

教育研究と政策
—RCTとメタアナリシスの発展—
Education Research and Policy
—The Development of RCTs and Meta-Analysis—

惣脇 宏*
SOWAKI Hiroshi

Abstract

Randomized controlled trials (RCTs) and the meta-analysis of their results typically provide the highest quality available evidence that can be used in evidence-based policymaking and practice. In the education arena, there is a problem that little primary research involves a high quality study design such as RCT, and evidence is not easily produced.

However, studies that are well designed, have conducted RCTs, and have produced robust findings in education have had a considerable impact. For example, the Perry Preschool Program and Tennessee STAR Project have influenced policies and subsequent research.

RCTs in the social sciences, including education research, developed in the 1960s with social legislation concerning the War on Poverty, and meta-analysis was proposed in 1976. Although RCTs in the education arena were rare, RCTs and meta-analysis have currently gained importance, as “scientifically based research” has gained prominence since the latter half of the 1990s.

In Japan, RCTs are almost non-existent. However a balanced research portfolio that includes more rigorous and robust quantitative research is needed. Clearly, a basis of statistics and longitudinal surveys is also required. Such research will be indispensable for improving schools and increasing the accountability of the education system and the quality of education research itself.

1. はじめに

エビデンスに基づく政策と実践において、活用することのできるエビデンスのうち最も強いものは、ランダム化比較試験 (randomized controlled trial: RCT)¹⁾ によるもの、とりわけ複数の RCT の結果をメタアナリシス (meta-analysis)²⁾ したものとされる。RCT を実施する場合には、コストや倫理的問題など解決すべき制約が一般的に存在するが、医療はもとより刑事司法や社会福祉などの分野と比べると、教育分野においては、RCT による質の高い一次研究が乏しく、エビデンスが産出されにくいという課題がある (岩崎 2010a p.19)。

しかしながら、これまでに実施されてきた教育分野の RCT の中には、優れた設計と実施によって頑健な結論が得られ、教育政策に大きな影響を与えているものがある。そのような RCT の例としては、幼児教育の効果に関するペリー就学前計画や、学級規模の縮小の効果に関するテネシー州の STAR プロジェクトが挙げられる。前者は小規模であるが約 40 年間にわたって追跡調査が行われ、

* 独立行政法人大学入試センター理事、前・国立教育政策研究所次長

また、後者は大規模な追跡調査であったために、研究で得られた成果が、政策や後続の研究に与える影響力は極めて強く、国際的にも大きな影響を及ぼしている。

教育を含む社会科学における RCT は、1960 年代から 1970 年代前半に米国の「貧困との闘い」(War on Poverty) のための社会政策立法と関連して発達してきたものであるが、実験デザインは 1920 年代以前の教育研究に遡る。また現在、医学を始め各分野で用いられているメタアナリシスは、1976 年に米国教育学会長のグラスが提唱したものである。このように、RCT を含む実験デザインとメタアナリシスは、どちらも教育分野において先行したのであるが、1970 年代後半以降は、質的研究の進展もあって、量的研究があまり活用されなくなっていた。

しかし、教育の質に対する要求の高まりに対応して、教育政策が基づくべき教育研究の質に対する批判も現れたこともあり、1990 年代後半以降の米国においては、連邦政府によって「科学的根拠に基づく研究」(scientifically based research) が強調され、RCT とメタアナリシスが改めて重視されるようになっていく。

本稿では、主として米国の教育研究における RCT とメタアナリシスの発展や、政策とのかかわりについて、主要な事例³⁾を概観することとしたい。

2. 「実験する社会」とペリー就学前計画

(1) キャンベルの「実験する社会」

社会科学における RCT を含む実験デザインの最初の体系的なテキストは、社会心理学者のキャンベル (Donald T. Campbell, 1916-1996) とスタンレー (Julian C. Stanley, 1918-2005) による 1963 年の『研究のための実験および準実験デザイン』(Experimental and quasi-experimental designs for research) である。これはもともと教育研究の方法論として書かれたものであるが⁴⁾、同書をきっかけに実験デザインは公共政策一般に広く普及し、現在でも基本文献として用いられている (佐々木 2010 p.65)。

同書は、自然科学における基本文献であるフィッシャー⁵⁾による 1935 年の『実験計画法』(The Design of Experiments) に対応するものと言えるが (Oakley 1998 p.1240)、キャンベルとスタンレーが、心理学者のマッコール (W. A. McCall, 1891-1982) による 1923 年の『教育における実験の方法』(How to Experiment in Education)』をモデルにしたと書いているように (pp.1-2)、教育研究における実験デザインは 1920 年代以前に遡る⁶⁾。

ところで、1999 年に準備会合が開催されたキャンベル共同計画は、その 3 年前に没したキャンベルの業績を記念して名付けられたものである。キャンベルは 1969 年の「実験としての改革 (Reforms as Experiments)」⁷⁾ や 1971 年の「実験する社会 (The Experimenting Society)」⁸⁾ において、公共政策は実験によって改良されるという考えを提唱した。興味深いことに、キャンベルはこのコンセプトをデューイ (John Dewey, 1859-1952) から得ている (Oakley 2000 p.321)。

(2) 貧困との闘いと教育政策研究

社会科学における実験が本格的に行われるようになったのは、1960 年代から 1970 年代前半であり、特にジョンソン大統領が 1964 年に「偉大な社会」構想のもとに提唱した「貧困との闘い」のための様々な社会政策立法や、それに伴うプログラム評価が背景にあった。

教育分野においては、貧困との闘いにおいて果たす役割が重要であることから、1964 年の経済機

会法に基づくヘッドスタート計画⁹⁾や、1965年の初等中等教育法に基づくプログラムなどが実施されるとともに、これらに対する実験デザインを含む評価研究が行われた。しかし、教育政策研究における実験は、時間的制約やデザインあるいは実施上の問題があった事例も多かったようである (Vinovskis 2009 p.53-54)。

この時期の教育政策研究としては、非実験デザインであるが、1964年の公民権法に基づく教育機会の均等に関する調査¹⁰⁾がコールマン (J. S. Coleman) によって行われ、1966年に発表されて教育社会学に大きな影響を与えた。また、1965年に教育局に国立教育統計センター (NCES) が設置され、1969年から児童生徒の学力を全国規模で客観的に測定するための学力調査 (National Assessment of Educational Progress: NAEP) が定期的に行われるようになるとともに、1972年から全国的な縦断調査として National Longitudinal Study of the High School Class of 1972: NELS-72 が始められ、その後各種の縦断調査が行われるようになった¹¹⁾ ことなどは、研究および研究基盤として重要である。なお 1972年には国立教育研究所 (National Institute of Education :NIE) が設立されている。

(3) ペリー就学前計画

こうした中で、幼児教育分野において RCT として行われたペリー就学前計画は、教育研究としても教育プログラムとしても極めて優れたものであり、国際的にも大きな影響を与えている。

ペリー就学前計画は、1962年からミシガン州イプシランティ (Ypsilanti) 学区の課長であったウェイカート (David Weikart, 1931-2003) によって始められた。いわば政策レベルと実践レベルの中間レベル (meso-level) の研究と言える。この研究のサンプルは小さく全国的状況を表すものではないが、デザインは RCT であり、かつ約 40 年間もの追跡調査が行われたことにより、極めて影響力の強いランダム研究となった。調査は、3歳から4歳のアフリカ系アメリカ人の恵まれない子供たち 123 人がランダムに実験群と対照群 (統制群)¹²⁾ に分けられ、実験群には、午前中の学校での教育と午後からの教師による家庭訪問を含む介入が行われた。そして、5歳時点での就学準備、14歳時点での学校への出席や成績、ハイスクールの卒業率、40歳時点での収入や犯罪率などが比較され、いずれも実験群が優れた結果となっている。なお、1970年にハイ・スコープ教育研究財団 (HighScope Educational Research Foundation)¹³⁾ が設立され、政府や助成財団などからの資金によって、研究やその成果の普及などが行われている。

幼児教育分野の実験的研究では、ペリーの他にノースカロライナ州チャペルヒルのアベセダリアン (Abecedarian) (1972-) とイリノイ州シカゴの CPC (Child Parent Centers) (1985-) が著名である。このうちアベセダリアンの研究デザインは RCT であるが、CPC は社会経済状況や人種構成が同等のものをマッチングする準実験デザインである。なおセサミストリート研究においても RCT が用いられた。

諸外国でも同様の研究が行われるようになり¹⁴⁾、1999年に開始された英国のシュアスタート (Sure Start) などの政策に影響を与えた。特に、ペリー就学前計画において行われた 27 歳時点の費用便益分析 (cost benefit analysis)¹⁵⁾ は、プログラムの費用 1 ドルあたり 7.16 ドルの公財政支出が節減されるというもので、この分析は経済学者のヘックマン (James J. Heckman, 1944-)¹⁶⁾ や OECD (2006 pp.36-8) などが根拠とし、わが国でも文部科学省の「今後の幼児教育の振興方策に関する研究会」が 2009 年の中間報告において引用している¹⁷⁾。

(4) 初期の重要な RCT

1960年代から70年代前半には、RCTが盛んに行われた。このうち特に注目すべきものには次のものがある。

i) ピグマリオン効果

心理学者のローゼンタール (Robert Rosenthal, 1933-) によって1964-66年に実施された。児童の成績に対する教師の期待が、実際に児童に効果を与えるかどうかに関する著名な研究であり、「教師期待効果」という考え方は教員養成などに大きな影響を与えている。

この効果を実験で確認するためには、実験群と対照群を設けるだけでなく、教師にその事実をわからないようにする必要があるので、工夫されたデザインで実施された。実験方法への批判もあったが、数多くの追試や類似の研究が行われ、メタアナリシスも行われている。なお、ローゼンタールによるメタアナリシスはグラスよりも早く行われており、メタアナリシスのような命名をしていれば、ローゼンタールが提唱者になっていたであろうとグラスは述べている (Glass 2000)。

ii) ハーバード・プロジェクト物理

ハーバード・プロジェクト物理は、ハイスクールにおける物理の履修者の急減への対策として実施され、理系の専門に進まない生徒向けのカリキュラムが開発された。カーネギー財団や連邦教育局などの助成による。

1967年から1年間、RCTによって検証が行われた。ランダムに選ばれた高校の物理教師のうち参加可能であった53人が実験群と対照群にランダムに分けられ、実験群はハーバード・プロジェクト物理カリキュラムのサマーコースを受けた。1年後に生徒を評価したところ、両群に成績の差は見られなかったが、物理に対する満足度や関心は実験群の方が高く、カリキュラムが意図した目的の達成が確認された。全国レベルで行われた初期のカリキュラム評価の代表である。

iii) 標準テストの影響

アイルランドの教育研究センター (Educational Research Centre) のケラガン (T. Kellaghan) と、米国のマダウス (G. F. Madaus)、エレイシャン (P. W. Airasian) によって1973-77年に実施され、この分野の古典的な研究となっている。

標準テストが実施されていなかったアイルランドにおいて、全国の小学校から175校が層別にランダムに選ばれ、①標準テストを実施し、結果を教師にフィードバックする実験群、②標準テストを実施するがフィードバックのない対照群、③標準テストを実施しない対照群が設けられた。学校、教師、児童 (第2~6学年)、保護者について調査が行われた。学校に対する影響はほとんどなく、成績との単純な関係は見られなかった。教師と児童はテストを公正・正確なものと受け止め、フィードバックを受けた教師はより積極的に受け止めた。保護者に対してはほとんど影響がなかった。また、実験群にもフィードバックを受けなかった対照群にも教師期待効果が認められた。この研究は人工的な条件のもとでの実験であり、教師はハイステイクスではないテストの結果はあまり利用しないということであろう (Gipps 1994 p.37, 鈴木訳 p.51)。総じて、この「標準テストの影響」は学校にとって有害な面はなく、有益であったとの結果であったが、この研究で測定されたデータは、教師の評定とテストのスコア、アンケートの回答に限られていたという批判もある (Alderson and Wall 1993 pp.122-3)。

3. 学級規模縮小の効果

(1) メタアナリシス

i) 米国教育学会長グラスの講演（1976）——メタアナリシスの提唱

医学統計学シリーズの一冊をひもとくと、「メタアナリシスという言葉の誕生とそれにとまなう方法論の研究が盛んであったのは社会科学の世界、特に、教育学の研究においてである」と書かれている（丹後 2002, p.3）。このように EBM（エビデンスに基づく医療）の関係者にはよく知られていることであるが、メタアナリシスの方法が提唱されたのは、教育心理学者のグラス（Gene V. Glass, 1940-）による 1976 年の米国教育学会（American Educational Research Association, AERA）の会長講演であった¹⁸⁾。その経緯については、サイエンスライターのアート（Morton Hunt）が一般読者向けに興味深く深く描いており（Hunt 1997 pp.20-44）、グラス自身も説明している（Glass 2000）。

講演は、‘Primary, Secondary, and Meta-Analysis of Research’（Glass 1976）と題するもので、primary analysis が収集されたデータの分析、secondary analysis がそれらのデータの再分析であるのに対して、meta-analysis はいくつもの分析を収集し、それらを「包括的に」分析すること、という意味で名付けられたものである。

グラスはスミス（Mary Lee Smith）とともに、論争のあった心理療法の効果について多数の研究を収集し、統計的に統合することで、心理療法が有効であるという結論を得ていた¹⁹⁾。グラスは講演でこの研究を紹介し、多数の量的研究を統計的に統合するメタアナリシスが教育研究において必要であることを提言した。

ii) 学級規模研究へのメタアナリシスの適用

グラスは講演で、メタアナリシスを必要とする課題の一つとして学級規模をあげていたが、グラスとスミスは、極西部地方教育開発研究所²⁰⁾のカーエン（Leonard S. Cahen）から「学級規模と指導プロジェクト」（The Class Size and Instruction Project）への参加を依頼されて、この課題に取り組むことになった。学級規模の縮小が生徒の成績等に与える効果は、教育政策研究の分野において最も重要な課題の一つとして長い間論じられてきたものである。

研究報告書は 2 冊作成されている²¹⁾。第 1 の報告書『学級規模と成績の関係についての研究のメタアナリシス』（Glass and Smith, 1978）は、1900 年以降の 77 の実験デザインによる研究²²⁾における、学級規模の大小と生徒の成績の高低の 725 の比較をメタアナリシスしたものである。得られた結果は、学級規模が縮小するほど成績は上昇する傾向にあり、1 クラス 20 人程度以下になると顕著に上昇するというもので、グラス・スミス曲線として知られる。第 2 の報告書『学級規模の学級プロセスに対する関係、教師の満足度と生徒への影響：メタアナリシス』（Smith and Glass, 1979）は、1925 年以降の 59 の研究における 371 の比較をメタアナリシスしたもので、児童生徒の感情的な側面への効果、教員に対する効果、教授過程への効果は、いずれも少人数学級の方が大きいという結果が得られた。

(2) 大規模 RCT によるテネシー州の STAR プロジェクト

i) プロジェクトの概要

メタアナリシスは、その対象となる一次研究によって制約されるものであり、グラス・スミスによるメタアナリシスの後も論争は続いた。実証研究やメタアナリシスも様々に行われたが、学級規

模の効果の問題に決着をつけるためには、大規模な実験が必要であると考えられるようになっていた。しかし、そのような実験は、巨額の資金と政治的合意が必要であるため、容易には実施できない。

これを実現したのがテネシー州の STAR (Student/Teacher Achievement Ratio) プロジェクトであり、同州は 1985-89 年の 4 年間に 1,200 万ドルを投入して実施した。この実験は、最も重要な教育調査であると言われている (Mosteller 1995 p.113)²³⁾。

この調査は、約 6,400 人の幼稚園児と 300 人の教員を、13-17 人の少人数クラス、22-25 人の普通クラス、普通クラスに補助教員を加えたクラス、の 3 つのグループにランダムに割り付け、クラスの分け方は小学校第 3 学年まで 4 年間維持された。読みと算数の成績を比較すると、小規模クラスの方が他の 2 グループよりも高いという結果が得られた。またこの効果は、小規模クラスへの在籍が早いほど、また長いほど大きく、さらに白人・非都市部よりもマイノリティ・都市部の方が大きかった。

プロジェクト終了後、州教育省の資金で、1989-90 年度から 1994-95 年度まで追跡調査 (Lasting Benefits Study) が行われた。その結果は、少人数学級の効果は 4 学年以上で通常学級となっても長期的に持続するというものであった。なお、その後の調査分析は研究財団 (Health and Education Research Operative Services: HEROS)²⁴⁾ によって行われてきた。

ii) 調査を実現可能にした要因

この調査を実現可能にした要因をとって、次の 4 つがあげられている (Ritter and Boruch 1999)。
①州の政治環境の機が熟していたこと。『危機にたつ国家』以降、各州で教育改革が進められていたが、同州では、後に連邦教育省長官となるアレグザンダー (Andrew Lamar Alexander) 知事が教育を最優先事項としていた。同州の 1984 年の教育改革法は、当時、最も包括的なものであったが、原案に含まれていた学級規模縮小は極めてコストがかかるため、審議過程で削除されていた。②研究のバックグラウンドのある提唱者が、教員組合の支持を得て熱心に活動したこと。元州教育協会会長で州立大学教授のベイン (Helen Pate Bain) は、自ら実施した小規模実験とインディアナ州の Project Prime Time における横断調査の結果をもとに、議会に対して熱心に学級規模縮小を働きかけていた。③社会科学のバックグラウンドのある政治家が関心を持ったこと。社会学の大学教授の経歴を持つ州議会議員のコブ (Steve Cobb) は働きかけを受けて、3 年間のインセンティブ・グラント法案を提出した。同法案はパイロット調査に縮減する修正を経て 1985 年に成立した。④複雑かつ困難な調査の設計・分析ができる研究者が存在したこと。州法にはランダム化実験の規定はなかったが、州教育省はヴァンダービルト大学の教育社会学者のフォルガー (John Folger) と同大学で RCT を手掛けていた心理学者・社会政策学者のビクマン (Leonard Bickman) を中心とする州内主要 4 大学の研究者に調査の設計・分析を依頼した。

これらの要因は、僥倖とも言うべき組合せが実現したものであり、一般化することはできないかもしれないが、政治学者のキングダン (John W. Kingdon) が言うように、問題 (学級規模縮小) と政策 (パイロット・プロジェクトと RCT) と政治 (教育改革) の 3 つの流れが合流して「政策の窓」が開いたと考えられる。この場合、最も重要な役割を果たしたのは、これらを仲介する役割を果たしたベインであった。

iii) STAR プロジェクトが政策に与えた影響

研究結果の政策への影響として、テネシー州では、1989-90 年度から 17 の小規模地方学区の K-3 の学級規模を縮小する Project Challenge が実施され、その効果も調査され確認されている。さらに 1992 年には、Education Improvement Act が成立し、学級規模が縮小された²⁵⁾。

テネシー州の STAR プロジェクトは他州にも大きな影響を及ぼし、半数以上の州で何らかの学級規模の縮小が行われた。ウィスコンシン州では 1996-7 年度から 5 年間の Student Achievement Guarantee in Education (SAGE) プログラムを開始した。プログラムは他の改革案を含むが、学級編制については、恵まれない地域の学校の低学年に 15 人学級を増やそうとするものである。これは社会経済状況や人種構成などが同等な学校を比較する準実験デザインで、15 人学級に教師 1 人のほか、30 人学級に教師 2 人など 6 つのバリエーションが試みられた²⁶⁾。結果が公表された後、プログラムは条件に合致する学区すべてに拡大されている (Biddle and Berliner 2008 p.91)。これに対し、1996 に開始されたカリフォルニア州の学級規模縮小 (Class Size Reduction: CSR) は議論を招いた。学力低下への対応の必要もあり、毎年 15 億ドルを投入して幼稚園から小学校 3 学年までのすべての学級を 20 人に縮小しようとするものであった。しかし、大量の教員採用が必要となったため、無資格を含む経験不足の教師の採用や教室の不足などが問題になり、プログラム評価によると、学級規模縮小の効果は限定的であった。

連邦レベルでは、クリントン大統領が 1998 年の一般教書演説²⁷⁾において、今後 7 年間で 10 万人の教員を増員し、小学校 3 学年までの学級規模を 18 人に縮小するという学級規模縮小計画 (Class Size Reduction Program) を提案した。連邦教育省は学級規模縮小の効果に関する研究や州の取組をレビューした『学級規模縮小：何がわかっているか』(U.S. Department of Education 1998) を発表した。学級規模縮小のための連邦政府の補助金は 1998 会計年度から措置されたが、ブッシュ政権成立後の 2002 会計年度から、NCLB 法 Title II の教員の質ブロックグラント (Teacher Quality block grant) に統合された。

国際的には、オーストラリア、ニュージーランド、スウェーデン、英国など多くの国で学級規模研究が実施されるようになってきている。我が国では国立教育政策研究所によって実証的研究が行われてきたが、近年は実験的要素を加味した研究も行われている。また都道府県独自の少人数学級も実施されている。これらの結果は、米国の動向に関する情報とともに文部科学省の審議会などに提出され、政策立案に活用されている (小川 2010)。

(3) 今後必要とされる研究

学級規模の教育効果に関しては、なお様々な議論が行われているが、現在では、STAR プロジェクトの結果が前提となっている。これは、STAR プロジェクトが数多くの研究事例の一つに留まるものではなく、大規模な RCT によって頑健なデータを提出したことによる。このように、この研究が決定的でランダムマーク的な研究であるという認識は、1990 年代に次第に高まってきた (Grissmer 1999)。もっとも、テネシー州で実施された大規模 RCT と追跡調査は、唯一の研究であり再現性に乏しい。しかし、これまでに行われてきた小規模 RCT や大規模な計量分析も同様の結果を示しており、ヘッジズは、このように相互に支えあっている 3 つの異なる研究方法による結果を統合すると、学級規模の 24 人から 15 人への縮小の効果は極めて大きいわけではないが、持続的な効果を持つと言える結論付けている (Hedges, 2000)。

今後も各種の条件 (学年、教科、指導方法等の相違) における実証研究、とりわけ RCT が必要で

あろう。また、少人数学級がなぜ教育効果を生み出すのかというメカニズムについても研究が必要である（小川 2010）。なお、準実験デザインの一つとしてキャンベルとスタンレーがあげていた回帰分断分析（regression discontinuity analysis）（Campbell and Stanley 1966 pp.61-64）は、強いエビデンスとして認められているが²⁸⁾、これが学級規模研究（Angrist and Lavy 1999）に用いられていることは注目される²⁹⁾。

4. メタアナリシスの発展

グラスの提唱の後、メタアナリシスは各分野で急速に活用されるようになり、技法も洗練されるようになった。また体系的な教科書は、1984年にローゼンタール³⁰⁾、同年に社会・発達心理学者のクーパー（H.M. Cooper³¹⁾、1985年に教育統計学者のヘッジズ（L.V. Hedges）とオルキン（Ingram Olkin³²⁾）によって書かれ、ハンドブックはクーパーとヘッジズによって1994年に編集された³³⁾。ここでは、ヘッジズとクーパーによるメタアナリシスの事例を紹介する。

(1) 学校財政支出と教育効果

1968年のワイズ（A.E. Wise）の“Rich Schools, Poor Schools”の影響や学校財政制度訴訟などにより、各州においては学区間財政格差の縮小が図られ、教育財政は増加を続けていた。しかし公立学校の教育に対する信頼の低下などから、特に1980年代から、学級規模の縮小を含めた財政支出に対する批判が出されていた。その代表的論客が教育経済学者のハヌシェク（Eric A. Hanushek）であり、教育生産関数に関する187の研究をレビューした結果、学校財政支出と児童生徒の成績との間には体系的な関係がみられないと結論付けた（Hanushek 1989）。

これに対しヘッジズは、ハヌシェクが研究を統合するにあたって採用した票集計（vote counting）の手法（一次研究の数で判断する）を批判し、ハヌシェクと同じレビュー対象をメタアナリシス（個々の研究の効果量（effect size）を重み付け平均）した結果、学校財政支出と児童生徒の成績には体系的で十分大きいプラスの関係があることを示している（Hedges et al. 1994）。この問題については、この後も様々な形で議論されている。

(2) 宿題の効果

クーパーは、国立科学財団（National Science Foundation: NSF）の助成を得て、米国で長く議論されてきた宿題の効果についてメタアナリシスを実施した。これは政策立案者や実践者だけでなく、保護者にも直接の影響を与えた。宿題をテーマとする約250の論文のうち、実証的な約120の研究を統合した結果、宿題を与えられた生徒の成績は、与えられなかった生徒よりも良好であり、長時間を要する宿題の方がそうでない宿題よりも良好であった。また効果は学校段階があがるほど高い（Cooper 1989）。さらにその後の研究についても分析が行われている（Cooper et al. 2006）。

この研究をもとにして、クーパーは1989年に『宿題』、1994年に『宿題をめぐる戦い：行政官、教師、保護者の共通の基盤（2007年第3版）』を出版している。また、連邦教育省は1995年に保護者向けの冊子を作成し（2002年、2005年に改訂）、ホームページ³⁴⁾も設けている。

5. キャリア・アカデミー研究

(1) 1980～90年代の動向

1980年代には、学校におけるプログラムのうち、児童生徒のメンタルヘルスの改善プログラムの効果や、暴力やたばこ・薬物・アルコールの防止に関するプログラムの効果など、健康教育や非行防止に関連した実験が、保健医療や刑事司法の研究として行われてきたが、教育研究としてはあまり行われてこなかった。これは、教育においては他分野よりもRCTの実施が困難ではないかと、教育研究者が考える傾向にあったことや、ヘッドスタートや初等中等教育法のプログラムの効果を検証するために、1970年代以前に実施された実験的な評価の結果が思わしくなかったことなどによると考えられる (Cook and Payne 2002 p.162)。この点でもテネシー州のSTARプロジェクトは際立っている。

また、1980年代の教育省の研究予算の削減が他省に比べて大きく、1985年に国立教育研究所(NIE)が廃止されたことなども影響していると思われる。しかし、教育省の政策評価としては、例えば中退防止プログラムの評価(1989-96)やUpward Bound(低所得層の家庭等の高校生の大学進学を支援するプログラム)の評価(1994-2009)がRCTによって行われている³⁵⁾。このほかニューヨーク市(1997-99)、オハイオ州デイトン市(1998-2000)、ワシントンDC(1998-2000)における民間財団によるスクールバウチャーが生徒の成績に与える影響について、RCTが行われている(Howell et al. 2002)。

(2) キャリア・アカデミー研究

このような中で教育分野の重要なRCTとしては、1993年から2008年までの間、労働力実証研究協会(Manpower Demonstration Research Corporation: MDRC)³⁶⁾によって実施されたキャリア・アカデミー研究(The Career Academies Study)がある。キャリア・アカデミーとは、ハイスクールの「学校内学校(a school-within-a-school)」として、募集に応じて選ばれた生徒に対し、職業科目とアカデミック科目とともに、夏季雇用の機会等が与えられるコースである。

この研究は、キャリア・アカデミーのコースが、生徒の成績、卒業や、進学・就職後の状況など長期的に与える影響を評価したものである。8都市の9校で約1,760人の第8・9学年の生徒がキャリア・アカデミー・コース959人と従来コース805人にランダムに分けられ、それぞれのコースにはハイリスクとローリスクのサブ・グループが含まれている。2年間のプログラムと8年間の追跡調査が行われた。研究費は、連邦教育省と連邦労働省と17の財団から提供された。調査結果のうち、例えば第12学年終了時における状況をみると、キャリア・アカデミー・コースのハイリスクの生徒は、従来コースのハイリスクの生徒に比べて、中退率の低下、出席率の向上、修業年限での卒業率の高さがみられたが、数学と読解の試験を受けた490人の成績については、両コースの差は見られなかった。

この研究は、志願者の中から選抜するというキャリア・アカデミーの仕組みの中で、選抜の方法を抽選とすることによって、対象者の同意を得つつ、ランダム割付による実験を可能にしたものである。社会政策の分野で数多くのRCTの実績のあるMDRCが、教育分野で初めて行った本格的なRCTであり、教育研究にとって意義は大きい。

6. 読みの研究から NCLB 法へ

(1) 読みの指導と「科学的根拠に基づく読みの研究」³⁷⁾

i) 「読みの戦争」(reading wars)

米国では長い間、読みの指導方法について論争があったが、実験デザインによる研究においては、単語の綴りと読みの関係を指導するフォニクス (phonics) などの方法に効果があるという結果が有力であった。しかし 1980 年代には、フォニクスなどを重視しないホール・ランゲージ (whole language) 運動が学校現場に広がり、論争は「読みの戦争」と呼ばれるようになった。カリフォルニア州の場合、1987 年のフレームワークでホール・ランゲージの考え方を採用したが、同州の 1992 年の NAEP の成績は全国レベルより大幅に低いことが判明し、1994 年にはさらに低下した。成績低下とホール・ランゲージの関係は明確ではないが、この戦争は政治の舞台に拡大した。カリフォルニア州は 1997 年にフレームワークを改訂し、同州をはじめ 18 の州で同年までにフォニクスなどを重視する法律が成立した。科学的根拠を持たなかったホール・ランゲージ運動が、政治の舞台で説得力を持たなかったためと言えよう (Kim 2008 pp.99-100)。

ii) 「科学的根拠に基づく読みの研究」

連邦レベルでは、1990 年代初めから、生理心理学者ライアン (G. Reid Lyon) が国立小児健康発達研究所 (NICHD)³⁸⁾ においてこの問題に取り組み、有効な介入に関する実験研究や縦断研究に対する助成プログラムが行われていた。

第 4 学年の児童の 40% が基礎レベルの読みができないという 1992 年の NAEP 調査の結果に続き、第 12 学年の reading の成績が有意に低下したという 1994 年の結果が明らかになると、読みの指導の問題が大きく取り上げられるようになった。クリントン大統領は、教育を最優先事項とした 1997 年の一般教書演説³⁹⁾ において、すべての 8 歳児が読めるようになることなど 3 つの目標を示すとともに、10 項目の行動方針の 3 番目で調査結果を引用して読みの指導の重要性を強調した。翌 1998 年には、全米研究評議会 (National Research Council: NRC)⁴⁰⁾ から報告書⁴¹⁾ が出され、ほとんどの読みの障害は防げることや、読みの指導の効果を高めるための研究の重要性などが提言された。同年には連邦議会において Reading Excellence Act が制定され、読みに関する教員研修や課外指導などに対する助成プログラムが定められた。この法律の中に「科学的根拠に基づく」(scientifically based) という言葉が初めて規定され、助成を受けるプログラムは「科学的根拠に基づく読みの研究」(scientifically based reading research: SBRR) に基づくことが要件とされた。ここでも問題 (読みの戦争と調査結果の発表) と政策 (科学的根拠に基づく研究) と政治 (教育改革) の流れが合流して「政策の窓」が開いたと言えよう。

iii) ナショナル・リーディング・パネルと Reading First

この間 1997 年には上院予算委員会において、読みの指導においてどの実践が有効かを検討するために、NICHD にナショナル・リーディング・パネル (National Reading Panel: NRP) を設けることが要請された。同パネルは 8 つの課題についてレビューを行い、メタアナリシス⁴²⁾ や質的分析による研究統合を行い、2000 年に報告書⁴³⁾ を提出した。

2001 年に成立した連邦法「落ちこぼれを作らないための初等中等教育法」(No Child Left Behind (NCLB) Act) は、Reading First および Early Reading First などの助成プログラムを規定し、助成を受

ける州や学区のプログラムが SBRR に基づくことを求めている。NRP の影響力は大きく、その知見は SBRR に該当すると考えられるため、多くの州や学区では、NRP 報告書が読みの指導において重要としている 5 つの要素⁴⁴⁾を用いてプログラムを組むようになっている (Kim 2008 p.103 p.276 note 64、海津 2006 p.23)。

さらに 2002 年には、国立リテラシー研究所 (National Institute for Literacy)⁴⁵⁾ が NICHD の協力を得て、NRP の対象とならなかった 0 歳から 5 歳までの子どもを対象とする National Early Literacy Panel を設け、NRP と同様にメタアナリシスなどによる科学的研究の統合を実施して、2008 年に報告書を提出している (Shanahan and Lonigan 2010)。

また、2004 年の Individuals with Disabilities Education Acts: IDEA の改正においては Response To Intervention: RTI という考え方が採用され、学区は LD の判定にあたって、「子どもが科学的かつ研究に基づいた介入に反応するかどうかを判定するプロセス」(a process that determines if the child responds to scientific, research-based intervention)を用いてもよいことが新たに定められた (海津 2006 pp.26-7)。

iv) 数学戦争

なお、算数・数学の分野でも数学戦争と呼ばれる論争が激化し、1998 年に全米科学財団 (NSF) と連邦教育省が米国科学アカデミーに調査を依頼して同様のパネルが設けられ、2001 年に報告書 “Adding It Up: Helping Children Learn Mathematics” が出された。さらに 2006 年には、ブッシュ大統領により、算数・数学の指導と学習に関するエビデンスをまとめ、算数・数学の学力改善のための提言をすることを主な目的として、全米数学諮問パネル National Mathematics Advisory Panel が設けられ、2008 年に報告書が提出されている。これらのパネルにおいても研究がレビューされ、メタアナリシスなどによる統合が行われた。

(2) NCLB 法と「科学的根拠に基づく研究」

クリントン大統領の科学技術諮問委員会の報告を受けて、NSF と教育省、NICHD は共同で 1999 年から 2005 年までの間、規模が大きい実験ないし準実験の方法論などによる科学・数学・読みに関する教育研究を助成する、教育研究イニシアティブ (Interagency Education Research Initiative: IERI) を実施していた (Boruch et al. 2009 pp152-3)。

2001 年にはブッシュ政権のもとで NCLB 法が超党派で成立した。同法は、「科学的根拠に基づく読みの研究」の考え方を他のプログラムに広く及ぼしており、同法に基づく助成を受ける州や学区のプログラムは、「科学的根拠に基づく研究」(scientifically based research: SBR) に基づくことが求められている。SBR とはどのような研究かについて、同法第 9101 条(37)(A)は、「教育活動及びプログラムに関連する信頼性が高く有効な知識を得るための厳密な、体系的な及び客観的な手続の適用を伴う研究」であるとし、かつ、同条(37)(B)に掲げる研究⁴⁶⁾が含まれると定義している。

連邦議会では、教育研究の質に対する批判に関連して、引き続き連邦教育省の教育研究に対する助成の在り方が議論され (Vinovskis 2009 pp.66-68)、2002 年には、連邦教育省の教育研究改善局を改組して教育科学研究所 (Institute of Education Sciences: IES) を設置する教育科学改革法 (Education Sciences Reform Act of 2002) が成立した。IES は、NCER や NCES などのセンターやプログラムを有しているが、新たに WWC 情報センター (What Works Clearinghouse: WWC) を開設するとともに、RCT による研究への助成、若手研究者の育成などのプログラムを実施している (Vinovskis 2009 p.69)。

RCTについては、連邦教育省の助成研究のみの数字であるが、2000年に実質的には1件のみであったのが、2001～06年には教育科学研究所（IES）から少なくとも60件以上が助成されている（Boruch et al. 2009 p.153）。また、WWC情報センターは、RCTを中心としてSBRに該当する質の高いエビデンスをレビューし、提供することを目的としている（田辺2006、OECD2007 pp.53-61（豊（訳）2009 pp.91-104））。

7. おわりに

米国の教育研究においては量的研究が大きな役割を果たしている。しかし学級規模研究の場合でわかるように、計量的方法はもとより、ランダムでない、あるいはコントロールの十分でない実験的方法では、必ずしも説得力のある結論が得られるわけではない。1985年にはじまるテネシー州のSTARプロジェクトが大規模RCTであることによって決定的な研究となって以降、RCTの強力さが教育研究者や政策立案者の間に広まったことが、1990年代後半からの「科学的根拠に基づく研究」につながったのである。

我が国の教育研究においては量的研究が少なく、RCTはほとんど実施されていない。教育分野のRCTについては、倫理上の問題などの課題があるが（岩崎2010b pp.15-19）、教育分野においてもキャリア・アカデミー研究のデザインや、その後の米国におけるRCTの増加に見られるように、RCTが可能な場合は少なくない。

もちろんRCTの実施が困難な場合も多いため、様々な量的研究の手法が必要である（Weiss2002 p.198-202）。また、これらの前提としてcross-sectionalな調査統計の充実はもとより、評価研究や測定研究の進展と相まって開発される学力調査や、我が国の教育行政による統計ではほとんど行われていない縦断調査などの調査研究を進める必要もある。資金も必要であるが、人材の育成や技術基盤の構築も求められる。一方、RCTをはじめとする量的研究を重視することは、質的研究や概念的研究を軽視するものではない。政策立案における判断にあたっては、厳密で頑健な量的研究が決定的であること、すなわち研究の厳密性や頑健性が政策との関連性・妥当性の重要な要素であることも多いが、関連性・妥当性には様々な要素があり、政策課題の把握や理解には、質的研究や概念的研究の成果も有効である。多様なデザインの質の高い研究が必要とされるのであって、リサーチ・ポートフォリオ（research portfolio）のバランスが重要である。

筆者は以前、エビデンスに基づく政策と実践について、研究者、政策立案者、実践者のそれぞれの課題や、保護者、住民、国民の理解などに簡単に触れたことがあるが（惣脇2009 pp.7-10、2010 pp.163-4）、今後、機会を改めてこれらの課題を整理、検討することとしたい。教育の質を高めるためには、研究が大きな役割を果たしうるが、教育の質に対する批判には、教育研究の質に対する批判が内在している。教育の改善を図るとともに、教育制度のアカウンタビリティを高め、教育研究の質を高めるために、米国の事例は大いに参考になると思われる。

注

- 1) RCTは「無作為化対照実験」などと訳されることも多いが、本稿では、エビデンスに基づく医療において、近年用いられるようになった「ランダム化比較試験」とした。
- 2) meta-analysisは「メタ分析」と訳されることも多いが、本稿では、エビデンスに基づく医療において用いられている「メタアナリシス」とした。

- 3) RCTの事例については、主として Nave et al. (2000)、Greenberg and Shroder (2004)、Stufflebeam and Shinkfield (2007)によるほか、調査実施機関等のホームページの情報を参照した。
- 4) このテキストは Gage, N. L. (ed.) (1963) “Handbook of research on teaching: a project of the American Educational Research Association”の第5章として執筆され、同章のリプリントが1966年に単行本として発行された。
- 5) フィッシャー (R. A. Fisher, 1890-1962) は、英国の統計学者・遺伝学者。1920年代にロザムステッド農事試験場でRCTを開発した。
- 6) Oakley (1998)によると、米国の教育心理学者ソーンダイク (Edward L. Thorndike, 1874-1949) とウッドワース (Robert S. Woodworth, 1869-1962) が1901年に対照(統制)群の必要性を見出し、英国のロンドンの視学官であったウィンチ (W. H. Winch, 1864-1935) による1908年の研究は、実験群と対照(統制)群に事前・事後のテストを行った最初の実験の一つであった。しかし、この時期に行われた実験は、完全なRCTとは言えないようである。
- 7) American Psychologist vol.24 issue 4に発表され、著作集に収録されている (Campbell, D. T. 1988 pp.261-289)。
- 8) 1971年にラッセル・セージ財団 (Russell Sage Foundation) にプロポーザルとして提出されたもので、著作集に収録された (Campbell, D. T. 1988 pp.290-314)。
- 9) 所管は経済機会局 (Office of Economic Opportunity) であったが、児童に対する包括的なプログラムであることから現在の所管は保健社会福祉省である。
- 10) 所管は法律の規定により連邦教育局であった。
- 11) NLS-72は、1972年当時の高校生を1986年まで追跡した調査である。次いで、1980年当時の高校生を1992年まで追跡した High School and Beyond :HS&B が行われ、この種の大規模縦断調査は2009年に開始された HSLs-09 まで、5本が行われている。このほか、幼児や初任教師などについての縦断調査も行われるようになってきている。
- 12) 社会科学では control group は「統制群」と訳されているが、本稿では、エビデンスに基づく医療において用いられている「対照群」とした。
- 13) <http://www.highscope.org>
- 14) オランダでは2000年に2つの実験的な育児施設プログラムの大規模な評価を完了し、英国ではプログラムの包括的な調査に従事していることが報告されている (OECD2002、御園生 (監訳) p.47)。
- 15) 費用便益分析 (cost benefit analysis) は、費用効果分析 (cost effectiveness analysis) のうち、効果を金銭に換算して評価するものである。
- 16) Heckman, J. (2000) ‘Policies to Foster Human Capital’, Research in Economics, vol.54 no.1, pp.3-56 など。なおヘックマンは2000年にノーベル経済学賞を受賞したことで著名である。
- 17) 同研究会『幼児教育の無償化について (中間報告)』は、「幼児教育は幼児の望ましい発達をもたらすという教育的効果のみならず、社会経済的効果を有しており、その波及効果は社会経済全体に及ぶ」「このことについては、近年、諸外国において、米国でのベリー就学前計画における研究・・・などで、質の高い幼児教育が、その後における成績の向上や進学率の上昇、所得の増大、犯罪率の減少をもたらすなど、教育的・社会経済的効果を有するとの実証的な研究成果が得られている。」「これらの研究成果では、小学校就学前に実施される幼児教育は、小学校就学後の教育投資の効果を増大させ、その効果は成人後まで及ぶと指摘されている。」と述べている。
- 18) もっとも、最初にメタアナリシスを行ったのはピアソン (Karl Pearson) で、1904年のことであった (Chalmers et al. 2002 p.14)。
- 19) 1977年に発表された論文 (Smith, M. L., & Glass, G. V. (1977) ‘Meta-analysis of psychotherapy outcome studies. American Psychologist’, 32, pp.752-760) は、精神分析や心理療法には効果がないと主張したアイゼンク (H. J. Eysenck) 以来の論争に決着をつける画期的なものであったとされる (丹野 2004 p.50)。
- 20) 極西部地方教育開発研究所 (Far West Laboratory of Education Research and Development: FWL) は、初等中等教育法に基づいて連邦教育局の助成を受ける研究所として、1966年に発足した地域教育研究所 (Regional Education Laboratory) の一つ。1995年に南西部の研究所と統合して WestEd となった。
- 21) 報告書に基づいて、1979年と1980年に論文が発表され、1982年に単行本『学校のクラスサイズ: 研究と政策』(School Class Size: Research and Policy) が出版されている。
- 22) コントロールされていない実験デザインは含まれているが、非実験デザインであるコールマン・スタディやIEA国際学力調査は含まれていない。
- 23) “one of the most important educational investigations ever carried out and illustrates the kind and magnitude of research needed

- in the field of education to strengthen schools”
- 24) <http://www.heros-inc.org>
 - 25) K-3 が 20 人、4-6 学年が 25 人、7-9 学年が 30 人となった。
 - 26) 学級規模研究では、STAR プロジェクトに次いで、このプログラムのデータセットがよく利用されている (山下 2008)。
 - 27) <http://clinton4.nara.gov/textonly/WH/SOTU98/address.html>
 - 28) 米国の WWC 情報センターにおいては、RCT のほか、準実験デザインのうち回帰分断分析と単一被験者モデル (single subject design) によるものが、エビデンスとして認められている (OECD 2007 p.57 (豊 (訳) 2009 p.97))
 - 29) 慶応義塾大学経済学部 赤林英夫教授のご教示による。
 - 30) “Meta-Analytic Procedures for Social Research”。1991 年に改訂版が出されている。
 - 31) “The Integrative Research Review, A Systematic Approach”。2010 年に第 4 版 “Research synthesis and meta-analysis: a step-by-step approach” が出されている。
 - 32) Hedges & Olkin (1985) “Statistical Methods for Meta-Analysis”。
 - 33) “The Handbook of Research Synthesis”。2009 年に編集に Valentine, J. C. を加え、第 2 版 “The Handbook of Research Synthesis and Meta-Analysis” が出されている。
 - 34) <http://www2.ed.gov/parents/academic/help/homework/index.html>
 - 35) この 2 つの評価は、マスマティカ政策研究所 (社会政策分野で RCT を実施することを主目的として 1968 年に設立されたシンクタンク) に委託して実施された。 <http://www.mathematica-mpr.com>
 - 36) MDRC は、1974 年にフォード財団などによって設立された非営利の研究所であり、低所得者のための社会政策・教育政策などについての調査研究を実施している。 <http://www.mdrc.org>
 - 37) 主として Kim 2008、Ravitch 2000 (末藤ほか (訳) pp.485-494) による。
 - 38) 国立小児保健発達研究所 (National Institute of Child Health and Human Development: NICHD) は、米国保健社会福祉省の研究機関・研究助成機関として著名な国立保健研究所 (National Institutes of Health: NIH) の研究所の一つ。同研究所で、読み障害 (reading disability, dyslexia) を含むこの分野の実験研究プログラムを実施していたライアンは、連邦政府の政策立案において重要な役割を果たした。
 - 39) <http://clinton4.nara.gov/textonly/WH/SOU97/>
 - 40) 科学のおよび技術的問題について連邦政府に勧告する役割を有する米国科学アカデミー (The National Academy of Sciences) の傘下の組織である。
 - 41) “Preventing Reading Difficulties in Young Children”
 - 42) 例えばフォニクスに関して確認された 1,373 の研究のうち、基準を満たす 38 の研究についてメタアナリシスが行われ、低学年においては体系的なフォニクスが他の指導方法よりも有意に効果的であることが明らかにされている。
 - 43) “Report of the National Reading Panel: Teaching Children to Read”
<http://www.nationalreadingpanel.orf/Publications/nrp/publications.htm>
 - 44) 音素認識 (phonemic awareness)、フォニクス (phonics)、流暢性 (fluency)、語彙 (vocabulary)、理解 (comprehension) の 5 つの要素である。
 - 45) National Literacy Act of 1991 によって 1991 年に設立された連邦教育省の研究所。
 - 46) 次の 6 つが掲げられている。(i) employs systematic, empirical methods that draw on observation or experiment; (ii) involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn; (iii) relies on measurements or observational methods that provide reliable and valid data across evaluators and observers, across multiple measurements and observations, and across studies by the same or different investigators; (iv) is evaluated using experimental or quasi-experimental designs in which individuals, entities, programs, or activities are assigned to different conditions and with appropriate controls to evaluate the effects of the condition of interest, with a preference for random-assignment experiments, or other designs to the extent that those designs contain within-condition or across-condition controls; (v) ensures that experimental studies are presented in sufficient detail and clarity to allow for replication or, at a minimum, offer the opportunity to build systematically on their findings; and (vi) has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review.

参考文献

- Alderson, J. C. and Wall D. (1993) 'Does Washback Exist?', *Applied Linguistics* vol.14 no.2, pp.115-129
- Angrist, J. D., Lavy, V. (1999) 'Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement', *The Quarterly Journal of Economics*, vol.114, no.2, pp.533-575
- Biddle, B. J. and Berliner, D. C. (2008) 'Small Class Size and Its Effects' in Ballantine J. H. and Spade, J. Z. (ed.) "Schools and Society: A Sociological Approach to Education 3rd Edition", Pine Forge, pp.86-95
- Boruch, R. F., Weisburd, D., Turner III, H. M., Karpyn, A., Littell, J. (2009) 'Randomized Controlled Trials for Evaluation and Planning' in Bickman, L. and Rog, D. (eds.) "The Sage Handbook of Applied Social Research Methods, 2nd edition", Sage Publications, pp.147-181
- Chalmers, I., Hedges, V., Cooper, H. (2002) 'A Brief History of Research Synthesis', *Evaluation & the Health Professions* vol.25 no.1, pp.12-37
- Campbell, D. T., edited by Overman, E. S. (1988) "Methodology and Epistemology for Social Science: Selected Papers", The University of Chicago Press
- Campbell, D.T. and Stanley, J.C. (1966) "Experimental and quasi-experimental designs for research", Houghton Mifflin
- Cook, T. D. and Payne, M. R. (2002) 'Objecting to the Objections to Using Random Assignment in Educational Research' in Mosteller, F. and Boruch, R. (eds.) "Evidence Matters: Randomized Trials in Education Research", Brookings Institution Press, pp.150-178
- Cooper, H. M. (1989) 'Synthesis of Research on Homework', *Educational Leadership*, vol.47, no.3, pp.85-91
- Cooper, H.M., (2007) "The Battle Over Homework: Common Ground for Administrators, Teachers, and Parents: 3rd edition", Corwin Press
- Cooper, H. M., Robinson, J. C. and Patall, E. A. (2006) 'Does Homework Improve Academic Achievement? A Synthesis of Research, 1987-2003', *Review of Educational Research*, vol.76, no.1, pp.1-62
- Gipps C. V. (1994) "Beyond testing: towards a theory of educational assessment", The Falmer Press (鈴木秀幸 (訳) (2001) 『新しい評価を求めて テスト教育の終焉』 論創社)
- Glass, G.V. (1976) 'Primary, Secondary, and Meta-Analysis of Research', *Educational Researcher* vol.5 no.10, pp.3-8
- Glass, G.V., and Smith, M.L. (1978) "Meta-Analysis of Research on the Relationship of Class-Size and Achievement", Far West Laboratory for Educational Research and Development (ERIC: ED168129)
- Glass, G.V. (2000) 'Meta-analysis at 25', College of Education, Arizona State University, (<http://www.gvglass.info/papers/meta25.html>)
- Greenberg, D. and Shroder M., (2004) "The Digest of Social Experiments Third Edition", The Urban Institute Press
- Grissmer, D. (1999) 'Conclusion: Class size effects: Assessing the evidence, its policy implications, and future research agenda', *Educational Evaluation and Policy Analysis*, vol.21, no.2, pp.231-248
- Hanushek, E. A. (1989) 'The Impact of Differential Expenditures on School Performance', *Educational Researcher*, vol.18, no.4, pp.45-62
- Hedges, L. V. (2000) 'Using Converging Evidence in Policy Formation: The Case of Class Size Research', *Evaluation & Research in Education*, vol.14, issue 3 & 4, pp.193-205
- Hedges, L. V., Laine, R. D., and Greenwald, R. (1994) 'Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes', *Educational Researcher*, vol.23, no.3, pp.5-14
- Howell, W. G., Wolf P. J., Campbell, D. E. and Peterson, P. E. (2002) 'School Vouchers and Academic Performance: Results from Three Randomized Field Trials', *Journal of Policy Analysis and Management*, vol.21, no. 2, pp191-217
- Hunt, M. (1997) "How Science Takes Stock: The Story of Meta-Analysis", Russell Sage Foundation
- Kim, J. S. (2008) 'Research and the Reading Wars' in Hess, F. M. (ed.) "When Research Matters: How Scholarship Influences Education Policy", Harvard Education Press, pp.89-111
- Mosteller, F. (1995) 'The Tennessee Study of Class Size in the Early Grades', *Future of Children*, vol.5 no.2, p.113-127
- Nave, B., Miech, E. J., & Mosteller, F. (2000) 'The role of field trials in evaluating school practices: A rare design' in D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (eds.) "Evaluation models: Viewpoints on educational and human services evaluation 2nd edition", pp.145-161

- Oakley, A., (1998) 'Experimentation and social interventions: a forgotten but important history', *BMJ* vol.317, pp.1239-1242
- Oakley, A., (2000) "Experiments in Knowing: Gender and Method in the Social Sciences", Cambridge UK, Polity Press
- OECD (2002) "Education Policy Analysis 2000 Edition", OECD Publishing (御園生純監訳 (2006) 『世界の教育改革 2 OECD 教育政策分析』 明石書店)
- OECD (2006) "Starting Strong II: Early Childhood Education and Care", OECD Publishing
- OECD (2007) "Knowledge Management: Evidence in Education: Linking Research and Policy", OECD Publications (OECD (編)、岩崎久美子、菊澤佐江子、藤江陽子、豊浩子 (訳) (2009) 『教育とエビデンス——研究と政策の協同に向けて』 明石書店)
- Ravitch, D. (2000) "LEFT BACK: A Century of Battles Over School Reform", Simon & Schuster (ダイアン・ラヴィッチ (著)、末藤美津子、宮本健市郎、佐藤隆之 (訳) (2008) 『学校改革抗争の 100 年 20 世紀アメリカ教育史』 東信堂)
- Ritter, G. W. and Boruch, R. F. (1999) 'The Political and Institutional Origins of a Randomized Controlled Trial on Elementary School Class Size: Tennessee's Project STAR', *Educational Evaluation and Policy Analysis*, vol.21, no.2, pp.111-25
- Shanahan, T. and Lonigan C. J. (2010) 'The National Early Literacy Panel: A Summary of the Process and the Report', *Educational Researcher*, vol. 39, no. 4, pp. 279-285
- Smith, M.L. and Glass, G.V. (1979) "Relationship of Class-Size to Classroom Processes, Teacher Satisfaction and Pupil Affect: A Meta-Analysis", Far West Laboratory for Educational Research and Development (ERIC: ED190698)
- Stufflebeam D. L. and Shinkfield A. J., (2007) "Evaluation Theory, Models, and Applications", Jossey-Bass
- U.S. Department of Education (1998) 'Reducing Class Size: What Do We Know?' <http://www2.ed.gov./pubs/ReducingClass/index.html> (「学級規模縮小：何が分かっているか」 西村和雄、戸瀬信之 (編訳) (2004) 『アメリカの教育改革』 京都大学学術出版会、pp.51-68)
- Vinovskis, M. A. (2009) 'A History of Efforts to Improve the Quality of Federal Education Research: From Gardner's Task Force to the Institute of Education Science' in Walters, P. B., Lareau, A. and Ranis, S. H. "Education Research on Trial: Policy Reform and the Call for Scientific Rigor", Routledge, pp.51-79
- Weiss, C. H. (2002) 'What to Do until the Random Assigner Comes' in Mosteller, F. and Boruch, R. (eds.) "Evidence Matters: Randomized Trials in Education Research", Brookings Institution Press, pp.198-224
- 岩崎久美子 (2010a) 「教育におけるエビデンスに基づく政策—新たな展開と課題」『日本評価研究 第 10 巻第 1 号』pp.17-29
- 岩崎久美子 (2010b) 「教育分野でのエビデンスの産出」『薬理と治療 第 38 巻第 1 号』 pp.11-25
- 小川正人 (2010) 「学級編制標準引下げの課題と中教審「提言」の意義」『教育委員会月報 平成 22 年 9 月号』 pp.2-8
- 海津亜希子 (2006) 「米国における読みの指導に関する研究の動向—The Three-Tier Reading Model」独立行政法人国立特殊教育総合研究所『世界の特殊教育 XX (20)』 pp.21-28
- 佐々木亮 (2010) 「エビデンスに基づく開発援助評価—援助評価の歴史、ランダム化実験の起源、スクリヴェンとバナージェの考え方の比較—」『日本評価研究 Vo.10 No.1』 pp.63-73
- 惣脇宏 (2009) 「より一層エビデンスに基づいた教育政策と実践を」 OECD (編)、岩崎久美子、菊澤佐江子、藤江陽子、豊浩子 (訳) (2009) 『教育とエビデンス——研究と政策の協同に向けて』 明石書店、pp.3-13
- 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』 pp.153-168
- 田辺智子 (2006) 「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」『日本評価研究 第 6 巻第 1 号』 pp.31-41
- 丹後俊郎 (2002) 『メタ・アナリシス入門 エビデンスの統合をめざす統計手法 (医学統計学シリーズ 4)』 朝倉書店
- 丹野義彦 (2004) 「エビデンス・ベースト・カウンセリングの理論的基礎」内山喜久雄、坂野雄二 (編) 『現代のエスプリ 別冊 エビデンス・ベースト・カウンセリング』 pp.48-58
- 山下絢 (2008) 「米国における学級規模縮小の効果に関する研究動向」『教育学研究 第 75 巻 第 1 号』 pp.13-23

—(受理日：平成 23 年 3 月 15 日)—