

「学習成果」論再考 —EBPMと因果推論を手がかりに—

中 尾 走

日本学術振興会特別研究員・広島大学大学院

樊 怡 舟

広島大学大学院

村 澤 昌 崇

広島大学高等教育研究開発センター
副センター長・准教授

西 谷 元

広島大学教授

松 宮 慎 治

広島大学大学院・神戸学院大学職員

【キーワード】因果推論、EBPM、反実仮想、自然実験、DAG

はじめに

本稿の目的は、近年その重要性が高まっているEBPM (Evidence-Based Policy Making: 証拠に基づく政策形成) とその根幹をなす因果推論の発想を手がかりに、①大学における「学習成果」に関する測定と分析に関する論理的な整理を行い、②学習成果の確認手段としてのEBPM的因果推論の重要性を提示する。その上で、③EBPM的因果推論を応用した先駆的事例の紹介を行い、④学習成果に関わる課題と展望を提起する。

I. 問題設定

大学における学習とその成果に関する近年の急激な関心の高まりは、周知の通り論を俟たない。そのテーマ自体は特段新しいものではないが、近年ではたとえ

ば大学IR、4つのポリシー、海外留学などと関連づけられたり、政府が全国学生調査へと着手したりしたことも重なり、ホットイシューとなっている。

この学習成果にとって重要な要素は「指標・調査・測定・評価」である。これら要素が殊更重要視された契機は、管見の限りでは90年代の「大学評価」必要論以降だろう。評価に際しては、その材料が必要となるので、指標の選定、調査・測定を経てようやく評価にこぎ着けられる。「質保証」議論も学習成果の議論も同様であり、「何」が質・成果なのかという問いのもとに、指標の選定に収斂しがちであった。その好例が『学習成果の把握と学習成果の評価についての具体的方策に関する調査研究』（文部科学省2014）であり、主たる関心が指標の種類に集中している。実はこうした一連の関心は、学術的には指標選択の妥当性の議論に留まっており、指標の精度の高さ、即ち指標の信頼性や、指標の使われ方・解釈の仕方⁽¹⁾、研究設計の妥当性に関する議論が不十分であった。

議論を深めるためには、「学習成果」に関する構造的

理解が必要である。以下簡単な数式で整理してみよう。

$$Y \text{ (成果の測定指標)} = \beta \cdot A \text{ (大学の作為} \equiv \text{処置介入)} + \gamma \cdot Z \text{ (A以外のYに与える要因群)} + e \text{ (AとZでは説明されないYの残余)} \quad \textcircled{1}$$

学習成果の議論は、指標すなわち①式の「Y」に関する議論となる。往々にして高等教育の政策や実践現場でこのYに関心が集中する。

しかしながら、Yが(正課の内外を問わず)変化・向上したとしても、それがAに相当する大学の作為によるものでなければ、大学での学習成果や、大学として質を保証したと見なすことはできない。言い換えれば、「大学に進学しなければ」「大学の授業を受けなければ」当該学生の成長や変化は見られなかったことを以て初めて、大学の文脈における「学習成果」と言っているのではないか、ということである。つまり、大学における成果を見極める上であらためて重視すべきは、大学固有の成果と見なしうるA、すなわち処置・介入の効果を厳密に検証する方法にある。

もちろん学習成果の研究では、Aに関する測定を行って回帰系の分析を適用する等の試みは数多ある。ただし実は、半ば慣習的に行われている“因果分析”では、厳密な因果推論(causal inference)の枠組みを踏襲していない場合が多く、因果効果を見いだすことが難しいと言われている。他にも理論的検討が不十分なまま慣習的に行われている分析や解釈があり、こうした“慣習化された因果分析”では、大学における学習成果の知見をミスリードしてしまう恐れもある⁽²⁾。

これらの点について、実はすでに他の領域では新たな取り組みが既に始まっている。いわゆるEBM(Evidence-Based Medicine: 証拠に基づく医療)およびEBPM(Evidence-Based Policy Making: 証拠に基づく政策形成)の名の下に、医療行為や政策の有無をより厳密に分析するための、因果推論の積極的な導入の事例である。残念ながら、高等教育領域におけるEBPMの議論や応用は限られている。

そこで以下本稿では、まず1) そもそもEBPMとはなにか、を論じる。次に2) EBPMが基礎とする因果推論による学習成果・効果検証の必要性を指摘した上

で、3) 因果推論による学習成果分析の先駆的事例を紹介する。

II. EBPMとは-その要は因果推論-

そもそもEBPMとは何か。内閣府(2017)によれば、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くということ」である。EBPMは、限られた資源の効果的・効率的な利用が可能(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。

さて、この説明だけを読むと、前世紀末からにわかに議論された政策評価(山谷2009)と大きく違わない。では何が異なるのか。近年この差異を整理した中尾(2021)、内山(2020)らによれば、政策評価とEBPMの違いは、エビデンスの形成に因果推論が推奨されている点、因果推論の方法に応じてエビデンスの質に格付けを行う(エビデンス・ピラミッド)点だとされる。

この因果推論の核となる考え方は、「反実仮想」(counterfactual)である。先の大学進学/非進学の場合だと、「進学者がもし非進学だった場合」「非進学者がもし進学者だった場合」をも想定し、それらの結果(潜在的結果: potential outcome)も得て比較を行うことで、初めて正確な因果が推定できる。しかしあくまでifであるがゆえにデータは得られない。その為因果推論は、不可能なものを推論する(大塚2020)とも言われる。そこで次善策として、ランダム化比較実験(Randomized Controlled Trial)のような実験的手法や、観察データから、反実仮想を想定した結果と理論的に一致するように導く統計的因果推論⁽³⁾などが編み出された。

この因果推論の発想は学習成果を考察する上でも有益である。以下では、テキストや先行研究ではあまり触れられないが、基本的且つ重要な論点に触れる⁽⁴⁾。第一に重要なのは「何の因果効果を明らかにしたいか」である。学生の学力を問いたい場合、その学力を変化させた原因が何かを明確にすることである。それは授業なのか、大学に進学したこと自体なのか、改革を施

した結果なのか—この観点は、冒頭でも触れたように、高等教育の文脈では指標（Y）に拘泥するあまり、見落とされがちであった。

第二に、「反実仮想」の中身を具体的に想定することである。例えば、国立大学の進学の効果を検討する場合、「もし国立大学に進学しなかったら」が「反実仮想」となるが、実はこのifは、複数想定され（就職、浪人、公立・私立への進学、専業主婦etc…）一意に定まらない。「反実仮想」の中身を具体的に想定したうえで、それとの比較に特化しない限り、得られた結果が解釈しにくく、現実世界への示唆もできないと言われている（Hernán and Taubman 2008；Rehkopf et al. 2016）。

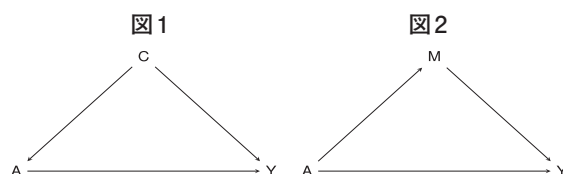
第三に、分析結果の汎用可能性、即ち外的妥当性（external validity）を考慮することである。EBPMでは、理念上は、因果推論に基づくエビデンスを構築する以上に、そのエビデンスを政策立案に繋げるところまでがセットである。その際、エビデンスとして得られた因果効果が、そのまま政策として実施された際に同様の効果が得られるかを改めて検討する必要がある⁽⁵⁾。実は、特にRCTを始めとした実験的因果推論は、実験デザイン上サンプルが限定されることが多く、その内的妥当性の高さと裏腹に、外的妥当性が低い場合がある⁽⁶⁾。

以上、本節ではEBPMの概要とEBPMで核となる因果推論を概説した。次節からは、学習成果と因果推論について考えていくこととする。

Ⅲ. 因果推論に基づく学習成果の評価・効果検証の必要性

高等教育において学習成果を論じる時によく用いられるのは、大規模な学生調査による分析であり、学生の状態や変化を基礎的な集計を通じて「記述」し、重回帰分析や共分散構造分析などの因果的な分析手法を適用した「説明」がよく試みられる⁽⁷⁾。

実はこのような試みは、因果推論の議論に位置づけると再考の余地がある。従来「説明」と称されてきたものは、厳密には狭義の「説明」と「因果」に分けられる。「因果」の効果推定は、計量経済学などでよく用いられる不偏性、一致性、有効性の検討であり、係数の量的



図注：①式のZは、図では因果関係上CとMに置き換えた

な評価となる。「説明」の文脈で重回帰分析を用いる場合は、決定係数などのモデルの予測性能を評価することに主眼がある。実は「因果」の評価には「説明」を必ずしも要せず、逆に「説明」の評価には、投入されたすべての変数の「因果」効果の評価は慎重ななければいけない。

ここでは、冒頭で示した①式を、因果推論の枠組みで理解し直すために、有向非巡回グラフ（Directed Acyclic Graph、以下DAG）を用いて書き直してみよう。①式で示した重回帰分析を、最も単純な被説明変数1つ、説明変数2つ（①式のZは、CとMに置き換える）で構成されるDAGで描くと図1、図2のような2つのケースが想定できる⁽⁸⁾。たとえば処置Aの純粋な因果効果を診る場合、交絡変数C⁽⁹⁾を統制変数として投入する必要がある（図1）。我々はCの効果も解釈しがちだが、図1からもわかるように、Cの係数は、Aを経由する効果が除かれており、且つ、CとYとの間の交絡要因が検討されていないことから、因果効果として解釈すべきではない（Keele et al. 2020）。図2のような因果が成立している場合も、Mを統制変数として投入してしまいがちであるが、図2の場合、Mは「中間変数」であり、Aの総合効果を制御してしまうため、過剰制御（over control）となる（狩野2002、岩崎2020）。

一方、狭義の「説明」が目的であれば、目的はYの予測（説明と同義）なので、予測を高めうる変数を可能な限り投入すればよく、このように分析者が構築したモデルのYに対する説明力の検討となる。ただし、この目的で用いた重回帰分析の説明変数の係数の値は、上述のような厳密な因果効果を想定しておらず、信頼できない。因果効果を解釈したいのであれば、図1、2のようなDAGを描いて適切に交絡要因を統制変数として投入し⁽¹⁰⁾、因果推論を目的とした分析をす

るべきである。

このように、これまでの学習成果の先行研究は、データを用いて狭義の「説明」をしてきたと理解できる。一方で、EBPMは反実仮定の発想に基づき因果推論によるエビデンスを求めている。この違いを理解することが学習成果のエビデンスを作る第一歩となる⁽¹¹⁾。

また、これまでの高等教育の諸改革は、その効果検証が不十分である。加えて学習成果は、「学生の学習や大学の教育的介入が、本当に成果をもたらしたのか」と問い直す必要がある。それらの検証法として因果推論が必要なのである。故に、これまでは説明的な問いを明らかにすることが主であったが、今後は因果的な問いを明らかにすることも必要となってくる。

Ⅳ. 分析事例

以上、EBPMと因果推論の必要性を提起してきたが、実際には、データの整備や利活用、分析方法の厳密性、大学「業務」に対するインセンティブの低さ等の諸々の壁があり、取り組みは不十分である。ただし、近年は大学IR (Institutional Research) の必要性が喧伝され、これを追い風にして学内の教学関連データの分析が加速化することも予想される。また、先駆的研究も散見されるようになった(田尻2018、武知ほか2020、川田・西谷2017等)。

(1) 短期留学効果の推定

ここで、川田・西谷(2017)の事例を概観してみよう。本研究は、広島大学の短期留学プログラム(以下START)がTOEIC®の得点に及ぼす効果を検証して

いる。具体的には、傾向スコア分析によりSTART参加者と非参加者を擬似的にランダム割付した状態を作り上げ、START参加前後の英語得点の変化をDID (Difference in Difference) 分析により検討している。分析の結果(表1)、STARTにはTOEIC®スコアを上昇させる効果があったと報告されている。

(2) 自然実験

つづいて、川田・西谷(2017)の派生として、一大学を対象に行った「自然実験」による因果効果の検討例(中尾ほか2021)を紹介する。

「自然実験」とは、偶然に実験的な状況が生じた環境を利用する(伊藤2017)手法である。中尾ほか(2021)では、A大学のXプログラムの効果検証を自然実験により行っている。Xプログラムは、留学参加者がTOEIC®の点数や面接など複数の評価項目に基づいて選抜され、毎年8月～9月の前期と2月～3月の後期の二回、各々約百名ずつの学生が約2週間の海外留学を行うものである。2019年度も百名程度の学生が、前期と後期に留学予定であったが、COVID-19禍により後期プログラムのみが中止となった。よって、参加予定の学生の中に、参加者(前期参加学生)と不参加者(後期参加予定学生)が混在することとなった。

これにより、前期参加学生と後期参加予定学生のTOEIC®得点の差分をとれば、留学プログラム参加の純粋効果に近いものを得られる。なぜなら、参加者も不参加者も同じ選抜による学生なので、学習意欲や語学力などの交絡要因が統制されていることが期待でき、COVID-19禍で外生的に参加・不参加が決まった

表1 STARTのTOEIC®に与える効果

VARIABLES	Difference in Means	DID	Abadie DID	
START	124.8*** (15.04)	44.28*** (12.36)	35.81*** (11.53)	38.97*** (10.41)
Control	No	No	学生の所属	学生の所属 + START前のTOEICスコア
Observations	2,242	2,242	1,681	1,666

※川田・西谷(2017)より引用

表2 XプログラムのTOEIC®に与える効果

	合計			リスニング			ライティング		
	Estimate	p. value		Estimate	p. value		Estimate	p. value	
Intercept	559.4 (10.9)	<.001	***	298.04 (5.76)	<.001	***	261.4 (6.24)	<.001	***
time	58.4 (19.1)	0.002	***	25.62 (10.07)	0.011	**	32.83 (10.92)	0.003	***
treat	16 (15)	0.284		7.49 (7.87)	0.342		8.56 (8.54)	0.317	
time:treat	-14.8 (24.3)	0.542		-6.85 (12.78)	0.592		-7.98 (13.86)	0.565	
R ²	0.049			0.035			0.048		
Ads. R ²	0.042			0.027			0.041		
n	346			346			346		

*** <.001; **<.01; * <.05

※中尾ほか(2021)より転載

ことで、それら二つの集団に学生がランダム割付されたに等しくなるからである。実際の分析結果では(表2)、川田・西谷(2017)の結果に反し、Xプログラム(表2のtreatに相当。time:treatは受験時期とXプログラムの交互作用)によるTOEIC®の点数の上昇が見出せないという結論に至っている。

(3) 新しいデータの発掘

学習成果を学内外で検討する場合、必ずしも測定したいデータが手元に無い場合もある。そのような場合、既存のデータの発掘や再定義・操作を通じて、代理となる指標を構築する必要もある。中尾・樊(2021)の研究では、このような意欲的な取り組みも行っている。

因果推論の内的妥当性を保証するための前提条件の内、処置群と統制群の間に相互干渉が起きていないというSUTVAの第一仮定がある(岩崎2015)。留学とTOEIC®得点を例に説明すると、学生AのTOEIC®得点は、その友人Bの留学の有無に影響されない、となる⁽¹²⁾。しかし、友人間での相互干渉が無いという仮定は非現実的である。友人同士であれば、互いに切磋琢磨する(もちろんその逆もあるが・・・)などの相互作用のあることが容易に想像できるからである。そうするとSUTVAの第一仮定が未充足となってしまう。この場合、先述の自然実験下だとしても、留学の処置効果が、友人間の相互作用から影響を受け、真の

値よりも低く推定される可能性もある。

そこで、学生の交友関係の統制が必要となるが、データが無い場合は既存の情報の活用により代理指標の作成で対応することになる。中尾・樊(2021)では、各学生の履修状況の情報から、学生間の距離を算出して代理指標とした。授業履修状況が似通った学生同士は、相互作用を受けるとして統制を試みている。

(4) 信頼性の高い指標の開発

余力があれば、専門家により開発された尺度の運用も、正確な分析に貢献する。因果分析には、交絡要因以外にも内的妥当性を脅かすものがあり、その一つが測定誤差(measurement error)である⁽¹³⁾。特に社会科学では抽象的な構成概念⁽¹⁴⁾を扱うことが多いため、その測定には常に誤差が付随する。心理学では信頼性の高い尺度が開発されているが、川田・西谷(2017)や中尾ほか(2021)の研究の一環で扱うBEVI⁽¹⁵⁾は、社会的望ましさバイアス(Social desirability bias)や努力の最小限化(Satisfice)等の統制機能が導入され⁽¹⁶⁾、学習成果や学生調査でよく使われる尺度や指標よりも測定誤差を抑える設計がされており、因果推論に用いる際に有利である。

おわりに

以上見てきたように、学習成果への関心が高まる

中、この課題に関する本稿の問題提起は、①大学の成果測定は、指標以上に大学固有の効果を見いだす調査分析デザインが重要である、②その鍵を握るのはEBPMで推奨される因果推論にある、③これまでの学習成果測定で利用された学生調査の分析は、厳密には「記述」「説明」「因果」等を丁寧に切り分け、文脈に応じた使い分けが必要である、④因果推論を応用した分析事例から、自然実験、学内既存データの再定義による利活用、専門家による精度の高い尺度の積極的利用が推奨される、である。

そして最も重要なのは、以上の取り組みには、個々の大学の文脈や経験知、専門知識の関与が欠かせない(林2019)という点である。あくまでEBPMや因果推論は、学習成果に関する従来の議論や、慣習化された分析手順を、改めて内省するためのきっかけを与えてくれるに過ぎない。猫も杓子も⁽¹⁷⁾EBPM・因果推論に飛びつく必要は無いし、機械的な乱用には、本稿で指摘したような落とし穴があるので注意が必要である。

いずれにせよ、このような野心的アプローチを投げかけることにより、大学の学習成果への取り組みに、科学のエッセンスがより付加された議論が勃興することを期待したい。

【注】

- (1) 指標や尺度自体の妥当性や信頼性の問題は、心理学を中心に長らく議論されてきた古典的・伝統的且つ継続的課題である。しかし高等教育においては、心理学の応用を装って根本的な禁忌を見過ごしている(例えばダブルバーレル問題)場合があり、実践や改革、政策上のミスリードにつながり兼ねない。重要な論点ではあるが、この問題は紙幅を超えるので本稿では扱わない。
- (2) 高等教育研究における計量分析手法の課題に関しては、併せて立石・村澤(2017)、立石ほか(2019)も参照のこと。
- (3) 因果推論の手法については、伊藤(2017)、森田(2014)、中室・津川(2017)、安井(2020)等を参照のこと。
- (4) 本稿における統計の専門用語などは、大久保(2019)、中尾(2021)に準拠している。
- (5) 関連する議論としては、ルーカス批判が挙げられる。EBPMとルーカス批判については、山名(2017)、立石ほか(2018)、竹田(2020)参照のこと。
- (6) ただしRCTであっても外的妥当性が満たされる場合がある。詳細はHernan and VanderWeele(2011)およびその日本語解説である中尾(2021)参照のこと。また、各種統計的因果推論の場合、元データの調整(中尾2019)、部分的に切り取る等の操作によって、母集団とのずれが生じているという理由で外的妥当性が満たされない場合もある。
- (7) 社会調査のテキストでは、前者を「記述」(事実はどうなのか: howの問い)、後者を「説明」(なぜその事実が生じているか: whyの問い)として区分している(高根1979; 盛山2004; 平沢2021)。
- (8) なお、ここでは簡略化のため、未観測要因 e は省略した。
- (9) 処置・介入Aと結果変数Yの両方に影響する変数のこと。これを制御しないとAの効果が誤推定される可能性がある。
- (10) ただし、DAGのみでは因果効果の推定が不十分であるとの指摘もある(Heckman 2005、Imbens 2020)。
- (11) 例えば、医学・疫学の分野では、統計分析は、記述(Description)、関連の探索(Exploring associations)、予測(Prediction)、因果推論(Causal inference)、効果の比較(Diagnostic performance evaluation)に分類される(佐藤・芝2021)。これらを踏まえると、説明的な問いの検証方法として因果推論を挙げる平沢(2021)の解説は、誤解を与えかねない。
- (12) 他の例だと、潰瘍や外傷の治療のために薬が投与されるケースを想定されたい。この場合、薬が投与された処置群は、未投与の統制群に干渉することはないことが仮定されている。ただし岩崎(2015)では、このSUTVAの未充足の例として、インフルエンザワクチン接種の効果評価

を取り上げている。具体的には、とある人物Aにワクチンが投与されれば、Aの周囲の人間は、ワクチンを接種してなくても、少なくともAからの感染リスクが低くなるので、ワクチンの厳密な効果が不透明になるケースである。

- (13) 例えば、処置変数の測定誤差 (Jerrim and Micklewright 2014 ; Engzell and Jonsson 2015 など)、従属変数の測定誤差 (VanderWeele and Hernan 2012)、交絡変数の測定誤差 (Frost 1979、Ogburn and VanderWeele 2012、中尾ほか2021) など。
- (14) 構成概念とは、直接それ自体を測定することは困難であるが、人の実際の言動の観察を通じて、それら言動を規定しうのような、想像上仮定される事象を指す。たとえば学力、能力、性格などである。
- (15) 詳細は<https://jp.thebevi.com/>を参照のこと。
- (16) 社会的望ましきバイアスとは、回答者が自分をよく見せようとする傾向 (Groves et al. 2004 = 2011) である。努力の最小限化とは、必要最小限の行動をとることであり、例えばアンケートにおいて同一の項目を選択する行動などを指す (三浦・小林2015)。
- (17) 中村 (2019) は、近年の傾向スコア分析の乱用を「猫も杓子も傾向スコア」と警笛を鳴らしている。

【参考文献】

- 浅野哲・中村二郎, 2009, 『計量経済学 第2版』有斐閣。
- Degtiar Irina & Rose Sherri, 2021, “A Review of Generalizability and Transportability” (<https://arxiv.org/pdf/2102.11904.pdf>, 2021.6.18 最終アクセス)。
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. and Tourangeau, R., 2004, *Survey Methodology* : John Wiley & Sons, Inc. (=2011, 大隈昇監訳『調査法ハンドブック』朝倉書店)。
- 林岳彦, 2019, 「因果推論の奥へ」『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』。
- Heckman, J. T., 2005, “The Scientific model of Causality.” *Sociological Methodology*, 35 : 1-97.
- Heckman, J. & Masterov, V. D., 2007, The Productivity Argument for Investing in Young Children. *Review of Agricultural Economics*, 29(3), pp.446-493.
- Hernán M. & Taubman S., 2008, “Does obesity shorten life The importance of well-defined interventions to answer causal questions,” *International Journal of Obesity*, 32 (3) : S8-S14.
- Hernán, M. A. & Vanderweele, T. J., 2011, Compound Treatments and Transportability of Causal Inference. *Epidemiology*, 22 (3), pp.368-377.
- Hernán, M. and J. Robins, 2019, *Causal Inference*, Chapman & Hall/CRC.
- 平沢和司, 2021, 「社会調査のデザイン」轟亮・杉野勇・平沢和司編『入門・社会調査法 第4版』法律文化社, 47-60頁。
- 星野崇宏, 2009, 『調査観察データの統計科学』岩波書店。
- 池田功毅・平石界, 2016, 「心理学における再現可能性危機：問題の構造と解決策」『心理学評論』59, 3-14頁。
- Imbens, G. W., 2020, “Potential Outcome and Directed Acyclic Graph Approaches to Causality-Relevance for Empirical Practice in Economics.” *Journal of Economic Literature*, 58 (4), pp.1129-1179.
- 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。
- 岩崎学, 2015, 『統計的因果推論』朝倉書店。
- 岩崎学, 2020, 「統計的因果推論の視点による重回帰分析」『日本統計学会誌』50 (2) : 363-379頁。
- 狩野裕, 2002, 「構造方程式モデリング, 因果推論, そして非正規性」甘利俊一・狩野裕・佐藤俊哉・松山裕・竹内啓・石黒真木夫『多変量解析の展開 隠れた構造と因果を推理する』岩波書店, 65-129頁。
- 川田恵介・西谷 元, 2018, 「短期留学プログラムが語学力到達度に与える影響について：広島大学

- START プログラムの事例から」<http://goo.gl/5C1hRY> (2021.6.18最終アクセス).
- Keele, L., Stevenson, R. T. and Elwert, F., 2020, "The causal interpretation of estimated associations in regression models," *Political Science Research and Methods*, 8 : pp.1-13.
- 三浦麻子・小林哲郎, 2015, 「オンライン調査モニタのSatisficeに関する実験的研究」『社会心理学研究』31 (1), 1-12頁.
- 三浦俊彦, 2017, 『改訂版 可能世界の哲学: 「存在」と「自己」を考える』二見書房.
- 文部科学省, 2014, 『学習成果の把握についての具体的方策に関する調査研究』(https://www.mext.go.jp/a_menu/koutou/itaku/1347643.htm, 2021.6.18最終アクセス).
- 森田果, 2014, 『実証分析入門』日本評論社.
- 中村知繁, 2019, 「統計的因果推論とデータ解析」(https://speakerdeck.com/tomoshige_n/causal-inference-and-data-analysis : 2021.6.18最終アクセス).
- 中室牧子・津川友介, 2017, 『「原因」と「結果」の経済学』ダイヤモンド社.
- 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.
- 中尾走, 2021, 「EBPMの動向と課題—高等教育政策への影響を見据えて」『大学論集』, 54, 39-52頁.
- 中尾走・樊怡舟・村澤昌崇, 2021, 「交絡変数の測定誤差が合流点バイアスをもたらす影響について: 合流点バイアスと交絡統制のジレンマに対する考察」『広島大学高等教育研究開発センター編 ディスカッションペーパーシリーズ』15, 1-18頁.
- 中尾走・樊怡舟・西谷元・村澤昌崇, 2021, 「自然実験環境を用いた留学効果の推定」『広島大学高等教育研究開発センター編『Advancement of Higher Education Research : RIHE Monograph Series』3, 1-10頁.
- 中尾走・樊怡舟, 2021, 「留学効果の検証」『広島大学高等教育研究開発センター 2020年度 第20回公開研究会発表資料』(<https://rihe.hiroshima-u.ac.jp/wp/wp-content/uploads/2021/04/1daab2921925f8da647727b0b1defa16.pdf>, 2021.6.18最終アクセス).
- 内閣府, 2017, 『第1回EBPM推進委員会 (平成29年8月1日) 議事要旨』(<http://www.kantei.go.jp/jp/singi/it2/ebpm/dail/gijiyoushi.pdf>, 2021.6.18最終アクセス).
- Nguyen, T. Q., Ebnesajjad, C., Cole, S. R., & Stuart, E. A., 2017, "Sensitivity analysis for an unobserved moderator in RCT-to-target-population generalization of treatment effects," *The Annual of Applied Statistics*, 11 (1) : pp.225-247.
- Nguyen, T. Q., Ackerman, B., Schmid, I., Cole, S. R., Stuart, E. A., 2018, "Sensitivity analyses for effect modifiers not observed in the target population when generalizing treatment effects from a randomized controlled trial- Assumptions, models, effect scales, data scenarios, and implementation details," *PLoS One*, 13 (12) : pp.1-17.
- 奥井亮, 2019, 「重回帰モデルの推定と検定」西山慶彦・新谷元嗣・川口大司・奥井亮『計量経済学』有斐閣, 141-207頁.
- 大久保翔貴, 2019, 「因果推論の工具箱」『理論と方法』65, 20-34頁.
- 大塚淳, 2020, 『統計学を哲学する』名古屋大学出版会.
- Rehkopf, D. H., Glymour, M.M. and Osypuk, T. L., 2016, "The Consistency Assumption for Causal Inference in Social Epidemiology-When a Rose is Not a Rose," *Current Epidemiology Report*, 3 : pp.63-71.
- 佐藤俊太郎・芝孝一郎, 2021, 「DAGと回帰分析を用いた因果推論入門」日本疫学会プレセミナー2021 発表資料.
- 盛山和夫, 2004, 『社会調査法入門』有斐閣ブックス.
- 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 (議論の整理と課題等)』(https://www.soumu.go.jp/main_content/000579366.pdf, 2021.6.18最終アクセス).
- 田尻慎太郎, 2018, 「SAを事例とした任意参加型プログラムの因果効果推計」, 大学教育学会第40回大会発表資料, 2018-6-3.

- 高根正昭, 1979, 『創造の方法学』講談社.
- 武知薫子・酒匂康裕・服部圭子, 2020, 「EQ指標を用いた留学の効果検証方法：質的研究の探索補助としての量的研究の提案」『教養・外国語教育センター紀要・外国語編』11 (1), 1-19頁.
- 竹田聡, 2020, 「LATE vs. 構造推定」『地域政策学ジャーナル』9, 63-72頁.
- 立石慎治・村澤昌崇, 2017, 「計量分析の新展開：過去10年間の経験を振り返って」『高等教育研究』20, 135-156頁.
- 立石慎治・丸山和昭・速水幹也・松宮慎治・中尾走・村澤昌崇, 2019, 「高等教育研究と計量分析」『教育社会学研究』104, 29-55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.
- 内山融, 2019, 「EBPMの行政への実装に向けて」『RIETI EBPMシンポジウム エビデンスに基づく政策立案を進展させるために 発表資料』(<https://bit.ly/3zs2c9F>, 2021.6.18最終アクセス).
- VanderWeele, T. J., 2019, “Principles of confounder selection” *European Journal of Epidemiology*, 34, 211-219頁.
- Westreich, D., Edwards, J. K., Lesko, C. R., Cole, S. R., & Stuart, E. A., 2018, Target Validity and the Hierarchy of Study Designs. *American Journal of Epidemiology*, 188 (2), pp.438-443.
- 山名一史, 2017, 「「エビデンスに基づく政策形成」とは何か」『ファイナンス』53 (5), 76-84頁.
- 安井翔太, 2020, 『効果検証入門：正しい比較のための因果推論／計量経済学の基礎』技術評論社.
- 山谷清志, 2009, 「公共部門における『評価』：政策評価とNPM型業績測定」『日本評価研究』9 (3), 3-16頁.

Rethinking the “Learning Outcome”—Using EBPM and Causal Inference as Clues—

※1 Ran NAKAO

※2 Yizhou FAN

※3 Masataka MURASAWA

※4 Hajime NISHITANI

※5 Shinji MATSUMIYA

[Abstract]

The purpose of this paper is to reconsider the theory of “learning outcomes” based on the idea of evidence-based policy making (EBPM), which has been gaining importance in recent years, and the causal reasoning that forms the basis of EBPM.

Specifically, we will (1) Logical organization of measurement and analysis of “learning outcomes” in universities, and (2) the importance of EBPM-based causal reasoning as a means of confirming learning outcomes. Then, (3) we introduced pioneering case studies in which EBPM causal reasoning was applied, and (4) we raised issues and prospects related to learning outcomes.

In the conventional discussion of learning outcomes, we have not been able to properly discuss learning outcomes because the discussion often converges on the appropriateness of outcome measures and there are challenges in the customary causal analysis of related studies. As a prescription for this problem, this paper proposes a reconsideration of recent causal reasoning using the ideas of “anti-real virtual” and “DAG”. In addition, in the case study, as a more rigorous method of educational intervention effectiveness, natural experiments, utilization of existing data in the university, and reliable scales under the supervision of experts are presented.

※1 JSPS Research Fellow, Graduate School of Education, Hiroshima University

※2 Graduate School of Education, Hiroshima University

※3 Research Institute for Higher Education, Hiroshima University, Associate Professor/Deputy Director, Research Institute for Higher Education

※4 Graduate School of Humanities and Social Sciences, Professor/Deputy Executive Director(SGU), Hiroshima University

※5 Graduate School of Education, Hiroshima University/Kobe Gakuin University, Staff