

広島大学 高等教育研究開発センター 大学論集  
第53集 (2020年度) 2021年3月発行：37-52

## EBPMの動向と課題

—今後の高等教育研究の蓄積のために—

中 尾 走



# EBPM の動向と課題

—今後の高等教育研究の蓄積のために—

中尾 走\*

## 1. はじめに

内閣府 (2017) によれば、科学的根拠に基づく政策立案 (Evidence-based policy making, 以下 EBPM) とは、「政策の立案の前提となる事実認識をきちんと行い、立案された政策とその効果を結びつけるロジックを踏まえ、その前提となるエビデンスをチェックすることで、合理的な政策立案に変えていこうということ」である。EBPM は、緊縮財政下において効率的で合理的な政策立案が可能 (総務省, 2018) という期待からわが国の政策立案全般に要請されている。

政策レベルで EBPM の推進が最も期待されているのは教育分野である (田中, 2020)。研究としても高等教育分野と隣接する諸学会では特集号が生まれ (日本教育学会, 2015; 日本評価学会, 2006, 2016 など), EBPM に高い関心を見せている。このように隣接分野が EBPM へ高い関心を寄せる一方、高等教育研究では一部の先行研究 (小方, 2014; 山本, 2016, 2018) があるにとどまる。しかしながら、①初中等教育に比べて高等教育は政策の影響を受けやすいという制度的文脈、②高等教育研究が政策と緊密に関わることで発展を遂げてきた (金子, 1993) という高等教育研究の特徴を踏まえると、EBPM は高等教育研究として重要なトピックであると考えられる。そこで本稿は、他分野で議論されている EBPM に関する先行研究のレビューを行い、今後高等教育研究としてどのような研究が必要かを導くことを目的とする。

## 2. レビューの枠組み

本節では、レビューの枠組みを示す。レビューを行う上で考慮しなければならない点は、EBPM をめぐる議論が噛み合っていないという指摘である。例えば、「EBPM の肯定と否定がエビデンスという言葉異なる意味で用いているために議論が噛み合っていない」(今井, 2015)、「EBPM をめぐる議論は、それが何なのか明らかでないまま批判することで混乱する」(Cairney, 2016) といった指摘がある。今井 (2015) は、否定派が用いられる文脈からエビデンスを捉えているのに対し、肯定派は、エビデンスを内的妥当性 (3節で解説) の質として捉えている、と指摘する。言い換えれば、否定派と肯定派の齟齬は、エビデンスを「E (Evidence) からみるか PM (Policy making) からみるか」(林, 2019b) というアプローチの違いにより生じている、と言えよう。そのため本稿では、エビデンス自体をエビデンスの質=内的妥当性 (以下、E からみたエビデンス) に

\* 広島大学大学院教育学研究科教育学習科学専攻 (高等教育学) / 日本学術振興会特別研究員

関する議論と、用いられる文脈からみたエビデンス（以下、PM からみたエビデンス）とに区別して、先行研究を整理することとする。

また、Cairney (2016) の指摘を踏まえると「EBPM とは何か？」をレビューから導き出すことが必要となる。この時、参考になるのは南島 (2019) による①政策分析 (policy analysis)、②プログラム評価 (program evaluation)、③業績測定 (performance measurement) という評価の理念に基づく類型から EBPM を捉える視点である。そこで、本稿では南島 (2019) の枠組みを踏襲し、政策分析、プログラム評価、業績測定それぞれと EBPM の共通点、相違点について先行研究を踏まえながら適宜言及する。このような整理を踏まえて、EBPM は評価の理念類型からどのように位置づけられるかを導き出すことが可能となり、今後の研究蓄積のために不可欠な作業であると考えている。以上のことを踏まえて、エビデンスを二つの視点、EBPM を三つの視点から捉え、レビューを行っていく。

### 3. エビデンスとは何か？

#### (1) E からみたエビデンス

はじめに、なぜエビデンスピラミッドの手法が推奨されているのだろうか。それは、交絡 (cofounder) や選択バイアス (selection bias) と呼ばれる問題に対処するためである。ランダム化比較試験 (Randomized control trial, 以下 RCT) は、交絡や選択バイアスがないことが確率的に期待され、処置を受けた／受けないという違い以外は同質な集団であるという交換可能性 (exchangeability) の仮定を満たす。観察データを用いた分析でも交絡や選択バイアスの問題に対処可能だが、交換可能性が満たされないため、条件付き交換可能性 (conditional exchangeability) が満たされるような調整が必要となる (大久保, 2019)<sup>1)</sup>。しかし、観察データを用いた因果推論では、観察されていない変数によって交絡・選択バイアスが生じ得るなどの限界がある。以上の理由によって、EBPM で推奨されている手法は、RCT, 準実験研究, 観察研究という順に序列化されている。

RCT は、倫理面・費用面が長きに渡って課題であったため、社会科学の分野では利用が減少していたが<sup>2)</sup>、開発経済学や労働経済学を中心に再び利用されてきている。また、倫理問題に対しては実験設計の工夫により倫理的な嫌悪感を軽減する仕組みづくり (横尾, 2019) や、Igarashi & Ono (2020) のように Web 調査を用いることで従来よりも少ない費用で実験研究を行うことが可能となってきた。加えて、自然に発生した実験的状况を用いることで交絡・選択バイアスの問題へ対処する自然実験と呼ばれる方法も倫理面や費用面のハードルを超えた実験研究の一つである (伊藤, 2017)。実験研究と言えば「子どもを実験台にするのか」(中澤, 2016) という倫理面からの批判が予想されたが、費用の問題も含めて課題に対処する努力がなされている。一方で、RCT の研究デザインは、裕福な高学歴の人が貧困者、低学歴の人に対して、介入を行うという構図になっており、その恩恵を受けるのは介入を行った裕福で高学歴な人であるという Deaton (2020) の指摘は新たな倫理問題として見逃せない問題である。

次に、政策立案の文脈で RCT や自然実験など実験研究全般の課題として指摘されるのが、内的

妥当性 (internal validity) だけでなく、外的妥当性 (external validity) も考慮しなければならない (井伊・五十嵐・中村, 2019, p.137) という点である。内的妥当性とは、因果推論の質をあらわすもので、外的妥当性とはあるサンプル (集団, 時代など) で得られた知見が、他の集団でも適用できるかを評価する指標である (津川, 2020, 99頁)。RCT は、内的妥当性が高い手法でありバイアスのない因果効果の推定が期待できる一方、他のサンプルからも同様の知見が得られるかは定かではない。そのため、外的妥当性の問題が蔑ろにされているとの指摘は多くある (中澤, 2016; 大久保, 2019)<sup>3)</sup>。つまり政策立案という文脈では、エビデンスピラミッド (=内的妥当性) だけでエビデンスの質の高さが保証されないのではないかという課題である。

逆になぜ、根拠に基づく医療 (Evidence-based medicine, 以下 EBM) では内的妥当性がエビデンスの質として捉えられてきたのだろうか。それは、EBM は生態学的斉一性が前提であるという林 (2019a) の指摘が一つの答えであろう。斉一性とは、対象が均質であるという原理である。一部のサンプルの知見が外的妥当性を満たすかどうかは、対象が均質であれば満たされて当然であるが、星野 (2009, 141頁) は、社会科学では斉一性のような仮定を置くことは出来ないとしている。この指摘を踏まえると、政策立案という文脈では RCT によるエビデンスが、そのまま質の高いエビデンスとなり得るとは限らないことが分かる。

外的妥当性が成立しえない事例として幼児教育無償化の議論を取り上げてみよう。幼児教育無償化のエビデンスとして言及される Heckman (2007) の研究は、RCT という内的妥当性の高い手法を用いたアメリカの事例である。けれども、50年前のアメリカで確認された教育機会に恵まれない就学前の子どもに対する効果が、幼児教育 (保育所を含む) をほとんどの人が受けている日本の現状でも当てはまるかは疑問が残る。実際に、幼児教育無償化はこのような社会背景の違いを踏まえていない暴論であると赤林 (2017) は批判している。これは、政策立案の文脈において、エビデンスピラミッドだけでエビデンスの質を評価できず、外的妥当性も考慮しなければならないという一つの事例であろう。ここまで RCT を含む実験研究全般に対して外的妥当性の問題を取り上げてきたが、観察研究でも同様であり、星野・岡田 (2006)、King & Nielsen (2019)、中尾 (2019) は観察研究に対する外的妥当性の問題を指摘した先行研究である<sup>4)</sup>。

ここまで実験研究、観察研究の共通課題である外的妥当性について考えてきたが、一貫していたのは「集団が異なる=外的妥当性」を満たさないという前提である。しかし、「集団が異なることによって、なぜ外的妥当性が満たされないのか」という問いには十分に答えていない。斉一性の原理が成り立たない社会科学の分野ではどのように考えるべきであろうか。この問いに対して、Hernan & VanderWeele (2011) の指摘が参考になる。Hernan & VanderWeele (2011) は、外的妥当性の問題は、根本的には効果修飾 (effect modification)、相互作用 (inference)、複合処置 (compound treatment) の三つの問題であると指摘する<sup>5)</sup>。

まず一つ目の効果修飾<sup>6)</sup>とは、共変量がとる値ごとに因果効果の大きさが異なる (黒木, 2017, 150頁) ことを言う。例えば、留学効果が学年で異なる場合、学年が効果修飾の因子である。この場合、実験研究や観察研究で交絡を統制した集団と現実の集団で、学年の比率が異なれば効果修飾によって外的妥当性が満たされることはない。逆に、集団間で学年の比率が異なっていたとしても

(その他の属性変数を含めて) 効果修飾がない(全員が同じ効果をもつ)のであれば外的妥当性が満たされることはある。

二つ目の相互作用とは、ある個体に対する処置が他の個体に間接的に影響を与えることである。例えば、現在のコロナ禍で大学1年生と大学2年生以上に、アンケートに回答してもらうためギフト券を配布することを考える。このとき、大学2年生以上は大学1年生よりも繋がりがあり、回答した学生が周囲の友人にギフト券がもらえることを宣伝し、回答率がより向上するかもしれない。けれども、大学1年生は繋がりが希薄であり、学生間での宣伝効果がないために大学2年生ほど回答率が期待できない可能性がある。このように、個人間の相互作用のレベルが異なることで、ある集団の因果効果(大学2年生以上に、ギフト券を配布したことによる回答率の向上)が他の集団(大学1年生)で異なり、外的妥当性が満たされない場合がある。

三つ目の複合処置とは、十分に定義されていない(ill-defined)処置である。これは例えば、学級規模の効果などが挙げられる。学級規模の効果を取ってきた先行研究では、小規模学級の効果が一貫しないクラスサイズパズル(Bosworsh & Caliend, 2007)と呼ばれる現象が生じている。それぞれの先行研究が、それぞれ異なる学級規模の人数で因果効果の推定を行うという複合処置の事例の一つである。山本・荻原(2016)が指摘するように、諸外国の小規模学級は日本では実現できない小ささである。もし仮にクラスサイズパズルが解消され、一貫した知見を得たとしても、海外と日本では、例えば20人と35人など異なる小規模学級の処置である可能性があり、人数を無視した小規模学級という複合処置によって、外的妥当性が満たされない恐れがある。

これらをまとめると外的妥当性の解決には、効果修飾や相互作用についての①対象の同質性と、複合処置ではないという②処置の同質性の二点を考えなければならない。社会科学では斉一性を前提とできない以上、①対象の同質性の担保は、困難な課題である。そこで、Sara et al. (2019)は、効果修飾の分布の違いによる影響を感度分析(sensitivity analysis)を用いて検証することを提案している。②処置の同質性の担保は、林(2019a)によればメカニズムの解明を含む質的な理解度に依存する。そして質的な理解度を高めるために、質的研究も重要であることを指摘している。

## (2) PM からみたエビデンス

「エビデンスに基づく」という考え方は、医療の臨床分野でEBMとして始まり、教育分野で最初に広がりを見せたのは英国である。米国でも2002年「落ちこぼれ防止法(No Child Left Behind Act of 2001, 以下NCLB法)」が契機となり(桐村, 2020)エビデンスに基づくことが重視されており、英米を中心としたEBPMの動向の整理(田辺, 2006; 大槻, 2018; 中泉, 2019など)は数多く存在する。また、政策立案という文脈からみたエビデンスについては、Weiss(1979)の研究活用モデルが有名であり、日本に限定すると、エビデンスの使われる文脈から整理した小野(2018)や政策過程の複雑性からエビデンス活用について考察した佐藤(2019)などがある。

日本よりも先にEBPMが広がった英米の先行研究で重要な知見は、PMからみるとエビデンスが非常に狭い概念であることを受けて反省の段階へ移行している点である。たとえば、英国では「エビデンスに基づく(Evidence-based)」ではなく、「エビデンスを踏まえた(Evidence-informed)」と

いう呼称が広がっている (Breckon & Gough, 2019)。こうした動向は、① EBPM で想定されているエビデンス概念の狭さと、② エビデンスと政策立案の直線的な結びつきに対する疑義から、より広い観点で捉え直そうという動きであり、EIPM の呼称が国際機関でも一般的になっている (加納他, 2020)。一方、米国では EIPM の呼称は用いられなくとも、エビデンス要求の柔軟化が生じている (桐村, 2020)。2015年にオバマ政権下で成立した「全ての生徒が成功する法 (Every Student Succeeds Act)」ではエビデンスの階層はあるものの NCLB 法に比べて曖昧化されており、このような変化は、エビデンスの定義の狭さと文脈から乖離した現実性の薄さという NCLB 法の失敗に対する反省から生じたと桐村 (2020) は説明している。

その他に PM からみたエビデンスへ着目した重要な先行研究として、EBPM の前提である合理的な意思決定に対する論点を投じた Botterill & Hindmoor (2012) がある。Botterill & Hindmoor (2012) は、そもそも政策立案者は限定合理性しか有しておらず、エビデンスの収集段階・伝達段階で歪曲され、矮小化されるリスクがあることを主張し、事例研究によってその過程を実証している。この論点は、エビデンスのもつ政治的バイアス (Parkhurst, 2017) を超えて、そもそも人間が限定合理性しか有していない以上、合理的な意思決定は困難であることを再確認させてくれる。たしかに、エビデンスの政治的利用は危惧されるべきだが、限定合理性が前提である以上、完全に政治的に利用されることもない。このことを踏まえると、EBPM の肯定派も否定派も暗黙のうちに政策立案者が完全合理性を有することを前提としていたということであろう。

## 4. EBPM とは何か？

ここからは「EBPM とは何か？」という問題を考えるためにレビューを行う。南島 (2019) によれば、EBPM は政策分析の「合理的選択モデル」、プログラム評価の「プログラムの明確化」と「アウトカム検証の技法」、業績測定 of 「計画・目標・指標」と親和的であると指摘している。そのため以下では、EBPM を①政策分析、②業績測定、③プログラム評価の三つの側面から捉えた上で、EBPM との共通点、相違点について先行研究を踏まえながら適宜言及するというレビューの仕方をとる。

### (1) 政策分析からみた EBPM

まず①政策分析とは、「明確な基準としっかりした成果および費用のデータにもとづいた政策代替案およびプログラム代替案の合理的な選択を可能にすること」(宮川, 2002, 39頁) とされる。この言葉の定義通り、データに基づく合理的な意思決定が中核に据えられている点が EBPM と共通している。けれども、分析の根底にある価値と評価のタイミングが少し異なる。EBPM がエビデンスピラミッドを用いて「何が有効か」(What works) を重視している (Bridge et al. 訳書, 2013, pp.17-48) のに対し、政策分析では、費用便益分析・費用効果分析を用いて 1 ドル当たりの価値 (効果) を最大限にする VFM (Value for Money) の考え方が強調されていた (宮川, 2002, 40頁)。そのため、代替手段を選択する価値基準が政策分析では効率性であったのに対し、EBPM では有効

性であるという違いがある。また政策分析が推計やシミュレーションを行うことで事前にアウトカムを予測する（南島，2020，60頁）のに対し，EBPMではこれまでの政策（施策，プログラム，プロジェクトも含む）の効果を明らかにするという事後評価であり，評価のタイミングが異なる<sup>7)</sup>。

政策分析の理念を最初に具体化したのは，1960年代アメリカのPPBS（Planning, Programming, and Budgeting System）である（南島，2018，186頁）。PPBSは，複数のプログラム群の中で費用と予想される効果から比較を行い，最適なプログラムに予算をつける（山谷，1997，41頁）ことを目指したが，アメリカでは10年足らずで頓挫することとなった。宮川（2002，41頁）は，その理由はプログラムの目標が曖昧で明確に定義されていなかったこと，必要なデータの収集や分析のための資金がなかったこと，十分なデータがなく政策分析のための専門的人材が不足していたことの三点を挙げており，それぞれのプログラムの比較が容易ではなかったということを指摘している。以上の三点は，教育政策におけるEBPMの課題とも共通している。つまり，EBPMを推進する上での課題の一部は，PPBSのアメリカでの失敗で，既に明らかになっていたことであり，この課題が解決されなければEBPMも日本で定着することは難しいことを示唆する。

日本でもPPBSは1960年代に紹介され，制度と手法の二つの側面から適用しようという動きがあった（南島，2020，53頁）。制度輸入は，1967年に経済企画庁システム分析室を発足し，PPBSのための研究が予算化されたため特定のテーマについてケース・スタディを行うなどの取り組みが行われた（加藤，2008，150-151頁）。手法摂取は，経済学，オペレーションズ・リサーチ，システムズ・アナリシスなどを学問的背景にもつ費用便益分析がメインである。この点，EBPMが経済学や医学（EBM）を学問的背景とするのと微妙に異なる。また，アメリカでPPBSを牽引したのは，社会科学以外の領域で訓練を受けた応用数学者，技術者，OR研究者，システムズ・アナリストなどであったが（宮川，1994，69頁），日本では経済学者，プログラム分析者などであった（加藤，2008，148頁）。このように，制度輸入と手法摂取の二つの側面から日本でもPPBSを適用しようとする動きはあったが，アメリカと同様に日本でも定着することはなかった。その理由を加藤（2008，145-161頁）は，アメリカと日本の予算編成の仕組みの違いが考慮されていなかったこと，そして分析が非常に抽象的で予算編成が遅れるなどの混乱があったことの二点を指摘している。政策分析の理念を最初に具体化したPPBSであったが，アメリカでも日本でも定着することはなく，岡本（2018）は合理的意思決定に対する批判が現実化したものと指摘している。けれども，PPBSが志向した合理的な意思決定のモデルは，その後も繰り返し目指され（南島，2020），EBPMでも同様に合理的な意思決定が目指されている。

## (2) プログラム評価からみた EBPM

次に，②プログラム評価とは，「社会的介入プログラムの効果性をシステムティックに検討するために，プログラムを取り巻く政治的・組織的環境に適合し，かつ社会状況を改善するための社会活動に有益な知識を提供しうる方法で，社会調査法を利用すること」（Rossi et al. 訳書，2005，p.15）とされる。プログラム評価は，政策分析のような事前分析ではなく，政策が実施された後に，政策効果を明らかにすることが主な目的である（南島，2018，188頁）。

プログラム評価がこのような形で用いられるようになったのは、1960年代に当時のアメリカの会計検査院（General accounting office, 以下 GAO）に PPBS の手法が持ち込まれたことに由来する（南島, 2020, 21頁）。ただし、複数のプログラムを事前に比較分析するという PPBS の困難を受けて、一つのプログラムの効果を事後的に評価するという修正を経て、プログラム評価として受け継がれた（山谷, 1997, 41頁）。PPBS からプログラム評価への修正は、事前評価から事後評価、効率性の分析から有効性の分析への移行であり、分析の根底にある価値観と評価のタイミングは EBPM と同じとなる。その後、アメリカでプログラム評価を進めていく中で簡潔に評価する方法が必要となり、その際に最も推奨されることとなったのが RCT である（西尾, 1976, 199頁）。つまり、EBPM のエビデンスピラミッドで最も質が高いとされる RCT であるが、1970年代には既にアメリカのプログラム評価の手法として応用が模索されていたということであろう。

日本でもプログラム評価を適用しようと、PPBS と同様に制度・手法の二つの側面から努力があった（南島, 2020, 54頁）。制度輸入は、1990年代の日本版 GAO（行政監視院）設置法案である。ただし、制度輸入の段階で既に GAO のプログラム評価の内容は適切に理解されていなかった。それは、この法案が挫折した政治過程を明らかにした窪田（2005）の以下の二つの理由から読み取れる。法案挫折の理由は、誰が監視を行うかという主体の問題と監視をどのような条件で発動するかという発動要件の二つの問題で折り合いがつかなかったこととされる（窪田, 2005, 48頁）。つまり、日本版 GAO（行政監視院）設置法案は、法案の協議プロセスで本来の GAO のプログラム評価とは異なる機能である監視に焦点が置かれ、協議が決裂した。また、手法摂取では、経済学、オペレーションズ・リサーチ、システムズ・アナリシスなどの数理系のアプローチの他に計量社会学が加わった（南島, 2020, 54頁）。しかしながら、プログラム評価を意識した日本の総合評価方式（小野, 2018）において推奨されている手法としては、クロスセクション分析や時系列分析が紹介されており（南島, 2020, 59頁）、政策効果を明らかにする方法として RCT が推奨されているわけではない<sup>8)</sup>。その理由もあってかプログラムの因果効果を明らかにするという本来の意味での GAO のプログラム評価は、日本ではほとんど行われていないのが現状である（小野, 2013）<sup>9)</sup>。

### (3) 業績測定からみた EBPM

最後に、③の業績測定とは、「サービスあるいはプログラム（施策）のアウトカム（成果）や効率を定期的に測定すること」（Hatry 訳書, 2013, p.3）とされる。業績測定は、アメリカで PPBS やプログラム評価を自治体への適用を模索する中で生み出され、アメリカの GPRA（Government Performance and Results Act, 1993）という法制度の下で展開されていった（南島, 2020）。元々は、民間企業で用いられている「目標管理」の仕組みをモデル化したものであり（南島, 2018, 189頁）、公共部門で用いられるようになったのは、世界中の公共部門が NPM を取り入れるようになってからである（小野, 2013）。政策分析、プログラム評価と比較した業績測定の特徴の一つとして、現場において取り組みやすいという実用性が挙げられる（南島, 2018, 189頁）。その反面、政策分析やプログラム評価と比べて科学的な方法論が学問的背景として強く存在しているわけではなく、多くの文献では指標または測定の質に対する注意喚起が促されるにとどまる（小野, 2013）。

業績測定の理念的な位置付けは、プログラム評価の一部という認識であり (Hatry 訳書, 2005), 理論上はお互い補完し合うもの (田辺, 2014) という理解が共有されている。一方現場では、プログラム評価と業績測定が混同し、異種混交 (heterogeneity) (山谷, 1997)<sup>10)</sup> と表現される状態となっている。そのため、理念上の政策評価制度<sup>11)</sup> はプログラム評価の側面が強かったにも関わらず、現実には業績測定が重要な役割を果たすこととなる (山谷, 2006, 299頁)。その原因として、accountability が日本語で説明責任と矮小化して理解されたことで説明さえすればよいというモラルハザードが生じたこと (山谷, 2006), 業績測定という概念が曖昧であることで、その外延が無限定になったこと (南島, 2020) の二点が指摘されている。政策評価制度のこのような失敗は、EBPM にも通ずる部分がある。小野 (2015) は、教育政策の審議会資料でエビデンスという言葉がどのように使われているかに着目している。その中で、高等教育政策に関する議論ではエビデンス (数値情報など) に基づいて PDCA サイクルを回すという意味で用いられていると指摘した。そもそも EBPM はプログラム評価と同様に政策の効果を明らかにすることを目的としており、そのための手法としてエビデンスピラミッドが存在していた。けれども、小野 (2015) の指摘は、現実の審議会の場面ではプログラム評価よりも業績測定に近い意味でエビデンスという言葉が用いられていることを示している。エビデンスという言葉のこのような用いられ方は、政策評価制度においてプログラム評価よりも業績測定が大きな役割を果たすようになったことを EBPM でも同様に繰り返す可能性を秘めている。政策評価制度で業績測定ばかりが現場で用いられるようになった経緯を踏まえると、前節で触れた EBPM から EIPM への動きは、EBPM を広い観点で捉え直すという修正を要求する反面、プログラム評価と業績測定の境界を曖昧化したままにする可能性もあり、適用には慎重を要するのかもしれない。もちろん、業績測定と EBPM が完全に相反するわけではなく、どちらも定量化を前提とし、成果を把握する段階に違いがあるに過ぎない (山本, 2018) との認識もあり、両者は不可分である。ただし、塚原 (2003) が注意を促すよう教育政策では成果を長期的にみる必要があり、成果の把握が難しいことを常に念頭におくべきであろう。

## 5. 結語

本稿では、エビデンスや EBPM についての議論が噛み合っていない現状を踏まえて、3節でエビデンスを二つの視点、4節で EBPM を三つの視点から先行研究の整理を行った。

まず3節での整理から、政策立案におけるエビデンスの質は、簡単には評価できないということが言えるのではないだろうか。「エビデンスの質=内的妥当性」が成り立つとは言い難く、用いられる文脈まで拡げるとそのプロセスの中で歪曲、矮小化されるリスクがある。つまり、エビデンスの質 = {内的妥当性, 外的妥当性, 文脈} という少なくとも3つの要素からなる関数であり、それぞれの要素を適切に考慮した上でエビデンスの質を考える必要がある。

また、4節で評価の理念に基づく類型から EBPM を捉えると、以下の三点に結論が集約される。

① EBPM は政策分析と同様に合理的な意思決定を支援するものとして期待されているが、両者では合理性の価値基準が異なる<sup>12)</sup>。②プログラム分析と同じように、因果関係を明らかにする手法と

密接に結びついており、理念上は古くから RCT が必要とされていたこと。けれども、目的と手法の乖離がみられたプログラム評価（総合評価方式）に対して、同様の目的を擁した EBPM は、エビデンスピラミッドという手法も同様に輸入し、手法を適用するという側面がより強く前面に出ていること。③政策評価制度では、プログラム評価を志向しながらも業績測定が大きな役割を果たすようになり、EBPM でも同様の失敗を繰り返す可能性があることである。このような整理から EBPM は、良いかどうかは別として3つの評価の理念類型を組み合わせた概念であるということが言えるのではないだろうか。

以上、他分野の先行研究を整理した上で、今後高等教育研究として何が求められるであろうか。まず一つ目に、高等教育政策を対象にエビデンスピラミッドの手法を応用し、これまで用いられてきた手法との結果を比較することでエビデンスピラミッドの手法がもつ可能性と限界について把握することであろう。3節のレビューから得られたように、政策立案の文脈では EBM と同様にエビデンスの質を捉えることは難しい。高等教育政策という対象に応用した上でどのような課題が析出されるのかを把握することは、エビデンスピラミッドの手法が「通俗化された経験主義」（中村，2007）として用いられることを防ぐことが可能であろう。

二つ目に、過去のエビデンス・ブーム（妹尾，2018）が現実の政策立案にどのような影響をもたらしたのかを明らかにすることである。4節での整理を踏まえると、EBPM はこれまでのエビデンス・ブームと共通する部分もあり、完全に新規なものとして捉えることは難しい。けれども、EBPM が求められるということはこれまでのエビデンス・ブームが定着していないからとも考えられる。また、3節での整理を踏まえるとエビデンスの質の高さは、それが使われる文脈と密接に関係する。それでは、過去のエビデンス・ブームは、現実の政策立案に対して、よりエビデンスを重視させるような変化をもたらしたのだろうか。もし仮に何の変化ももたらしていないのであれば、過去のエビデンス・ブームがなぜ形骸化を繰り返してきたのかを明らかにすることで、最も新しいエビデンス・ブームである EBPM に対する重要な示唆を与えるだろう。

このような研究を通して、EBPM という一つのエビデンス・ブームを相対化して捉えることが今後必要になるのではないだろうか<sup>13)</sup>。

## 【付記】

本研究は JSPS 科研費20J14673の助成を受けた成果の一部である。また、本研究推進に際し、JSPS 科研費 JP18K18651・JP19H00621・JP20H01643の資金提供を受けた。ここに記して感謝申し上げる次第である。

## 【注】

- 1) 重回帰分析の文脈で言えば交絡や選択バイアスの変数を適切に統制するということである。
- 2) 末吉・片岡（1957）など、RCT を用いた研究も以前は一部存在した。

- 3) RCT などの実験研究を行えば外的妥当性が必ず低くなるわけではない。外的妥当性まで考慮に入れた実験研究として善教・坂本（2020）参照のこと。
- 4) King & Nielsen（2019）、中尾（2019）は、エビデンスピラミッドで推奨されている手法の一つである傾向スコアで、外的妥当性の問題が生じることを実証している。
- 5) 相互作用、複合処置については因果効果の識別仮定である SUTVA 問題（岩崎，2015，77-79 頁）と同様であり、因果効果の識別のためだけでなく外的妥当性を満たす意味でも重要であるということであろう。そして、相互作用の例のように宣伝効果を含む因果効果を推定した場合、内的妥当性は低いが個人間の相互作用の度合いが同じであれば、宣伝効果を含む因果効果の外的妥当性は高いということもあり得る。
- 6) 効果修飾と交互作用の違いについては、VanderWeele（2009）参照のこと。
- 7) ただし、過去の政策の有効性を明らかにした上で、現在の政策の効果を推計することも考えられ、明確に切り分けられるわけではない。
- 8) これは、「政策評価に関する基本方針」（総務省，2015）で「初めから高度かつ厳格な手法の適用を画一的に行うより、簡易な手法であっても、その有用性が認められているものがあれば当該手法を適用し、政策評価の実施の過程を通じ知見を蓄積して手法の高度化を進めていくこと」という文章からも伺うことができる。
- 9) そもそもアメリカの GAO のプログラム評価は、立法府であり、日本の行政のプログラム評価と単純比較できない点には注意が必要である。
- 10) ただし、山谷（1997）ではプログラム評価と業績測定という言葉を用いてはいない。また、高等教育の評価で同様のことを指摘した論文として羽田（2006）参照のこと。
- 11) 政策評価と行政評価の関係で言えば、政策効果を基礎とするものが行政評価の中の政策評価と位置付けられ（南島，2020，81頁）、どちらも業績測定と親和的である。しかし、EBPM が行政の有効性ではなく、政策効果の文脈でエビデンスピラミッドの使用が推進されているため、本稿では取り上げない。
- 12) エビデンスピラミッドの手法（実験研究、観察研究両方含む）で得られる集団はそれぞれ異なり、置かれている仮定も違うため、それぞれの比較は容易にできないという Ellicot et al.（2020）の指摘は、有効性を基準に比較検討することの困難性を示唆する。PPBS が比較困難を理由に頓挫したことを踏まえると有効性の比較による合理的な意思決定を目指す EBPM も同様に頓挫する可能性を含んでいる。
- 13) EBPM も一種のブームであり、動きが早く目の前の現状を捉えきることが困難であるという課題が本稿にはある。

## 【参考文献】

赤林英夫（2017）「幼児教育の無償化はマジックか？－日本の現状から出発した緻密な議論を」『SYNODOS』（<https://synodos.jp/education/19911>）<2020年8月25日アクセス>。

- 井伊雅子・五十嵐中・中村良太（2019）『新医療経済学』日本評論社。
- 伊藤公一朗（2017）『データ分析の力 因果関係に迫る思考法』光文社新書。
- 今井康雄（2015）「教育にとってエビデンスとは何か」『教育学研究』82(2), 188-201頁。
- 岩崎学（2015）『統計的因果推論』朝倉書店。
- 大久保翔貴（2019）「因果推論の道具箱」『理論と方法』65, 20-34頁。
- 大槻大輔（2018）「英国、米国における EBPM の動向」『統計』69(7), 28-39頁。
- 小方直幸（2014）「専門知と政策形成」『大学論集』第47集, 73-88頁。
- 岡本雪乃（2018）「政策科学の来歴と政策過程論」佐藤満編『政策過程論』20-36頁。
- 小野達也（2013）「政策評価と実績測定」『日本評価研究』13(2), 21-36頁。
- 小野達也（2018）「エビデンスに基づく政策評価」『統計』69(7), 8-13頁。
- 小野方資（2015）「教育政策形成における「エビデンス」と政治」『教育刷新委員会』82(2), 55-66頁。
- 金子元久（1993）「高等教育制度・政策の研究」『大学論集』第22集, 187-208頁。
- 加藤芳太郎（聞き手：納富一郎）（2008）『予算論研究の歩み』敬文堂。
- 加納寛之・林岳彦・岸本充生（2020）「EBPM から EIPM へ」『環境経済・政策研究』13(1), 77-81頁。
- 桐村豪文（2020）「米国連邦教育政策におけるエビデンス要求の現在」『地域連携教育研究』5, 14-26頁。
- 窪田好男（2005）『日本型政策評価としての事務事業評価』日本評論社。
- 黒木学（2017）『構造的因果モデルの基礎』共立出版。
- 小林信一（2018）「科学的根拠に基づく政策」『科学』88(11), 1149-1156頁。
- 佐藤仁（2019）「教育政策においてエビデンスを「つかう」とはどういうことか」杉田浩崇・熊井翔太編『「エビデンスに基づく教育」の闕を探る』42-68頁。
- 末吉梯次・片岡徳雄（1957）「学習指導の実験研究」『教育社会学研究』11, 1-14頁。
- 善教将大・坂本治也（2020）「サーベイ実験の再現可能性と外的妥当性—オンラインフィールド実験による追検証」『ノモス』46, 1-15頁。
- 妹尾渉（2018）「教育経済学におけるエビデンス」『第70回大会日本教育社会学会発表要旨集録』462-463頁。
- 総務省（2015）『政策評価に関する基本方針』。
- 総務省（2018）『EBPM（エビデンスに基づく政策立案）に関する有識者との意見交換会報告』。
- 田中隆一（2020）「教育政策における EBPM」大橋弘編『EBPM の経済学』61-87頁。
- 田辺智子（2006）「エビデンスに基づく教育」『日本評価研究』6(1), 31-41頁。
- 田辺智子（2014）「業績測定を補完するプログラム評価の役割」『日本評価研究』14(2), 1-16頁。
- 塚原修一（2003）「教育政策と評価のダイナミズム」『教育社会学研究』72, 5-20頁。
- 津川友介（2020）『世界一わかりやすい「医療政策」の教科書』医学書院。
- 内閣府（2017）『第1回 EBPM 推進委員会議事要旨』。

- 中泉拓也 (2019) 「英国の EBPM (Evidence Based Policy Making) の動向と我が国への EBPM 導入の課題」『経済経営研究所年報』41, 3-9頁。
- 中尾走 (2019) 「外部資金獲得の効果は？」『広島大学大学院教育学研究科紀要 第三部 (教育人間科学関連領域)』68, 151-159頁。
- 中澤渉 (2016) 「教育政策とエビデンス」志水宏吉編『岩波講座 教育 変革への展望2 社会のなかの教育』73-101頁。
- 中村高康 (2007) 「高等教育研究と社会学的想像力」『高等教育研究』10, 97-109頁。
- 南島和久 (2018) 「評価」石橋章市朗・佐野亘・土山希美枝・南島和久『公共政策学』185-209頁。
- 南島和久 (2019) 「政策評価の総論—政策評価と EBPM の現在—」『総務省政策評価に関する統一研修発表資料』([https://www.soumu.go.jp/main\\_content/000607589.pdf](https://www.soumu.go.jp/main_content/000607589.pdf)) <2020年8月28日アクセス>。
- 南島和久 (2020) 『政策評価の行政学』晃洋書房。
- 西尾勝 (1976) 「効率と能率」辻清明・編集代表『行政学講座3 行政の過程』167-219頁。
- 日本教育学会編 (2015) 『教育学研究』82(2)。
- 日本評価学会編 (2006) 『日本評価研究』6(1)。
- 日本評価学会編 (2016) 『日本評価研究』17(1)。
- 羽田貴史 (2006) 「大学評価, 神話と現実」『大学評価研究』5, 6-13頁。
- 林岳彦 (2019a) 「因果推論駅の奥へ」『広島大学高等教育研究開発センター 2019年度 第4回公開研究会発表資料』。
- 林岳彦 (2019b) 「EBPM, “E” から見るか? “PM” から見るか?」『研究者/研究所として“EBPM”とどう関わるとよいのか? 国立環境研究所研究集会発表資料』。
- 星野崇宏・岡田謙介 (2006) 「傾向スコアを用いた共変量調整による因果効果の推定と臨床医学・疫学・薬学・公衆衛生分野での応用について」『保健医療科学』55(3), 230-243頁。
- 星野崇宏 (2009) 『調査観察データの統計科学』岩波書店。
- 宮川公男 (1994) 『政策科学の基礎』東洋経済新報社。
- 宮川公男 (2002) 『政策科学入門』東洋経済新報社。
- 山谷清 (1997) 『政策評価の理論とその展開』晃洋書房。
- 山谷清 (2006) 『政策評価の実践とその課題』萌書房。
- 山本清 (2016) 「証拠に基づく大学政策の推進の実態と課題」『大学経営政策研究』6, 1-16頁。
- 山本清 (2018) 「『証拠に基づく政策立案』の課題と展望」『大学経営政策研究』8, 217-230頁
- 山森光陽・荻原康仁 (2016) 「学級規模の大小と学年学級数の多少による児童の過去と後続の学力との関係の違い」『教育心理学研究』64, 555-568頁。
- 横尾英史 (2019) 「経済学者による RCT は倫理的に問題か?」『RIETI Discussion Paper Series 19-J-004』独立行政法人経済産業研究所。
- Bosworth, R., & Caliendo, F. (2007). Educational production and teacher preferences. *Economics of Education Review*, 26, 487-500.
- Botterill, L.C. & Hindmoor, A. (2012). Turtles all the way down: bounded rationality in an evidence-based

- age, *Policy Studies* 33(5), 367–379.
- Breckon, J., & Gough, D. (2019). Using evidence in the UK. Boaz, A., Davis, H., Fraser, A., and Nutley, S., eds., *What Works Now? : Evidence-informed policy and practice*, 285-319.
- Bridge, D., Smeyers, P., & Smith, R. (2009). *Evidence-Based Policy : What Evidence? What Basis? Whose Policy?*. Philosophy of Education Society of Great Britain. (=2013拓植雅義・葉養正明・加治佐哲也訳『エビデンスに基づく教育政策』勁草書房).
- Cairney, P. (2016). *The Politics of Evidence-Based Policy Making*. Palgrave Pivot.
- Deaton, A. (2020). Randomization in the Tropics Revisited: A Theme and Eleven Variations. *NBER WORKING PAPER SERIES*.
- Ellicott, C. M., Hagan, E., Gottlieb, I. M., Tan, M. L., Vlahov, D., Adlera, N. E., Glymour, M. M. (2020). Alternative causal inference methods in population health research- Evaluating tradeoffs and triangulating evidence. *SSM-population Health*, 10, 1-9.
- Hatry, H. P. (1999). *PERFORMANCE MEASUREMENT : Getting Result*. The Urban Institute. (=2013上野宏・上野真城子訳『政策評価入門』東洋経済新報社).
- Heckman, J.& Masterov, V. D. (2007). The Productivity Argument for Investing in Young Children. *Review of Agricultural Economics*, 29(3), 446-493.
- Hernan, M. A. & Vanderweele, T. J. (2011). Compound Treatments and Transportability of Causal Inference. *Epidemiology*, 22(3), 368-377.
- Igarashi A., & Ono Y. (2020). The Effects of Negative and Positive Information on Attitudes toward Immigration. *RIETI Discussion Paper Series*, 20-E-023.
- King G., & Nielsen R. (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*, 27(4), 435-454.
- Parkhurst, J. (2017). *The Politics of Evidence : From Evidence Base Policy to the Good Government of Evidence*. Routledge.
- Rossi, P.H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation : A System Approach SEVENTH EDITION*. Sage publications, INC. (=2005大島巖・平岡公一・森俊夫・元永拓郎『プログラム評価の理論と方法』日本評論社).
- Sara Lodi, et al. (2019). Effect Estimates in Randomized Trials and Observational Studies — Comparing Apples With Apples. *American Journal of Epidemiology*, 188(8), 569-577.
- VanderWeel, T. J. (2009). On the distinction between interaction and effect modification. *Epidemiology*, 20(6), 863-871.
- Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review*, 39(5), 426-431.

## **Trends and Issues of EBPM: Looking at the Impact on the Higher Education Sector**

Ran NAKAO \*

In this study, we review the trends and issues of EBPM. Based on the argument that previous discussion on this topic has not been in harmony, we reviewed the evidence from two perspectives and EBPM from three. The findings of the study are as follows. In Japan, higher education is more easily influenced by policy than primary and secondary education. In addition, higher education research has developed by being closely related to policy. Therefore, EBPM is a particularly important topic in the field of higher education, and more policy-focused research is expected to be accumulated in the future. Finally, this paper presents issues that need to be addressed in higher education research.

---

\* Doctoral Student, Graduate School of Education, Hiroshima University / JSPS Research Fellow