

【研究ノート：依頼原稿】

エビデンスに基づく教育 アメリカの教育改革とWhat Works Clearinghouseの動向

田辺 智子

国立国会図書館

s-tanabe@ndl.go.jp

要 約

アメリカの「子どもを一人も落ちこぼれにしないための2001年法」(NCLB法)は、教育現場にエビデンスに基づく施策の採用を求めている。エビデンスとして特に重視されるのが、実験デザインを用いた評価研究である。教育省は、評価研究に関する信頼できる情報を提供することを目的に、What Works Clearinghouse (WWC) というウェブサイトを開設した。評価研究の系統的レビューを行い、その結果をエビデンス・レポートとして掲載している。WWCはまだ構築の途上であるが、今後、どのようにコンテンツが充実され、教育現場に採用されていくのか、動向が注目される。

日本の教育界にとっても、エビデンスを重視するNCLB法の発想は示唆的である。わが国でも、信頼性の高い評価研究を蓄積し、系統的レビューを行っていく意義は大きい。

キーワード

エビデンス、プログラム評価、実験デザイン、教育、What Works Clearinghouse

はじめに

エビデンスに基づく施策の採用を目指す最近の評価研究の流れにあって、新しい動きとして注目されるのが、アメリカの教育改革とWhat Works Clearinghouse (WWC) の動向である。ブッシュ政権は、州や学校現場に対し、エビデンスに基づいた意思決定を求める法改正を行っており、そうした意思決定を支援するために、評価研究の系統的レビュー (Systematic Review) を提供する情報サイトであるWWCが開設された¹⁾。

エビデンスの系統的レビューの試みとしては、保健医療分野のコクラン共同計画や、社会科学の

キャンベル共同計画が先行している。WWCも両計画に大きな影響を受けているが、コクランやキャンベルとの重要な違いは、WWCは開設の時点ですでに、コンテンツに対する需要が用意されているという点にある。WWCによって今後、充実が図られていくコンテンツは、アメリカの教育現場の意思決定に、ダイレクトな影響を与えることが予想される。

本稿では、アメリカの教育改革におけるエビデンスの重視とWWCの現状について、関連する議論の動向を含めながら紹介する。

1. アメリカの教育改革とエビデンスの重視

(1) 「子どもを一人も落ちこぼれにしないための2001年法」

2001年に誕生した現ブッシュ政権は、教育を最優先の政策課題と位置づけ、政権発足後いち早く教育改革に着手した。2002年1月には、超党派の支持を受け、「子どもを一人も落ちこぼれにしないための2001年法」(No Child Left Behind Act of 2001: NCLB法²⁾)が成立した。NCLB法の基本的目的は、アメリカの教育水準を引き上げることに置かれている。そのために、州や学校の裁量を拡大し、代わりに子どもの学力向上という「結果」に対する責任を厳しく問うアカウンタビリティ・システムが導入された。その他、具体的な施策として盛り込まれたのは、読解力の向上、親への情報と選択権の付与、教師の質の向上、学校の安全確保、テクノロジーを活用した教育の促進などである(松山2001)。

以上に加え、NCLB法は、それまでにない重要な特徴を持っていた。科学研究によって有効性が裏付けられた施策や指導法の重視である。NCLB法の条文には、100回以上も「科学的根拠のある研究」(Scientifically-Based Research)という表現が登場する。たとえば、「科学研究に基づいて総合的な読解プログラムを構築した州政府に対し予算を与える」、「不利な条件にある生徒への教育サービスの計画にあたっては、科学的根拠のある研究結果を考慮しなければならない」、「若者の暴力やドラッグを防止すると科学的に示された施策に予算を振り向ける」といった具合である(Reyna 2005; U.S. Department of Education 2002)。

NCLB法のベースにあるのは、連邦政府はこれまで教育に多額の予算を投入してきたにもかかわらず、学力向上という成果につながっていないという問題意識であった。そして、成果を上げるための手段の一つと位置づけられたのが、科学的に有効性が確認された施策の採用、つまりエビデンスの重視である。

(2) 何がエビデンスとされるのか

では、NCLB法のいう「科学的根拠のある研究」とは、具体的にどのようなものを指すのだろうか。同法の第9101条では、その定義を次のように与えている。「教育活動や施策に関する妥当で信頼できる知識を得るための、厳格・体系的かつ客観的な手続きの適用を含む研究」であって、それには、観察や実験に基づく体系的・実証的研究、仮説の検証を伴うデータ分析、実験デザイン・擬似実験デザインを用いた研究などが含まれる。

この定義からは、実験デザインや擬似実験デザインが「科学的根拠のある研究」に含まれることは読み取れるが、定義を満たす研究の要件は必ずしも明らかではない。また、一般の教育関係者には、こうした分析手法は必ずしも馴染みのある概念ではなかった。このため、「科学的根拠のある研究」をどのように解釈し運用するか、教育現場に戸惑いが生じることとなった。

こうした状況を受け、教育省は、2003年12月に「厳密なエビデンスによって裏付けられた教育実践の識別と実施：ユーザーフレンドリー・ガイド」と題したガイドラインを発表した(Coalition for Evidence-Based Policy 2003)。このガイドラインでは、教育にかかわる介入(intervention)の有効性を、どのように見極めるかを説明している。

それによれば、最も厳密なエビデンスが得られるのは、ランダム化比較試験(Randomized Controlled Trial: RCT)である。生徒やクラス、学校などをランダムに実験グループと比較グループに分けるこの方法によれば、他の要因による影響を排し、純粋な介入の効果を測定することができる。教育分野でのRCTの適用例としては、学力面で遅れをとるリスクのある子供に対する個別指導、学級規模の縮小、英語習得のための発音練習などを対象としたものがあり、それぞれ有効性が確認されている。

教育分野でより広く使われているのは、ランダム化以外の方法で比較グループを設定する擬似実験デザインである。この方法では、グループ間に存在する見かけではわからない差異のために、正しい結果が得られない場合がある。同ガイドラインは、RCTと擬似実験デザインで正反対の結果が

出たケースを紹介し、この方法だけで最終的な結論を出すことはできないとしている。さらに、学校現場でしばしば行われる事前事後比較では、意味のある評価結果は得られないとされた。比較グループがないままに施策を適用し、実施前と実施後のデータを比較しても、その変化が施策によるものなのか、施策が無くても生じたものなのかを区別できないためである。

このように、教育省は、施策のエビデンスとしてRCTを最も上位に位置づけている。RCTは、医学や農学などの自然科学で一般的に用いられてきた手法である。それを教育分野にも採用し、施策の有効性をあらかじめ実験的に確かめることが意図されている。

(3) RCTをめぐる論争

このRCTの重視は、教育や評価にかかわる研究者の間に、大きな論争を巻き起こすこととなった。その論点は多岐にわたるが、ここでは二つの主要な論点に絞って紹介する。

第一の論点は、RCTを他の評価手法より上位に位置づけることの是非である。一部の評価研究者からは、擬似実験や相関分析などでも因果関係の検証は可能であり、RCTが現実社会の複雑な因果関係を検証するのに適さない場合もあるなどの反論が提起された（Donaldson and Christie 2005, p.63）。こうした見方は、上記の教育省の立場と真っ向から対立する。

第二の論点は、教育現場でRCTの広範な適用が可能かどうかという問題である。社会科学におけるRCTの実施は、倫理的に難しい場合があることが指摘されている。加えて、学校や生徒を取り巻く状況は多様かつ複雑であるため、医療分野で行われるような統制された実験は難しく、実験結果の一般化もしにくいという議論がある（Glenn 2004）。これに対し、RCTを擁護する立場からは、医療においても患者の個人差や症状の多様性を織り込んだ上で一般化可能な実験を行っているという反論が示されている（Slavin 2004）。

多くの場合、RCTに批判的な論者であっても、必ずしもRCTの有用性自体を否定しているわけではない。RCTは強力な評価手法ではあるが、現

実に適用できるケースは限られており、RCT以外の研究にも重要な役割があるというのがその主張である。反対に、RCTを擁護する論者も、RCT以外の手法の有用性を否定しているわけではない。NCLB法のいう「科学的根拠のある研究」には様々な研究が含まれるが、施策の有効性、すなわち "What works?" という問いに答えるためには、RCTが最も有用だというのがその主張なのである（Mageau 2004, p.2）。

アメリカの評価関係者の間には、実験デザインに代表される定量的評価の優位性を主張する立場と、定性的評価を重視する立場とがあり、長い間、論争を繰り返してきた歴史がある（龍・佐々木 2000, pp.12-14）。上記の意見対立には、こうした従来からのパラダイム論争が色濃く反映されている。いいかえれば、NCLB法を契機に初めてRCTをめぐる論争が巻き起こったわけではなく、専門家間で続いてきた大きな論争が、NCLB法によってあらためて表面化したということになる。

(4) 労働・福祉分野の経験

以上のような議論があるにもかかわらず、RCTの有用性が、ここにきて強調されているのはなぜなのだろうか。一つの重要な背景となっているのは、コクラン共同計画とEBM（Evidence-Based Medicine）の発展、キャンベル共同計画やイギリスにおけるエビデンス重視の動き（Evans and Benefield 2001）であろう。こうした一連の活動の中では、RCTは最もレベルの高いエビデンスを産出するものと位置づけられている。これに加えアメリカの場合、労働や福祉など他の社会科学分野において、RCTがすでに広く実施され、政策決定に影響を与えてきた経緯があることを指摘しておきたい。

アメリカで、社会科学におけるRCTの適用が始まったのは1960年代である。その時点では、現実社会における実験が果たして可能なのかどうか、また有用な情報が得られるのかどうかは未知数であった。それ以降、労働、社会福祉、刑事司法などの分野で、数多くのRCTが実施されてきたが、特に事例が多いのは、福祉政策が雇用に与える影

響や、職業訓練を対象としたRCTである (Greenberg and Shroder 2004)。この結果、RCTは十分に適用可能であり、政策決定に有用な情報を与えることがわかってきた。

RCTの結果が、政策決定に直接的な影響を与えた例の一つに福祉改革がある。1996年までの社会保障法では、州が新しい福祉改革アプローチを試行する際に評価を義務付けていたため、州の施策を対象とした数万人規模のRCTが実施された (Gueron 2002, p.33,42)。1988年には、福祉受給者の就職を支援する法改正が行われたが、そこでは、RCTによる評価結果が重要な役割を果たしたとされる (Szanton 1991)。

RCTは、評価結果の信頼性が高いことから、他の手法に比べ、政策決定への影響力が大きいと指摘されている (Burtless 2002, p.183)。たとえば、1980年代の労働省の職業訓練プログラムでは、擬似実験による評価結果にバラつきがあり、正しい施策の効果が判然としなかった。このため、専門家の助言を受けた労働省がRCTによる評価を行ったところ、若年層を対象とした職業訓練には効果がないことが判明し、この結果、効果が認められた他の年齢層に大きく予算が振り替えられた。RCTによる評価結果に説得力があったため、議会も抵抗なく行政府の決定に従ったという。

こうした経験を通じ、労働や福祉の分野では、RCTはもはや特別なものではなくっており、プログラム評価の一環として、ごく普通に実施されている。Abt Associates、Mathematica Policy Research、MDRC (Manpower Demonstration Research Corporationから2003年に改称) など、RCT実施のノウハウを持つコンサルタントも成長した。MDRCは、2002年までの時点で30の大規模なRCTを実施しており、200以上の地域において30万人近い人々を対象としたという (Gueron 2002, p.16)。

これに対し教育分野では、RCTが行われる場合も、施策の有効性評価よりは理論的な検証が多く、実験期間も短かったと指摘されている (Slavin 2002, p.16)。教育施策のRCTの事例としては、1960年代に早期教育の有効性を評価したペリー就学前教育プロジェクトや、1980年代に学級規模と学力の因果関係を検証したテネシー州STARプロ

ジェクトなどが有名である。しかし、全体として見れば実施は限定的であった。

労働・福祉分野と教育分野で、このような違いが生まれた背景として、次のような要因が指摘されている (Burtless 2002 ; Gueron 2002)。まず、労働省や保健福祉省では、経済学者など厳密な評価に積極的な専門家が影響力ある地位に就いたが、教育省では、そうした状況が生まれなかった。また、福祉では連邦政府の政策への影響力が大きいのに対し、教育は原則として州の所管であり、連邦の影響力が小さかったことがある。連邦のほうの評価予算や人材面で恵まれているため、連邦の影響力が大きい分野のほうが、RCTが実施されやすい傾向がある。さらに、労働や福祉分野では、短期間で実施できる介入が多いこと、実験グループによる比較グループへの波及的な影響を遮断しやすいことなど、教育と比較して実験の実施に適していた要素もあった。

果たして、このような過去の傾向をくつがえし、教育分野でのRCTの産出と利用は進むのだろうか。その答えは、今後、実際にNCLB法を運用していく中で検証していくほかはないように思われる。そして、NCLB法の成否の一つのカギを握っているのが、次に述べるWWCの動向である。

2. What Works Clearinghouse (WWC) の概要と現状

(1) 目的と位置付け

NCLB法によって、教育界には、エビデンスを求める強い需要が生まれた。しかし、膨大な教育分野の研究の中から、質の高いエビデンスを見分けることは容易ではない。NCLB法を運用していくためには、科学研究をわかりやすい言葉で教育関係者に伝える仕組みが必要となる。この役割を期待されているのが、WWCによる系統的レビューである。

WWCは、教育にかかわる政策、施策、活動、製品の有効性について、最新で質の高いレビューを提供し、エビデンスに関する信頼できる情報源となることを目的としている。2004年7月には、

第一弾のレビュー結果を掲載したウェブサイトが公開された。WWCの運営は、教育省の教育科学研究所が、American Institute for Research とキャンベル共同計画によるジョイント・ベンチャーへの契約を通じて行っている。

(2) 系統的レビューのプロセス

WWCにおける系統的レビューのプロセスは、以下のとおりである。

レビュー・トピック

まず第一に、レビュー対象となるトピックが選定される。WWCは、ウェブサイトや専門家のネットワークを通じ、トピックの候補を募集している。選定にあたっては、生徒への成果が期待できるか、現場での適用可能性があるか、教育関係者にレビューに対する需要があるか、科学的研究が存在しそうかどうか、などが基準とされる。表1に、WWCがこれまでに選定したレビュー・トピックをまとめた。

表1 WWCで選定されたレビュー・トピック

- ・ 中学校の数学カリキュラム
- ・ 小学3年生までの初期の読解
- ・ 人格（道徳）教育
- ・ 高校の退学防止策
- ・ 英語を母語としない小学生の英語学習
- ・ 小学校の算数カリキュラム
- ・ 3～5歳児の就学前教育
- ・ 中学校・高校の非行・暴力的行動の防止策
- ・ 成人の読解力向上策
- ・ 小学生同士の相互指導とグループ学習

(出所) WWC ウェブサイトをもとに
筆者作成 <http://w-w-c.org/topics/current_topics.html>

レビュー・チーム

続いて、あるトピックを担当するレビュー・チームが編成される。それぞれのチームには、当該分野の専門家と分析手法の専門家がアドバイザーとして含まれる。彼らの指導のもとで複数のリサーチ・アナリストがレビューを行い、プロジェクト・コーディネーターがレビュー・プロセス全体の監督を行う。

文献検索とスクリーニング

レビュー・チームはまず、体系的かつ包括的な文献検索を行い、当該トピックにかかわる研究を収集する。検索対象には、査読誌だけでなく、非査読誌や非出版物も含まれる。教育関係者や出版社などが、関連する研究を推薦することもできる。集められた研究は、トピックとの関連性や新しさなどの観点からスクリーニングされる。

エビデンス基準

続いてレビュー・チームは、個々の研究のエビデンスとしての信頼性をWWCエビデンス基準に基づいて判定する。その結果、それぞれの研究は、以下の3カテゴリーにランク付けされる。

「エビデンス基準を満たす」...RCT、回帰非連続デザイン

「留保つきでエビデンス基準を満たす」...擬似実験デザイン、ランダム化・サンプルの脱落・実施過程等に問題のあるRCT、サンプルの脱落・実施過程等に問題のある回帰非連続デザイン

×「エビデンス・スクリーンを通らない」...上記以外の研究

「エビデンス基準を満たす」、「留保つきでエビデンス基準を満たす」と判定された研究は、介入の正確さ、アウトカム指標、統計分析などの重要な側面について、さらなるレビューを受ける。

エビデンス・レポート

レビューの結果は、WWCの最終成果物であるエビデンス・レポートにまとめられる。これに含まれるのは、「エビデンス基準を満たす」、「留保つきでエビデンス基準を満たす」と判定された研究のみである。エビデンス・レポートには、個々の研究の要約である「研究レポート」、介入ごとのエビデンスをまとめた「介入レポート」、トピック全体のレビュー結果を総合した「トピック・レポート」の3種類がある（表2）。WWCは、新しい研究結果を含めるために、エビデンス・レポートを定期的にアップデートする予定である。

表2 WWCのエビデンス・レポート

レポート	内 容
研究レポート	エビデンス基準を満たす個々の研究を要約したもの。当該研究の概要、研究結果、エビデンス基準を満たすかどうか、研究の長所と短所などが記載される。
介入レポート	エビデンス基準を満たす研究（留保つきであってもよい）が一つ以上ある介入について作成される。当該介入についての説明、個々の研究結果の要約、全体としてのエビデンスの強さなどが記載される。付録として、個々の研究から得られた統計データの概略が掲載される。
トピック・レポート	当該トピックに関連するすべてのレビュー結果をまとめたもの。エビデンス基準を満たす全研究結果が要約され、総合的なエビデンスの状況が説明される。

(出所) WWC ウェブサイトをもとに掲載されているエビデンス・レポートをもとに筆者作成
http://w-w-c.org/topics/current_topics.html

品質管理

WWCには、リサーチ・デザインや評価研究の専門家からなるテクニカル・アドバイザー・グループ（TAG）が儲けられている。レビュー・プロセスのあらゆる段階で、このTAGやその他の専門家によるピア・レビューが行われる。また、レビューを受けている研究の実施者や、レビュー対象の教育ソフトウェアの開発者などにも、エビデンス・レポートをレビューする機会が与えられる。最後に、教育省がレビュー・プロセスが適切だったかを検証し、承認されたものがウェブサイトにアップされる。

(3) 系統的レビューの実際

現時点で、WWCのウェブサイトに掲載されているエビデンス・レポートは、「中学校の数学カリキュラム」というトピックに関するもののみである。ここでは、その概要を紹介する。

アメリカでは、国際的な学力調査でアメリカの子どもの数学力が低位に甘んじていることもあり、数学力を向上させねばという問題意識が強い。そのための方策の一つとして、画一的な内容だった従来の教科書に代わる、新しいタイプの数学カリキュラムの開発が進んでいる。新種の教材、コンピュータ・ソフト、ビデオなどを用い、指導方法を工夫することによって、生徒の数学的理解を深めることを意図したものである。こうした数学カリキュラムは、すでに一部で普及が始まっているが、その有効性については専門家の間でも見解が分かれていた。

WWCは、様々な情報源をもとに、中学生を対象とした44種類の数学カリキュラムを特定した。このうち、関連する評価研究が存在した20種類のカリキュラムに対するレビュー結果が、WWCウェブサイトに掲載されている（表3）。これらカリキュラムに対する評価研究のうち、文献検索の段階でのスクリーニングを通過したものは78件であった。それらがエビデンス基準に基づいてレビューされた結果、4件が「エビデンス基準を満たす」、6件が「留保つきでエビデンス基準を満たす」と判定され、残る68件は「エビデンス・スクリーンを通らない」とされた。この結果、10件の研究レポートと5件の介入レポートが作成された。表4に、介入レポートの要約表の一例を示す。

トピック・レポートは、以上のレビュー結果を総合したものである。エビデンス基準を満たす研究が存在したのは、表5に示す5種類のカリキュラムである。このうち、統計的に有意な有効性が示されたのは、Cognitive Tutor®とI CAN Learn® Mathematics Curriculumの2種類であった。それぞれについて有効性を示す研究は1件のみである。こうした結果から、トピック・レポートでは、中学校の数学カリキュラムについてのエビデンスは貧弱であるとして、評価手法の改善を呼びかけている。

ここで注意してほしいのは、ある介入について十分なエビデンスがあるということと、その介入が有効であるということとは別だという点である。信頼性の高い研究が存在したとしても、その結果が「当該介入は有効でない」という場合もある。同様に、ある介入についてのエビデンスが存

表3 トピック「中学校の数学カリキュラム」のレビュー結果

カリキュラムの名称	エビデンス基準			介入レポートの有無
	満たす	留保つきで満たす	満たさない	
Adventures of Jasper Woodbury Series	0	0	1	-
Algebra Project	0	0	2	-
Algebraic Thinking	0	0	1	-
Cognitive Tutor®	1	0	5	
CompassLearning	0	0	2	-
Connected Mathematics Project (CMP)	0	3	12	
Connecting Math Concepts (CMC) mathematics program	0	0	2	-
CORD Applied Math	0	0	1	-
Destination Math	0	0	1	-
I CAN Learn® Education Systems	1	2	4	
Integrated Mathematics, Science, and Technology (IMaST)	0	0	1	-
Mathematics in Context (MiC)	0	0	3	-
MATHThematics	0	0	3	-
Model Mathematics Program (MMP)	0	0	1	-
Moving with Math	0	0	12	-
Partnership for Access to Higher Mathematics (PATH Mathematics)	0	0	1	-
PLATO	0	0	1	-
Saxon Math	1	1	4	
Successmaker	0	0	9	-
The Expert Mathematician	1	0	0	
Studies not categorized by intervention	0	0	2	-
合計	4	6	68	
レポート数	4	6	-	5

(出所) WWC ウェブサイトをもとに筆者作成 <<http://w-w-c.org/topics/TopicInterventions.asp?EvidenceRptID=03>>

在しないということは、その介入が有効でないことを意味しない。まだ科学的に有効性が検証されていないだけで、今後の研究しだいで有効性が明らかになる可能性もある。

(4) WWCと意思決定

冒頭でも述べたように、WWCの場合、エビデンスの使い道がNCLB法によってあらかじめ用意されている点が、キャンベル共同計画やコクラン共同計画との大きな違いである。今後、WWCで収集され、レビューされたエビデンスが、どのように教育現場で活用されていくかが注目される。

教育省の関係者は、WWCはあくまで意思決定を助ける情報を提供するものであり、特定の施策や製品を支持するものではないという点を強調し

ている (Megeru 2004, p.35)。NCLB法はエビデンスに基づいた施策の採用を求めているが、実際にどのような施策を科学的根拠があると判断するかは現場に委ねられており、WWCのエビデンス基準を満たす施策のみを採用せよということではない。つまり、WWCはあくまで参考情報であるというのが教育省の立場である。

しかし、教育現場は、WWCをもっと拘束力のあるものと捉え、過剰に反応するおそれも指摘されている。これは、連邦の補助金を受ける際に、特に問題となる。NCLB法のいくつかの条文では、連邦の交付する資金を「科学的根拠のある研究」で有効性が確認されたものだけに使用するように求めている。このため、現場が補助金を受けられなくなることを恐れ、WWCの基準を満たさない施策の採用を自粛する可能性が出てくる。WWCが

表4 I CAN Learn® Mathematics Curriculumの介入レポート要約表

介入	I CAN Learn® Mathematics Curriculumは、基礎代数・代数のための双方向のソフトウェア・システムである。初歩的な数学から高等代数までの326のレッスンから成る。生徒が自分のペースで学習するためにデザインされており、主に7年生から10年生の人種的に多様なインナー・シティの生徒を対象としている。
対象	中学生
評価結果	<ul style="list-style-type: none"> ・1件のRCT（1校、254人）によれば、I CAN Learn を利用した生徒は、州の数学テストにおいて、通常のカリキュラムに従った比較グループよりも統計的に高い点を取ったことが示された。 ・1件の擬似実験によれば、I CAN Learn を用いた生徒が比較グループより高い点数をとったが、その差は統計的に有意ではなかった。 ・もう1件の擬似実験によれば、I CAN Learn を使った生徒が比較グループより高い点数をとったが、分析に欠陥があるため、結果の有意性を正確に見極めることができない。
エビデンスの基盤	エビデンス基準を満たす...1（RCT） 留保つきでエビデンス基準を満たす...2（擬似実験） × エビデンス・スクリーンを通過しない...4
エビデンスの限界	I CAN Learn® Mathematics Curriculumのエビデンスの基盤は3件の研究であり、そのうち1件のみがRCTによるものである。その他は擬似実験によるもので、2グループ間の測定されていない差異が結果に影響を与えている可能性があり、エビデンスとしてはRCTより弱い。擬似実験のうちの1件は、推計されたインパクトの信頼区間を計算していないという分析上の欠陥がある。
利用状況	このカリキュラムは1995年に最初に導入された。2004年10月までに、23州501校で利用されている。典型的な利用例は、タイトル ^(注) 指定校や、都市部の大規模学区における学力面で遅れをとるリスクのある生徒、マイノリティーの生徒である。
開発者と連絡先	JRL Enterprises, Inc., 400 Poydras Street, Suite 100, New Orleans, LA 70130; telephone (504) 263-1380; www.icanlearn.com.

（注）タイトル とは、社会経済的に恵まれず学力が低位の生徒に対し、連邦から補助金を支給する施策である。

（出所）Institute of Education Sciences, *What Works Clearinghouse Intervention Report: I CAN Learn® Mathematics Curriculum*. Updated December 1, 2004, p.1

どの程度、現場へのインパクトを持つかは、地域、施策内容、分野により異なると予想されている（Schneiderman, 2004, p.32）。

（5）エビデンスの需要と供給

現在のところ、教育分野でのRCTの事例は非常に限られており、エビデンスに対する需要が、供給を大きく上回っている状況にある。中学校の数学カリキュラムのレビューでも、WWCが特定した44のカリキュラムのうち、24については、評価が一件も行われていなかった。この需要と供給の大きなギャップは、エビデンスを積極的に生み出させる強いインセンティブとなる。

教育省は、プログラム評価を実施する際に、実験デザインによる研究に優先順位を与える方針を発表している⁴。この結果、2004年に行われた同

省の施策の有効性評価のうち、90%でRCTが採用されている（U.S. Department of Education 2005）。その中には非常に大規模なものもあり、たとえば、Mathematica Policy Researchは、16種類の商業ソフトウェアの英語と数学への効果を評価するために、560万ドルの資金を受けている（Glenn 2004）。

評価研究への需要と同時に、評価研究のレビューに対する需要もまた増加している。もともとRCTの実施例が限られていることを反映し、WWCがレビュー対象に選定しているトピックはまだ限定的なものである。WWCによるレビューがさらに進展するまでの間は、教師や地方政府の行政官など、教育関係者自身に、エビデンスを見極める目が求められる。WWCのレビュー・トピックの中でも、英語読解などはRCTが比較的多く存在する分野であり、そのレビュー結果が待たれるところである。

表5 トピック「中学校の数学カリキュラム」に関するエビデンスの要約

介入	サンプルと環境	エビデンスの強さ		
		研究の強さ	サンプル数	効果の推計結果
Cognitive Tutor®	・ 9年生	RCT	<u>1RCT</u> ・ 360人	<u>RCT</u> ・ Percentile rank: 9.1 ・ Standard scores: 0.23
Connected mathematics Project	・ 6-8年生 ・ 中西部、西部、東部、北東部、南部 ・ 農村部、都市部、郊外 ・ 社会経済状況が高位の地域と低位の地域	擬似実験	<u>3擬似実験</u> ・ 1,095人 ・ 50校 ・ 36-42校	<u>擬似実験</u> ・ 有意ではない
The Expert Mathematics Curriculum	・ 8年生 ・ 低所得層	RCT	<u>1RCT</u> ・ 70人	<u>RCT</u> ・ 有意ではない
I CAN Learn® Mathematics Curriculum	・ 7-10年生 ・ 都市部のタイトル指定校	RCT 擬似実験	<u>1RCT</u> ・ 254人 <u>2擬似実験</u> ・ 169学級 ・ 116学級	<u>RCT</u> ・ Percentile rank: 15.9 ・ Standard scores : 0.41 <u>擬似実験</u> ・ 有意ではない
Saxon Math	・ 8年生 ・ 都市部	RCT 擬似実験	<u>1RCT</u> ・ 36人 <u>1擬似実験</u> ・ 78人	<u>RCT</u> ・ 有意ではない <u>擬似実験</u> ・ 有意ではない

(出所) Institute of Education Sciences, *What Works Clearinghouse Topic Report: Curriculum-based interventions for increasing K-12 math achievement - middle school*. Updated December 1, 2004, p.4

今後、アメリカの教育界で、どれだけRCTが生み出され、レビューされ、活用されていくのか。WWCの息の長い取り組みは、ようやく動き始めたばかりである。

おわりに～日本への示唆～

今、わが国では、学校評価への関心が高まっている。それぞれの学校が、目標を設定した上で教育活動を評価し、その結果を教育の改善に役立てることをねらったものである。しかし、学校を評価するという発想はあっても、学校で行われる施策や指導方法の一つ一つを評価するという発想は、希薄なように感じられる。日本の教育現場でも、習熟度別授業、ITの活用、学級規模の縮小など、多くの熱意ある取り組みが行われている。こうした取り組みが、学力向上など意図した効果を上げ

ているか科学的に評価し、効果のある取り組みを採用していこうというのがNCLB法の発想なのである。

NCLB法は、まだ緒に付いたばかりであり、その成否は今後の検証を待たなければならない。また、RCTを特に重視することに関しては、アメリカにおいても議論のあるところである。しかし、エビデンスに基づく教育を目指すNCLB法の方向性は、わが国にとっても示唆に富むものといえるだろう。

数年前に話題になった学力論争が浮き彫りにしたのは、日本では、そもそも議論のベースとなる学力のデータが十分存在せず、学習指導要領の変更などの重要な政策変更が、科学的に評価されていないという問題であった。定性的な評価と定量的な評価はいずれも大切なものであるが、わが国の場合、定量的な評価が不足している感は否めない。今後はわが国でも、定量的な有効性評価の蓄

積が必要といえよう。アメリカのようなRCTの実施が日本でどの程度可能かどうかは議論の余地があるが、大切なのは、それぞれの施策に応じ、適切なデザインを用いた評価を行うことである。その上で、評価の質をレビューし、意思決定につなげるためのインフラとして、WWCのような系統的レビューが求められる。文部科学省は、学力向上フロンティア事業で各県の実践研究を支援しているが、こうしたモデル事業の評価から始めるのも一案であろう。

注記

- 1 <http://w-w-c.org>
- 2 正式名称An act to close the achievement gap with accountability, flexibility, and choice, so that no child is left behind. P.L. 107-110.
- 3 実験的手法には、ランダム化バイアス、代替バイアスなどの手法上の問題があることが指摘されているが（労働政策研究・研修機構 2004、p.11）アメリカの議論においては、こうした点は主たる論点とはなっていない。
- 4 Department of Education, RIN 1890-ZA00, "Notice of proposed priority," *Federal Register*, 68 (213), Nov. 4, 2003; "Notice of final priority," *Federal Register*, 70(15), Jan. 25, 2005. 教育分野では、従来、定性的な研究が主流であったため、評価予算の一部は、労働経済学者などRCTを頻繁に行ってきた他分野の研究者に流れているという（Glenn 2004）。この評価予算の配分問題は、RCTをめぐる研究者間の論争にも影響を与えている。

参考文献

- 労働政策研究報告書 No.L-4）
- Boruch, R., de Moya, D., and Snider, B. (2002). The importance of randomized field trials in education and related areas. In F. Mosteller & R. Boruch (Eds.), *Evidence Matters: Randomized Trials in Education Research*. Washington, DC; Brookings Institution Press.
- Burtless, G. (2002). Randomized field trials for policy evaluation: why not in education? In F. Mosteller & R. Boruch (Eds.), *Evidence Matters: Randomized Trials in Education Research*. Washington, DC; Brookings Institution Press.
- Carlson, J. S., and Levin, J. R. (Eds.). (2005). *The No Child Left Behind Legislation: Educational Research and Federal Funding*. Greenwich, Conn: Information Age Publishing.
- Coalition for Evidence-Based Policy. (2003). *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.
- Cook, T. D., and Payne, M. R. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller & R. Boruch (Eds.), *Evidence Matters: Randomized Trials in Education Research*. Washington, DC; Brookings Institution Press.
- Donaldson, S. I., and Christie C. A. (2005). The 2004 Claremont Debate: Lipsey vs. Scriven. *Journal of MultiDisciplinary Evaluation*, 60-77.
- Evans, J., and Benefield, P. (2001). Systematic reviews of educational research: does the medical model fit? *British Educational Research Journal*, 27 (5), 527-541.
- Fletcher G. H. (2004). Scientifically based research: guidelines or mandates for product purchasing? *T.H.E. Journal*, 31 (8), 22-25.
- Glenn, D. (2004). No classroom left unstudied. *The Chronicle of Higher Education*, 50 (38), A12-A15.
- Greenberg, D., and Shroder, M. (2004). *The Digest of Social Experiments, 3rd Edition*. Washington, DC: Urban Institute Press.
- Gueron, J. M. (2002). The politics of random assignment: implementing studies and affecting policy. In F. Mosteller & R. Boruch (Eds.), *Evidence Matters: Randomized*
- 教育政策の評価手法研究会 (2004)『プログラム評価の手法とその活用』(平成15年度文部科学省委嘱研究:教育行政の評価手法の在り方に関する調査研究報告書)
- 松山幸弘 (2001)「アメリカの教育改革から学ぶこと」、『Economic Review』、5 (4) : 93-109
- 龍慶昭・佐々木亮 (2000)『「政策評価」の理論と技法』、多賀出版
- 労働政策研究・研修機構 (2004)『労働市場政策の効

- Trials in Education Research*. Washington, DC; Brookings Institution Press.
- Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, 114, 497-532.
- Mageau, T. (2004). Determining 'what works'; an interview with Dr. Grover 'Russ' Whitehurst. *T.H.E. Journal*, 31 (6), 32-37.
- Mervis, J. (2004). Meager evaluations make it hard to find out what works. *Sciences*, 304, 1583.
- Schneiderman, M. (2004). What does SBR mean for education technology? *T.H.E. Journal*, 31 (11), 30-36.
- Slavin, R. E. (2002). Evidence-based educational policies: transforming educational practice and research. *Educational Researcher*, 31 (7), 15-21.
- Slavin, R. E. (2004). Education research can and must address "what works" questions. *Educational Researcher*, 33 (1), 27-28.
- Szanton, P. L. (1991). The remarkable "quango": knowledge, politics, and welfare reform. *Journal of Policy Analysis and Management*, 10 (4), 590-602.
- U.S. Department of Education (2002). *No Child Left Behind: A Desktop Reference*.
- U.S. Department of Education (2005). *FY2006 Program Performance Plan*.
- Viadero, D. (2004). Ed. Dept. issues practical guide to research-based practice. *Education Week*, 23 (16), 12. (2006.2.16受理)

Evidence-Based Education What Works Clearinghouse in the United States

Satoko Tanabe

National Diet Library
s-tanabe@ndl.go.jp

Abstract

The No Child Left Behind Act, which was enacted in the United States in 2002, includes challenging provisions that require states and local schools to adopt programs and practices based on scientifically-based researches. The act focuses especially on studies using experimental designs. In order to provide the education community reliable information on "what works," Department of Education opened a web site called What Works Clearinghouse (WWC). WWC review teams conduct systematic reviews of studies and the results are disseminated as WWC evidence reports. WWC is now under development, and is likely to influence U.S. education programs by its contents.

The idea of evidence-based education is suggestive for the Japanese education community. It is desirable that we accumulate reliable evaluation studies and conduct systematic reviews.

Keywords

evidence, program evaluation, experimental designs, education, What Works Clearinghouse