

## 【論文】

# 米国連邦教育政策におけるエビデンス要求の変遷 —エビデンス要求の厳格化から緩和に向けて—

弘前大学 桐村 豪文

## 1. はじめに

近年、我が国を含め、教育をめぐる公の議論において「エビデンス」という言葉が当たり前のよう  
に用いられるようになった。「エビデンスに基づく教育 (Evidence-Based Education: EBE)」という概念  
の定着化が図られながら、教育には、その有効性—すなわちすべての子どもの成功—を保証するため、エ  
ビデンスという基盤を持つことが要求されるのである。

しかしながらその要求に関してはいまだ議論の余地が多分に残されている。一つは、そもそもここで  
要求される「エビデンス」とは何か、どのような条件を満たすものであれば「エビデンス」と呼称する  
ことが許されるのか、ということ。併せて、教育現場の日々の実践との整合性や親和性という観点から、  
どのようなものを「エビデンス」と見なすべきかという論点が残されている。事実、教育再生実行会議  
第八次提言素案をめぐる審議（7月1日第3分科会）の中で河野達信委員はこのように発言している。

「科学的手法がどのようなものか分かりにくいと感じました。数値だけに頼ったエビデンスに基づく手  
法であれば、教育現場においては数値に表れにくい様々な重要な点があります。そうしたものが見過ご  
されるのではない心配になる記述ではないかと感じます。」つまり、教育現場の流動的で複雑性をもつ  
た日々の営みに対して、簡潔で明快に表されるエビデンスとの整合性や親和性について、何らかの疑問  
が抱かれるのである。

重要なことは、こうした疑問の根底にはある価値的対立が伏在しているということである。今井康雄  
は、ハーグリーブズやハマスレイらの議論を踏まえながら次のように述べる。「教育の場でエビデンス  
を強調することは、専門家としての教師の自由を狭める結果をもたらすのではないか」<sup>1</sup>と。すなわち、  
有効性に関するアカウントビリティを前に、教師による教育の自由とエビデンスの介入との間に軋轢の  
可能性が常に横たわって存在しているのである。しかし今井は続けてこのように言う。「EBEの主張は  
教師の判断の自由空間を狭めるかもしれない。しかし、教師の自由裁量の余地を確保することが教育の  
最終目的なのではあるまい。」<sup>2</sup>真に確保すべきは、教育の「望ましさ」だという。

本稿は、今井と同じく教育の「望ましさ」を求め、教育とエビデンスとの友好的関係を構築すべく、  
基づくべき「エビデンス」がどのようなものであるべきかについて、米国の事例をもとに検討したい。  
米国では、連邦法を通して教育に「エビデンスに基づくこと」を明確に要求しているわけだが、しかし  
政策の変遷(表1参照)を追うと、その概念が少しずつ変化してきていることがわかる。本稿では、NCLB  
法とCSRプログラム、i3、ESSAを取り上げ、その変化を明らかにしたいと思う。この間の政権交代の  
影響、また2008年にWWCによってスタンダードが策定され、3度改訂されてきた中で、エビデンス  
概念が醸成されてきた影響が想像される。そうした点からもそれら4者間の差異を考察したいと思う。

表1 エビデンスにかかわる連邦教育政策年表

年	出来事
1997	包括的学校改善デモンストレーション (Comprehensive School Reform Demonstration: CSRD) プログラムが成立。
2002	どの子も置き去りにしない法 (No Child Left Behind Act of 2001: NCLB) が成立。 CSRД は包括的学校改善 (Comprehensive School Reform: CSR) に名称変更。その後 2005 年度をもって終了。
2002	教育科学改革法 (Education Sciences Reform Act of 2002) が制定され、教育科学研究機構 (Institute of Education Sciences ; IES) が創設され、その中に教育エビデンス情報センター (What Works Clearinghouse: WWC) が設置される。
2007	学校改善補助プログラム (School Improvement Grants: SIG) が開始。
2009	米国再生・再投資法 (American Recovery and Reinvestment Act: ARRA) のもと Investing in Innovation Fund (i3) と Race to the Top (RTTT) が開始。
2015	全ての生徒が成功する法 (Every Student Succeeds Acts: ESSA) が成立。 Investing in Innovation Fund (i3) は Education Innovation and Research に名称変更。SIG は廃止され Title I に統合。

【出典】 筆者作成

## 2. エビデンス要求の厳格化

### (1) NCLB 法における「エビデンス」の概念

表2 NCLB 法における「科学的基盤を持つ研究」についての定義

<p>「科学的基盤を持つ研究」の用語は、</p> <p>(A) 教育活動や教育プログラムに関する信頼できる確実な知識を獲得するために、厳格で体系的で客観的な手続きを適用することを含む研究を意味する。</p> <p>(B) そして以下のような研究を含む</p> <p>(i) 観察または実験を参考とする、体系的で経験的な方法を用いている</p> <p>(ii) 測定された仮説をテストし、そして導き出された一般的な結論が正しいことを証明するため適切で厳密なデータ分析を含んでいる</p> <p>(iii) 評価者や観察者、多様な測定や観察、そして同一または異なる者による研究、それらのいたるところで、信頼できかつ確実なデータを提供する、測定または観察の方法に依拠している</p> <p>(iv) ランダム化比較試験 (random assignment experiments) が好まれるなかで、個人、団体、プログラム、活動を異なる条件に割り当て、そして関心ある条件の効果を評価するために適切な統制を行う実験または準実験の研究デザインを用いて評価を受けている、または条件内または条件を横断して統制を行うものである範囲で他の研究デザインを用いて評価を受けている</p> <p>(v) 再現性を確保するため、または少なくともその実験調査で明らかになった結論をもとに体系的に研究を進める機会を提供するため、十分詳細かつ明快に実験調査を提示している</p> <p>(vi) ピアレビューを受けるジャーナルでアクセプトされている、または比較的厳格で客観的で科学的なレビューを通して独立した専門家委員会によって認められている</p>
--

【出典】 The No Child Left Behind Act, Section 9101.

共和党ブッシュ政権下の 2002 年に成立したどの子も置き去りにしない法 (No Child Left Behind Act of 2001 : NCLB 法) は、それまでにないほど強力なアカウンタビリティ政策をとったことで有名だが、またそれまでにないほど強力に“エビデンスに基づくこと”を要求するものでもあった。教育実践や実践される教育プログラムは有効性に関する研究的裏付けを得るべきことが要求され、そしてその裏付けを与える研究は厳格に科学的でなければならないとされた。そしてその「科学的」とする基準が、表 2 に示す、NCLB 法に規定する「科学的基盤を持つ研究 (scientifically based research)」の定義である。

この定義によれば、実験を用いることが「科学的」であるための「厳格で体系的で客観的な手続き」であり、中でも特にランダム化比較試験 (randomized controlled trials: RCT) が最良の方法論とされる。RCT とは、「エレガントなデザインを持つシンプルな研究方法である。ランダム割付けによって、2 つまたはそれ以上の数の群を作る。1 つまたはそれ以上の数の群に対してはある介入を行い (介入群)、他の群 (統制群) に対しては別の措置を行うか、何もしない。介入の効果は、両群に生じた結果を比較す

ることによって観察する。もし、ランダム化によってできた群のサイズが十分に大きなものであれば、両群の間に観察された差は、介入以外の、既知あるいは未知の変数の影響によるのではなく、介入の結果であることに確信が持てる。」<sup>3</sup>、そのような研究方法である。連邦教育省によれば「多くの個人を介入群または統制群に無作為に割り当てるプロセスによって、どんな特性（観察されるものも観察されないものも）においても、群の間には体系的な違いは存在しないということが保証され」、それゆえ「介入群と統制群との間で生じた結果の違いは、信頼性をもって介入に起因しており、他の要因には起因していないといえる」<sup>4</sup>という。こうして RCT は、無作為割当と統制によって様々な雑音、バイアスを取り除き、最も純度の高い因果関係を抽出することのできる最良の方法として称揚されるのである。

併せて、NCLB 法では準実験法 (quasi-experimental studies) も称揚の対象となる。準実験法とは、「比較グループ研究」とも呼ばれ、無作為とは別の手法を通じて選ばれた比較群と介入群とを対比させ、その両群の間に見られる結果を比較するものである。連邦教育省によれば、「よく調整された調査は、医学やその他分野と同様、教育の分野でも、介入の有効性について『ありうる (possible)』エビデンスを打ち立て、RCT によって確証を得るに値する仮説を構築するという点で、価値ある役割を果たしうると考える。」<sup>5</sup>という。ただし、その扱いには注意を要すると言ひ、「しかしそのエビデンスは、十分に調整された比較グループ研究でさえ、何が有効か否かの最後の裁定者 (final arbiter) として、あるいは効果の強さに対する信頼あるガイドとして活用されることに対しては、注意する必要がある」<sup>6</sup>。つまり RCT は、何が有効であるかを決定する「最後の裁定者」の身分を持つ一方、準実験法はその位には及ばず、許されるのは、最後の裁定者に対して有益な仮説を提示するという役割に限られるのである。

NCLB 法では、以上のとおり、実践やプログラムが基づくべきとする「科学的基盤を持つ研究」の概念は、実験法に限定された中で、RCT が最良の研究法として称揚され、そして準実験法が、その補助の役割としてその位置を許される、そのような特徴を有している。

このような特徴を有する「科学的基盤を持つ研究」の概念について、ではそれを、実践と接する場面において、いかにして具現化するすることができるか、その具体的手続きを次に探る必要がある。次項では、CSR プログラムに例に、連邦教育省がどのようにその具現化に取り組んだかを見ていこうと思う。

## (2) CSR プログラムにおける「エビデンス」の定義：「科学的基盤を持つ研究」の具現化の手続き

スラヴィン (Robert E. Slavin) によれば、連邦教育補助金が有効性にかかわるエビデンスと直接結び付けられたのは包括的学校改善デモンストレーション (Comprehensive School Reform Demonstration: CSR) プログラムが歴史上初めてであるという<sup>7</sup>。

CSR プログラムは、1997年に成立したもので (1998年7月1日施行)、その趣旨は、有効性が立証された方法を用いて学校改善を包括的に行い (表2②に示すように、教科指導、評価、クラス運営、専門職開発、保護者参加、学校運営のいずれかに限定するのではなく、それらを網羅した改革デザインを実施)、児童生徒の学力を向上させることにある。そしてこの目的のため、採択される学校に対して、「信頼性のある研究や効果的实践に基づき、そして多種多様な特徴を持つ学校において首尾よく再現される、児童生徒の学習、指導、学校経営において革新的な方略やその有効性が証明された方法」を用いること、つまり、「エビデンスに基づくこと」を要求したのである。

表3 Comprehensive School Reform の11つ構成要素

<p>① 科学的基盤を持つ研究 (scientifically based research) に基づき、有効性が立証された (proven) 方法と方略— 包括的学校改善プログラムは、科学的基盤をもつ研究や効果の実践例に基づき、そして学校において首尾よく再現されてきた、児童生徒の学習や教科指導、学校運営に関する、その有効性が立証された方略や方法を活用する。 第1構成要素は、学校が包括的改善プログラムを設計する上で、主要教科、特に数学とリーディングにおいて科学的基盤をもつ研究に根ざし、有効性が立証された方略や方法を採用する必要性を強調する。学業成績へ焦点化しつづけること、それを支える包括的プログラムを構築すること、授業において「何が有効に作用するか (what works) 」を重視すること、これらは成功する包括的デザインの重要な要素である。</p> <p>② 包括的デザイン— 効果的に学校を機能させるための包括的デザインは、教科指導、評価、クラス運営、専門職開発、保護者参加、学校運営を総合化するものである。学校のニーズアセスメントを通じて同定されたニーズに取り組むことによって、包括的デザインは、カリキュラム、テクノロジー、専門職開発を、スクールワイドな変革のための計画の中に組み入れる。このデザインの最終目標は、すべての児童生徒が州の定める、内容及び学業成績に関する挑戦的基準を満たすことができるようにすること。</p> <p>③ 専門職開発— プログラムでは、質の高い、そして継続性を持った、教職員のための専門職開発及び訓練を提供する。専門職開発は、その有効性が立証され、革新的で、費用効果が高く、容易に利用可能な方略を含むもので、そして教員が州のアセスメントと州の定める内容に関する挑戦的スタンダードを用いて、教科指導の実践と児童生徒の学業成績を改善させることが可能となるようにするものである。(略)</p> <p>④ 測定可能な目標及びベンチマーク— 包括的学校改善プログラムは、児童生徒の学業成績に関する測定可能な目標を含み、その目標を達成するためのベンチマークが設定される。連邦教育省は、地方学区がこれらの目標を、初等中等教育法 ESEA Section 1111(b)(2)の下、州の定める AYP の定義と連結させることを奨励する。</p> <p>⑤ 学校内の支援体制— 学校を通じて教員、校長、管理職その他の職員は、CSR 学校におけるプログラムを支援するものである。教職員は、この支援を、学校の包括的改善プログラムを理解し、喜んで応ずることにより、そして授業での指導の継続的改善を重点的に取り扱うことにより、そして専門職開発へ参加することにより、この支援があることを立証すること。</p> <p>⑥ 教員、校長への支援 (Added in 2001)— CSR プログラムは、教員、校長、管理職その他の職員に対し、共有されるリーダーシップと、改善努力に対する応答責任の広範な基礎を構築することによって、支援を提供するものである。CSR プログラムは、チームワークと功績を祝賀することを奨励する。これらを含む支援の方法は、学校の包括的デザインの役目である。</p> <p>⑦ 保護者と地域住民の参加— プログラムは、学校改善活動を計画、実施、評価する上で、保護者や地域住民の有意義な参加を認めるものである。この構成要素を扱う際には、学校は Title I, Part A (ESEA Section 1118) の保護者参加の要件と矛盾しない方略を構築する。学校は、保護者が参加する能力を形成することに特別な注意を払い、保護者が教科指導プログラムの話し合いの場に参加でき、自分の子どもの学業成績へ貢献できるような方法を設計するものである。</p> <p>⑧ 外部の技術支援及び援助— プログラムは、スクールワイドな学校改善における専門的知識や経験を有した主体 (高等教育機関を含む) から、質の高い外部支援及び援助を活用する。CSR の法律上においては、補助金を受けたプログラムが、成功を取ってきた過去の業績や、財政的安定性をもち、そして改善実施期間中、質の高い道具、学校職員のための専門職開発、現地支援を提供する能力をもった、資格を有する技術援助プロバイダーによって支援を受けることを州教育当局が保証するよう、要求している。</p> <p>⑨ 毎年の評価— 学校改革の実施や児童生徒の学力結果を評価する毎年の方策を立てること。(略)</p> <p>⑩ 資源の調和— 他の活用可能な資源 (連邦、州、地方、民間) がどのようにして、学校の改革をサポートし継続させるためのサービスを学校が調和させることができるか、その方法を同定する。(略)</p> <p>⑪ 学業成績を改善させる方略 (Added in 2001)— プログラムは、以下の要件のうち1つを満たさなければならない。</p> <ul style="list-style-type: none"> <li>■ プログラムは、科学的基盤を持つ研究 (scientifically based research) を通じて、参加する児童生徒の学業成績を改善させることについて、有意な研究結果をもつものであること。</li> <li>■ プログラムは、参加する子どもの学業成績を改善させることについて、有意な強力なエビデンス (strong evidence) をもつことが判明していること。</li> </ul>
---

【出典】 U.S. Department of Education, Comprehensive School Reform (CSR) Program Guidance, 2002

ただし本節で取り上げたいのは、CSR プログラムの後身に当たる包括的学校改善 (comprehensive school reform: CSR) プログラムである。CSR プログラムは、NCLB 法の中に再編され、科学的裏付け

をより強く、またより明確に要求するようになったものである。

CSR プログラムで補助を受けるには、表3に示す11の構成要素をすべて満たさなければならない。この中でエビデンスに関わるのは①と⑩である。①は教科指導の方略に適用される規定で（学業成績に関わる場合のみ学校運営の方略もこれに含まれる）、⑩は①～⑨の要素すべてに適用されるものである。

前身の CSRD プログラムでは①の教科指導のみエビデンスに基づくことが求められていたのが、NCLB 法のもとでは、専門職開発、学校内の支援体制、保護者と地域住民の参加など、すべての要素にエビデンスに基づくことが要求された。それほどに NCLB 法ではより強力にエビデンス志向が前面に出されたのである。

ただしその要求は、その当時の実情を踏まえると、そのままの形で徹頭徹尾具現化することには困難を孕んでいたことから、次のように配慮が施された。つまり、①～⑩の要素すべてに適用される⑩では、「科学的基盤を持つ研究」と「強力なエビデンス」という2つの基準が敷かれたのである。

教育省 CSR 課は2002年8月に、「科学的基盤を持つ研究と CSR プログラム (Scientifically based research and the Comprehensive School Reform Program)」という冊子を発行している。それによれば、「ある場合では、これらの基準を満たす研究がごく少ないか全くないことがある。(略) この場合、スクールリーダーは、プログラムを作るうえで、最も有用な経験的エビデンスとある程度の専門的判断に依拠する必要がある。経験的で、ランダム化された研究の質・量・利用可能性が高まるにつれ、学校は、プログラムの設計の決定と、『何が有効に作用するか』に関するエビデンスとをより強固に結びつけることができるようになるだろう」<sup>8</sup>という。つまり、法律で要求する「科学的基盤を持つ研究」という高い基準を満たす研究が、ある分野においては殆どないか全くない場合が存在する。そのため、「科学的基盤を持つ研究」が進展することを期待しながらも、緩和されたより低い基準が別に設定されるのである。それが「強力なエビデンス (strong evidence)」である。

では、その基準の差異はどこにあるのだろうか。冊子の中に示される「研究の質を判断するためのガイドライン (Guidelines for Judging the Quality of a Study)」<sup>9</sup>によれば、次のとおりである (図1)。

要約すると、ある実践またはプログラムの有効性を支持するうえで、示される基準のすべてを満たす「質の高い研究 (high-quality study)」が一定数あり、かつそれらの多くで学業成績に対してポジティブな効果が有意に見られた場合、その研究の集合を「科学的基盤を持つ研究」と呼び、他方で、基準のすべてではないが、そのほとんどを満たす「まあまあの研究 (reasonable quality study)」および「質の高い研究」が一定数あり、かつそれらの多くで学業成績に対してポジティブな効果が有意に見られた場合、その研究の集合を「強力なエビデンス」と呼ぶ。そして、教科指導による学業成績に対する効果が「科学的基盤を持つ研究」によって支持される時、その実践またはプログラムを「有効性が立証されたプログラム (proven program)」と呼び、他方、教科指導による学業成績に対する効果が「強力なエビデンス」によって支持される時、その実践またはプログラムを「見込みのあるプログラム (promising program)」と呼ぶのである。

CSR プログラムでは、有効性が立証されたプログラムを同定するため、その実践またはプログラムの有効性を立証しようとする研究の中から「質の高い研究」を選別することからまず始められる。そこで用いられる基準が表5に示す6つの条件である。「質の高い研究」と認められるためには、体系的で経験的な方法を用いていること、厳密なデータ分析を行っていること、信頼性のある妥当なデータの収集を

行っていること、強固な研究デザインをとっていること、再現性を見越して詳細な結果を示していること、研究の結果が精密な吟味を受けていることの6つの条件すべてを満たすことが求められる。そのうえで、例えば強固な研究デザインであるランダム化比較試験を用いていなかったり、再現性を考慮していなかったり、ある1つの条件を満たさない場合は「まあまあの研究」と位置づけられる。

そしてより高い信頼性をもって実践またはプログラムの有効性を支持するためには、「質の高い研究」が5つ以上なくてはならない、とされる(表4問い3)。そして「質の高い研究」が5つある中の4つで学業成績に対してポジティブな効果があること、そして4つある中の3つで有意でポジティブな効果が示されることが求められる。こうして四重のハードル(研究の質、量、効果、有意性)をすべてクリアするとき、それら研究の集合をCSRプログラムでは「科学的基盤を持つ研究」と呼ぶのである。

しかしCSRプログラムでは、先述のとおり、基準を満たす研究が、殆どないか全くない分野があることが考慮される結果、「まあまあの研究」を含んだうえで、研究の量、効果、有意性のハードルを同様にクリアする研究の集合を「強力なエビデンス」というのである。ただしそのエビデンスとしての質が「科学的基盤を持つ研究」に劣ることは、CSRプログラムのガイダンスにも述べられるところであり、「『強固なエビデンス』とは、『科学的基盤に基づく研究』よりも、より厳密さを欠く基準として定義される。『強固なエビデンス』として積極的効果を示す実践や方策、プログラムは、科学的基盤を持つ研究の定義で設定された基準を満たす、広範な研究的基盤を欠いている」<sup>10</sup>という。

表4 研究の質を判断するために用いられる3つの問い

問い1：検討対象の実践またはプログラムについて理論的基盤は存在するか？

理論的基盤に関する問い	理論的基盤の質を評価する
<input type="radio"/> この実践またはプログラムの背後にある考えは何か？ <input type="radio"/> その指導原理は何か？ <input type="radio"/> どのようにしてそれは有効に作用するのか？ <input type="radio"/> なぜそれは有効に作用するのか？	1) 実践またはプログラムの中心的考えと目的について、明快で、専門用語を用いない説明があるか？ 2) 実践またはプログラムの中核に位置する教科指導の活動について、明快な説明はあるか？ 3) 実践は、例えば子どもの発達や言語習得といった確立された学習理論とはっきり結びつけられているか？

問い2：この実践またはプログラムが首尾よく実施され、様々な状況の下でポジティブな結果を産出したというエビデンスは存在するか？

この実践を検討する学校と類似する文脈において成功してきたか？

実施と再現可能性に関する問い	実施と再現可能性の質を評価する
<input type="radio"/> この実践またはプログラムは広く用いられてきたか？ <input type="radio"/> この改革はどのような状況でうまく作用すると期待できるか？ <input type="radio"/> どのような状況であれば、最も効果的か？	1) 何校でこの実践またはプログラムが用いられたか？ 2) それら学校では、実践またはプログラムを十分に実施したか？ 3) どのような環境で実施されたか？ 4) 児童生徒の学業成績の改善は、様々な環境で、説得力をもって証明されたか？

問い3： この実践またはプログラムが、児童生徒の学業成績に対してポジティブな効果をもつというエビデンスは存在するか？

実践またはプログラムに関する問い	エビデンスの質を評価する	
	科学的基盤を持つ研究	科学的基盤を持つ研究に向けた開発
○実践かつ／またはプログラムが児童生徒の学業成績の改善を示す厳密な研究に基づくエビデンスは存在するか？	実践またはプログラムについて 1) 実践またはプログラムの児童生徒に対する影響に着眼した研究は存在するか？ 2) それらの研究は質の高い研究か？ 3) 質の高い研究が少なくとも5つあるか？ 4) 質の高い研究が5つある中の4つで、児童生徒の学業成績の改善を示しているか？ 5) 「はい」の場合、4つある中の3つの研究で、有意な結果が出力されているか？	実践またはプログラムについて 1) 実践またはプログラムの児童生徒に対する影響に着眼した研究は存在するか？ 2) それらの研究はまあまあの質の研究か？ 3) 質の高い研究またはまあまあの質の研究が少なくとも5つあるか？ 4) まあまあの質の研究が5つある中の4つで、児童生徒の学業成績の改善を示しているか？ 5) 「はい」の場合、4つある中の3つの研究で、有意な結果が出力されているか？
	これらの問いの全てに対する回答が「はい」の場合、この実践またはプログラムに関して科学的基盤を持つ研究は存在する。 The "Gold" Standard	これらの問いの全てに対する回答が「はい」の場合、たとえ実践またはプログラムに基づく研究が、法律に規定する科学的基盤を持つ研究の条件の全てを満たすものではないとしても、この実践またはプログラムに関して「強力なエビデンス」は存在する。 The "Silver" Standard

【出典】 U.S. Department of Education, Comprehensive Scientifically based research and the Comprehensive School Reform (CSR) Program, 2002, pp.5-6.

表5 科学的基盤を持つ研究が否か評価するために用いられる問い

<ol style="list-style-type: none"> <li>1. 体系的で経験的な方法                             <ul style="list-style-type: none"> <li>○ 研究は、しっかりした理論的基盤を持っているか？</li> <li>○ 方法論、被験者、研究者ははっきり示されているか？</li> <li>○ 研究は、首尾一貫し、規律付けられ、秩序立った方法を用いて実行されたか？</li> <li>○ データは、観察や実験を用いて得られたか？</li> <li>○ 研究は、個人見解ではなく事実即ちデータに基づいているか？</li> <li>○ 研究で明らかになったことは、明白で、測定可能なエビデンスによって支持されているか？</li> </ul> </li> <li>2. 厳密なデータ分析                             <ul style="list-style-type: none"> <li>○ 研究は、示された仮説をテストし、そして得られた一般的結論の正しさを証明したか？</li> <li>○ 研究は、サンプルサイズや用いた統計的手続きを報告しているか？</li> <li>○ データは、タスクに適した方法を用いて分析されたか？</li> <li>○ 方法は、データの特徴や構造に合致しているか？</li> <li>○ 研究は、観察された効果に対して他の説明が生じないように設計されているか？</li> <li>○ 研究で明らかになったことは、観察された結果が介入によって引き起こされたものであると納得させる証拠を示しているか？</li> </ul> </li> <li>3. 信頼性のある妥当なデータの収集                             <ul style="list-style-type: none"> <li>○ データは、多くの場所で、多数の調査者が関わる研究から得られたものか？</li> <li>○ 研究のバイアスは最小化されたか？</li> <li>○ データは、母数に一致するように測定されたか？ 同じ条件下で被験者に繰り返し行った測定では同様の結果を示したか？</li> </ul> </li> <li>4. 強固な研究デザイン                             <ul style="list-style-type: none"> <li>○ 研究デザインは、適切な統制の下、被験者を異なる条件に無作為に割り当てるランダム化比較試験を説明しているか？</li> <li>○ 統制は、関心ある条件の評価を見越したものになっているか？</li> <li>○ 研究は、調査者がリサーチアクションに答えることができるよう最適化した設計になっているか？</li> </ul> </li> <li>5. 再現性を見越した詳細な結果                             <ul style="list-style-type: none"> <li>○ 研究で明らかになったことは、明快に記述され報告されているか？</li> <li>○ 研究の結果は、デザインの再現が可能となるよう、十分詳細に示されているか？</li> <li>○ 研究で明らかになったことは、さらなる研究を行うことで、さらにその強度を増すか？</li> </ul> </li> <li>6. 研究の結果を精密な吟味にかける                             <ul style="list-style-type: none"> <li>○ 研究は、ピアレビューが行われる雑誌または第三者の専門家によって受容されたか？</li> <li>○ 研究は、その研究に関わりのないバイアスのかかっていない専門家によってレビューされたか？</li> <li>○ レビューを行う者は、学問としての厳格な水準を適用して、レビューする研究に対して質の高い統制を提供したか？</li> <li>○ 研究は、外部の検証を受けたか？</li> </ul> </li> </ol>
--

【出典】 U.S. Department of Education, Comprehensive School Reform (CSR) Program Guidance, 2002, pp.8-9.

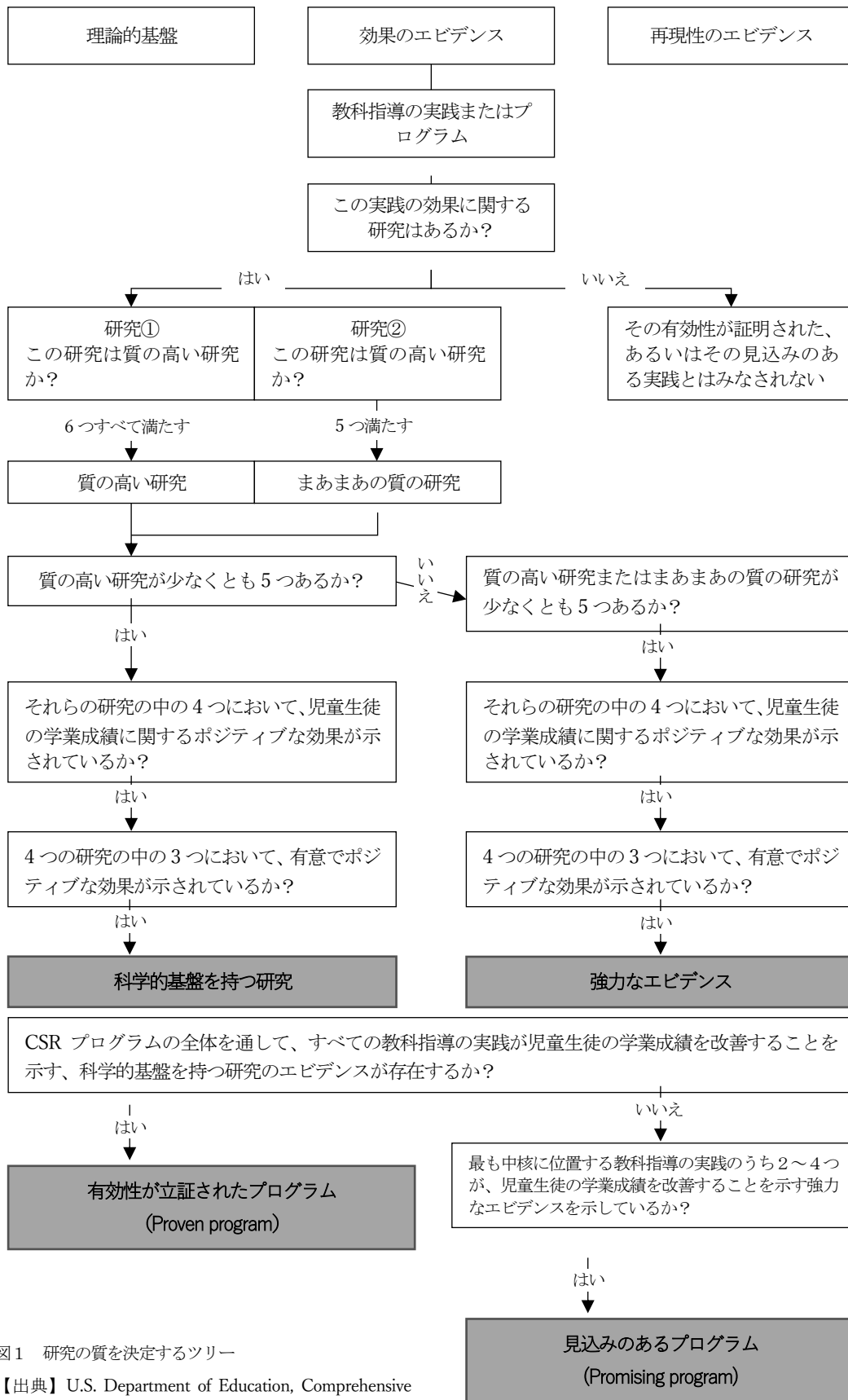


図1 研究の質を決定するツリー

【出典】 U.S. Department of Education, Comprehensive Scientifically based research and the Comprehensive School Reform (CSR) Program, 2002, p.17.



ここに読み取れることは、法律に規定する「科学的基盤を持つ研究」の概念を具現化するにおいて、より現実的な視点に立って2つの基準が敷かれたなかにあっても、「RCTを頂点とするエビデンスの位階秩序」<sup>11</sup>を前提に、また再現性（一般化可能性）を求める、“高度”な基準を維持しようとする意志、である。

この意志はその後も生き続け、民主党オバマ政権下においても、2009年に創設されたイノベーション投資補助プログラム（Investing in Innovation Fund :i3）では、その特徴を見ることができる。

### （3）i3プログラムで求められる「エビデンス」の基準

イノベーション投資補助プログラム（Investing in Innovation Fund :i3）は、米国再生・再投資法（American Recovery and Reinvestment Act of 2009 :ARRA）によって創設されたもので、（1）地方教育当局、または（2）（a）1つ以上の地方教育当局と連携する非営利組織、または（b）複数の学校から成るコンソーシアムと連携する非営利組織に対して支援する補助金を提供するものである。その目的は、児童生徒の学業成績またはその成長を改善すること、学業成績における格差を是正すること、退学率を軽減すること、高校卒業率を向上させること、または大学進学率・卒業率を向上させることに対して効果を有することが立証されている革新的実践の実施及びそれに対する投資を拡大するため、児童生徒の学業成績を改善させる証拠書類と共に申請する者に対して競争的資金を提供するものである（ARRA section 14007）。

このプログラムでは、「革新的実践」をより着実に進めるため、「開発（Development）」「正当性の立証（Validation）」「普及拡大（Scale-up）」の3つの段階にしたがって補助金のタイプが用意されており、そして、その段階ごとに要求されるエビデンスの内容には、明確な差異が示されるのである。

表6に示すように、申請者は、開発の段階では「もっともらしい仮説（Reasonable hypotheses）」が、正当性の立証の段階では「まあまあのエビデンス（Moderate Evidence）」が、そして最後の段階「普及拡大」では「強力なエビデンス（Strong Evidence）」の提出が要求される。着眼すべきは最後の段階で、すなわち最も広く活用されることを期待する中にある最高水準のエビデンスが要求されるのである。

ここで、「強力なエビデンス」とは、その研究デザインが因果的結論を支持することができ（すなわち高度な内的妥当性をもった研究）、そして州、地区、全国のレベルまで普及拡大することを支持できるだけ十分な範囲の参加者と環境を全体として含んでいる研究（すなわち高度な外的妥当性をもった研究）であることを意味する。次のものは、強力なエビデンスの例である。（1）実践、方略またはプログラムの有効性を支持する、2つ以上の、良く設計され良く実施される実験法を用いた研究または良く設計され良く実施される準実験法を用いた研究、（2）実践、方略またはプログラムの有効性を支持する、1つの大規模な、良く設計され良く実施される、複数の場所で試されたランダム化比較試験。

なおこの定義では、教育エビデンス情報センター（What Works Clearinghouse: WWC）が作成するエビデンスの基準が採用されている。表7に示すように、「強力なエビデンス」と認められるためには、その基準を留保なしで満たしているランダム化比較試験を用いた研究が1件以上ある場合、または基準を留保付きで満たすランダム化比較試験か準実験を用いた研究が2件以上ある場合、かつ表に示す他の条件を満たす場合に限定されている。

【論文】米国連邦教育政策におけるエビデンス要求の変遷

表6 i3プログラムで要求されるエビデンスの内容

タイプ	普及拡大	正当性の立証	開発
研究の強靭さ	強力なエビデンス (Strong evidence)	まあまあのエビデンス (Moderate evidence)	もっともらしい仮説 (Reasonable hypotheses)
内的妥当性 (因果的結論の強靭さ) と外的妥当性 (一般化可能性)	高度な内的妥当性と高度な外的妥当性	(1) 高度な内的妥当性とまあまあの外的妥当性、または (2) まあまあの内的妥当性と高度な外的妥当性	少なくとも何人かの参加者やいくつかの環境で効果の潜在性を、理論とレポートされた実践が示唆している
提示する実践、方略、プログラムの有効性や効力を支持する先行研究	(1) 2つ以上の、良く設計され良く実施される (well-designed and well-implemented) 実験法または準実験法の研究、または (2) 1つの大規模な、良く設計され良く実施される、複数の場所で試されたランダム化比較試験	(1) サンプルサイズが小さかったり他の実施や分析の条件のために、一般化可能性を制約してしまう、少なくとも1つの良く設計され良く実施される実験法または準実験法の研究 (2) 介入群と比較群が、介入の開始時に同等であることを証明できていないが、その他の点で内的妥当性に関して大きな不備を孕んでいない、少なくとも1つの良く設計され良く実施される実験法または準実験法の研究 (3) 選択バイアスを処理し、内部要因の影響を識別するため、強力な統計的統制を用いる相関関係の研究	(1) 提示する実践、方略、プログラム、あるいはそれに類するものが、限られたスケールで、または限られた環境で行われたものであっても、以前に試みられ、より公的で組織的な研究を行うに値することを示唆する見込みあるの結果を産み出したというエビデンス、そして (2) 教育や他の分野の中の関連する研究または理論を含め、研究で明らかになったことやもっともらしい仮説に基づいた提示する実践、方略、プログラムに対する論理的説明
先行研究における実践、方略、プログラム	普及拡大の補助を受ける、提案するものと同じもの	正当性立証の補助を受ける、提案するものと同じもの、あるいはまさしく同様のもの	開発の補助を受ける、提案するものと同じもの、あるいは同様のもの
先行研究における参加者と環境	先行研究における参加者と環境には、普及拡大の補助のもと処置を受けるよう提案される参加者と環境と同じ種類の参加者と環境が含まれている	先行研究における参加者と環境は、正当性立証の補助のもと処置を受けるよう提案される参加者や環境より、制限されたものになるかもしれない	先行研究における参加者と環境は、開発の補助のもと処置を受けるよう提案される参加者や環境より、制限されたものになるかもしれない
効果の有意性	先行研究における効果は統計的有意であり、普及拡大の補助金を受けるよう提案するサイズのサンプルで統計的有意であることが予期されうる	先行研究における効果は、正当性立証の補助金を受けるよう提案するサイズのサンプルで統計的有意であることが予期されうる	実践、方略、プログラムは、効力を調査するためのさらなる研究の必要性を認める
効果の大きさ	先行研究に基づき、普及拡大のプロジェクトがターゲットとする人々にとって本質的で重要	先行研究に基づき、正当性立証のプロジェクトがターゲットとする人々と同じ潜在性をもって、本質的で重要	先行の実施の取り組みに基づき、開発のプロジェクトがターゲットとする人々にとって見込みのあること
期待される規模	全国、地区、州	地区、または州	さらなる開発と規模

【出典】 U.S. Department of Education, Federal Register, Vol. 75, No. 48, March 12, 2010.

表7 i3プログラム 普及拡大の補助金におけるエビデンスの基準「有効性を示す強力なエビデンス (Strong evidence)」

	選択肢1	選択肢2
介入を対象とする研究の数	少なくとも1件	少なくとも2件
関連するアウトカムの研究結果	統計的有意で最重要の好ましくない影響を一つも伴わず、統計的有意で好ましい影響	
WWCのエビデンスの基準	留保なしで基準を満たす (摩擦 (attrition) に問題がなく良く実施されるランダム化比較試験のみこれに相当する)	留保付きで基準を満たす (摩擦に問題があるか良く実施されていないランダム化比較試験、または準実験がこれに相当する)
研究のサンプルサイズ	大規模	
研究が実施される場所の数	多数の場所でのサンプル	
研究対象の人々の類似性	i3プログラムの介入を受けるよう提案される人々と環境と部分的に重なる	

【出典】 U.S. Department of Education, Investing in Innovation (i3) Fund Evidence & Evaluation Webinar, 2016.

このように、i3プログラムにおいても、実践、方略、プログラムを学校に普及する段階にあつては、かなり厳しい基準をクリアした研究から成るエビデンスを示すことが要求されることが示される。

ただし、CSRプログラムとの比較の観点からすると、CSRプログラムでは、「科学的基盤を持つ研究」

と認められるためには「質の高い研究」が5件以上あることが必要条件であったのが、i3プログラムでは、留保なしでWWCの基準を満たすランダム化比較試験を用いる研究が1件あれば良しとされる。しかし一方でCSRプログラムでは、摩擦（attrition）にも言及しながら研究デザインが詳細に設計されていないのに対して、i3プログラムでは、WWCの基準に準拠したうえで詳細にそれが設計されている。この点において、i3プログラムはCSRプログラムに対して、量的に要求が軽減されながらも質的には要求が厳格化された、という特徴の変化を読み取ることができよう。

米国連邦政策ではNCLB法の成立以降、こうしてエビデンス要求が厳格化してきたのだが、しかし、昨今興味深いさらなる変化が見られるのである。それは、エビデンス要求の緩和である。

### 3. エビデンス要求の緩和

オバマ政権下の2015年12月10日に初等中等教育法の改正法として「全ての生徒が成功する法（Every Student Succeeds Act: ESSA）」が成立した。そのESSAではそれまでと同様に、エビデンスに基づく実践や方略、介入の活用を求めている。ただしその用語の用いられ方が、それまでと異なるのである。それは、エビデンスの基準に対して設定されてきた階層性が曖昧になったことである。

ESSAでは、表8に示すように「エビデンスに基づく」の用語が法的に定義されており、そこでは「強力なエビデンス（strong evidence）」「まあまあのエビデンス（moderate evidence）」「見込みのあるエビデンス（promising evidence）」「原理的説明を立証するもの（demonstrates a rationale）」の4つの階層が設定されている。問題は、その定義の内容に加えて、活用の在り方にある。

表8 ESSAに定める「エビデンスに基づく」の定義

<p>「エビデンスに基づく（evidence-based）」の用語は、州、地方教育当局または学校の活動に関して用いられるとき、以下の活動、方略または介入を意味する。</p> <p>(i) (I) 少なくとも1つの良く設計され良く実施された実験法を用いた研究から得られる強力なエビデンス（strong evidence）、または (II) 少なくとも1つの良く設計され良く実施された準実験法を用いた研究から得られるまあまあのエビデンス（moderate evidence）、または (III) 少なくとも1つの良く設計され良く実施された、選択バイアスに対して統計的統制を用いる相関関係の研究から得られる見込みのあるエビデンス（promising evidence）に基づき、児童生徒のアウトカムや他の関連するアウトカムの改善に統計的有意に効果があることを立証するもの、または</p> <p>(ii) (I) 質の高い研究で明らかになったことまたはポジティブな評価に基づき、前述の活動、方略または介入が児童生徒のアウトカムまたは他の関連するアウトカムを改善することが予期されることの原理的説明を立証するもの（demonstrates a rationale）、かつ (II) 前述の活動、方略または介入の効果を検証する進行中の取り組みを含む</p>
--

【出典】 section 8101(21)(A) of the ESEA

NCLB法では、「科学的基盤を持つ研究」の定義に照らして事前に活用が許される介入が選定され、州や地方教育当局はそれらの中から選択する体制がとられた。それがESSAでは、州や地方教育当局は自らの裁量をもって介入を選択したり開発する権限が与えられたのである。そのうえタイトルI補助金では、「強力なエビデンス」「まあまあのエビデンス」「見込みのあるエビデンス」のいずれかの裏付けが求められる一方、それ以外の補助金プログラム（タイトルIIなど）では、4番目の「原理的説明を立証するもの」でもよいとされる。これにより、活用が許される介入の幅が一気に広がるのである<sup>12</sup>。

#### 4. おわりに

本稿では、米国連邦教育政策において求められる「エビデンス」の内容の変遷を辿った。NCLB法で強力なアカウントビリティ政策がとられたことと軌を一にして、エビデンス要求が強く押し出された。有効性が立証されたプログラムの活用を促した CSR プログラムでは、より詳細にエビデンスの概念が整理され、その後 i3 プログラムでは、WWC の基準に準拠しながら、普及拡大においては「強力なエビデンス」を要求することが示された。しかしながら 2016 年に成立した ESSA では、むしろエビデンス要求が緩和される嫌いがあり、研究者の中にはこれに疑義を呈するものも存在する。

次稿では、ESSA に示されるエビデンス要求の緩和をめぐる背景や諸課題について、より詳細に検討したいと思う。

<sup>1</sup> 今井康雄「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—」『教育学研究』82(2)、2015年、192頁。

<sup>2</sup> 同上、192頁。

<sup>3</sup> David J. Torgerson, Carole J. Torgerson (原田隆之他訳)『ランダム化比較試験 (RCT) の設計』日本評論社、2010年、2頁。

<sup>4</sup> Coalition for Evidence-Based Policy, *Bringing Evidence-Driven Progress To Education: A Recommended Strategy for Report of the U.S. Department of Education*, 2002, p.2.

<sup>5</sup> *ibid.* p.4.

<sup>6</sup> *ibid.* p.4.

<sup>7</sup> Robert E. Slavin, *Design competitions: A proposal for a new federal role in educational research and development. Educational Researcher*, 26(1), 1997, pp.22–28..

<sup>8</sup> Comprehensive School Reform Program Office, Office of Elementary and Secondary Education, U.S. Department of Education, *Scientifically based research and the Comprehensive School Reform (CSR) Program*, 2002, p.3.

<sup>9</sup> *ibid.* pp.4-10.

<sup>10</sup> U.S. Department of Education, Office of Elementary and Secondary Education, *Comprehensive School Reform (CSR) Program Guidance*, 2002, p.9.

<sup>11</sup> 今井康雄、前掲論文、191頁。

<sup>12</sup> Rebecca Herman, et.al., *School Leadership Interventions Under the Every Student Succeeds Act: Evidence Review*, RAND, 2017.

## “Evidence” Requirement Change in the Federal Educational Policy in the United States: A Shift from a Strict Evidence Request Toward a Flexible Standard

Takafumi KIRIMURA

In recent years, the word “evidence” has come to be commonly used in the public debate over education in Japan.

In this paper, in relation to the question of what kind of “evidence” should be required, and in what kind of content should “evidence” be praised in the federal education policy of the United States, the cases of the No Child Left Behind Act of 2001 (NCLB), Comprehensive School Reform (CSR), Investing in Innovation Fund (i3), and Every Student Succeeds Act (ESSA) were examined, and the transition of the policies were overviewed.

NCLB gave the definition of “scientifically based research,” and practices and programs called for scientific support by providing evidence on its effectiveness. CSR tried to embody the concept of its definition, where a lower standard called “strong evidence” was prepared separately from “scientifically based research” from a realistic point of view. Still CSR tried to maintain the hierarchical order of evidence with the randomized controlled trial at the top. This feature is also seen in subsequent policies, and can also be seen in the i3 founded in 2009.

However, the Every Student Succeeds Act, established in 2015, granted substantial discretion to state and local education authorities in the context of the use of evidence. Moreover, it is said that only the lowest-rated evidence should be presented for grants other than Title I. Therefore, it can be said that the policy of “evidence” in the federal education policy of the United States has undergone a major change.