉がないまま評価実践が行われてきた研究評価と政策評価の交錯している問題構造の解明をテーマとして実施された。冒頭趣旨説明の後に、基調講演として「責任ある研究イノベーション（RRI）と研究評価」と題し、標葉隆馬氏（大阪大学）から、科学技術・イノベーション政策での近年の政策アジェンダである責任ある研究イノベーションと研究評価について、英国の大学評価（REF）や計量書誌学的な評価の弊害・留意点など実践例を網羅的に紹介した。続く発表2「科学技術イノベーション政策における研究評価と政策評価：国プロジェクトと政策過程における知識生産の実務的視点から」では、高谷徹会員（三菱総合研究所）から、日本の大規模研究開発プロジェクト（SIP）などの評価の事例から、現状の課題を析出した。次いで、発表3「欧州連合（EU）におけるSmart Specialisationの取組に関する一考察 - 地域政策と科学技術イノベーション政策の包摂 -」と題して、野呂高樹会員（政策研究大学院大学）から、日本の政策推進・評価システムの比較・ベンチマーク対象として欧州委員会の評価と政策推進その他行政分野との施策推進の一体化の進展と総合計画・評価との一体的な推進体制が確立され推進されている状況を明らかにすることで、日本の抱える政策推進と評価の実施体制の問題を照射した。次いで、座長を兼ねる白川展之会員（新潟大学）の発表4「政策評価と研究評価の交錯：責任ある研究イノベーション（RRI）と根拠に基づく政策（EBPM）をどう読み解くか？」を通じて、発表の論点を整理したうえで、指定討論者の南島和久会員（龍谷大学）のコーディネートののもと、今後の日本での科学技術に関する評価研究の課題を整理し構造化した。

## 共通論題4 『学校評価マップ』の開発：学校評価学の構築のために

座長・司会 小澤 伊久美（国際基督教大学）  
 討論者 中谷 美南子（評価コンサルタント）

本共通論題では、学校評価分科会有志が試作した学校評価マップ（全体俯瞰図的な用語集）の概要を説明した上で、各発表者がマップに即して評価実践や研究事例を発表した。学校評価マップは、過去15年の学校評価の実践や議論を踏まえて学校評価を巡る主要概念を整理し、学校評価のリフレーミングを試みたものである。その目的は、学校評価を行う際に、あるいは学校評価を対象とした研究を行う際に知っておくと良い知識・知見を共有し、それらを体系的に理解したり、活用したりすることにある。

まず、座長が学校評価マップ制作に至る経緯や開発の方法について概要を報告した。次に、橋本昭彦会員（日本女子大学）が「マッピング（マップを制作する行為、また、マップを活用する行為）」により、学校評価制度の諸課題と学校評価学の研究者が向き合う可能性を、実例を示しつつ論じた。続いて佐々木保孝会員（天理大学）が学校関係者評価を軸として地域関係者が学校評価マップを用いるプロセスで学校評価マップを洗練するとともに、地域の人びとの学習過程をも学校評価学の体系に位置づける可能性を指摘した。

石田健一会員（計画と評価専門家）は、学校評価マップを踏まえて、第三者評価を行った体験型の学習を再分析し、当事者の自主性・主体性に委ねる評価法について考察した上で、評価者がソムリエ的な関係性で学校評価に関わる、学校外の学びの場における評価を行うことを提案した。小澤伊久美会員（国際基督教大学）は、学校外の学びの場を支える評価を行う実践例をもとに学校評価マップと評価の実践がいかに相互に寄与し得るかを報告した。

最後に中谷美南子会員（評価コンサルタント）より、学校評価に関わってこなかった評価者・評価研究者にとっての学校評価マップ活用の有効性、学校評価において課題とされている実効性とは何かなどが問われ、全体でこれを討議した。

## 共通論題5 「政策形成・評価に関する改革の動向～総務省行政評価局における最近の取組～」

座長・司会 渡邊 洋平（総務省）  
 討論者 小野 達也（鳥取大学）

本セッションでは、風早勇作会員（総務省）より「政策評価制度の運用見直しと最新の状況について」、菊池明宏会員（総務省）より「政府におけるEBPM推進の取組～EBPMの実証的共同研究を中心に～」、玉置賢会員（総務省）より「一人暮らしの高齢者に対する見守り活動」及び「遺留金等」に関する調査について報告が行われた。

風早会員からは、政策評価制度の概要、政策評価審議会答申、「政策評価に関する基本方針」の一部変更等の報告があり、それを踏まえて各行政機関が機動的かつ柔軟な政策展開のため工夫して政策評価に取

り組んだ事例等について報告された。

菊池会員からは、令和4年度に実施した総務省と法務省の実証的共同研究「刑務所における受刑者の就労支援の申し出促進策」の検証結果や今後の取組の方向性等が報告された。また、これまでの実証的共同研究で培ってきた知見をさらにブラッシュアップさせつつ、様々なタイプの実例を創出し、各府省間の横の繋がりを意識しながら、政策効果の把握・分析の取組を推進していきたい旨が報告された。

玉置会員からは、令和5年度に実施した行政運営改善調査2件「一人暮らしの高齢者に対する見守り活動に関する調査」及び「遺留金等に関する実態調査」の、調査内容や各府省への通知勧告内容が、公表資料を用い具体的に報告された。

討論者の小野会員からは、風早報告に対しては、各府省が政策評価を柔軟に行えるようになったことにより、一覧性や比較可能性が失われるのではないかと、EBPMは政策評価制度をベースに進められるべきではないか等の質問が行われた。菊池報告に対しては、本報告の事例について、重要な最終的アウトカムの評価検証はこれから行われるのか、試行期間の短さは初年度と同様の課題が残っているのか等の質問が行われた。玉置報告に対しては、全国の介護弱者について実態把握が必要ではないかと、また、行政運営改善調査によって、政策評価の枠外で政策立案等に用いられるエビデンスの点検ができないかとの指摘があった。

## 共通論題6 「What DE in Action Looks Like: 発展的評価の実践における鍵概念と事例研究」

座長・司会 今田 克司（ブルー・マーブル・ジャパン）  
 討論者 米原 あき（東洋大学）

本セッションは、2023年1月より活動を本格化させている、日本評価学会発展的評価分科会が企画したセッションである。同分科会は、昨年の全国大会の共通論題にて取り上げた『A Developmental Evaluation Companion』（Jamie Gamble, Kate McKegg and Mark Cabaj 共著、2021年、The McConnell Foundation）とその日本語版『DE Companion ～発展的評価の旅のおともに～』を輪読しながら、今年、発展的評価（以下、DE）について議論を重ねてきた。分科会において、ある程度DEの基本的特徴についての理解が進んでいるが、本セッションでは、より実践的なDEの鍵概念の整理をした上で、日本国内でのDEの実践事例を紹介した。

冒頭、座長よりセッションの趣旨説明やDEの定義について説明があった後、千葉直紀会員（ブルー・マーブル・ジャパン）より、報告1「発展的評価（DE）の鍵概念の整理：実践にむけて」において、DEの概要とDE評価者に求められる力量についての概説があった。続けて、津崎たから会員（ウエスタン・ミシガン大学大学院）より、報告2「ウィズコロナを機にDE的評価の導入に取り組んだ自治体と中間支援組織の挑戦事例」において、福岡市NPO・ボランティア交流センター「あすみん」の運営組織による自己評価ワークショップがDEの実践例として紹介された。さらに中谷美南子会員（チームやまびこ）より、報告3「発展的評価（DE）の実践としてのアウトカム・ハーベスト手法の導入：子ども支援におけるコレクティブ・インパクトの評価事例より」において、休眠預金活用事業の第三者評価事例がDEの実践例とし

て紹介された。報告2、3において、DE評価者の役割と力量に焦点が当てられ、DE評価の特徴が浮き彫りになった。

これらの報告を受けて、討論者の米原あき会員からは、DEは事業実施者の能動性を起動させるOS（オペレーティングシステム）ではないかという提起があり、DE評価だからできることは何か、DE評価者はいかに育成できるのか、などの質問が出された。加えて、会場からの複数の質問も交え、報告者との討議が行われた。

本セッションにおいて、DEに関する疑問にすべて答えが出たわけではないが、今後も分科会活動や大会等でのセッションを重ね、DEの射程や守備範囲、DE評価者の役割に関する理解が進み、DEの実践者が増えていくことが期待される。

## 共通論題7「日本における業績測定の展開とその意義の再評価： ハリー・ハトリの業績と関連づけて」

座長・司会・討論者 田中 啓（静岡文化芸術大学）

業績測定の方法論の確立と公共部門への普及において大きな役割を果たしたハリー・ハトリが2023年2月に亡くなった。そこで、ハトリの業績測定に関する主張を改めて確認した上で、日本の業績測定の実践への示唆を得ることを目的として、本セッションを企画・実施した。

最初に田中啓会員（静岡文化芸術大学）は、「ハトリによる業績測定の方法論—その功績を偲んで—」において、ハトリの1970年代から最近までの著作にみられるハトリの主張の一貫性や特徴、さらには変化した点について報告をおこなった。

続いて左近靖博会員（浜銀総合研究所）からは、「米国の業績測定から我が国取組が学ぶべきこと—日本の行政評価の取組を振り返る—」において、ハトリに直接インタビューした経験から、ハトリが強調していた業績測定に関する8つの要点が示された。

小野達也会員（鳥取大学）は、「業績測定に何を求めるか—EBPMと評価疲れの間隙を巡って—」と題し、ハトリの主要な著作を10篇取り上げた上で、それらから読み取れる内容のうち、日本の政府・自治体の業績測定型評価で大方欠落していると考えられる大小8点の事項について報告をおこなった。

最後に上野宏会員（アジア都市コミュニティ研究センター）は、「政策評価における業績測定型評価の役割—プログラム評価型（含RCT型）、費用便益型評価との比較—」と題した報告において、まず政策プロセスの中にプログラム評価をはじめとする各種の評価手法を位置づけた上で、それに基づきプログラム評価と対比しながら、業績測定型評価の役割を考察した結果が示された。

その後、討論者の田中啓会員から、「ハトリの主張のうち特に重要だと思うのはどのようなことか」「ハトリの主張を踏まえた上で、業績測定の領域において今後どのようなことを優先的に実施していくべきか」という2つの論点が提示され、報告者間で意見交換をおこなった。

## 共通論題8 「Overview of Evaluation Education in the Asia Pacific Region and Cases of Graduate Programs in Australia and Japanese Universities (アジア太平洋地域における評価教育の概要と豪州及び日本の大学院教育の事例)」

座長 佐藤 由利子 (日本学生支援機構)

司会 佐々木 亮 (国際開発センター)

討論者 石田 洋子 (広島大学)

本セッションでは、アジア太平洋地域における評価教育の現状を概観した後、オーストラリアと日本の大学院における評価教育の実際を紹介し、同地域の評価教育の特徴や課題について活発な討議が行われた。

第1発表者Shyam Singh氏は「モニタリング及び評価教育の主流化にむけて－世界の大学の教育課程の比較分析」と題し、APEAの報告書に基づき、世界26の大学院におけるモニタリング及び評価 (M&E) の教育課程を分析した結果を報告し、いずれの課程においても理論より実践、地域文脈よりグローバルな視点が重視され、前者の教育が弱いことを指摘した。

第2発表者Amy Gullickson氏とGhislain Arbour氏は、メルボルン大学大学院の評価学プログラムを紹介した。修士課程は1年間であるが、仕事を持ちながら学ぶ者が2年かけて履修することも多い。卒業生は、政府やNGO勤務者が最も多く、次いで大学、民間企業で、分野別では、教育、国際／コミュニティ開発分野で働く者が多い。

第3発表者は、日本社会事業大学の贅川信幸会員で、社会福祉分野では、幸福を追求して生きることが困難な状況にある人々に対する“相談援助”を中心とする個別・直接的支援 (ミクロ実践) に留まらず、そうした状況を生み出す社会や構造への働きかけ、新たな制度やサービスの開発、既存の制度やサービスの改善 (マクロ実践) が求められることを示した後、日本社会事業大学大学院に2009年に開設された福祉プログラム評価履修コースの概要と教育成果、今後の課題について報告した。

続いて指定討論者の石田洋子会員より、①アジアにおいてM&Eの教育課程が少ない理由、②M&Eの教育のlocalization (地域適応) において何が重要か、③オンライン教育のメリット／デメリットなどについて質問があり、回答が行われた。

またArbour氏よりSingh氏に、評価学において理論的基盤が欠けているのはどうしてかという質問が寄せられ、会場からも、評価学における理論的基盤構築の必要性について共感の声寄せられた。

## 共通論題9「日本評価学会『評価倫理ガイドライン』の改定案に関する 会員意見交換会（RTD）」

座長・発表者	小林 信行（OPMAC）
司会	中谷 美南子（評価コンサルタント）
討論者	佐々木 亮（国際開発センター）
討論者	西野 桂子（関西学院大学）

評価倫理・スタンダード分科会は、2021年より評価を取り巻く近年の環境変化を踏まえ、日本評価学会「評価倫理ガイドライン」の改定に関する作業を進めている。本セッションでは、当分科会が策定した同ガイドラインの改定案を発表し、指定討論者との意見交換を経て、セッション参加者も交えた改定案の討議を行った。

小林信行会員の発表では、改定案の概要として、①改定の論点抽出と課題分野のカテゴリー分類、②各カテゴリーの内容（IT技術の進展、調査対象者の尊重、参加型アプローチへの対応、依頼する者の適切な行動、評価の重複防止、その他）、③カテゴリー別、評価のフェーズ別に見た改定の特徴を説明した。その上で、課題分野の6つのカテゴリー毎に現行の記載と改定案を対比して改定内容を提示した。

討論者の佐々木亮会員から、改定案に関し、①生成AIの利用に制約をかけること、②RCTにおいて対照群に意図的に割り付けをしないこと、の二点につき是非を問う意見が出された。また、討論者の西野桂子会員から、アジア太平洋評価学会（APEA）の倫理規程策定時に現行のガイドラインが参考となった一方、記載が詳細すぎるとの意見があったことに言及があった。また、ガイドライン利用者の理解を助ける観点から、改定案の修正が望ましい点（評価に携わる者、割り付け等）も提案された。

改定案に関するフロアとの意見交換では、①長期的な観点から生成AI利用への過度な制約は避けるべき、②ガイドラインで言及されている行為者を整理し読み手の理解を助ける記載を期待する、等の指摘があった。また、いくつかの論点（評価の重複防止、介入の割り付け等）に関しては、改定案の背景や文脈への十分な説明がないと、改定の必要性自体が理解しにくいとの意見があった。

## 共通論題10「国際協力機構（JICA）の事業評価における様々な評価手法の活用について」

座長・司会	佐藤 真司（国際協力機構）
討論者	正木 朋也（国際協力機構／北里大学大学院）

本セッションは、JICAの事業評価における最新の取り組み及び検討状況について紹介し、セッション参加者との議論から得られる示唆に基づき、今後のJICA事業評価のより良い運用・改善につなげることを目的に開催した。

まず報告1では、最新のJICAや開発援助を取り巻く潮流と、それに対応したJICA事業評価の取り組みを俯瞰して紹介した。次に報告2では、世界銀行で開発された簡易的な調査手法であるSWIFT（瞬間・高頻度追跡によるウェルビーイング調査）を、技術協力「マラウイ国市場志向型小規模園芸農業推進プロジェ

クト（MA-SHEP）」に適用した事例について、計4回の家計調査を終えたことから、昨年度に続き同プロジェクトに関する事業モニタリング・評価の結果を共有した。最後に報告3では、JICA新評価基準において「整合性」の一部として整理しているJICA内外、特に他機関事業との連携・調整が事業効果の長期的な持続を促す要因となっている可能性について、質的比較分析（QCA）の手法を用いて因果推論を試みた事例について紹介した。

その後討論者からの論点提示として、JICA事業のインパクト及び動的な変化の確認・モニタリングを実施するうえでのToCの活用の重要性に関する説明があり、これを踏まえて各報告につき参加者からの質問をもとに総合議論が行われた。報告2では、数ある評価手法の中でSWIFTの活用に至った背景や今後のJICA内でのSWIFTの活用方針に関する質問と、ToCシーケンスと対応する集団を踏まえた解析が必要ではないかとの指摘があった。また報告3では、今後のJICA内でのQCAの分析結果の活用に関して、中でも特に今回の調査のテーマとも関連し、他機関との連携によって生まれる相乗効果を今後QCAや定性調査を組み合わせてどのように確認していくのかという質問があり、今後定性調査を通じて得られる教訓等は現在JICA全体で取り組んでいるグローバル・アジェンダやクラスター事業戦略にも活用していくことを想定している点を説明した。

## 自由論題セッション

### 自由論題1「開発協力」

座長・司会 石田 洋子（広島大学）

本セッション最初の発表は、丸山隆央会員（JICA）による「Using data to improve and expand operations for educational development: An example of Indian NGO Pratham」で、インドのNGO「Pratham」が行う活動現場からのデータに基づく意思決定並びに透明性強化のための取組の概要と成果について報告があった。指標の選定やデータ収集のための様々な組織の連携、データ統合のプロセスや人材育成の進捗と課題などが共有された。

2番目の発表は、宗像朗会員（JICA）、志賀圭会員・杉山卓氏（コーエイリサーチ&コンサルティング）による「バングラデシュの郡自治体ガバナンスのパフォーマンス評価結果について」で、JICA有償資金協力「郡自治体借款事業（UGDP）」において、全郡自治体を対象に毎年行っているガバナンス・行政評価（PA）について報告があった。PAの前提条件をパスしたのは1回目217郡であったが、5年後の5回目は475郡に増加し、介入によりガバナンス能力が底上げされつつあるものと推測される。

3番目の発表は、佐々木亮会員・藪田みちるの会員（国際開発センター）、Maree Soos氏（Ministry of Education, Palestine Authority）による「Impact Evaluation – Palestine Education Program (PIQMAS)」で、JICA技術協力「パレスチナ理数科教育質の改善プロジェクト」のインパクト評価により数学及び理科で正の変化が確認された。また、重回帰分析では、ジェンダー、携帯・PC所有の有無、ネット環境等の要因が成績

に影響する可能性が確認された。

最後の発表は、牟田博光会員（国際開発センター／大妻女子大学）による「ミャンマー連邦共和国における大学合格基準変更が合格者の男女比に及ぼす影響に関する研究」であった。ミャンマーでは、基礎教育修了のGrade11が終わる時点でマトリキュレーション試験（大学入学資格試験）が行われている。一般的に女子が高得点で、女子の合格率が高い状況にある。男女の合格率を近づけるために同国政府は大学合格基準を見直しており、この基準変更が男女比にもたらす影響に関する分析結果が報告された。

## 自由論題2「コロナとICT」

座長・司会 佐藤 由利子（日本学生支援機構）

本セッションでは4つの報告があり、それぞれ活発な質疑応答が行われた。

小西敦会員（静岡県立大学）による「COVID-19における都道府県間格差はなぜ生じたのか」では、COVID-19の影響が都道府県間で異なる要因を解明するため、人口関連（人口密度、高齢者率等）、医療保健体制（医師数、救急搬送困難率等）、意識や社会状況（規範意識、生活保護率）、ワクチン接種率など13の指標を選び、COVID-19の陽性率、死亡率との関連性を分析した結果を報告した。質疑応答では、データマイニング的手法を採用した理由、政策的な示唆、予防効果検証の必要性、統計的手法を事例研究と組み合わせる必要性などについて、質疑応答が行われた。

中谷大洋会員（滋賀大学）他による「コロナ禍における宿泊飲食業の正規・非正規雇用調整の効果検証」では、コロナ禍により大きな影響を受けた宿泊飲食業における従業員の雇用形態と業績の関係性について、企業行動についてのモデルを作成した上で、帝国データバンクの企業データを用いて統計分析した結果が報告され、宿泊飲食業の立地性の制約（従業員募集が困難）という仮定の是非、欠測値の代入方法、非正規雇用が多いと業績が向上するという分析結果が長期的にも妥当か、といった点について質疑応答が行われた。

周鵬会員（麗澤大学大学院）他による「ICT産業における受託開発企業の規模の経済の基礎的研究」では、Krugmanのモデルを応用し、あるICT企業の財務データを用いて、ICT受託開発業における規模の経済を推定した結果が報告され、ICT以外からICT産業への労働力移動に関する仮定の是非について、質疑応答が行われた。

白井宏隆会員（麗澤大学大学院）他による「循環型発電事業による地域振興への影響評価に関する検討」では、製紙工場の跡地にユーグレナ培養施設を設置し、その粉末を用いた火力発電を行う釧路市の戦略をバランス・スコアカードを用いて評価する計画が報告され、「釧路モデル」の内容、戦略策定に先立ってのSWOT分析の実施とその内容について質疑応答が行われた。

### 自由論題3 「データとエビデンス」

座長・司会 田辺 智子（早稲田大学）

本セッションでは、4名の会員からデータや指標、エビデンスに関連する報告がなされた。いずれの報告についてもフロアから積極的な発言・コメントがあり、活発な議論が行われた。

村上裕一会員（北海道大学）からは、「「アジャイル・ガバナンス」と政策評価」と題した報告があった。スピーディーかつ柔軟に政策の見直しを行うアジャイル・ガバナンスについて、事例を混じえつつ理論的整理が提示された。フロアとは、具体的な運用のあり方、アジャイル・ガバナンスとEBPMとの関係等について議論が行われた。

西村健会員（日本公共利益研究所）からは、「行政経営・各種政策にかかわる指標・KPIの問題～EBPMの観点から」と題した報告があった。自治体の各種政策における成果指標・KPIについて、問題点の指摘と考察が報告された。フロアとは、予算獲得のためのKPI利用や、KPIへの関心を高める必要性等について議論が行われた。

本田正美会員（関東学院大学）からは、「デジタル庁の提示する「オープンデータ取組の質評価指標」を拡張したオープンデータ施策の評価のあり方」と題した報告があった。デジタル庁による「オープンデータ取組の質評価指標」を踏まえ、オープンデータ施策の評価のあり方が提案された。フロアとは、海外の事例について、またデジタル庁に対する評価について議論が行われた。

森俊郎会員（養老町立東部中学校／名古屋大学博士課程）からは、「ニュージーランド the Iterative Best Evidence Synthesis programmeの概要と動向」と題した報告があった。ニュージーランド教育省が主導する教育エビデンス支援プログラムの活動内容について、英米との比較や我が国への示唆が報告された。フロアとは、“iterative”と“synthesis”の意味合いについて、また組織の性格やPISAとの関係等について議論が行われた。

### 自由論題4 「CBAと社会的インパクト」

座長・司会 今田 克司（ブルー・マーブル・ジャパン）

本セッションでは、社会的インパクトに関する評価を中心課題に、4つの報告があった。

まず、石田健一会員（計画と評価専門家）より、報告1「ソーシャル・イノベーション活動を評価する」において、国際教育フィールドプロジェクトにおける伴走型の評価事例の報告があった。

次に、桑原亜希子会員（日立製作所）他より、報告2「社会イノベーション事業における社会的インパクト評価方法の検討」において、営利企業の社会イノベーション事業の取り組みを促進させる方策として社会的インパクト評価方法の検討についての概説があった。

続けて、宮本幸平会員（神戸学院大学）より、報告3「CBAによる事務事業のインパクト測定の妥当性」において、政府・事務事業のアウトカム・インパクト値につき、SROI手法を参酌して測定出来るかが考

察された。

最後に佐藤夢乃会員（関西学院大学大学院）より、報告4「国際協力における社会的インパクト評価のあり方検討～財源基盤のない組織が評価を実践するために～」において、国際協力NGOの評価活動に関する実態調査と評価マネジメントの枠組みの提案に関する考察が報告された。それぞれの報告後、若干の質疑応答がなされた。

これらの報告は、イノベーションや社会的インパクトをキーワードに、評価実践をより効果的な事業実践に結びつけようとする試行錯誤が様々な分野で進んでいることを物語るものであり、日本評価学会の特徴である理論と実践の往還が重要であることを指し示していた。これらの取り組みが今後どのように発展していくのかについて、継続的に報告があることが望まれる。

## 自由論題5「国立大学と文化芸術」

座長・司会 源 由理子（明治大学）  
 討論者 米原 あき（東洋大学）

本セッションでは「国立大学」と「文化芸術」という二つの異なる領域の組合せから成る4名の発表があった。

第1報告では、押海圭一会員（人間文化研究機構）より「国立大学法人評価における研究の社会的インパクト評価はどうあるべきか」と題し、新たに導入される研究の社会的インパクト評価のあり方について海外の事例比較をとおした報告があった。討論者・フロアから、事例から分類されたインパクトの種類をもとに検討することの妥当性への疑義、エビデンスの評価デザインの違いに関する質問、評価結果が研究予算の配分に与える影響等の質疑があった。

第2報告は、洪井進会員（大学改革支援・学位授与機構）他による「国立大学法人評価の影響に関するアンケート調査の縦断的分析」と題する報告であった。法人の達成度評価による法人への効果・影響について、過去20年間の法人対象のデータを4期に分け2限配置分散分析により縦断的に分析したものである。討論者・フロアから属性による違いの比較の必要性、データから読み取れる内容（法人評価のあり方の検討内容）等について質疑があった。

第3報告では、中村美亜会員（九州大学）より「文化事業におけるプログラム評価の課題」と題し、文化事業におけるロジックモデルの活用に関し、芸術活動とマネジメントの混合から生まれる課題の指摘とそれを踏まえての活用のあり方について報告があった。討論者・フロアから、報告内容におけるEvaluability Assessmentやセオリー評価の視点、福祉領域との連携の可能性等の質疑があった。

第4報告は、金谷重朗会員（国立美術館）による「アウトカム：文化政策へのプログラム評価のアプローチ」と題した報告であった。芸術団体に対する文化芸術振興費補助金の評価のあり方について、ロジックモデル導入の観点から論じたもので、討論者から、ロジックモデルを新たに導入するにあたり直面した困難な点について質疑があった。

いずれもこれからの更なる研究・理論化が期待される報告で、今後の展開が待たれるテーマであった。

～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。投稿の締め切りは9月末日（翌年3月刊行）及び3月末日（9月刊行）です。

ご興味をお持ちの方は投稿規定・執筆要領・査読要領、並びに原稿見本をご参照のうえ、奮ってご投稿ください。<http://evaluationjp.org/activity/journal.html>

投稿先：特定非営利活動法人日本評価学会 編集委員会事務局

E-mail：journal@evaluationjp.org

『日本評価研究』第24巻第1号

2024年3月31日

編集・発行 特定非営利活動法人日本評価学会 編集委員会  
〒113-0034 東京都文京区湯島3-31-1  
中川ビル5階  
一般財団法人行政管理研究センター内

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. 24, No. 1, March 2024

## CONTENTS

### Preface

International Cooperation of Japan Evaluation Society Keiko Nishino

### Special Issue: Diversity and polysemy of evidence in evaluation research and practice

Editor's Note: Diversity and polysemy of evidence in evaluation research and practice  
Aki Yonehara

Reconsideration of Contemporary Debates Around Evidence:  
Including Credibility and Actionability As Key Concepts  
Katsuji Imata, Takara Tsuzaki, Minako Nakatani  
<Comment> Kunpei Nishimura, Shuya Wu  
<Reply> Katsuji Imata

Theory of Realistic Evaluation: Implications for Evidence-Based Policy Making in Japan  
Kunpei Nishimura, Shuya Wu  
<Comment> Hiroshi Tsutomi  
<Reply> Kunpei Nishimura, Shuya Wu

How to grasp the evidence in the formative evaluation methodology  
aimed at constructing an EBP program  
Kenta Shindo  
<Comment> Keitaro Aoyagi  
<Reply> Kenta Shindo

Evidence Utilization in the Japanese Nonprofit Sector  
- Analysis on a Survey of Child and Family Welfare Organizations -  
Junko Shimizu  
<Comment> Kenta Shindo  
<Reply> Junko Shimizu

### Article

A Participatory Evaluation Method using incomplete matrices of the AHP  
-Toward a Meaningful Relative Evaluation of Projects- Yoichi Iida

### Report

Evidence in policy making on small class sizes  
-Did the evidence influence the decision? Shintaro Araki

### Report of the 24<sup>th</sup> Annual Conference