

教育と経済・社会を考える  
第10回 教育の費用対効果

福田光宏

目次

1. 費用対効果の測定方法	2
2. 教育の効果	3
3. 学力と所得の関係	4
4. 機会費用	8
5. 割引計算	8
6. 実験	9
7. 準実験的な方法	12
8. 統計分析	13
9. 学級規模縮小の費用対効果の推計方法	14
10. 教員の能力向上策の費用対効果の推計方法	16
(1) 教員の能力	16
(2) 教育生産関数の修正	17
(3) 教員能力生産関数	19
(4) 研修の効果	20
(5) 教職大学院での教育の費用対効果	22
11. 学級規模縮小の効果に関する調査研究の動向	24
(1) 実験	24
(2) 準実験的な方法	29
(3) 統計分析	31
(4) 統合的な研究	37
(5) その他の研究	39
(6) 総括	41
(7) 私見	41
12. 教員の能力が児童生徒に与える影響についての調査研究の動向	43
(1) 既存研究	43
(2) 総括	46
13. 就学前教育の費用対効果に関する調査研究の動向	46
14. 教育の費用対効果に関する統合的な研究	47
(1) 既存研究	47
(2) 総括	49

## 1. 費用対効果の測定方法

教育の費用対効果を話題にすることは、教育界ではタブーに近いことのようにある。「第2回 教育経済学の基本 1. 教育経済学の重要性」で述べたように、教育は「人格の完成」を目指すべきものであり、教育投資に対する収益、教育投資の効率性などというように考えるべきではないというイデオロギーが教育界を支配している。未来を担う子どもたちの「人格の完成」のためには、お金を惜しんではならないことであろう。しかし、「第1回 はじめに 3. 教育関係者にとっての夢の時代の終わり」で述べたように、日本経済の衰退は、所得の減少と財政悪化を招き、教育費の負担が家計と財政の重荷になってきている。「人格の完成」という名目で、掴み金を教育に使う余裕はもうない。教育投資の費用対効果を検証して、効率的な教育投資を行わなければ、家計と国・地方自治体が財政的に破綻する時代になってしまった。年々乏しくなっていくお金を、各分野の人々が自分たちの生き残りをかけて奪い合うという時代が始まり、教育関係者は、教育の経済的・社会的な効果を立証しなければ生き残れなくなったのである。自分の今日・明日のことで精一杯で、日本の遠い未来のことなど知ったことではないと考える人たちが増えてきたということである。

費用対効果の測定方法には、費用便益分析、費用効果分析、費用効用分析の3種類がある。

費用便益分析では、施策の効果を金銭的な価値である「便益」で測定し、費用を上回る便益を生み出すことができるのか、いくつかの代替案の中で費用に対する便益の比率が最も高いものはどれかなどを調べる。

費用効果分析では、施策の効果を金銭的価値以外の何らかの尺度である「効果」で測定し、いくつかの代替案の中で費用に対する効果の比率が最も高いものはどれかを調べる。

費用効用分析では、施策の効果を各個人の満足度である「効用」で測定し、いくつかの代替案の中で費用に対する効用の比率が最も高いものはどれかを調べる。

費用便益分析では施策の効果を何らかの方法で金銭的価値に換算することが必要であるが、それができない場合には、費用効果分析や費用効用分析を用いざるを得ない。費用効果分析では、いくつかの施策の「効果」を同一尺度で比較する必要があるため、それができない場合には、費用効用分析を用いることになる。費用効用分析には、「効用」の客観性の担保が困難であるという欠点がある。

なお、「第2回 教育経済学の基本 2. 人的資本論」で説明した教育投資の収益率の推計は費用便益分析の一種であるが、「第2回 教育経済学の基本 8. 人的資本論とシグナリング理論の問題点 (2) 能力差に影響を与える要因、(3) 能力差と所得差の関係」で指摘したように、所得の上昇による便益を過大に推計している。

## 2. 教育の効果

伝統的な教育観では、教育は陶冶（知識・技能の習得）と訓育（道徳性・社会的行動様式の形成）による「全人格の発達」を目指すものとされているから、教育の効果は「全人格の発達」であるということになる。費用対効果の分析では、効果を何らかの方法で数量化する必要があるので、「全人格の発達」の程度を数量化できるのかが、例えば、「あなたの人格偏差値は 30 です」というようなことができるのかが、問題になる。「あなたの人格偏差値は 30 です」と言われて、「なるほど、私の人格はかなり劣っているのですね」と納得する人はほとんどいないであろう。「ふざけるな。お前は神様か。何を根拠にして人格偏差値なんてものを決めたのだ。どのようにして計ったのだ。おれは、そんなものは認めないぞ」というのが普通の反応であろう。

物事の全ての側面を数量化することなどできない。何らかの方法で計測できる側面しか数量化できない。何かを計測するということは、特定の方法・基準を用いて、物事の特定の側面を数量化するということである。数量化できる側面が複数あったり、計測の方法・基準が複数あったりする場合には、そのどれを用いるかに関して、恣意性が介入する。したがって、費用対効果の分析は、計測可能な効果しか扱えないという点で、物事の一面しか捉えることができないものであり、また、どの効果に着目するか（逆に言えば、どの効果を見捨てるか）や、どのような計測の方法・基準を用いるかによって、結論が異なってしまうという点で、恣意的なものである。どの効果に着目するかや、どのような計測の方法・基準を用いるかによって、ある人が有利になったり、不利になったりする場合には、そこに争いが生じる。どれだけ厳密な費用対効果の分析を行っても、それは、効果の特定の側面を特定の計測の方法・基準で計測するという枠組み内で客観的なものに過ぎず、効果の特定の側面と特定の計測の方法・基準という枠組みの選択に関しては主観的なものなのである。

教育の費用対効果の分析に関して言えば、数量化できるのは「全人格の発達」の一部にしか過ぎない。陶冶に関しては、どのような知識、技能を習得すべきかに関して、学習指導要領などの基準があり、知識、技能の習得の程度を計る方法として学力テストが実施されてきた長い歴史があるので、ペーパーテストで計れるものが学力の全てとは言えないという批判、さらには、学力の本質に関する論争があるとは思いますが、一応、数量化できる（ただし、どのような学力試験を選ぶかに関して恣意性がある）。しかし、訓育に関しては、どのような道徳規範が正しいのか、どのような社会的行動様式が好ましいのかに関して争いがあるので、数量化が困難である。数量化できるのは、学校で暴力行為やいじめを何回行ったか、社会に出てから犯罪行為を何回行ったか（あるいは、どの程度重い罪か）程度であろう。なお、理論的には、高所得を得るために望ましい社会的行動様式を就労後の所得を基準にして計測することができるが、社会的行動様式と就労後の所得の関係に関するデータがほとんど無いので、実用的ではない。

結局、教育の費用対効果の分析では、学力試験で計測できる学力の向上を教育の効果とするか、教育を受けたことによる将来の所得の増加を教育の便益とするしか良い方法がなく、その方法では、教育の効果の一部しか計測していないことになるので、教育の効果を過小評価していることになる。教育の費用対効果の分析では教育の効果の一部しか捉えることができないので、教育の費用対効果の分析結果によって教育を論じるのは不当であるという批判があると思うが、自分が見える範囲内だけで主観的な基準を用いて教育の効果を把握して教育を論じるのも不当であると思う。

### 3. 学力と所得の関係

教育に関する費用便益分析では、教育成果を金銭に換算しなければならない。教育投資の収益率の推計（「第2回 教育経済学の基本 2. 人的資本論」参照）のように、教育成果を直接、金銭で計測する場合もあるが、それが困難な場合もある。そのような場合には、教育の成果として計測容易な学力の向上を、所得の上昇などの金銭的価値に換算する必要がある。

学力の向上がどれだけの所得の上昇をもたらすかを調査分析するためには、賃金関数を用いるのが一般的である。最もポピュラーな賃金関数はミンサー型賃金関数で、

$$\ln W = \alpha + \beta S + \gamma_1 X + \gamma_2 X^2 + \varepsilon \quad (\text{ただし、} W \text{は賃金、} S \text{は教育年数、} X \text{は労働経験年数、} \alpha, \beta, \gamma_1, \gamma_2 \text{は定数、} \varepsilon \text{は誤差項})$$

というような形をとる（「第2回 教育経済学の基本 2. 人的資本論」参照）。このミンサー型賃金関数では、教育年数だけが説明変数として取り上げられ、学力が無視されているので、学力Aを説明変数に加えて修正する必要がある（社会的行動様式、親の社会経済的地位、運なども賃金に影響を与えているはずなので、これらも説明変数に加える必要があると思う）。例えば、

$$\ln W = \alpha + \beta S + \delta A + \gamma_1 X + \gamma_2 X^2 + \varepsilon$$

理論的には、この式を回帰式にして、回帰分析を行えば（「第2回 教育経済学の基本 10. 回帰分析」参照）、学力Aが賃金Wに与える影響が分かるはずだが、日本では、学力と賃金に関する詳細なデータがほとんど無いので、現実には無理である。もし、学力と賃金に関する詳細なデータがあったとしても、それは、過去の教育における学力と現在の賃金の関係を示すものであり、現在の教育における学力と未来の賃金の関係を示すものではない。例えば、今40歳の人々のデータは、20年以上前の教育における学力と今の賃金の関係に関するデータであり、今10歳の人々の学力とその人の30年後の賃金の関係を示すものではない。今10歳の人々の30年後の正確な賃金が分かるのは30年後のことである。将来発生する便益に関する費用便益分析は、未来予測の性質を併せ持っており、今後、経済・社会構造や生産技術・生産構造がどのように変化するかを正確に予測することが誰にもできない以上、未来は過去何年間かの傾向の延長線上にある（外挿する）という非現実的な仮定をした上でしか、学力と賃金に関するデータから教育の便益を推計することが

できないのである。この問題は、「第2回 教育経済学の基本 8.人的資本論とシグナリング理論の問題点 (4)教育投資についての意思決定」で述べた教育投資の収益率の問題と同様の問題である。過去何年の傾向を用いて未来を予測するかに関して特に決まりはないので、そのことを利用して、ある程度、教育の便益の推計結果を操作することができる。

なお、経済が成長している国では、経済成長に伴って、将来、賃金がどの程度上がるかを予測して推計する必要があるが、日本のように賃金が低下傾向にある国では、将来、賃金がどの程度下がるかを予測して推計する必要がある。将来、賃金がどの程度上がるか下がるかは誰にも分からないので、賃金上昇（下降）率の予測値を操作することによって、ある程度、教育の便益の推計結果を操作することができる。

財やサービスの生産のために必要な能力は、その時々での産業技術の水準や生産構造に左右されるので、どのような教育、そして学力・社会的行動様式が所得の上昇につながるかは、その時々での産業技術の水準や生産構造との関係で決まる。そのため、教育の世界の基準で、高い学力と優れた社会的行動様式を身につけているとされても、産業の世界の基準では、能力が低いとされることもありうる。基準が変われば、能力に対する評価も変わる。学力の向上を「効果」とする費用効果分析では、教育の世界での自己満足に陥る危険性があるので、所得の上昇を「便益」とする費用便益分析の方が望ましい。こういう主張をすると、それは産業の世界の自己満足に過ぎないという批判があると思うが、学力の向上という「効果」と、所得の上昇という「便益」のどちらを選ぶかと問われれば、所得の上昇という「便益」を選ぶ人の方が多数派であろう。山田昌弘氏は『希望格差社会』（P.159-160）で次のように指摘している。

従来の教育界主流の考え方では、教育の目的は、人格の完成とか、学ぶこと自体が楽しいとか、文化の伝達などで「あるべき」であって、投資とか金銭的リターンなど「経済的利益」、ましてや「社会的効率」の視点から教育を考えることは良くないことであるという意見が強かった。……それこそ、教育を受ける側の欲求や、社会全体の要請とは、「隔たり」がある。……教育は、「手段」なのであって、それ自体が目的ではないことを肝に銘じるべきである。何事も、それをすること自体が目的となってしまうと、一種の「宗教」となってしまうのだ。

欧米には、教育成果の金銭への換算に用いることができる調査研究がいくつか存在する。

Alan. B. Kruegerは“Understanding the magnitude and effect of class size on student achievement”（“Mishel, Lawrence and Richard Rothstein, eds., *THE CLASS SIZE DEBATE.*” 所収）（P.24～25）で、生徒の学業成績と仕事に就いた後の所得との関係についての研究を紹介している。

それによると、Richard J. Murnane、John B. Willett、Frank Levyは“The Growing

Importance of Cognitive Skills in Wage Determination”で、“the High School and Beyond”（アメリカで、1980年に高校3年生であった11,500人に関するデータ）のデータを分析して、1980年時点で高校3年時の数学の成績が1標準偏差分（偏差値に換算すると、10）高いと、6年後には、男性で7.7%、女性で10.9%高い所得を得ていると述べている。Janet CurrieとDuncan Thomasは、“Early Test Scores, Socioeconomic Status, and Future Outcomes”で“the British National Child Development Survey”（1958年のある1週間にイギリスで生まれた子どもに関するデータ）のデータを分析して、7歳の時に、読解のテストで上位四分位数（100人中25位であるということ）に属する人は、下位四分位数（100人中75位であるということ）に属する人に比べて、20%高い所得を33歳時点で得ていて、数学のテストで上位四分位数に属する人は、下位四分位数に属する人に比べて、19%高い所得を得ていると述べている。この結果は、読解のテストの成績が1標準偏差分（偏差値に換算すると、10）高いと所得が8%高くなり、数学のテストの成績が1標準偏差分（偏差値に換算すると、10）高いと所得が7.6%高くなることを示唆している。Derek NealとWilliam Johnsonは“The Role of Premarket Factors in Black- White Wage Differentials”で、“the National Longitudinal Survey of Youth”（1957～1964年にアメリカで生まれた12,686人に関するパネルデータ）から1961年以降に生まれた人のデータを分析して、15～18歳の時に、1980年に実施された“the Armed Forces Qualification Test (AFQT)”（アメリカ軍入隊のためのテスト）を受験して（対象者の90%以上が受験している）、成績が1標準偏差分（偏差値に換算すると、10）高いと、男女ともに26～29歳の時の所得が20%高くなると述べている。

（注）「標準偏差の〇〇倍」という場合、その数字を10倍すると、偏差値での差に換算できる。偏差値は標準偏差を10にしているからである。

Alan. B. Kruegerは、Derek NealとWilliam Johnsonの研究の方がJanet CurrieとDuncan Thomasの研究よりも、所得への大きな影響を示している理由は、第一に、Janet CurrieとDuncan Thomasが7歳の時のテストを利用したのに対し、Derek NealとWilliam Johnsonが15～18歳の時のテストを利用したことにあると指摘している。小さいときに非常に高い得点をとった子どもはその後の成績の伸びが、小さいときに非常に低い得点をとった子どもに比べて小さい。第二に、Derek NealとWilliam Johnsonが一つのテストの得点を利用しているのに対して、Janet CurrieとDuncan Thomasは相互に関連性のある読解と数学のテストの得点を利用している。第三に、イギリスとアメリカの労働市場の違いがある。

Alan. B. Kruegerは、これら三つの研究結果から、数学または読解のいずれかの成績が1標準偏差分（偏差値に換算すると、10）高いと、所得が約8%高くなると言えるとしている。

Eric A. HanushekとLei Zhangは“Quality-Consistent Estimates of International Schooling and Skill Gradients”で、“the International Adult Literacy Survey”の結果

をミンサー型賃金関数を修正した関数  $\ln(y_{ikc}) = \delta_{ik} \cdot \tilde{S}_{ikc} + \delta_{2k} \cdot L_{ikc} + Z_{ikc} \cdot \theta_k + v_{ikc}$ 、

ただし、 $y$  は年間所得、 $\tilde{S}$  は教育の質によって修正した教育年数、 $L$  は成績、 $Z$  は統制変数（労働経験年数、性別、居住地域等）で、添え字の  $i$  は個人別、 $k$  は国別、 $c$  は年齢階層別を表している）を用いて分析したところ、成績が 1 標準偏差分（偏差値に換算すると、10）高いと、所得がどれだけ高くなるかは国より異なるが、最高はアメリカで 20% 高くなり、最低はスウェーデンで 3.8% 高くなることを述べている。なお、“the International Adult Literacy Survey”（略して、IALS）とは、1994年、1996年、1998年に、23カ国が参加して（日本は参加していない）、16~65歳を対象にして行われた、日常の暮らしに関わる各種の情報をを用いる能力と基本的計算能力の調査であり、各個人の出生国、年齢、学歴、職業、所得、親の学歴も併せて調べられた。

Eric A. Hanushek と Ludger Woessmann は “Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation” で、同一国からアメリカへ移民した人々で、移民前に母国で教育を受けた人と移民後にアメリカで教育を受けた人のアメリカでの所得の差は、母国とアメリカの国際的な学力調査（TIMSS、PISA等）の成績の差を反映していると考えて、教育年数と労働経験年数の影響も勘案した上で、データを分析すると、成績が 1 標準偏差分（偏差値に換算すると、10）高いと、所得が 14% 高いことになると述べている。

私が知る限りでは、日本には、教育成果の金銭への換算に用いることができる調査研究は存在しない。

北條雅一氏は「大学在学中の学習成績・学習への取り組みと卒業後の所得」で、ある大学の人文系、社会科学系、教育系学部の卒業生を対象にしたアンケート調査結果から、専門科目の成績が「とても良かった」「やや良かった」と答えた卒業生は、そうでなかった卒業生に比べて、現在の所得が 13.7% 高くなっていると指摘しているが、特定大学の特定学部のデータに過ぎず、成績も自己評価なので、この結果からは確たることは言えないと思う。

出身高校ランク（大学進学実績等）別の所得差を調査した研究があるので、これらの研究から、学力の向上がどれだけの所得の上昇をもたらすかについてのおおざっぱな推計を行うことはできる。

橘木俊詔氏と八木匡氏は『教育と格差』（P.28-33）で、2004~2006 年に行ったインターネット調査の結果に基づいて、30~65 歳の有職者男性の平均年収を、大学進学実績を基に大学通信社が作成したランキング表によって比較すると、高校トップ校出身者は 592.5 万円、中堅上位校出身者は 540.2 万円、中堅下位校出身者は 479.8 万円、下位校出身者は 483.8 万円であると述べている。各校の在籍者の平均成績が、高校トップ校は 87.5 パーセントイル（100 人中 12.5 位であるということ）、中堅上位校は 62.5 パーセントイル（100 人中 37.5 位であるということ）、中堅下位校は 37.5 パーセントイル（100 人中

62.5位であるということ)、下位校は12.5パーセンタイル(100人中87.5位であるということ)の位置にあると仮定すると、成績が1パーセンタイル高くなる(100人中の順位が1位高くなること)と、年収が、62.5～87.5パーセンタイルの間では20,920円、37.5～62.5パーセンタイルの間では24,160円、12.5～37.5パーセンタイルの間では1,600円、それぞれ上昇することになる。なお、「第2回 教育経済学の基本 3.大学教育の収益率」で指摘したように、この結果には、インターネット調査によったことによるサンプリングの偏りの影響があると思う。

中西祐子氏は近藤博之編『日本の階層システム3 戦後日本の教育社会』「3章 学校ランクと社会移動」(P.43)で、1995年と1985年のSSM調査(社会階層と社会移動全国調査)を分析した結果によると、出身高校別の年収は、普通科Aが607万円、普通科Bが599万円、普通科Cが564万円、職業科が559万円(1995年のデータについては、普通科Aはほぼ全員が大学・短大進学、普通科Bは半分以上が大学・短大進学、普通科Cは大学・短大進学者が半分以下、1985年のデータについては、1980年4月20日発行の『サンデー毎日』に掲載された「全国1300高校大学合格者調査」を用い、そこに記載のない高校が普通科C、旧帝国大学・東京工業大学・一橋大学の合格者数が計5名以上いる高校が普通科A、記載のあるその他の高校が普通科Bとされている)であると述べている。1995年の大学・短大進学率(浪人を含む)が45.2%であることから、各校の在籍者の平均成績が普通科Aは87.5パーセンタイル、普通科Bは62.5パーセンタイル、普通科Cは37.5パーセンタイルの位置にあると仮定すると、成績が1パーセンタイル高くなると、年収が、62.5～87.5パーセンタイルの間では3,200円、37.5～62.5パーセンタイルの間では14,000円、それぞれ上昇することになる。

#### 4. 機会費用

ある施策を実行する際の費用を推計する際に、「機会費用」が無視されることがある。何かあることをすると、そのためにできなくなることがあるが、そのできなくなったことを、あることをするための費用であると考えるのが「機会費用」という考え方である。

例えば、小・中・高等学校の教員になるためには修士以上の学位を要するようにするという施策の費用を推計する際、修士課程の2年間は働けなくなるので、その間、働いていたら得られたであろう賃金(放棄所得)が「機会費用」になる。なお、賃金は税引き前の金額にする必要がある。国や地方公共団体が税収の減少という「機会費用」を負担しているからである。

#### 5. 割引計算

今の100万円と2年後の100万円では価値が異なる。今100万円もらうのと、2年後に100万円もらうのでは、どちらかを選ぶかと問われて、2年後に100万円もらう方を選ぶ人はほとんどいないであろう。100万円を投資すれば収益が生じ、2年後には、100

万円よりも大きな価値を持つようになる（現在の日本のように超低金利だと実感がわかないかもしれないが、超低金利が異常な状態なのである）。そのため、費用便益分析では、将来発生するであろう便益や費用を割り引いて、現在価値に換算してから、便益と費用を比較する必要がある。

新古典派経済学やケインズ経済学では、資本市場（お金を貸したり、借りたるする市場）では一つの利率が支配し、その利率で、各経済主体（人や会社のこと）は、自分が望む金額の資金を自由に貸したり、借りたりできると仮定する。この利率で、将来の費用や便益を割り引いて、現在価値を求める（割引計算）。例えば、市場利率が1%の時に、100万円を投資すると、2年後には

$$\{100 \times (1 + 0.01)\} \times (1 + 0.01) = 100 \times (1 + 0.01)^2 = 102.01 \text{万円}$$

になる（要するに、複利計算をする）から、市場利率が1%の時の2年後の102.01万円の現在価値は、 $\frac{102.01}{(1 + 0.01)^2} = 100 \text{万円}$  となる。

一般化すると、市場利率が*i*の時の*t*年後の*M*円の現在価値は、

$$\frac{M}{(1 + i)^t} \text{円} \quad \text{となる。}$$

資本市場では一つの利率が支配するというのは非現実的な仮定で、現実には、各経済主体の信用力によって、利率や借金できる上限は異なる。例えば、銀行に預金する（預金するとは、銀行にお金を貸していることである）際の金利と借入する際の金利には大きな差がある。そもそも、将来、市場利率がどう変動するかは誰にも分からない。したがって、割引計算でどの利率を用いるべきかはあいまいであり、割引率（割引計算で用いる利率のことを割引率という）の選び方によって、費用便益分析の結果を左右することができる。例えば、Alan. B. Kruegerは“Understanding the magnitude and effect of class size on student achievement”で学級規模縮小の費用便益分析を行っているが、その結果は割引率によって大きく異なっている（「12.学級規模縮小の効果に関する調査研究の動向」参照）。

なお、この割引計算では、名目利率（借金や預金の際に約定される金利）ではなく、物価変動を考慮に入れた実質利率（＝名目利率－物価上昇率）を用いる必要がある。将来の物価変動がどうなるかは誰にも分からないので、将来の物価変動をどのように想定するかによって、費用便益分析の結果を左右することができる。

つまり、費用便益分析では、割引率や物価変動の予測値を操作することによって、ある程度、その結果を操作することができるので、割引率や物価変動の予測値に十分な注意を払う必要があるということである。

## 6. 実験

実験 (experimental design) とは、実際にある処置を施してみて、その効果を測定するということである。実験を厳密に行うためには、多数の人々を、ある処置を施されるグループ (実験群、experiment group、処理群、処置グループ、treated group などと呼ばれている) と、施されないグループ (統制群、制御群、control group、対照群、comparison group、非処置グループなどと呼ばれている) に無作為 (random) に振り分け、特定期間経過後に両者を比較して、処置の効果を測定する必要がある。この方式は、古典的実験計画法、統制群法、ランダム化比較試験などと呼ばれている。この方式の利点は、多数の人々 (標本、sample) を無作為に振り分けることにより、各個人の属性の違いが効果に与える影響の違いを無視できるようになることにある。

教育生産関数 (「第 2 回 教育経済学の基本 9.教育生産関数」参照) を用いて説明すると、特定の教育方法  $X$  が施された実験群  $\alpha$  と、施されなかった統制群  $\beta$  では、特定の教育方法以外の学校教育の平均値  $\bar{S}$ 、学校外教育の平均値  $\bar{T}$ 、自学自習の平均値  $\bar{L}$ 、仲間集団の平均値  $\bar{P}$ 、家庭環境の平均値  $\bar{F}$ 、社会環境の平均値  $\bar{E}$ 、生徒の生得的な能力・気質の平均値  $\bar{I}$ 、攪乱項 (観察できない、あるいは未知の要因)  $u$  はほぼ等しくなるので、

$$\bar{A}_\alpha = a + b\bar{S}_\alpha + c\bar{T}_\alpha + d\bar{L}_\alpha + e\bar{P}_\alpha + f\bar{F}_\alpha + g\bar{E}_\alpha + h\bar{I}_\alpha + jX + u_\alpha$$

$$\bar{A}_\beta = a + b\bar{S}_\beta + c\bar{T}_\beta + d\bar{L}_\beta + e\bar{P}_\beta + f\bar{F}_\beta + g\bar{E}_\beta + h\bar{I}_\beta + u_\beta$$

から、

$$\bar{A}_\alpha - \bar{A}_\beta \approx jX$$

となり、教育方法  $X$  の効果だけを計測できるようになる。

ただし、この方法で計測できる教育方法  $X$  の効果は、平均的な生徒に対して教育方法  $X$  を施した場合の効果であり、平均から外れた生徒に教育方法  $X$  を施した場合の効果は計測できない。平均から外れた生徒に教育方法  $X$  を施した場合の効果の計測するためには、個人の属性 (例えば、家庭環境  $F$  や生徒の生得的な能力・気質  $I$ ) によって、実験群と統制群をいくつかのグループに分けて、実験を行う必要がある。

実験には、実験前と実験後の 2 回データを計測し (例えば、学力調査をする)、実験群と統制群のデータの変化の度合いの差を効果であるとする (例えば、実験期間中の実験群と統制群の成績の伸びの差を効果とする) 方法 (前後比較デザイン) と、実験後の 1 回だけデータを計測し、実験群と統制群のデータの差 (例えば、実験後の実験群と統制群の成績差) を効果であるとする方法 (事後比較デザイン) がある。標本数が多く、実験群と統制群への振り分けが完全に無作為であれば、実験前にデータを計測しなくても、実験群と統制群のデータに差はないはずなので、事後比較デザインで十分であるということになるが、標本数が少なかったり (社会的な実験では、それぞれのグループが 200 ~ 300 人以下だと危ないと言われている)、無作為性に疑問があったりする場合には、実

験前にデータを計測して、実験群と統制群と同質性を確認し（ただし、計測したデータ以外の同質性は確認できない）、問題がある場合には統計的方法等で修正を加えた方がよい。

なお、実験では、実験で試されたこと以外のことは分からないということに注意する必要がある。例えば、40人学級から30人学級に縮小する実験が行われ、30人学級の成績は40人学級の成績よりも偏差値で1高いという結果が得られたとしよう。この結果から、学級の人数を1人減らすと成績が偏差値で0.1高くなる（つまり、学級規模と成績は比例関係、線形関係にある）と言えそうに思えるが、そうではない。学級の人数が少なくなるほど、1人減らすことによる成績向上の効果が増加したり、減少したりする（つまり、学級規模と成績は非線形関係にある）かもしれないからである。

また、実験では、着目している事象以外の事象が発生しても、それを見落としてしまうという危険がある。新しい教育方法を実験すると、期待している効果ばかりに目が向き、副作用が発生しても、そのことに気づかないことがある。

古典的実験計画法による実験を行うには、様々な困難がある。例えば、「12.学級規模縮小の効果に関する調査研究の動向」で紹介する「STAR プロジェクト」（1985年～1989年にアメリカのテネシー州で実施）という学級規模縮小の効果に関する実証実験に対しては、通常規模の学級と少人数学級への無作為な振り分けに対して保護者から不満が出て、実験途中で一部入れ替えが行われたり、一部の児童が他の公立学校や私立学校に転校したりし、実験途中での転居による流出・流入もあったので、無作為性が損なわれているという指摘がある（例えば、教育熱心な保護者ほど、少人数学級への入れ替えを要求したり、私立学校に転校させたりする可能性が高い）。また、少人数学級を担当した教員には、学級規模縮小の効果を示すという期待がかかり、その期待に応えようががんばったことが実験結果に影響を与えたのではないかという批判もある。さらに、約1200万ドルという多額の経費を費やしている。実験に要する時間も問題になるかもしれない。「STAR プロジェクト」では4年かかっている。

日本で、教育に関する厳密な実験を行おうとする際の最大の障害は、おそらく保護者からの抗議であろう。効果があると思われている教育方法の実験を行う場合には、統制群にされた児童・生徒の保護者から差別だという抗議がくるだろうし、効果が疑われている教育方法の実験を行う場合には、実験群にされた児童・生徒の保護者から子どもを実験台にするなという抗議が来るだろう。保護者からの抗議に応じて、実験群と統制群の間で入れ替えを行うと無作為性が損なわれてしまい、実験にならない。保護者からの抗議を無視して実験を強行すると、自分の子どもが不利な立場に立たされた考える保護者は、子ども他の学校に転校させたり、学習塾に通わせたりするであろう。その結果、大きなバイアス（偏り）が生じてしまう。実験によって損害を受ける可能性があると考えられる保護者が実験の差止めを求める訴訟を起こしたり、実験によって損害を受けたと考える保護者が損害賠償請求を行ったりする可能性もある。なお、実験対象になることに

関して本人や保護者の同意を得るようにしたり、志願制にしたりすると、バイアスが生じ、実験にならない。例えば、チャレンジ精神に溢れているか、そうでないかによって、教育効果に差が出る可能性があるからである。日本で、教育に関する厳密な実験を行うことは不可能であると言っても、間違いではないだろう。

## 7. 準実験的な方法

古典的実験計画法による厳密な実証実験を行うことができない場合には、準実験的な（擬似実験的）な方法（quasi-experimental design）が用いられる。準実験的な方法にはいくつかのタイプがある。例えば、「12.学級規模縮小の効果に関する調査研究の動向」で紹介する「SAGE プログラム」（1996～2000年度にアメリカのウィスコンシン州で実施）ではマッチング法（matching method）が用いられ、貧困レベル以下の家庭の子どもが多い学校30校（各学区から1校ずつ応募できるとされ、応募してきた学校は全て実験対象校になった）を対象にして学級規模が縮小され、同一学区内にあり、類似した世帯収入、読解の成績、人種構成の児童がいる学校14～17校と、幼稚園年長組から小学3年生までの4年間、学業成績が比較された。

マッチング法では、実験開始前に、実験対象とされたグループと比較対照されるグループ（擬似統制群、擬似的対照群、比較対照グループなどと呼ばれる）の間にどの程度と同質性があるかが鍵になるが、観察困難な要因や未知の要因についての非同質性による影響を排除することはできない。マッチング法でも、実験対象にされたり、されなかったりすることに対する不満が出たり、多額の経費がかかったりすることに関しては、古典的実験計画法とあまり変わらない。実験対象になることに関して応募制にして、くじ引きで選ぶようにすれば、不満が少なくなるかもしれないが、その場合、応募してきた家庭や学校が、子どもの教育に熱心であったり、新しい試みに挑戦することを好むタイプであったり、情報収集に熱心であったりし、逆に、応募してこない家庭や学校が、子どもの教育に関心が乏しかったり、新しい試みに挑戦することを嫌うタイプであったり、情報収集に無関心であったりする（情報収集に無関心で、募集されていることに気づかないということ）、というようなバイアスが生じてしまい、実験結果の信頼性を損ねてしまう。

“Regression-discontinuity design”（非連続回帰デザイン、非連続回帰分断などと訳される）という方法もある。例えば、「12.学級規模縮小の効果に関する調査研究の動向」で紹介する赤林英夫氏と中村亮介氏の調査研究『学級規模縮小が学力に与えた効果の分析－横浜市公開データにもとづく実証分析－』で、この方法が用いられており、日本における学級規模は40人を標準とし（小学校1年生では35人）、学級規模は在籍児童生徒数に応じて決まるので、在籍児童生徒数が40人から41人に変化すれば学級規模は40人（1学級）から20～21人（2学級）へと変化するということを利用して、学級規模縮小の効果を推計している。このように、制度的な不連続性により2つのグループ（こ

の場合は、40 人学級と 20～21 人学級)に分けられる場合には、意識的な選択が働く余地がほとんどない(学級編制の弾力化により意図的に少人数学級化している場合は別)ので、無作為による振り分けに近い状態が生じ、分かれた時に、2つのグループの間にはある程度同質性があると考えられる。ただし、制度的な不連続性が生じた要因(例えば、入学者数の増減)が測定したい効果(例えば、学力)に影響を与えている場合には推計が不正確になる。

“Regression-discontinuity design”は自然実験(Natural Experiment)的な方法であり、調査費用があまりかからない、保護者からの抗議が少ない(学力を調査し、分析することに反対される可能性はある)といったメリットがあるが、制度的な不連続性が生じないと調査できないので、調査できる事柄には限界がある。

## 8. 統計分析

実験や準実験的な方法がとれない場合には、実験や準実験を行わなくても収集できるデータや既存のデータを回帰分析することにより、教育生産関数(「第2回 教育経済学の基本 9.教育生産関数」参照)の係数を推計し、特定の教育方法を採用すると、教育生産関数の特定の説明変数の値が変化し、教育生産関数上の教育成果も変化するので、その変化分を特定の教育方法の効果であると推計するという方法が用いられる。

回帰分析の方法については、「第2回 教育経済学の基本 10.回帰分析」で説明しているが、学級規模縮小による学力向上効果を推計するという例を用いて、再度説明する。まず、個人の能力をA、学校教育(ただし、学級規模を除く)をS、学校外教育をT、自学自習をL、個人が属する仲間集団(による影響)をP、家庭環境をF、社会環境をE、生得的な能力・気質をI、学級規模をC、観察できない、あるいは、未知の要因(攪乱項、誤差項などと呼ばれる)をuとして、教育生産関数が次のような形をとると、大した根拠もなく、仮説を立てる(ある意味、決めつける)。この場合、Aが被説明変数、S、T、L、P、F、E、I、Cが説明変数になる。なお、実際に行われている分析では、説明変数のうちのいくつかが無視されることが多い。

$$A = a + bS + cT + dL + eP + fF + gE + hI + jC + u$$

(ただし、a、b、c、d、e、f、g、h、jは定数)

次に、A、S、T、L、P、F、E、I、Cに関する実際のデータを学級単位などで多数集める。質的なデータは何らかの方法で数量化する。学級*i* ( $i = 1, 2, \dots, n$ )のデータは、 $(A_i, S_i, T_i, L_i, P_i, F_i, E_i, I_i, C_i)$  というようにベクトルの形をとる。そして、これらのベクトル全てに最も良く当てはまるような教育生産関数の形を割り出す。具体的には、教育生産関数による予測値(理論値)を $\hat{A}_i = a + bS_i + cT_i + dL_i + eP_i + fF_i + gE_i + hI_i + jC_i$  とすると、実測値(観測値) $A_i$ と

予測値  $\hat{A}_i$  の差である残差（回帰残差）  $e_i = A_i - \hat{A}_i$  が小さいほど、教育生産関数が実際のデータに良く当てはまっていることになるので、残差の2乗の和  $\sum_{i=1}^n e_i^2$  が最も小さくなるような a、b、c、d、e、f、g、h、J（回帰係数）の値を推定するという最小2乗法が用いられる。

回帰分析で推定された教育生産関数を用いると、学級規模 C の値を変化させると能力 A の値がどれだけ変化するかが全て分かるように思えるが、そうではない。例えば、収集されたデータが学級規模で 20～40 人のものである場合、それらのデータから推計された教育生産関数を用いて、20 人未満に学級規模を縮小した場合や 40 人超に学級規模を拡大した場合の教育効果を推計する（外挿する）ことは危険である。20 人未満や 40 人超になると教育生産関数の形が変わるかもしれないからである。回帰分析の結果は、収集されたデータから見える範囲内での現実もどきのものを示しているに過ぎないのである。比喩的に言えば、目に見える部分が直線だからと言って、全体が直線であるとは限らない、目に見えない部分で突然、折れ曲がっているかもしれないということである。

## 9. 学級規模縮小の費用対効果の推計方法

小学校の学級規模を 40 人から 20 人に縮小する場合に、1 年単位での費用対効果を推計するという例で説明する。

まず、「費用」について考える。

学級規模を 40 人から 20 人に縮小すると、学級の数約 2 倍になるので、教員数を約 2 倍に増やさなければならなくなり、人件費が約 2 倍に増加するということになりそうであるが、年功賃金になっている場合には、話は複雑になる。現在の教員の年齢構成を維持できるように、教員を採用すれば、人件費は約 2 倍になる。しかし、実際には、若年者からの採用が多いということになりそうなので、当面は、人件費の増加が 2 倍よりも少なくすむが、新採用者の高齢化とともに、人件費は増加していき、いずれは、2 倍よりも大きくなる。この場合、どの時点の人件費を、「費用」と考えるべきかという問題が起こる。費用対効果を大きく推計したいという意図が働いている場合には、若年者からの採用が多くなるから、人件費の増加は 2 倍よりも小さくなると推計して、将来の人件費の増加を無視することになるであろう。

教員数を約 2 倍にする場合、現在の教員資格（教育職員免許状）保有者数と教員養成数で足りるかという問題も生じる。教員養成数が足りない場合には、教員養成数を増やすための費用が必要になる。

学級数を約 2 倍にすると、教室数を約 2 倍にしなければならないので、校舎の改築・増築のための費用や設備の購入のための費用が必要になる。改築・増築のための費用の推計では、校舎の耐用年数を考慮して、1 年間に必要な費用を推計することになるの

で、既存の校舎の耐用年数次第で、必要な費用が大きく変わってくる。設備に関しても、耐用年数を考慮して、1年間に必要な費用を推計する必要がある。

次に「効果」について考える。

一般に期待される学級規模縮小の効果は、学力の向上と生徒指導の充実であるが、生徒指導の充実度の測定と数量化は困難である。客観的に測定できるのは、暴力行為、いじめ、不登校の件数程度であり、後は、教員や児童の主観的評価に頼らざるを得ない。そのため、学級規模縮小の効果に関する既存の調査研究（教員の主観的評価に頼った調査研究は別）は、学力の向上のみに着目したものがほとんどである。

単純に考えれば、今ある40人学級と20人学級で、学力試験を実施して平均成績を比較すれば、その差が学級規模縮小による学力向上効果（あるいは、学力低下効果）であるということになるように思えるが、話はそう簡単ではない。40人学級と20人学級では、教員の能力、指導方法、児童がそれまでに受けてきた教育内容、自学自習の質と量、仲間集団からの影響、家庭環境、生得的な能力・気質などに違いがあり、これらの要因が学力に影響している可能性があるからである。

$$\bar{A}_{20} - \bar{A}_{40} = b(\ddot{S}_{20} - \ddot{S}_{40}) + (u_{20} - u_{40}) \quad (\text{この場合の } \ddot{S}_{40} \text{ は 40 人学級における調査時}$$

点での教育、 $\ddot{S}_{20}$  は 20 人学級における調査  
時点での教育)

という分析をしてはならないのである。

教員の能力、指導方法、児童がそれまでに受けてきた教育内容、自学自習の質と量、仲間集団からの影響、家庭環境、生得的な能力・気質などに関するデータが全て入手可能であれば、既存のデータを統計分析して、教育生産関数を推計することにより、学級規模を40人から20人に縮小した場合の効果をかなり正確に割り出すことができる。しかし、プライバシー、測定方法、数量化の方法などの問題から、これらのデータはほとんど入手できないというのが現実である。これらのデータが入手できないのであれば、これらの要因による影響が無視できるような状態で、40人学級と20人学級の学力を比較する必要がある。これらの要因による影響が無視できるような状態を人為的に作り出すのが「7. 実験」で説明した実験であり、これらの要因による影響が無視できるような状態が制度的に生まれる機会を利用するのが「8. 準実験的な方法」で説明した“Regression-discontinuity design”である。

「7. 実験」で述べたように、日本では、教育に関する厳密な実験を行うことは不可能に近いので、“Regression-discontinuity design”を用いるのが最も良い方法であるということになる。ただし、2001年度からの学級編制の弾力化によって、意図的に少人数学級化や少人数指導（少人数の学習集団をつくる方法）ができるようになったことが、“Regression-discontinuity design”を用いた調査分析にバイアスを与える可能性がある。

なお、学級規模縮小による学力向上効果（あるいは、学力低下効果）があらわれるには、どの程度の時間がかかるのか、また、どの程度、継続するのかという問題がある。この問題を解明するためには、長期にわたる個人単位での追跡調査が必要である。また、費用便益分析を行うためには、学級規模縮小による学力向上効果の価値を金銭に換算しなければならないが、「3. 学力と所得の関係」で述べたように、日本には、児童・生徒の学業成績と仕事に就いた後の所得との関係を正確に調べた調査研究が存在しないので、金銭への換算は無理である。

## 10. 教員の能力向上策の費用対効果の推計方法

### (1) 教員の能力

先ず、教員の能力とは何か、どのようにして計測するのかという問題がある。教育は陶冶（知識・技能の習得）と訓育（道徳性・社会的行動様式の形成）による児童・生徒の「全人格の発達」を目指すものだとする、児童・生徒の「全人格の発達」を支援する能力が教員の能力であるということになる。しかし、「2. 教育の効果」で述べたように「全人格の発達」の全てを数量化することはできないので、教員の能力向上策の費用対効果の推計では、児童生徒の学力をどの程度向上させることができるかが、教員の能力であるとせざるを得ない。児童生徒の学力をどの程度向上させることができるかが、教員の能力であるとしても、その能力そのものを計測することは不可能に近い。どのような能力がどの程度、児童生徒の学力向上に有益であるのかが分からないからである。

文部省教育職員養成審議会の『養成と採用・研修との連携の円滑化について（第3次答申）』（1999年12月10日）は、「昭和62年12月18日付け本審議会答申「教員の資質能力の向上方策等について」……において示されているとおり、教育者としての使命感、人間の成長・発達についての深い理解、幼児・児童・生徒に対する教育的愛情、教科等に関する専門的知識、広く豊かな教養、そしてこれらを基盤とした実践的指導力といった能力がいつの時代にも教員に求められる資質能力であると考え」と述べているが、これらの資質能力と児童生徒の「全人格の発達」との因果関係も寄与度も明らかではない。「教育者としての使命感」や「教育的愛情」に溢れていても、使命感や愛情が空回りするだけで、児童・生徒の学力を向上させることができなければ、その教員は無能と言うべきであろう。そもそも、使命感や愛情を客観的に計測する方法がない。「広く豊かな教養」の具体的内容が明らかではない。児童・生徒の「全人格の発達」を支援するために必要な教養とは何かを明らかにすることが必要であるが、そのためには大量のデータを集めて緻密な調査分析を行う必要があり、容易なことではない。結局、教員の能力は、児童・生徒の学力をどの程度向上させたのかという実績に基づいて計測するしかない。児童・生徒の学力を向上させたという実績こそが「実践的指導力」の客観的な証拠である。

(2) 教育生産関数の修正

教員の能力を児童・生徒の学力をどの程度向上させたのかという実績に基づいて計測するためには、教員の能力を説明変数の一つとした教育生産関数（「第2回 教育経済学の基本 9. 教育生産関数」参照）を用いることが有益である。

$$A = a + b \sum B_{it} (S_{i+1} - S_i) + cT + dL + eP + fF + gE + hI + u \quad (11-1)$$

ここで、 $B_{it}$  は、第  $i$  期（時点  $i$  から時点  $(i+1)$  の直前まで）の教育を担当した教員  $t$  の能力で、 $(S_{i+1} - S_i)$  は、第  $i$  期の学校教育である。第  $i$  期の学校教育が児童生徒の学力  $A$  の向上に与える影響は、第  $i$  期の教育を担当した教員の能力に左右されるという趣旨である。 $B_{it} = 1$  であれば、学力の向上は標準的なものであるが、 $B_{it} > 1$  であれば、学力の向上は標準を上回り、 $B_{it} < 1$  であれば、学力の向上は標準を下回る。児童・生徒の生得的な能力・気質  $I$  は変化しないと仮定すると、(11-1)式から、第  $i$  期の教育を担当した教員  $t$  の能力は、

$$B_{it} = \frac{1}{b(S_{i+1} - S_i)} \left[ (A_{i+1} - A_i) - c(T_{i+1} - T_i) - d(L_{i+1} - L_i) - e(P_{i+1} - P_i) \right. \\ \left. - f(F_{i+1} - F_i) - g(E_{i+1} - E_i) - (u_{i+1} - u_i) \right] \quad (11-2)$$

となる。

なお、「第2回 教育経済学の基本 9. 教育生産関数」で述べた、教育生産関数が線形関数であると仮定することの問題点に注意する必要がある。特に、学校教育  $S$  の効果は家庭環境  $F$  や児童・生徒の生得的な能力・気質  $I$  などに影響されるという関係がある（ $b = f(F, I)$ ）のではないかということが問題である。 $b = f(F, I)$  であるとする、家庭環境  $F$  や生得的な能力・気質  $I$  に恵まれた児童・生徒の教育を担当した教員の能力が過大評価され、家庭環境  $F$  や生得的な能力・気質  $I$  に恵まれない児童生徒の教育を担当した教員の能力が過小評価される危険性がある。(11-2)式は不完全な教育生産関数に基づいているため、不完全なものなのである。

また、「第2回 教育経済学の基本 10. 回帰分析」で述べたように、児童・生徒の学校外教育  $T$ 、自学自習  $L$ 、仲間集団の影響  $P$ 、家庭環境  $F$  に関するデータは、その一部しか入手できないというのが現実である。各種データの入手が困難な状況では、学校外教育  $T$ 、自学自習  $L$ 、仲間集団の影響  $P$ 、家庭環境  $F$ 、社会環境  $E$  の影響を無視した

$$B_{it} = \frac{1}{b(S_{i+1} - S_i)} [(A_{i+1} - A_i) - (u_{i+1} - u_i)]$$

というような粗雑な分析が行われる危険性がある。このような粗雑な分析では、学力向上に有利な状況にある児童生徒の教育を担当した教員の能力が過大評価され、学力向上に不利な状況にある児童生徒の教育を担当した教員の能力が過小評価されてしまう。

ただし、学級への児童・生徒の割り振りが無作為になされているのであれば、各学級の児童生徒の学校外教育  $T$ 、自学自習  $L$ 、仲間集団の影響  $P$ 、家庭環境  $F$ 、社会環境  $E$  が

ほぼ等しくなるので、同一学校内での教員の能力の相対評価としては妥当性がある。例えば、同一学校の同一学年内で、児童・生徒の学級への割り振りが無作為になされているならば、学級  $\alpha$  と学級  $\beta$  では、 $A_i$ 、 $(S_{i+1} - S_i)$ 、 $(T_{i+1} - T_i)$ 、 $(L_{i+1} - L_i)$ 、 $(P_{i+1} - P_i)$ 、 $(F_{i+1} - F_i)$ 、 $(E_{i+1} - E_i)$ 、 $(u_{i+1} - u_i)$  がほぼ等しくなるので、

$$B_\alpha - B_\beta \approx \frac{A_{\alpha(i+1)} - A_{\beta(i+1)}}{b(S_{i+1} - S_i)} \quad (11-3)$$

となり、学級  $\alpha$  と学級  $\beta$  の第  $(i+1)$  期での成績差から、学級  $\alpha$  を担当した教員の能力  $B_\alpha$  と学級  $\beta$  を担当した教員の能力  $B_\beta$  の差を推計することができる。ただし、分かるのは教員

間の能力の相対的な差（相対評価）だけであり、教員が児童・生徒の学力向上にどれだけ貢献したのか（絶対評価）、寄与の割合はどれだけかは分からない。教員の能力に偏差値を付けて、偏差値で 10 違うと 1 年間の教育で児童・生徒の学力が偏差値で 1 違ってくるというようなことは示せるが、偏差値 50 の教員が児童・生徒の学力向上にどの程度貢献しているかは分からないということである。偏差値 50 の教員が担当した学級の学力向上の度合い（例えば、1 年間の教育で偏差値が 5 違ってくるというようなこと）を測定することはできるが、その値から、学校外教育 T、自学自習 L、仲間集団の影響 P、家庭環境 F、社会環境 E の影響を除去して、教員の貢献だけを抽出すること（例えば、偏差値 5 のうち、教員の貢献は 2 であって、残り 3 は他の要因の影響によるものであると示すようなこと）ができないのである。

なお、事前の学力テストの点数において、各学級が同レベルであっても、児童・生徒の学校外教育 T、自学自習 L、仲間集団の影響 P、家庭環境 F、社会環境 E、生得的な能力・気質 I に違いがあれば、その後の学力の伸びが違ってくるので、各学級への児童・生徒の割り振りが無作為になされた場合と同等とみなしてはならない。

教員間の能力の相対的な差しか分からなくても、それを用いて、教員の能力向上策の費用対効果を分析することはできる。ある施策の実施によって、教員の能力を偏差値で 10 引き上げることができれば、児童・生徒の学力向上の度合いが 10% 向上するというようなことが分かるからである。

しかし、教員を雇うこと自体の費用便益分析、端的に言うと、教員が給料以上の働きをしているのか、それとも、給料以下の働きしかしていないのかを知るためには、教員が児童・生徒の学力向上にどれだけ貢献したのか、寄与の割合はどれだけかという絶対評価が必要である。統計分析で、教員が児童生徒の学力向上にどれだけ貢献したのかを割り出すためには、児童・生徒の学校外教育 T、自学自習 L、仲間集団の影響 P、家庭環境 F、社会環境 E、生得的な能力・気質 I に関する完全なデータが必要であり、そのようなデータの収集は不可能に近い。実験で、教員が児童・生徒の学力向上にどれだけ貢献したのかを割り出そうとすると、教師による教育を行わない（つまり、自習する）学級と教師による教育を行う学級に児童・生徒を無作為に割り振り、1 年後の成績を比較する

というような非人道的なことをすることになってしまう。非人道的な実験を強行しても、保護者は児童・生徒を転校させたり、学習塾に通わせたりするなどの対抗手段をとって、大きなバイアスが生じてしまい、実験にならないだろう。費用便益分析で教員の給料の適正水準を割り出すことは、理論的には可能だが、現実には不可能ということである。

### (3) 教員能力生産関数

教員の能力の向上方策の効果を推計するために、教員能力生産関数というものを考えてみる。教員の能力を  $B$ 、教員が受けてきた学校教育を  $S_T$ 、教員経験を  $X$ 、教員が受けてきた研修を  $W$ 、自己研鑽を  $L_T$ 、職場環境を  $P_T$ 、家庭環境を  $F_T$ 、社会環境を  $E_T$ 、教員の生得的な能力・気質を  $I_T$ 、観察できない、あるいは、未知の要因（攪乱項、誤差項）を  $\varepsilon$  とすると、

$$B = f(S_T, X, W, L_T, P_T, F_T, E_T, I_T) + \varepsilon \quad (11-4)$$

単純化のために、線形関数であると仮定すると、

$$B = \alpha + \beta S_T + \chi X + \delta W + \phi L_T + \gamma P_T + \eta F_T + \iota E_T + \kappa I_T + \varepsilon \quad (11-5)$$

（ただし、 $\alpha$ 、 $\beta$ 、 $\chi$ 、 $\delta$ 、 $\phi$ 、 $\gamma$ 、 $\eta$ 、 $\iota$ 、 $\kappa$  は定数）

理論的には、(11-5)式を用いて教員の能力に影響を及ぼす要因を変化させた時、例えば、教員資格を取得するために必要な教育年数を延ばしたり、教員に対する研修を増やしたりした時にどのような効果があるかを推計することになる。しかし、教員の生得的な能力・気質  $I_T$  に関するデータの入手は不可能に近く、教員の自己研鑽  $L_T$ 、職場環境  $P_T$ 、家庭環境  $F_T$  に関するデータはその一部しか入手できないというのが現実である。「9. 統計分析」で述べたように、Value Added モデルを使えば、時間を通じて変化しない属性による影響（固定効果、個人効果等と呼ばれる）を除去できるので、時点1と時点2を比較して、

$$\begin{aligned} B_2 - B_1 = & \alpha + \beta(S_{T2} - S_{T1}) + \chi(X_2 - X_1) + \delta(W_2 - W_1) + \phi(L_{T2} - L_{T1}) + \gamma(P_{T2} - P_{T1}) \\ & + \eta(F_{T2} - F_{T1}) + \iota(E_{T2} - E_{T1}) + (\varepsilon_2 - \varepsilon_1) \end{aligned} \quad (11-6)$$

とすると、生得的な能力・気質  $I_T$  の影響を無視できるようになる。

しかし、このような単純な線形関数を用いることには、教育生産関数に線形関数を用いた場合と同様の問題がある（「第2回 教育経済学の基本 9. 教育生産関数」参照）。第一に、 $\alpha$ 、 $\beta$ 、 $\chi$ 、 $\delta$ 、 $\phi$ 、 $\gamma$ 、 $\eta$ 、 $\iota$ 、 $\kappa$  の値は、個人毎に異なるのではないかという問題がある。例えば、学校教育  $S_T$ 、教員経験  $X$ 、研修  $W$ 、自己研鑽  $L_T$  の効果は職場環境  $P_T$ 、家庭環境  $F_T$ 、生得的な能力・気質  $I_T$  などに影響されるという関係がある（ $\beta = f(F_T, I_T)$ 、 $\chi = f(P_T, I_T)$ 、 $\delta = f(P_T, I_T)$ 、 $\phi = f(P_T, F_T, I_T)$ ）のではないかということである。 $\alpha$ 、 $\beta$ 、 $\chi$ 、 $\delta$ 、 $\phi$ 、 $\gamma$ 、 $\eta$ 、 $\iota$ 、 $\kappa$  の値が個人毎に異なると、実験や統計分析によって、上述の形の教員の能力の生産関数を推計しても、それは、実験対象や調査対象になった教員の中で平均的な教員に適用できるものに過ぎず、平均から外

れた教員には適用できないことになる。第二に、研修 X の量が職場環境  $P_T$ 、家庭環境  $F_T$ 、社会環境  $E_T$ 、生得的な能力・気質  $I_T$  の影響を受けたり ( $T = f(P_T, F_T, E_T, I_T)$ )、自己研鑽  $L_T$  の量が、研修 W、職場環境  $P_T$ 、家庭環境  $F_T$ 、個人の生得的な能力・気質  $I_T$  などの影響を受けたり ( $L_T = f(W, P_T, F_T, I_T)$ ) するというように、要因 (説明変数) 同士に影響関係があるのではないかという問題がある。第三に、学校教育  $S_T$  を増やせば (教員資格取得のために必要な教育年数を延ばせば、ということ)、それに比例して、教員の能力が向上するというような線形性が本当にあるのかという疑問がある。学校教育を増やしていくと、次第に一単位当たりの学校教育  $S_T$  の効果が増えたり (収穫逓増)、減ったり (収穫逓減) するというような非線形の関係があるのでないかということである。研修 W、自己研鑽  $L_T$  についても同様である。なお、Steven G. Rivkin、Eric A. Hanushek、John F. Kain は “Teachers, Schools, and Academic Achievement” で、修士の学位の有無は教員の能力とは無関係であることと、教員経験が教員の能力を向上させる働きは 1 年目が大きく、2 年目以降は小さくなり、3 年経過後にはほとんどなくなることを指摘している (「12. 教員の能力が児童生徒に与える影響についての調査研究の動向」参照)。

つまり、教員能力生産関数は非線形の連立方程式にすべきであり、しかも、全ての教員に適用できる教員能力生産関数にするためには、非常に複雑な形のものになるということである。しかし、現時点では、教員の能力 B、学校教育  $S_T$ 、研修 W、自己研鑽  $L_T$ 、職場環境  $P_T$ 、家庭環境  $F_T$ 、社会環境  $E_T$ 、生得的な能力・気質  $I_T$  の関係について判明していることは非常に少ないので、具体的にどのような教員能力生産関数にすれば良いのかは分からない。「7. 実験」と「9. 統計分析」で述べたように、教育分野では、実験を行うことは極めて困難であり、個人別の正確なデータ、特に時系列のデータや生得的な能力に関するデータの入手も非常に困難なので、正確な形の教員能力生産関数を割り出すことは不可能に近い。教育分野では正確なデータの入手が非常に困難であるという現状からすれば、単純な教員能力生産関数を用いた分析に甘んじざるを得ないであろう。

#### (4) 研修の効果

ある研修 Y の効果を実験によって確かめる方法を説明する。多数の教員を無作為に研修 Y を受ける実験群 a と研修 Y を受けない統制群 b に振り分ける。無作為に振り分けた結果、実験群 a と統制群 b では、研修 Y 以外の研修の平均値  $\bar{W}$ 、学校教育の平均値  $\bar{S}_T$ 、自己研鑽の平均値  $\bar{L}_T$ 、職場環境の平均値  $\bar{P}_T$ 、家庭環境の平均値  $\bar{F}_T$ 、社会環境の平均値  $\bar{E}_T$ 、生得的な能力・気質の平均値  $\bar{I}_T$ 、攪乱項 (観察できない、あるいは未知の要因)  $\varepsilon$  はほぼ等しくなるので、

$$\bar{B}_a = \alpha + \beta \bar{S}_{Ta} + \chi \bar{X}_a + \delta \bar{W}_a + \phi \bar{L}_{Ta} + \gamma \bar{P}_{Ta} + \eta \bar{F}_{Ta} + \iota \bar{E}_{Ta} + \kappa \bar{I}_{Ta} + \lambda Y + \varepsilon_a$$

$$\bar{B}_b = \alpha + \beta \bar{S}_{Tb} + \chi \bar{X}_b + \delta \bar{W}_b + \phi \bar{L}_{Tb} + \gamma \bar{P}_{Tb} + \eta \bar{F}_{Tb} + \iota \bar{E}_{Tb} + \kappa \bar{I}_{Tb} + \varepsilon_b$$

から、

$$\bar{B}_a - \bar{B}_b \approx \lambda Y \quad (11-7)$$

となり、研修 Y の効果だけを計測できるようになる。

次に、 $\bar{B}_a$  と  $\bar{B}_b$  を計測する必要がある。無作為に選んだ学校に、実験群 a と統制群 b の教員の分布が均等になるように、各学校の学級数に比例した人数を無作為に割り振る。例えば、学級数 4 の学校には、実験群 a から 1 名、統制群 b から 1 名割り振り、学級数 8 の学校には、実験群 a から 2 名、統制群 b から 2 名割り振り、学級数 12 の学校には、実験群 a から 3 名、統制群 b から 3 名割り振るようにするということである。各学校では、児童生徒を無作為に学級に割り振り、それらの学級に実験群 a の教員と統制群 b の教員を無作為に割り振り、1 年間ほど授業を行わせた後（能力を計測する期間が 1 年間で良いのかという問題がある。長期間経過後に効果が表れてくる教育があるかもしれないからである）、児童生徒の成績を共通の学力テストによって測定する。実験群 a が担当した学級の平均成績を  $\bar{A}_{a(i+1)}$  とし、統制群 b が担当した学級の平均成績を  $\bar{A}_{b(i+1)}$  とする。実験群 a と統制群 b の教員の分布が均等になるように、各学校の学級数に比例した人数を無作為に割り振り、各学校で児童生徒を無作為に学級に割り振った結果、実験群 a が担当する学級と統制群 b が担当する学級の児童生徒の学校教育 S、学校外教育 T、自学自習 L、仲間集団の影響 P、家庭環境 F、社会環境 E がほぼ等しくなるので、(11-3)式が成り立ち、

$$\bar{B}_a - \bar{B}_b \approx \frac{\bar{A}_{a(i+1)} - \bar{A}_{b(i+1)}}{b(\bar{S}_{i+1} - \bar{S}_i)} \quad (11-8)$$

となり、(11-7)式を代入して、

$$\lambda Y \approx \frac{\bar{A}_{a(i+1)} - \bar{A}_{b(i+1)}}{b(\bar{S}_{i+1} - \bar{S}_i)} \quad (11-9)$$

となる。 $b(\bar{S}_{i+1} - \bar{S}_i)$  は、平均的な学校における 1 年間の教育であるから、平均的な学校において教員に対して研修 Y を行くと、児童生徒の学力が  $(\bar{A}_{a(i+1)} - \bar{A}_{b(i+1)})$  向上するということである。

費用便益分析を行うためには、 $(\bar{A}_{a(i+1)} - \bar{A}_{b(i+1)})$  という児童生徒の学力向上の価値を金

銭に換算しなければならないが、「3. 学力と所得の関係」で述べたように、日本には、生徒の学業成績と仕事に就いた後の所得との関係を正確に調べた調査研究が存在しないので、金銭への換算は無理である。

#### (5) 教職大学院での教育の費用対効果

教職大学院での教育の費用対効果も同様の方法の実験で推計することができるが、現職教員に教職大学院での教育を行う実験と、教員未経験者に教職大学院での教育を行う実験に分けて考える必要がある。以下の説明では小学校の教員を例にする。

現職教員に対する実験の場合には、まず、実験を実施する都道府県を無作為にいくつか選ぶ。選ばれた都道府県は、実験対象者になる教員を無作為に選び出す。実験対象者に本人が選んだ教職大学院の入学試験を受けさせ、合格しなかった者は実験対象者から外す。次に、合格した実験対象者を教職大学院に行く実験群と行かない統制群に無作為に振り分け、実験群を合格した教職大学院に行かせる。教職大学院での教育で落第した者は、実験群から外す。教職大学院での実験群の教育終了後に、無作為に選んだ小学校に、実験群と統制群の教員の分布が均等になるように、各学校の学級数に比例した人数を無作為に割り振る。例えば、学級数4の学校には、実験群と統制群から1名ずつ割り振り、学級数8の学校には、実験群と統制群から2名ずつ割り振り、学級数12の学校には、実験群と統制群から3名ずつ割り振るようにするということである。各学校では、児童生徒を無作為に学級に割り振り、それらの学級に実験群の教員と統制群の教員を無作為に割り振り、1年間ほど授業を行わせた後、児童生徒の成績を共通の学力テストによって測定する。実験群が担当した学級の平均成績と統制群が担当した教員の平均成績の差が、教職大学院での2年間の教育の効果と2年間の実務経験の効果の差となる。

教員未経験者に対する実験の場合には、まず、実験を実施する都道府県を無作為にいくつか選ぶ。選ばれた都道府県は、教員志望者に対して教員採用試験を行い、合格者の中から無作為に実験対象者を選び出す。実験対象者に本人が選んだ教職大学院の入学試験を受けさせ、合格しなかった者は実験対象者から外す。次に、実験対象者を無作為に教職大学院に行く実験群と行かない統制群に振り分け、実験群を合格した教職大学院に行かせ、教職大学院での教育終了後に実験群を教員として採用し、無作為に選んだ小学校で3年間、教育を担当させる。教職大学院での教育で落第した者は、実験群から外す。統制群は教員として採用し、無作為に選んだ小学校で5年間、教育を担当させる。5年経過後に、無作為に選んだ小学校に、実験群と統制群の教員の分布が均等になるように、各学校の学級数に比例した人数を無作為に割り振る。各学校では、児童生徒を無作為に学級に割り振り、それらの学級に実験群の教員と統制群の教員を無作為に割り振り、1年間ほど授業を行わせた後、児童生徒の成績を共通の学力テストによって測定する。実験群が担当した学級の平均成績と統制群が担当した教員の平均成績の差が、教職大学院での2年間の教育の効果と2年間の実務経験の効果の差となる。なお、実験群に3年間の

実務経験をさせてから、能力を計測するのは、Steven G. Rivkin、Eric A. Hanushek、John F. Kain が “Teachers, Schools, and Academic Achievement” で、教員経験が教員の能力を向上させる働きは 1 年目が大きく、2 年目以降は小さくなり、3 年経過後にはほとんどなくなると指摘していることを踏まえたものである。

これらの実験には多くの問題がある。実験対象になるかならないか、また、教職大学院に行くか行かないかを無作為で決める（例えば、抽選で決める）ことに対する抵抗は、非常に強いと思われる。教職大学院の学費とその間の生活費を誰が負担するのかという問題もある。実験をスムーズに行うためには、実験の実施者が教職大学院の学費とその間の生活費を負担することが必要であろう。そうすると、実験に要する費用が相当な金額になってしまう。また、実験期間中に教員が不足する、教職大学院への入学希望者が急増・急減する、入学辞退者が増えるという問題もある。これらの影響を小さくするためには、実験規模を小さくする必要があるが、実験規模が小さ過ぎると、信頼性が低下する。実験に 3 年または 6 年という期間を要するという問題もあるが、これはどうしようもない。実験対象者の授業を受ける児童やその保護者から不満が出る可能性もある。以上述べたような問題があることからすると、実験を行うことは不可能に近いであろう。

次に、準実験的な方法を考えてみる。まず、準実験を実施する都道府県を無作為でいくつか選ぶ。選ばれた都道府県は、教職大学院を終了しているか否か以外の属性の分布が、当該都道府県内の教員の属性の分布とほぼ等しくなるように、教職大学院を修了した教員のグループ A と、教職大学院に行っていない教員のグループ B を作る。しかし、教員の属性で調査可能なことは少ないので、同じ大学の同じ学部を卒業している者の割合が同じであるとか、教員経験年数（教職大学院修了者については、教員経験年数に 2 年プラスする）の分布が同じであるとかいう程度のことしかできないであろう。そのため、準実験による教職大学院の教育の効果の推計はかなり不正確なものになる。そもそも、教職大学院は 2008 年度にできたばかりなので、教職大学院修了者の数が少なく、グループ A の教員の属性の分布を当該都道府県内の教員の属性の分布とほぼ等しくなるようにするというのには無理があると思う。準実験の精度が落ちるが、グループ A の教員の属性の分布とほぼ等しくなるようにグループ B を作るという方法を取らざるを得ないであろう。次に、無作為に選んだ小学校に、グループ A とグループ B の教員の分布が均等になるように、各学校の学級数に比例した人数を無作為に割り振る。各学校では、児童生徒を無作為に学級に割り振り、それらの学級にグループ A の教員とグループ B の教員を無作為に割り振り、1 年間ほど授業を行わせた後、児童生徒の成績を共通の学力テストによって測定する。グループ A が担当した学級の平均成績とグループ B が担当した教員の平均成績の差が、教職大学院での 2 年間の教育の効果と 2 年間の実務経験の効果の差となる。

なお、教員の能力  $B$ 、自己研鑽  $L_T$ 、職場環境  $P_T$ 、家庭環境  $F_T$ 、生得的な能力・気質  $I_T$  に関する網羅的で正確なデータと児童生徒の能力  $A$ （正確な統計分析のためには、全国共

通のテストの個人別の成績が必要である)、学校外教育 T、自学自習 L、仲間集団 P、家庭環境 F、生得的な能力・気質 I に関する網羅的で正確なデータの入手は不可能に近いので、統計分析では、教職大学院の教育の効果のまともな推計はできない。入手容易だが不完全で不正確なデータに基づいて統計分析を行っても、不正確な推計結果が出るだけで、害悪にしかならない。

次に費用の推計であるが、教職大学院の教育費用の推計には困難を伴う。個人が払う学費だけでは不十分で、公的補助や学部から大学院への内部補助、そして放棄所得（教職大学院に入学すれば、大学を卒業して直ぐに就職していれば得られていたであろう 2 年間の賃金を得ることができない）を含める必要がある。大学全体への公的補助から、どのようにして、教職大学院分を切り分け、教育費用と研究費用を分離するのかというのは難問である。例えば、大学院と学部の両方を担当している教員の人件費をどのようにして切り分けるのか。担当コマ数で切り分けるのか。管理費（事務費）をどのように割り振るのか。大学院と学部で施設・設備を共用している場合にはどのようにして費用を切り分けるのか。利用時間で切り分けるのか。国立大学法人や学校法人が公開している財務諸表、計算書類のみから教職大学院の教育費用を計算することは不可能であるので、国立大学法人や学校法人の協力を得る必要がある。放棄所得に関しては、大学学部卒業後に教員経験なしに教職大学院に進学した者に関しては、大学卒業後直ぐに教員になった場合の 1 年目と 2 年目の給料ということになる。現職教員が教育委員会から派遣されて教職大学院に行く場合は、教育委員会がその間支払った給料を教育費用に加える必要があるが、経験年数によって給料に差があるので、どの人を基準にして計算するのは難問である。

## 11. 学級規模縮小の効果に関する調査研究の動向

### (1) 実験

1985 年~1989 年にアメリカのテネシー州で、「STAR プロジェクト」(the Tennessee Student/Teacher Achievement Ratio experiment、略して、Project STAR) という、学級規模縮小の効果に関する実験が行われた。STAR プロジェクトでは、参加校の児童を、幼稚園（日本の幼稚園での年長組に相当）から小学校 3 年生になるまでの 4 年間、13~17 人の少人数学級、22~26 人の通常規模学級、常勤の補助教員付きの通常規模学級の 3 種類に無作為に振り分け（どの学校に入るかは無作為ではなく、特定の学校内のどの学級に入るかが無作為であるということ）、それぞれの学級に教員を無作為に割り当て、同じカリキュラムで教育を行い、学業成績（毎年 3 月末か 4 月はじめに行われる読解、言語認識、算数に関する標準化されたテストで測定）が調査された。少人数学級と常勤の補助教員付きの通常規模学級が実験群、通常規模学級が統制群ということになる。教員に特別な訓練を施すことはなく、カリキュラムは特別なものではなく、通常のもので使われた。調査期間終了後、児童は通常規模の学級に戻されたが、その後も学業成績等の追

跡調査が行われた (the Lasting Benefits Study)。1年目の参加校は79校で、参加児童は6,324人、2年目の参加校は76校で、参加児童は6,829人、3年目の参加校は75校で、参加児童は6,840人、4年目の参加校は75校で、参加児童は6,802人である。調査の途中で参加校に入学・転校してきた児童も調査対象に加え、他の学校に転校した児童は調査対象から外したため、小学校1年生からの参加が2,314人、2年生からの参加が1,791人、3年生からの参加が1,389人いる。このため、最初に調査に参加した児童で、調査の最後までいた児童は、48%である。

(注) アメリカの小学校には、付属の幼稚園があるところが多い。

STAR プロジェクトは実験の困難性を示している。無作為なクラス分けに父母から不満が出て、不満をなだめるために、小学校1年生になるときに、通常規模学級と補助教員付きの通常規模学級との間で無作為な入れ替えが行われた(ただし、少人数学級では入れ替えは行われていない)。児童の行動上の問題や父母の不満から、約10%の児童が学年途中で少人数学級と通常規模学級の間を移動した。住民の転居の結果、実際には、少人数学級は11~20人、通常規模学級は15~30人というサイズになってしまった。一部の児童が他の公立学校や私立学校に転校した。これらのことから、実験群と統制群への振り分けの無作為性が損なわれ、バイアスが生じてしまった。また、STAR プロジェクトの実施には約1,200万ドルの経費を要した。

STAR プロジェクトの調査結果に対しては評価が分かれている。

Alan. B. Krueger は “Experimental Estimates of Education Production Functions” で、STAR プロジェクトの調査結果を分析して、少人数学級の児童の成績は、初年度に4パーセント分上昇し、その後、毎年1パーセント分上昇した、この上昇は、マイノリティや低所得層(正確に言うと、給食費を免除されている児童)の児童で大きかった、常勤の補助教員付きの通常規模学級には、学力向上の効果はほとんどなかった(ただし、パートタイムの補助教員を頻繁に利用したので、このような結果になったのかもしれない)と指摘している。パーセントというものは、値の小さい方から大きい方へ順番にデータを並べて、何パーセント目にあるかということである。また、少人数学級に入ったことによる学力向上の効果は1年後にあらわれ、その後継続して少人数学級にいても、より小さな学力向上効果しかないことから、低学年で少人数学級に入ると、その後の学業達成のレベルを引き上げる、一回限りの「学校による社会化の効果」が働くのではないかと指摘している。

Alan. B. Krueger は、STAR プロジェクトのおおざっぱな費用便益分析も行っており、少人数学級に4年間在籍することに要した1人当たりの費用は約7,400ドル(1999年での価値に換算した金額)と推計され、少人数学級に4年間在籍したことによる学力向上は約0.22標準偏差分(偏差値にすると、約2.2)で、この学力向上による生涯賃金(20歳から65歳まで働くと仮定)の増加は、現在価値(1999年、実質利子率3%で計算)に換算して、男性で9,603ドル、女性で7,851ドルと推計されるとしている。なお、学力

向上による賃金増加の推計には、Richard J. Murnane、John B. Willett、Frank Levy が “The Growing Importance of Cognitive Skills in Wage Determination” で示している調査結果によると、1980年の時点で高校3年の数学の成績が0.22標準偏差分（偏差値に換算すると、約2.2）高いと、6年後には、男性で1.7%、女性で2.4%高い賃金を得ることになるということを利用している。

Alan. B. Krueger は “Understanding the magnitude and effect of class size on student achievement” (“Mishel, Lawrence and Richard Rothstein, eds., *THE CLASS SIZE DEBATE*.” 所収) でも、STAR プロジェクトの費用便益分析を行っている。STAR プロジェクトでは約22人から約15人に学級規模を縮小しており、1997-98会計年度に児童一人当たり教育支出は全米平均で7,502ドルなので、学級規模の縮小に要する1人当たりの費用は7,502ドルの15分の7の3,501ドルになり（22人学級でも15人学級でも、一学級当たりには費用は変わらないと仮定している）、STAR 計画で少人数学級に属した児童の平均在籍年数は2.3年なので、1年目と2年目に3,501ドルがそれぞれ必要で、3年目に3,501ドルの30%が必要で、4年目には費用を要しないと仮定して、STAR プロジェクトで少人数学級に在籍するために必要な費用を推計する。Richard J. Murnane、John B. Willett、Frank Levy などの研究結果から、数学または読解のいずれかの成績が偏差値で10高いと、所得が約8%高くなること、及びSTAR プロジェクトで少人数学級に属した児童は数学と読解の成績が0.2標準偏差分（偏差値に換算すると、2.0）高くなったことから、18歳から65歳まで働いた場合の生涯賃金の増加、つまり、便益を推計する。費用と便益は、実質利子率（＝名目利子率－物価上昇率）と実質賃金上昇率に左右され、下表のようになる。

(1998年の現在価値に換算)

実質利子率 (割引率)	費用	生涯賃金の増加額（つまり、便益）		
		実質賃金上昇率 0%	1%	2%
2%	\$7,787	\$21,725	\$31,478	\$46,294
3%	7,660	15,174	21,667	31,403
4%	7,537	10,784	15,180	21,686
5%	7,417	7,791	10,819	15,238
6%	7,300	5,718	7,836	10,889

なお、この費用便益分析では、学級規模が、テストでは計れない非認知的な能力に影響を与え、所得に影響を与える可能性を無視している。また、外部性や、非経済的な便益も無視している。

Eric A. Hanushek は “Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects” で、

要約すると次のような指摘をしている。

実験における不確実性とバイアスの可能性を無視しても、STAR プロジェクトは、かなり大規模で費用のかかる学級規模の縮小を幼稚園または小学校 1 年で行った場合に効果があることを示しているに過ぎず、小学校 2 年生以降でも効果があるかどうかについてや、STAR プロジェクト以上の学級規模の縮小を行った場合にどのような効果があるかについては何も示していない。

STAR プロジェクトの無作為性にはいくつかの問題があり、バイアスが生じている可能性がある。第一に、参加校は自ら希望した学校であり、無作為には選ばれていない。少人数学級、通常規模学級、常勤の補助教員付きの通常規模学級の 3 種類の学級を作ることができない小規模校は参加できない。テネシー州の公立学校では児童の 23% が黒人であるが、参加校では児童の 33% が黒人である。低所得層やマイノリティは、学級規模の大小の影響を受けやすいので、学級規模縮小の効果は、対象児童の構成の影響を受ける。第二に、教員配置は無作為になされたとされているが、具体的にどのような方法で配置されたのかに関しては、ほとんど説明がない。Alan. B. Krueger は “Experimental Estimates of Education Production Functions (Working Paper 1977)” で、人種、経験年数、学位のチェックによって、教員配置に偏りがないと指摘しているが、人種、経験年数、学位だけでは教員の質は分からないので、教員配置に偏りがないとはいえない。第三に、毎年、20～30%の児童が転校等で調査対象から外れている。第四に、毎年、9～12%の児童が、通常規模学級から少人数学級に移っているが、少人数学級から通常規模学級に移った児童は 1～2%に過ぎない。校長が、自分の子どもを少人数学級に入れて欲しいという父母の圧力にさらされていたことを示唆している。第五に、毎年、児童の 3～12%が標準化されたテストを受けていない。第六に、教員や校長は、少人数学級で成果を示して欲しいとの期待にさらされていた。なお、Eric A. Hanushek は “Evidence, politics, and the class size debate” で、参加校は費用の一部を自己負担しなければならなかったという問題もあったことを指摘している。

下表に、全児童と、最初から調査に参加して最後までいた 48%の児童（4 年間継続して調査に参加した児童）の平均成績を示す。

全児童	(偏差値)							
	幼稚園 読解	幼稚園 算数	1 年 読解	1 年 算数	2 年 読解	2 年 算数	3 年 読解	3 年 算数
少人数学級	51.2	51.2	51.7	51.9	51.4	51.3	51.5	51.2
通常規模学級	49.5	49.5	49.3	49.2	49.4	49.4	49.3	49.4
差	1.7	1.7	2.3	2.6	2.0	1.9	2.2	1.8

4年間継続して調査に参加した児童

(偏差値)

	幼稚園 読解	幼稚園 算数	1年 読解	1年 算数	2年 読解	2年 算数	3年 読解	3年 算数
少人数学級	53.8	54.0	54.1	54.4	53.4	53.2	53.2	52.9
通常規模学級	51.9	52.0	52.6	52.2	52.3	52.0	51.8	51.9
差	1.8	1.9	1.5	2.2	1.1	1.3	1.4	1.0

Eric A. Hanushek は点数を標準得点（得点から平均値を引いて、標準偏差で割った数字）で示しているが、それを偏差値に換算した（標準得点を10倍して、50を加える）。

4年間継続して調査に参加した児童は、バイアスの影響を免れているので、これらの児童について見ると、少人数学級と通常規模学級の学力差は、学年が進むにつれて、縮小していることが分かる。少人数学級に在ることに累積的な効果があるならば、学力差は拡大していくはずである。この結果は、少人数学級には一回限りの効果しかないという見解と矛盾しない。少人数学級に居続けないと、効果が消えていく可能性もある。STAR プロジェクトの実験デザインでは、どちらが正しいかは分からない。“the Lasting Benefits Study” の小学校6年生時の報告書（1993年）では、STAR プロジェクト実施時に少人数学級にいた児童は、通常規模学級にいた児童に比べて、読解のテストでは偏差値で2.1、算数のテストでは偏差値で1.6、それぞれ成績が良かったとされているが、“the Lasting Benefits Study” の元データが公開されていないので、それが本当かどうか分からない。

幼稚園段階で、少人数学級の成績が通常規模学級と常勤の補助教員付きの通常規模学級の成績を上回ったのは、79校中40校に過ぎない。これは、学級規模以外の要因が学力を決める上で重要であることを示している。

STAR プロジェクト終了後の追跡調査（the Lasting Benefits Study）の結果は明快ではない。

Barbara Nye, Larry V. Hedges, Spyros Konstantopoulos は “The Long-Term Effects of Small Classes: A Five-Year Follow-up of the Tennessee Class Size Experiment” で、“the Lasting Benefits Study” の結果を分析して、学級規模の縮小による学力向上の効果は、4年生になって通常規模学級に戻った後、減少するが、8年生になっても、効果の70%以上は残っていると指摘している。

これに対し、Alan B. Krueger と Diane M. Whitmore は “The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR” で、“the Lasting Benefits Study” の結果を分

析して、4年生になって通常規模学級に戻った後、8年生になるまでに、学級規模の縮小による学力向上の効果の少なくとも半分は消えると指摘している。また、少人数学級に在籍した生徒は、アメリカの大学入学のための試験である ACT や SAT を受験する割合が 2.7%高くなり、特にマイノリティや低所得層（正確に言うと、給食費を免除・減額されている生徒）の生徒には影響が大きい。黒人の場合は、ACT または SAT を受験する割合が 5.9%高くなる。少人数学級に在籍した生徒も、通常規模の学級に在籍した生徒も ACT や SAT での成績の平均値はほぼ等しいが、少人数学級に在籍したことによって ACT や SAT を受験する人が増えることを考慮すれば、少人数学級に在籍したことによって、0.1 標準偏差分（偏差値に換算すると、1.0）高い成績（黒人の場合は、偏差値で 2.0～2.6 高い成績）を取っていると言える。と述べている。

## (2) 準実験的な方法

準実験的な方法による調査研究では、1996～2000 年度（ウィスコンシン州の学校の年度、1 学年は 10 月に始まり 5 月に終わる）に、アメリカのウィスコンシン州で実施された SAGE プログラム（the Student Achievement Guarantee in Education (SAGE) Program）が有名である。貧困レベル以下の家庭の子どもが多い学校 30 校を対象にし、類似した世帯収入、読解の成績、人種構成の児童がいる学校 14～17 校と、幼稚園年長組から小学 3 年生までの 4 年間、学業成績（プロジェクト開始時と各学年末に実施されたテストで測定）が比較された。SAGE プログラムでは、児童と教員を無作為に各学級に割り当てることができず、準実験的なマッチング法がとられたということである（「8.準実験的な方法」参照）。実験対象校には、貧困レベル以下の家庭の子どもが 50%以上いる学校が 1 校以上ある学校区が、貧困レベル以下の家庭の子どもが 30%以上いる学校を 1 校選んで応募できるとされ（人口の多いミルウォーキー学校区のみ 10 校まで応募できるとされた）、応募してきた学校は全て実験対象校に選ばれた。実験対象校には、低所得の家庭の子ども一人当たり最大 2,000 ドルが支給された。実験対象校では、1996 年度に幼稚園年長組と小学校 1 年生で、1997 年度に幼稚園年長組から小学校 2 年生まで、1998 年度以降は幼稚園年長組から小学校 3 年生まで、12～15 人の学級が編成され（児童と教員の比率を 15 対 1 にすることを目的にしていたので、30 人以下の学級に教員 2 人などのタイプもあった）、比較対照校での学級の規模は 21～25 人であった。

SAGE プログラムの結果に関しては、University of Wisconsin–Milwaukee の “the SAGE Evaluation Team” (Alex Molnar, Philip Smith, John Zahorik 等) が、“1997-98 Evaluation Results of The Student Achievement Guarantee in Education (SAGE) Program”、“1999-2000 Evaluation Results of The Student Achievement Guarantee in Education (SAGE) Program”、“2000-2001 Evaluation Results of The Student Achievement Guarantee in Education (SAGE) Program ”などで報告している。

これらの報告をまとめて、Phil Smith、Alex Molnar、John Zahorik は “Class Size

Reduction in Wisconsin: A Fresh Look at the Data” (Arizona State University の刊行物の方) で、学級規模縮小による成績向上の効果は小学校 1 年生で現れ、2 年生、3 年生でも維持された、つまり、2 年生、3 年生になっても、実験対象校と比較対照校の成績差は広がることも縮まることもなかった、また、実験対象校では、アフリカ系アメリカ人と白人との成績差が縮小していったが、比較対照校では成績差が拡大していったと述べている。さらに、Phil Smith 等は、1997 年度に 1 年生になって、その後継続して SAGE プログラムに参加した児童を対象を絞って、SAGE プログラムの結果を再分析して、次のように指摘している。第一に、学級規模縮小の効果は、一般的な児童の 1 学年での成長の 25~30%に相当する。第二に、学級規模縮小による成績向上の効果は、1 年生で現れ、2 年生、3 年生でも維持されるが、SAGE プログラムが終わって、通常規模学級に戻った後も維持されるかどうかは、利用可能なデータからは何とも言えない。第三に、少人数学級の利益は全ての児童が受けるが、アフリカ系アメリカ人はより大きな利益を得る。特に、学級内に社会経済的地位 (SES) が上位 (給食費の減額や免除を受けていない児童を上位、減額されている児童を中位、免除されている児童を下位としている) の児童がいると、その利益はさらに大きくなる。第四に、少人数学級には、アフリカ系アメリカ人と白人との成績差を、1 年生では縮小させ、2 年生、3 年生ではその拡大を防ぐ効果がある。通常規模学級では、アフリカ系アメリカ人と白人との成績差は毎年、拡大していった。第五に、少人数学級の欠席率の高い (1 年に 16 日以上欠席) 児童の成績は、通常規模学級の欠席率の低い (1 年での欠席が 5 日以下) 児童の成績とほぼ同じなので、学級規模の縮小は、貧困地域の児童は欠席率が高いため成績が低いという問題の対策になる。

SAGE プログラムでは、学級規模の縮小が教員や児童の行動に与える影響を調べるために、教員へのアンケート調査、インタビュー、学級の観察なども行われた。この結果、分かったことに関して、Phil Smith、Alex Molnar、John Zahorik は “Class Size Reduction in Wisconsin: A Fresh Look at the Data” (EDUCATIONAL LEADERSHIP, 掲載の方) で、少人数学級の教員は、授業により多くの時間を使うようになり、児童個々により多くの注意を払うようになり、児童のニーズや関心をより多く知るようになった、また、学級の運営管理や事務的作業に使う時間が少なくなり、現場で、個別的な活動をする時間が増えた、その結果、指導上の問題を抱えることが少なくなり、職務への満足度が増したと述べている。

赤林英夫氏と中村亮介氏は『学級規模縮小が学力に与えた効果の分析 —横浜市公開データにもとづく実証分析—』で、日本の義務教育段階の学級編制制度の非連続性を利用して、学級規模縮小の学力向上効果の識別を行っている (“Regression-discontinuity design” による分析)。日本における学級規模は 40 人を標準とし (小学校 1 年生では 35 人)、学級規模は在籍児童生徒数に応じて決まるので、在籍児童生徒数が 40 人から 41 人に変化すれば、学級規模は 40 人 (1 学級) から 20~21 人 (2 学級) へと変化すること

を利用したものである。分析の対象となったのは、横浜市への情報公開請求によって入手した 2008～2009 年度の「横浜市学習状況調査」と 2007～2009 年度の「全国学力・学習状況調査」の学校別平均正答率である。横浜市以外にも自治体独自の学力テスト結果を求めて情報公開請求を行ったが、「過度な競争が生じるおそれ」などの理由により不開示の決定がなされたとのことである。1 年間での変化を見る Value Added モデルによる分析の結果、1 学級あたりの児童生徒数が 1 人減ると、小学校 6 年生の国語の成績が 0.0112 標準偏差分（偏差値に換算すると、0.112）高くなること分かったが、小学校 6 年生の算数及び中学校 3 年生の国語と数学では、統計的に有意な効果は観察されなかった（学級規模と成績は無関係であるということ）。

### (3) 統計分析

1996 年 7 月、アメリカのカリフォルニア州で、学級規模の縮小を促進するための法律が制定され、同年秋の学期始めから実施された。当時のカリフォルニア州の小学校の学級規模は平均 29 人で、アメリカで一番大きかったが、この法律では、幼稚園年長組から小学校 3 年生までの学級規模を 20 人以下にすることが目指され、そのために、学級規模の縮小を行う学校区に対して、生徒 1 人あたり、1996 年度で 650 ドル、2004 年度で 928 ドル支給された（1998 年度までは教室を新設する際の補助金もあった）。この法律の実施に要した総額は、1996 年度で約 10 億ドル、2003 年度で約 16 億ドルである。1996 年度には、幼稚園年長組の 14%、小学校 1 年生の 88%、小学校 2 年生の 57%、小学校 3 年生の 18%が小規模学級に属するようになり、1999 年度には、それぞれ、96%、99%、97%、91%が小規模学級に属するようになった。

カリフォルニア州は、この学級規模縮小の効果を検証する計画を持っておらず、効果の検証は、1998～2002 年に、5 つの研究機関が組織するコンソーシアム（the CSR Research Consortium）によってなされ、その結果は、G. W. Bohrnstedt and B. M. Stecher eds. “What We Have Learned About Class Size Reduction in California”などで報告されている。“What We Have Learned About Class Size Reduction in California”は、学級規模縮小と学業成績の関係については結論が出なかったと述べている（1998 年の試験成績が 3 年生では少人数学級の方が少し高く、この差が 4 年生になって通常規模の学級に移った後も継続したことが確認できただけ）。また、教員数が 1995 年度の 62,226 人から、1998 年度の 91,112 人まで、42%増えたため、十分な資格のない（インターンや臨時の資格などの）教員が多くなった（プログラム開始前に 1.8%であったものが、2 年目には 12.5%に）こと、少人数学級では、教員は、個々の児童により多くの注意を払うようになり、児童の行動上の問題が少なくなり、1 日の授業目標を達成しやすくなるが、大人数学級と類似のカリキュラムで、同じような教え方をしていること、学級規模縮小のために州政府から交付される金額が少なかったため、各学校区では、不足額を補うために、施設設備費や管理費を減らしたり、職業開発、コンピュータ教育、図書のための

予算を減らしたりしていること、教室の不足を補うために、特別支援教育、音楽、芸術、体育などのための部屋を一般の教室に転用していることも指摘している。

Fatih Unluは“California Class Size Reduction Reform: New Findings from the NAEP”で、the CSR Research Consortiumの報告書は、カリフォルニア州で学級規模縮小に参加した学校と参加しなかった学校との比較に基づいたものであり、児童の属性の違い（マイノリティや低所得層の多い都心部の学校の参加が遅れたなど）による影響を除去できておらず、また、学級規模縮小実施前の児童の学業成績に関するデータを欠いているという問題があると指摘し、1996年と2000年のthe National Assessment of Educational Progress (NAEP) の児童個人単位の成績データを利用して、小学校4年生のNAEPの数学の成績を比較すると、少人数学級にいた児童の方が、他州の同じ属性の学校の同じ属性の児童に比べて、0.2～0.3標準偏差分（偏差値に換算すると、2～3）高いと述べている。

国立教育政策研究所の研究グループによる「学級規模に関する調査研究」は、2000年度に、日本全国から無作為に選んだ公立小学校（156校）の5年生と、公立中学校（143校）の2年生の算数・数学、理科の学力（1回実施されたテストの得点）と学級規模（20人以下、21～25人、26～30人、31～35人、36～40人に区分）の関係を調査し、学級規模による学力の差はほとんどないという結果を得ている。また、「教師の指導法の調査において、学級人数により指導法に大きな差が見られなかったことを考えると、この結果はある意味で一貫した結果となっている。米国における調査では、学級人数による成績の差が報告される場合があるが、これは実際には指導法などの差を反映しているのかもしれない」と分析している。教師の指導法の調査というのは、この調査の一環として小学校と中学校の教員に対して行われた調査である。

この国立教育政策研究所の調査に対して、小塩隆士氏と妹尾渉氏は『日本の教育経済学：実証分析の展望と課題』で、「残念ながら、この分析は学級規模以外の要因を制御していない。そのため、ここから明確な結論や政策的な意味合いを導き出すことは、まったくできない形になっている」と指摘している。確かに、国立教育政策研究所の調査は自然にできた学級規模の大小を利用したものであるため、無作為抽出とは言っても、学級規模の大小のでき方次第では、学級規模によるグループ間で、学級規模以外の学校教育の平均値 $\bar{S}$ 、学校外教育の平均値 $\bar{T}$ 、自学自習の平均値 $\bar{L}$ 、仲間集団の平均値 $\bar{P}$ 、家庭環境の平均値 $\bar{F}$ 、社会環境の平均値 $\bar{E}$ 、児童生徒の生得的な能力・気質の平均値 $\bar{I}$ に違いが出て、これらの違いが学力に影響を与えている可能性がある。日本で学級規模の大小ができる最大の原因は、学級規模は40人（小学校1年生では35人）を標準としていることである。例えば、一つの学校の在籍児童生徒数が40人から41人に増加すれば、40人学級1クラスから、20人学級と21人学級の2クラスへと変化する（当然、この逆のケースもある）ということである。ということは、地域の児童生徒数の増減という社会環境の変化が児童生徒の学力に与える影響が軽微であれば、「7.実験」で説明した

“Regression-discontinuity design”に近いものとなり、自然にできた学級規模の大小を利用した調査でも、ある程度の信頼性があるということになる。ただし、在籍児童生徒数が 80 人から 81 人に変化すれば学級規模は 40 人（2 学級）から 27 人（3 学級）へと変化し、在籍児童生徒数が 120 人から 121 人に変化すれば学級規模は 40 人（3 学級）から 30～31 人（4 学級）へ変化するというように、大規模校になるほど学級を分割した後の規模が大きくなるので、学校の規模が学力に影響を与えているならば、このことがバイアスになっている可能性がある。例えば、20 人以下の学級は 1 学年 2 学級以下の規模の学校にしかできないので、学校の規模が小さいほど学力が低くなる傾向があるとすると（学校の規模と学力の関係については、はっきりしたことは分かっていない）、このことが、少人数学級化による学力向上を相殺してしまう可能性がある。なお、1 年以上の期間を置いて学力テストが 2 回以上行われていれば、Value Added モデルを用いて、調査の精度を増すことができたはずである。

この国立教育政策研究所の調査の一環として行われた、適正な学級規模に関する意識の実態調査では、36～40人の学級規模を多すぎると感じていると回答した教職員が多い。また、「児童生徒の実態把握」に関する教員の意識の調査も行われており、小学校では児童数26人から28人を境に「大変苦勞した」とする割合が上昇に転じる傾向がみられ、中学校では生徒数31人を境に苦勞の割合がやや上昇傾向に転じているという結果が出ている。他方、「学級の運営や児童・生徒の生活・生徒指導」に関する教員の意識の調査では、どの規模の学級でも指導の実態は類似しているようである。

須田康之氏、水野孝氏、藤井宣彰氏、西本裕輝氏、高旗浩志氏は「学級規模が授業と学力に与える影響 —全国 4 県児童生徒調査から—」で、小学校では学級規模が小さいほど学力偏差値が高く、中学校では「21・25 人学級」での学力偏差値が高かったとしている。

この「全国 4 県児童生徒調査」には問題がある。調査対象となったのは、北海道、広島県、島根県、沖縄県の小学校 5 年生と中学校 2 年生であるが、小学校 58 校中 45 校、中学校 29 校中 18 校が広島県の学校であり、サンプルが偏っている。調査が行われた時期には、広島市では少人数教育推進のための段階的プランが実施されており、教育方法の改善も行われていたので、教育方法の改善が学力に影響を与えた可能性がある。

篠崎武久氏は、千葉県検証改善委員会『平成 19 年度「全国学力・学習状況調査」分析報告書』「第 7 章 教育資源と学力の関係」で、2007 年 4 月に文部科学省が実施した「全国学力・学習状況調査」の千葉県内の公立校に関するデータを学校単位で分析して、教育生産関数を推計し、シミュレーションすると、小学校 6 年生の国語では、授業研究を伴う研修回数が 0 回の場合は、学級規模が 28～29 人の場合に平均正答率が最も低くなり、これよりも規模が小さい、または大きい場合に平均正答率が高くなり、38 人の場合と 19 人の場合がほぼ同じ正答率となるが、研修回数が 15 回の場合は、逆に、学級規模が約 29 人の場合に平均正答率が最も高くなり、その前後では平均正答率が低くなっている、学級規模を 19 ～38 人の間で、研修回数を 0～15 回の間でそれぞれ変化させたとき、

最も平均正答率が高いのは学級規模が 29 人で研修回数が 15 回の場合であるが、平均正答率の変化の程度は非常に小さく、学級規模が 29 人で研修回数が 15 回の場合と学級規模が 38 人で研修回数が 0 回の場合の差は 1.17%ポイントである、小学校 6 年生の算数でもほぼ同様である、他方、中学校 3 年生の数学では、研修回数が 0 回の場合、学級規模が 30 ～38 人の間では平均正答率の変化はほとんど観察されない、研修回数が 15 回の場合学級規模が 38 人だと平均正答率が少し高いが、これよりも点数が高いのは学級規模が 19 人の場合であると述べている。

野崎祐子氏、平木耕平氏、篠崎武久氏、妹尾渉氏は『学力の生産関数の推定：底上げをどう図るか』で、2007 年から 2009 年に文部科学省が実施した「全国学力・学習状況調査」の公表された結果を用いて、3 カ年分の都道府県別パネルデータを作成し、公立小学校 6 年生のテスト結果を分析したところ、教育予算の高低と全国学力テストの平均点との間には必ずしも明快な関係を見つけないことができないが、下位成績者層に着目すると、教育予算が多いほど、テストスコアのばらつきが少ないという関係がある。教員一人あたり児童数と平均点および下位成績者層のちらばりとの関係は複雑で、授業研修の回数など他の要因によって結論が変わってくる。授業研修回数が少ない、あるいは全国平均レベルであれば、教員一人あたり児童数が多いほど平均点が低く、下位成績者層のちらばりが大きいという関係が観察されるが、授業研修回数が全国平均レベルよりも多い場合には、教員一人あたり児童数が多いほど平均点が高く、下位成績者層のちらばりが小さいという関係が見られる。この結果は、教員一人あたり児童数、一般には学級規模のみを取り出して、その効果を議論することの難しさと、学級規模以外の政策的に変更可能な要因がテストスコアに与える影響が小さくない可能性を示していると指摘している。

北條雅一氏は『国際学力調査結果を用いた教育生産関数の推計 一少人数指導・習熟度別指導・ピアグループの効果一』で、国際教育到達度評価学会 (IEA) が行っている「国際数学・理科教育動向調査」(TIMSS) の 2007 年調査の結果を分析して (理科の成績を使用せず、算数・数学の成績だけを使用、私立学校を除外)、小学 4 年、中学 2 年ともに、学級規模は児童・生徒の成績に有意な影響を与えていないことが確認されたと述べている。また、この論文では、本人の生まれ月、家庭の蔵書数や両親の学歴、経済的に恵まれない生徒の割合といった要因が、試験の点数に強い影響をもたらしていること、教師や学校に対する物的および人的資源投入の効果は限定的で、一貫した傾向は確認されないこと、習熟度別授業の実施校では中学 2 年生の数学の点数が有意に高くなっていること、習熟度別授業の実施によって、本人の生まれ月や親の学歴、経済的に恵まれない生徒の割合といった、本人固有の要因や社会経済的要因が学力に及ぼす影響が縮小するという副次的な効果があることが明らかとなったこと、ピア・グループ効果 (同一の学習集団で学ぶ生徒の影響や刺激を受けることによって、本人の学習成果に変化をもたらすような効果のこと) の存在が確認されたこと、教職経験年数が成績に有意な影響を与え

ていることを確認できたのは、小学 4 年で習熟度別授業を実施している学校のみであることも指摘している。なお、北條雅一氏は、児童の授業理解度が停滞している学校において少人数指導や習熟度別授業が積極的に導入されている可能性があることから、上述の結果を「少人数指導や習熟度別授業には得点を上昇させる効果がない」と解釈することには慎重でなければならないと述べている。

国立教育政策研究所の研究グループ（研究代表者 工藤文三氏）による『学級編制と少人数指導形態が児童の学力に与える影響についての調査 報告書』は、京都府内（京都市を除く）の小学校 110 校の児童の第 4 学年時（4 月）と第 6 学年時（4 月）の学力診断テスト（国語と算数）の成績を比較分析した結果、従前の学力が同程度の児童でみると、国語では、「現行の基準によって編制された 30 人を超える学級規模の学校」の児童より、京都式少人数教育により「現行を下回る基準による学級編制を継続的に実施した学校」の児童の方が、その後の学力が高く、算数では、「現行の基準によって編制された 30 人以下の学級規模の学校」の児童より、京都式少人数教育により「現行を下回る基準による学級編制を継続的に実施した学校」の児童の方が、その後の学力が高いという結果が得られ、この結果は、「現行を下回る基準による少人数学級編制を継続的に実施した学校の優位性を示唆」しているとのことである。

この報告書には不自然な点がいくつかある。

なぜ、「現行を下回る基準による少人数学級編制を継続的に実施した学校の優位性を示唆」しているというような持って回った言い方をしているのか。なぜ、当初の学力が同程度であれば、少人数学級に継続的に在籍した方が学力が高くなると、はっきり言えないのか。また、「本研究の結果では、規模の小さい学級に在籍すること自体の優位性は示されず、現行を下回る基準によって編制された学級に在籍し続けることの優位性が示された」（P.54）とも書かれているが、意味不明瞭である。これでは、「規模の小さい学級」と「現行を下回る基準によって編制された学級」を対比したいのか、それとも、「在籍すること自体の優位性」と「在籍し続けることの優位性」を対比したいのか、分からない。

国語では、「現行の基準によって編制された 30 人を超える学級規模の学校」と、「現行を下回る基準による学級編制を継続的に実施した学校」を比較して、後者の優位性を指摘しておきながら、算数では、「現行の基準によって編制された 30 人以下の学級規模の学校」と、「現行を下回る基準による学級編制を継続的に実施した学校」を比較して、後者の優位性を指摘するというのは、いかにも不自然である。なぜ、算数では、大人数学級と少人数学級を比較せずに、少人数学級同士を比較しているのか。

この報告書には、階層的線形モデルの分析の結果、切片または傾きに差がみられた学級編制の類型間の比較を図示するという旨の記述と、「なお、他の類型に関する係数の 90%信用区間には 0 が含まれていた」という記述があるが（P.39～44）、これは、この報告書に比較が図示されていない類型間では、切片にも傾きにも、統計的に有意な差がみられなかったということである。つまり、算数で「現行の基準によって編制された 30 人

を超える学級規模の学校」と、「現行を下回る基準による学級編制を継続的に実施した学校」を比較したり、国語と算数で「現行の基準によって編制された 30 人を超える学級規模の学校」と「現行の基準によって編制された 30 人以下の学級規模の学校」を比較したりするというような方法では、少人数学級に継続的に在籍することの優位性を示すことができなかつたということである。このことを報告書や調査結果概要に明示せず、専門的知識を持つ者が精読しなければ分からないような表現で書くというのは、不誠実ではないだろうか。

第 4 学年時と第 6 学年時の学力診断テストの平均成績を下表に示す。「伸び」は私が計算したものである。なお、第 4 学年時の成績が同程度の者同士で「伸び」を比較するのが正しい方法であるが、この報告書に記載されているデータからでは、そのようなことができなかつたので、やむを得ず、このような比較を行った。また、第 4 学年時と第 6 学年時の学力診断テストは異なったテストなので、その得点差を「伸び」と呼ぶのは正確な表現ではないが、他に適当な言葉がないので、「伸び」と呼ぶことにした。

#### 国語（テスト得点）

	4 学年	6 学年	伸び
現行を下回る基準による学級編制を継続的に実施した学校	66.0 点	78.3 点	1.19 倍
現行の基準によって編制された 30 人を超える学級規模の学校	70.2 点	78.7 点	1.12 倍
現行の基準によって編制された 30 人以下の学級規模の学校	68.3 点	78.1 点	1.14 倍

#### 算数（テスト得点）

	4 学年	6 学年	伸び
現行を下回る基準による学級編制を継続的に実施した学校	77.7 点	77.7 点	1.00 倍
現行の基準によって編制された 30 人を超える学級規模の学校	79.6 点	78.1 点	0.98 倍
現行の基準によって編制された 30 人以下の学級規模の学校	79.4 点	76.7 点	0.97 倍

「現行を下回る基準による学級編制を継続的に実施した」学級の平均人数は、4 年時 25.6 人、5 年時 25.5 人であるが、「現行の基準によって編制された 30 人以下の学級」の平均人数は、4 年時 26.0 人、5 年時 25.9 人であり、ほとんど差がない。それにもかかわらず、学力の「伸び」にはかなりの差がある。これは、京都式少人数教育で「現行を下回る基準による学級編制を継続的に実施した学校」では、特別な指導がなされた、成果を示そうと教員が張り切った、優秀な教員が配置されたなど、学級規模以外の要因によって、学力が向上したことを示唆している。

なお、文部科学省の「公立義務教育諸学校の学級規模及び教職員配置の適正化に関する検討会議」（第 11 回）において、国立教育政策研究所の工藤文三氏と山森光陽氏から、この報告書の説明があり、同検討会議の副主査の小川正人氏から「政策的につくられた

少人数学級と、自然な状態で作られた少人数学級との間で、差異が生じるというのは、  
「どういうふうに理解していいのか」という質問があったのに対して、山森光陽氏は「こ  
この調査を行った京都府の先生方に聞いたときに、単に自然にできてしまった少人数だか  
ら頑張らないとか、逆に、意図的に少人数学級をつくったから頑張るといったようなこ  
とはないだろう。どんな条件であっても、きちんと頑張っていると思うので、先生方の  
動きとして何か劇的に違ふとか、または学校の取り組みとして何か全く違うことをやっ  
ているということはどうも考えにくいということでした。……現行を下回る基準によっ  
て学級編制を実施するという事は、学年の学級数は増えるということになりますよね。  
なので、もしかしたら、学年の学級数が増えるというところの良さが効いているのかな  
というのが私たちの現時点での解釈です。……学級規模を小さくして学年の学級数が多  
いと、安定的な学級づくりに寄与するのではないか……そういったことももしかしたら  
働いて、このような結果になっているのではないか……ただし、こういったことについ  
ての先行研究も乏しい状態にあるので、積極的に、これが理由ですといった理論的な根  
拠をつけることは現時点では難しい」と答えている。

ちなみに、「現行の基準によって編制された 30 人を超える学級」の平均人数は、4 年時  
33.8 人、5 年時 33.6 人である。

また、「現行の基準によって編制された 30 人を超える学級規模の学校」の児童と、「現  
行の基準によって編制された 30 人以下の学級規模の学校」の児童の学力の「伸び」を比  
較すると、大きな差がなく、国語では後者の方が「伸び」が大きい、算数では前者の  
方が「伸び」が大きいという相反する結果が出ている。

つまり、この調査結果では、少人数学級に継続的に在籍することが学力向上の面で有  
利であるとはとても言えないのである。この調査結果から言えることは、「京都式少人数  
教育によって」「現行を下回る基準による学級編制を継続的に実施した学校」の“国語に  
おける”優位性だけである。

「本研究の結果では、規模の小さい学級に在籍すること自体の優位性は示されず、現  
行を下回る基準によって編制された学級に在籍し続けることの優位性が示された」とい  
う意味不明瞭な言葉の裏には、現行の基準の下でたまたま少人数になった学級の優位性  
を示すことはできなかったということがほめかされているのではないだろうか。

#### (4) 統合的な研究

山下絢氏は「米国における学級規模縮小の効果に関する研究動向」で、アメリカにお  
ける学級規模縮小の効果に関する研究動向を分析した結果、第一に、通常の学級規模の  
児童生徒に比べて、小さな学級規模の児童生徒のテストの点数は高いことが指摘可能で  
あるが、学級規模の縮小効果によるテスト得点の差の大きさについては、論者によって  
結果が異なり、一致した結論には至っていない、第二に、学級規模の縮小が、児童生徒  
の学力向上政策として最も効率のいい費用の使い方か否かは今後の検証の課題とされて

おり、費用対効果の観点からの議論は結論に達していない、第三に、学級規模の縮小の効果が顕著にみられるのは、マイノリティや低所得家庭の児童生徒である、第四に、学級規模の縮小が児童生徒の学力向上につながる明確な因果メカニズムは明らかにされていないと指摘している。

杉江修治氏は「学級規模と教育効果」で、日本における学級規模縮小の効果に関する過去の研究を総括して、「学級を単位とした学習指導法による場合は、ほとんどの教科で20～30名規模の学級が児童生徒の学習内容の習得の上で効果的である。この規模はまた、児童生徒の相互作用から得られる多様な経験の習得をも見込めるものである」という結論を導くことができるとしている。

この杉江修治氏の結論には疑問がある。日本の過去の研究は、学級規模に関する教員の意見の調査が多く、また、実証的研究であってもその調査方法に問題のあるものが多いので、これらの研究から一定の結論を導くことはできないと思う。杉江修治氏は、「主観的な資料として個々には排除されがちな、現場教師の体験に基づく学級規模への意見、態度調査も、多様な指導態度、指導経験を持つ者を対象としているのであるから、一概に軽視して良い資料ではない。多数の調査に比較的一貫した傾向性が見られるとしたら、それは十分考慮に値するものといえよう」と指摘しているが、歴史を振り返れば、多くの人々が長年にわたり間違ったことを信じてきたという例は数多くある。

Gene V. Glass と Mary Lee Smith は“Meta-Analysis of Research on the Relationship of Class-Size and Achievement”で、学級規模の縮小の効果に関する過去の調査研究 77 件をメタ分析した結果、学級規模と学力との間には下表のような非線形の関係があり（学級規模縮小による学力向上効果の大きさは、何人から減らすかによって変わるということ）、学級規模縮小の主たる効果が現れるのは 20 人以下からである、また、規模の異なる学級への児童生徒の振り分けが無作為に行われている調査研究の方が、自然にできた学級規模の違いを利用した調査研究に比べて、学級規模縮小の効果をより明瞭に示していると指摘している。

少人数学級	大人数学級	成績差（偏差値）
1 人	40 人	5.65
10 人	40 人	2.68
20 人	40 人	0.51
30 人	40 人	-0.48
1 人	25 人	5.52
5 人	25 人	4.09
10 人	25 人	2.56
15 人	25 人	1.33
20 人	25 人	0.39

少人数学級	大人数学級	小学生・成績差	中学・高校生・成績差
1人	40人	4.90	7.49
10人	40人	2.41	3.57
20人	40人	0.63	0.57
30人	40人	-0.11	-1.02
1人	10人	3.87	7.16
3人	10人	3.24	6.19
5人	10人	2.65	5.27

Gene V. Glass 等は、成績差を標準得点の差で示しているが、それを偏差値に換算した。

Harvey Goldstein、Min Yang、Rumann Omar、Rebecca Turner、Simon Thompson は“Meta-Analysis Using Multilevel Models with Application to the Study of Class Size Effects”で、①少人数学級と大人数学級への配置が無作為に行われているか、いろいろな基準で同水準になるように試みられていること、②学業成績で比較されていること、③1年以上の期間を使って比較されていること、④少人数学級は15人以上、大人数学級は40人以下であること、という条件を満たしている過去の研究9件をマルチレベルモデル（階層線形モデル）を使ってメタ分析した結果、線形性（学級の人数を1人減らしたことによる学力向上効果は、何人から減らした場合でも一定であるということ）を前提にすると、平均して、1人減少につき、+0.022 標準偏差分（偏差値に換算すると、+0.22）、成績が向上すると述べている。

橋野晶寛氏は「準実験的手法による学級規模効果研究のメタアナリシス」で、準実験的手法を用いて学級規模縮小の効果进行分析した諸外国の22論文を階層線形モデルを使ってメタ分析した結果、学級規模縮小による学力向上は、線形性を前提にした研究では、1人減少につき、-0.01 から+0.07 標準偏差分（偏差値に換算すると、-0.1 から+0.7）まで結果が分かれているが、平均すると、+0.013 標準偏差分程度（偏差値に換算すると、+0.13 程度）であるとしている。また、学級規模縮小による学力向上効果は非線形であるが、必ずしも学級規模が小さいほど効果が大きいとは言えないこと（学級規模が27人以上では、学級規模が小さいほど効果が小さくなっている）、中等教育段階になると、学級規模の影響は小さくなること、教員の給与水準が高い国では、学級規模の影響が小さくなる可能性があることも指摘している。

##### (5) その他の研究

Jennifer King Rice は“Making the evidence matter: Implications of the class size research debate for policy makers”（“Mishel, Lawrence and Richard Rothstein, eds., *THE CLASS SIZE DEBATE.*” 所収）で、学級規模の縮小の実施には次のような問題が

あると指摘している。第一に、過去の研究成果から、学級規模縮小の効果はマイノリティや低所得家庭の児童生徒に対して大きいことが分かっているので、これらの児童生徒を対象を絞って、学級規模を縮小した方が教育投資の効果が高い。第二に、有能な教員と教室用のスペースが確保できるかという問題がある。これらが確保できないと、学級規模縮小の効果は減殺されてしまう。第三に、関連するさまざまな施策が学級規模縮小の費用と効果に影響を及ぼす。補完的な施策としては、教員の能力を開発したり、学級規模縮小の効果を最大限に発揮するようなインセンティブを教員に与えたりすることがある。競合する施策としては、リスクのある子どもたちのためのプログラム、教員の給料を上げること、授業時間数を増やすことなどがある。学級規模の縮小がどのようにして、また、なぜ、効果を上げるのかについての洞察が必要である。少人数学級で、教員が何をすれば効果が上がり、何をすれば効果が下がるのか知る必要がある。第四に、費用対便益を考慮する必要がある。学級規模の縮小は、教育効果を上げるための一つの選択肢に過ぎず、他にも、教員の給料を上げる、授業時間数を増やす、早期教育の機会を増やすなどの方法がある。どのような状況では学級規模を小さくすべきであり、どのような状況では学級規模が大きいままでも良いのかを選び分ける必要がある。

Simon Dobbelsteen、Jesse Levin、Hessel Oosterbek は“*The causal effect of class size on scholastic achievement: distinguishing the pure class size effect from the effect if changes in class composition*”で、オランダの小学4年生、6年生、8年生（日本の中学2年生にあたる）の学級規模、学業成績、知能指数に関する1995年度のデータを分析して、学級規模を縮小しても学業成績が向上しなかったり、低下したりする場合があるが、これは、同一学級内にいる同程度の知能指数の子どもの数が減ることによって、学業成績にマイナスの影響を与えるためであると指摘している。

この研究結果を受けて、荒井一博氏は『学歴社会の法則 教育を経済学から見直す』（P.207-208）で次のように指摘している。

学級規模は小さいほどよいといえない要因が……あります。生徒が同級生から受ける……効果で、ピア効果（peer effect）と呼ばれるものです。……生徒は教師だけでなく同級生からも多くを学びます。同級生を観察したり一緒に行動したりして、知識・技能・態度などを多く学びます。「同級生が知っていることなので、自分も習得しなければならない」とか、「同級生がこうしているから、自分もそうしなければならない」という動機が学習を強く促進します。……右で指摘したような効果の多くは、同レベルの能力の生徒の間で最も大きくなるでしょう。たとえば、同程度の能力の同級生がよい成績を上げれば、自分もできるはずだと発奮する可能性が高いはずです。あまりに能力差があれば、諦めてしまっただけで効果が生まれにくいことも少なくありません。……このようとき、生徒が複数の少人数クラスに無作為に割り振られると、同水準の能力の級友が平均的に少なくなります。多人数クラスのほうが自分と能力の近い同級生

にめぐり合える確率が高くなり、学力が高まる可能性があるのです。

「ピア効果」は「集団社会化」によって生じると考えられる（「第8回 能力の個人差 14. 集団社会化」参照）。したがって、勉強が得意な子どもたちが集団を作って、競い合って勉強するようになり、学業成績にプラスの影響を及ぼすこともあれば、勉強が苦手な子どもたちが集団を作って、勉強が得意な子どもの集団を「ガリ勉」とばかにして、ますます勉強しなくなることによって、学業成績にマイナスの影響を及ぼすこともある。ただし、「集団社会化」は7、8歳頃から始まり、小学校5年生頃にその力を増すので、それ以前であれば、同一学級内にいる同程度の知能指数の子どもの数が減ることによって、学業成績にマイナスの影響を与えるということはあまりないと思われる。このことが、学級規模縮小による学力向上効果は低学年の方が大きいということの原因の一つになっているのではないだろうか。

#### (6) 総括

上述の調査研究を総括すると、第一に、学級規模縮小による学力向上効果はあるようだが、あまり大きなものではない。Harvey Goldstein、Min Yang、Rumann Omar、Rebecca Turner、Simon Thompson の“Meta-Analysis Using Multilevel Models with Application to the Study of Class Size Effects”と橋野晶寛氏の「準実験的手法による学級規模効果研究のメタアナリシス」によると、1人減少による学力の向上は、平均すると、偏差値で0.1~0.2程度の大きさである。

第二に、学級規模縮小による学力向上効果は、Gene V. Glass と Mary Lee Smith が“Meta-Analysis of Research on the Relationship of Class-Size and Achievement”で指摘しているように、非線形のものである。つまり、学級規模縮小による学力向上効果の大きさは、何人から減らしたかによって違ってくる。学級規模縮小によって、学力が低下する場合もある。

第三に、学級規模縮小による学力向上効果の大きさは、対象となる児童・生徒の学年、学力、家庭の状況、教員の指導方法、能力に左右されるようであるが、児童生徒と教員の属性の違いに着目した調査研究が少ないために、よく分からない。STAR プロジェクトと SAGE プログラムの結果によると、幼稚園と小学校1年生、低所得層とマイノリティに対しては、学力向上効果が大きい。なお、高学力の子どもの場合には、学級規模縮小によって学力が低下する可能性がある。

#### (7) 私見

Alan. B. Krueger が“Experimental Estimates of Education Production Functions”で、①低学年で少人数学級に入ると、その後の学業達成のレベルを引き上げる、一回限りの「学校による社会化の効果」が働くのではないかとということと、②学級規模縮小に

よる学力向上効果はマイノリティや低所得層の子どもで大きいことを指摘しているが（「(1) 実験」参照）、これらの指摘は問題の本質を突いていると思う。

「第 4 回 教育の経済効果（その 2）」で述べたように、私は、一般に「社会化」と呼ばれている学校の機能は、「「事実の認識と思考の方法」の改造」、「価値観の押し付け」、「情報共有・連結の経済性」の基盤形成」に分類されると考えているが、この中で、学力と深く関係するのは、「「事実の認識と思考の方法」の改造」である。「「事実の認識と思考の方法」の改造」とは、「声の文化」の「人間の生活世界に密着した」「状況依存的であって、抽象的ではない」認識と思考（具体的思考）を、「文字の文化」の「外界を記号によって抽象化した理論モデルを用いた」「状況依存的ではない、形式論理的な」認識と思考（抽象的思考）に改造することである。学校で教えられる学問は、「抽象的思考」の産物であり、「抽象的思考」ができなければ理解できないものである（「第 8 回 能力の個人差 7.知能検査で測られる「知能」の正体 (6) 知能検査は近代啓蒙主義的な偏見に基づいて作られている」参照）。

「声の文化」に親和的な親は、「抽象的思考」能力が低く、学力も低いので、低所得になることが多い、つまり、「声の文化」に親和的な家庭は低所得層に多い。他方、「文字の文化」に親和的な親は、「抽象的思考」能力が高く、学力も高いので、高所得になることが多い、つまり、「文字の文化」に親和的な家庭は高所得層に多い（「第 8 回 能力の個人差 8.「意識的汎用知能」の優劣が職業能力の優劣を決める社会」参照）。

「第 8 回 能力の個人差 17.認知的文化の差」で述べたように、「声の文化」に親和的な家庭では「具体的思考」の訓練になるような子育てを行い、「文字の文化」に親和的な家庭では「抽象的思考」の訓練になるような子育てを行うという違いが見られる。つまり、「文字の文化」に親和的な家庭では、就学前から、「抽象的思考」への改造が行われており、子どもはスタートラインから有利な位置にいることになる（「第 8 回 能力の個人差 7.知能検査で測られる「知能」の正体 (3) 読み書きによる脳の変化」参照）。

この結果、小学校の教員は、1年生の段階から「抽象的思考能力」に大きな差がある子どもたちを相手にしなければならないことになる。大規模な学級、画一的なカリキュラムの下では、このことは困難であり、「声の文化」に親和的な家庭で育った子どもたちに「抽象的思考」を身につけさせることが手薄になってしまい、これらの子どもたちは低学力に陥ってしまう。学級規模が縮小されれば、ある程度、個別的な指導を行うことも可能になるので、「声の文化」に親和的な家庭で育った子どもたちに「抽象的思考」を身につけさせて、学力を向上させることが、ある程度可能になってくる。ただし、「声の文化」に親和的な家庭で育った子どもたちの学力を限界近くまで向上させるためには、学級規模の縮小だけでは不十分で、特別な教育的支援を行う必要があると思われる。

「具体的思考」から「抽象的思考」への改造は、脳の可塑性が高い幼少期に行う必要があり、また、「抽象的思考」は学校での勉強の基礎となるものなので、学級規模縮小が学力向上に及ぼす効果は低学年の大きい。ただし、「第 8 回 能力の個人差 18.就学前

教育の効果」で述べたように、「抽象的思考」能力を限界近くまで高めるためには、就学後では遅すぎて、就学前に「抽象的思考」を身につけるための環境を整えてやる必要がある。例えば、アメリカで実施された“Carolina Abecedarian Project”という、親の教育年数、所得、職業、知能指数などから、知的発達に遅れが生じるおそれがあると判断された子どもに対して、0歳から5歳まで続けられた教育・医療・栄養支援プログラムでは、子どもの知能指数が上昇し、入学後の学校での成績も上がることが立証されている（「第8回 能力の個人差 18.就学前教育の効果」参照）。つまり、学級規模縮小に使うお金があるならば、そのお金を就学前教育の充実に振り向けた方が、学力は向上するのである。

「具体的思考」から「抽象的思考」への改造は「1回限り」のものであるが、「声の文化」に親和的な家庭で育つ子どもは「具体的思考」へ逆戻りさせよう、あるいは、勉強嫌いにさせようとする親や友だちからの誘惑にさらされ続けており、その誘惑に抵抗して、「抽象的思考」を維持させ続けるために、教員が盾になる必要があるので、少人数学級に居続けないと学力が低下してしまう。

「文字の文化」に親和的な家庭で育った子どもは、「抽象的思考」を身につけやすい環境の中で育ってきているので、「抽象的思考」を身につけるための特別の環境を整えてやる必要性は少ない。したがって、就学前教育や学級規模の縮小が学力向上に及ぼす影響は小さい。「(5)その他の研究」で述べたように、学級規模を縮小し過ぎると、ライバルが減ることによって、学習意欲が衰え、学力が低下する可能性もある。

なお、勉強することを習慣づけるということが関係している可能性もある（「第8回 能力の個人差 13.才能の差異を増幅する本能」参照）。学級規模が小さいほど、教員の目が行き届きやすくなり、子どもたちに勉強することを習慣づけることが容易になるということである。

以上に述べた私の仮説は、いくつかの状況証拠からの推測に過ぎず、客観的な証拠に基づくものではない。実験ができれば、私の主張が裏付けられると信じているが、そのような実験を行うことは不可能であろう。小学校入学前に「抽象的思考」能力を測る検査（知能検査がその代用品となる）を行い、「抽象的思考能力」が低いグループと高いグループに分けて、それぞれを、少人数学級と通常規模学級に振り分けて、数年間教育を行い、「抽象的思考」能力と学力の伸びを調査するというような「差別的な」実験の実施が許されるはずがないし、ましてや、その実験結果に基づいて、「抽象的思考」能力が低い子どもは少人数学級に入れて特別な教育的支援を行い、「抽象的思考」能力が高い子どもは通常規模学級に入れるというような「差別的な」学級編制を行うことが許されるはずもない。

## 12. 教員の能力が児童生徒に与える影響についての調査研究の動向

### (1) 既存研究

Steven G. Rivkin、Eric A. Hanushek、John F. Kainは“Teachers, Schools, and Academic Achievement”で、“the University of Texas at Dallas”が“the Texas Education Agency”の協力を得て行った“the UTD Texas Schools Project”の成果の中から、1993年～1995年にテキサス州の小学校4年生であった児童・生徒（3,000校以上の600,000人以上の児童・生徒）の小学校3年から中学校1年までの“the Texas Assessment of Academic Skills”（毎年春にテキサス州の小学校3年生から中学校2年生を対象に行われている学力調査）の成績のデータと教員の職務経験・教員資格・学歴・人種等のデータを分析して（ただし、児童・生徒の成績データと彼・彼女らを担当した教員のデータとを結び付けることができたのは、“the Lone Star district”内の学校のデータのみ）、①教員の能力（児童生徒の学力を向上させるという意味での能力）が、同一学校内の教員同士の比較で1標準偏差分（偏差値に換算すると10）違うと、児童・生徒の算数・数学の成績が0.11標準偏差分（偏差値に換算すると1.1）、読解の成績が0.095標準偏差分（偏差値に換算すると0.95）違ってくこと、②修士の学位の有無は教員の能力とは無関係であること（正確に言うと、修士の学位を持っていると教員の能力が高くなるという証拠は得られなかったこと）、③教員経験が教員の能力を向上させる働きは1年目が大きく、2年目以降は小さくなり、3年経過後にはほとんどなくなること（正確に言うと、3年経過後も能力の向上が継続するという証拠は少ししか得られなかったこと）、④学級規模の縮小が学力向上に及ぼす影響力は、小学校4年生では、10人減少につき0.1標準偏差分（偏差値に換算すると1）程度であるが（学級規模が10人から35人の間での検証）、その影響力は学年が上がるにつれて小さくなっていくことを指摘している。

なお、Eric A. Hanushek、John F. Kain、Daniel M. O'Brien、Steven G. Rivkinは“The Market for Teacher Quality”で、上述の“the UTD Texas Schools Project”の成果の中から、児童・生徒の個人別データ（“the Texas Assessment of Academic Skills”の成績、家族、人種・民族、性別、貧困度等）が入手できる“the Lone Star district”（テキサス州内の“school district”（学区）の一つ、アメリカでは、学区という組織が地域内の複数の公立学校を運営している）の小学校3年から中学校2年に関する児童・生徒と教員の1995～2000年度のデータ（児童・生徒の成績のデータは算数・数学だけを利用）を分析して、①教員資格試験に合格しているか否か、あるいは1回で合格したか否かは教員の能力とは無関係であること、②同一人種の児童・生徒を教えた方が教員は能力をより多く発揮できることを指摘している。

また、Eric A. Hanushek と Steven G. Rivkin は “Constrained job matching: does teacher job search harm disadvantaged urban schools?” で、上述の “the UTD Texas Schools Project” の成果の中から、“the Lone Star district” の小学校 3 年から中学校 2 年に関する児童・生徒と教員の 1995～2000 年度のデータを分析して、同一学区内の比較では、教員の能力が 1 標準偏差分（偏差値に換算すると 10）違うと、児童・

生徒の算数・数学の成績が 0.17 標準偏差分（偏差値に換算すると 1.7）違い、同一学校内の比較では、教員の能力が 1 標準偏差分違うと、児童・生徒の算数・数学の成績が 0.11 標準偏差分（偏差値に換算すると 1.1）違うので、その間には 50%ほどの差があると述べている。

Eric A. Hanushek と Steven G. Rivkin は “Generalizations about Using Value-Added Measures of Teacher Quality” で、アメリカで 2004～2010 年に行われた 10 件の調査研究結果を平均すると、教員の能力が 1 標準偏差分（偏差値に換算すると 10）違うと、児童・生徒の読解の成績が 0.11 標準偏差分（偏差値に換算すると 1.1）、算数・数学の成績が 0.15 標準偏差分（偏差値に換算すると 1.5）違ってくると述べている。なお、この計算には間違いがあったようで、Eric A. Hanushek は後述の “The economic value of higher teacher quality” で、教員の能力が 1 標準偏差分（偏差値に換算すると 10）違うと、児童・生徒の読解の成績が 0.13 標準偏差分（偏差値に換算すると 1.3）、算数・数学の成績が 0.17 標準偏差分（偏差値に換算すると 1.7）違ってくると修正し、さらに、これらの調査研究結果は、一つの研究を除いて、同一学校内の教員同士の能力の比較に基づいたものであり、学校間の教員の能力の差が無視されていると指摘している。

Eric A. Hanushek は “The economic value of higher teacher quality” で、過去のいくつかの調査研究結果からすると、教師の能力が 1 標準偏差分（偏差値に換算すると 10）高いと児童・生徒の認知的能力が 0.2 標準偏差分（偏差値に換算すると 2）向上するが、その後、学習を全くしないと、認知的能力が毎年 30% ずつ低下していくこと、また、成績（認知的能力）が 1 標準偏差分上がると将来の所得が 13% 増え、アメリカの平均的な生涯所得の現在価値（割引率 3% で計算）は 116 万ドルなので（2010 年の第 1 四半期の平均所得のデータから、所得の将来の伸び率を年率 1% と仮定して推計）、生涯所得の現在価値が 15 万ドル増えることになると言えることから、平均的な教員に比べ 1 標準偏差分だけ能力が高い教員が毎年生み出す価値は、20 人学級を担当している場合には 40 万ドル以上（当該教員の授業を受けた児童・生徒の将来の所得の増加を現在価値に換算）と推計され、それよりも規模の大きい学級を担当している場合はさらに大きくなると述べている。

篠崎武久氏は、千葉県検証改善委員会『平成 19 年度「全国学力・学習状況調査」分析報告書』「第 7 章 教育資源と学力の関係」で、2007 年 4 月に文部科学省が実施した「全国学力・学習状況調査」の千葉県内の公立校に関するデータを学校単位で分析して、小学 6 年生、中学 3 年生ともに、教員に対する実践的な研修、および授業研究を伴う研修が学力と有意な関係にあり、中学校 3 年生では、教員平均年齢が学力と有意に関係している（教員平均年齢が 47～48 歳の場合に平均正答率が最も高くなり、その前後では低くなる）と指摘している。

亀井慶二氏は「小中学校の教育生産関数の推計」で、2009 年度に実施された『全国

学力・学習状況調査』のデータを都道府県単位で分析して、小学6年生、中学3年生ともに、教員の平均勤続年数が長いほど、児童・生徒の成績が良いという関係があること、教員研修参加度は成績とは無関係であること（正確に言うと、統計的に有意ではないこと）、校長による指導は、中学3年の数学のみ、生徒の成績を向上させること、中学3年では、教員養成系高等教育機関卒業者の全教員に占める割合が高いほど、生徒の成績が良いことを示している。なお、地方自治体から支出される公立学校費（児童・生徒一人当たり教育費）、学校規模、IT技術導入度と児童・生徒の成績とは無関係であることも示されている。

## (2) 総括

上述の調査研究を総括すると、教員の能力の違いは児童・生徒の学力に影響を与えるが、何が教員の能力の違いを生み出しているのかはよく分からないということになる。修士の学位の有無は教員の能力とはほとんど無関係である。教員経験は3年を超えると、教員の能力とはほとんど無関係になる。実践的な研修は教員の能力向上に役立っているようだが、それ以外の研修の効果は疑問である。

既存研究によっては教員の能力の違いを生み出す要因が分からないということは、既存研究が無視している要因が教員の能力の違いを生み出しているということである。おそらく、それは素質と熱意であろう。しかし、素質と熱意の影響を客観的に調査することは極めて困難である。

なお、教員の能力の違いが児童・生徒の学力に与える影響の大きさは、児童・生徒の能力の違いに左右されるということは十分考えられることだが、この問題を調査した研究はないようである。

## 13. 就学前教育の費用対効果に関する調査研究の動向

「ペリー就学前プログラム」、「Carolina Abecedarian Project」などの就学前教育の効果に関しては、「第8回 能力の個人差 18.就学前教育の効果」で説明したが、ここでは、就学前教育の費用便益分析に関する研究を紹介する。

L. J. Schweinhart, J. Montie, Z. Xiang, W. S. Barnett, C. R. Belfield, M. Nores, は “The High/Scope Perry Preschool study through age 40: Summery, Conclusion, and Frequently Asked Question” で、「ペリー就学前プログラム」の費用便益分析を行い、1人当たり 15,166 ドルの費用に対して、教育費の削減 7,303 ドル、所得税の増加 14,078 ドル、社会福祉費の削減 2,786 ドル、犯罪の減少 171,473 ドル、合計 195,621 ドルの社会的便益があり、実験対象者個人には 49,190 ドルの便益があったとしている（割引率 3%で 2000 年の現在価値に換算）。

W. S. Barnett と Leonard N. Masse は “Comparative benefit-cost analysis of the Abecedarian program and its policy implications” で、“Carolina Abecedarian Project”

の費用便益分析を行い、1人当たり63,476ドル（子育てに一般的に要する費用を差し引くと、35,864ドル）の費用に対して、実験対象者の所得の増加37,531ドル、実験対象者の子どもの所得の増加5,722ドル、実験対象者の母親の所得の増加68,728ドル、小・中学校の特殊教育学級在籍者の減少8,836ドル、喫煙率の減少・健康の改善17,781ドル、高等教育進学率の上昇8,128ドル、“Aid to Families with Dependent Children”（1996年までアメリカ連邦政府が行っていた、片親や低所得の家庭の子どもへの経済的支援）受給率の減少196ドル、合計130,666ドルの便益があったとしている（割引率3%で2002年の現在価値に換算）。なお、“Carolina Abecedarian Project”では、犯罪の減少に関するデータがないので、犯罪の減少による便益は推計されていない。

James J. Heckman、Flavio Cunha等は“Skill Formation and the Economics of Investing in Disadvantaged Children”、“Interpreting the Evidence on Life Cycle Skill Formation”などで、教育投資の収益率に関する調査研究（「第2回 教育経済学の基本 2. 人的資本論」参照）や就学前教育の費用便益分析に基づいて、教育投資の収益率は、就学前教育が最も高く、教育段階が進むほど低下していく、スキルはスキルを産み、学習は学習を産むという増幅機能があり、就学後の教育の効率性は就学前教育によって決まる、10代以降の教育で欠点を直そうとしても効率が悪いと指摘している。また、小学校3年生までに、家庭の社会経済的地位の差による学力格差が固定してしまう、格差を減らすためには就学前教育が重要であるとも述べている。

#### 14. 教育の費用対効果に関する統合的な研究

##### (1) 既存研究

Eric A. Hanushekは“Measuring Investment in Education”で、1994年までに発表された教育の生産関数に関する実証分析377件（90件の文献に掲載）の結果を調査して、下表のようにまとめ、多くの場合、教育投資の多寡は、教育成果に影響を及ぼさず、教育の質の指標としてふさわしくないと結論づけている。また、①どのような場合に教育投資が効果的になり、どのような場合に教育投資が効果的でなくなるかを明らかにできた研究はないこと、②学業達成や人的資本の蓄積には学校の影響だけでなく、家族や仲間の影響もあるので、それらの影響を無視した分析は偏ったものになること、③学校の非効率性のゆえに教育資源の投入量は教育の質の指標としてふさわしくないと、④公教育において最小限の競争しかなく、業績向上へのインセンティブがほとんどないため、資源の有効活用への動機付けが乏しいことを指摘している。

	分析 事例数	統計的に有意		統計的に有意でない		不明
		プラス	マイナス	プラス	マイナス	
教員・生徒比率	277	15%	13%	27%	25%	20%
教員の学歴	171	9	5	33	27	26
教員の経験	207	29	5	30	24	12
教員の給料	119	20	7	25	20	28
生徒1人当たりの支出額	163	27	7	34	19	13
管理費	75	12	5	23	28	33
設備	91	9	5	23	19	44

(有意水準は5%)

David Card と Alan B. Krueger は “Labor Market Effects of School Quality: Theory and Evidence” で、1968 年から 1986 年の間に行われた 11 件の研究に出てくる 24 件の推計をレビューして、生徒 1 人当たりの教育支出を 10%増やすと、生徒の将来の所得（生徒が働くようになってから得る所得）が 1～2%（正確に言うと、平均値は 1.52%、中央値は 1.25%、四分位範囲は 0.85%～1.95%）増えると述べている。また、1975 年から 1995 年の間に行われた、生徒・教員比率（生徒数÷教員数）と生徒の将来の所得との関係についての 5 件の研究をレビューして、4 件の研究では、生徒・教員比率が大きくなると、生徒の将来の所得が減少する（平均すると、生徒・教員比率が 1%大きくなると、所得が 0.053%減少する）という結果を得ているが、そのうちで統計的に有意なのは 2 件だけであると述べている（他の 1 件の研究では、生徒・教員比率が大きくなると、生徒の将来の所得が増えるという逆の結果を得ているが、統計的に有意ではない）。

Julian R. Betts は “Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature” で、学校への資源投入量（以下、投入と略す）と生徒の成績及び将来の所得（以下、産出と略す）との関係に関する過去の研究をレビューして、①投入と産出との間に有意な関係を見出すことができなかった研究は、個々の学校での投入に着目し、投入と産出との間に有意な関係を見出した研究は、州単位での平均値に着目している傾向があること、②投入と産出との間に有意な関係を見出すことができなかった研究は、1960 年代から 1980 年代初頭に教育を受けた人々の 32 歳以下での賃金を調査対象にし、投入と産出との間に有意な関係を見出した研究は、20 世紀前半に教育を受けた人々を調査対象にしている傾向があると指摘し、①州単位での研究は、集計バイアス、変数の漏れ、ランダムではない転居などによる影響を受けている、②収穫逓減（投入量を増やすにつれて、1 単位の投入による産出量が減少すること）、公立学校の官僚制化と中央集権化、教員の労働組合の組織化の結果、投入と産出との関係が時を経るにつれて弱くなってきている、という仮説を提起している。また、Julian R. Betts は、教育投資の収穫逓減性を示す研究を紹介している。

## (2) 総括

教育投資と教育成果の関係についての調査研究結果の混乱の原因は、Julian R. Betts が“Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature”で指摘している教育投資の収穫逓減性にあるのではないかと思われる。教育投資を増やしストックが充実するにつれて、次第に教育投資の効果が薄れていくので、ストックが不十分な分野の教育投資を増やせば成果がでるが、ストックが十分な分野に投資しても成果はでない。そのため、ストックが不十分な分野、国、時代に着目して調査すれば、教育投資には効果があるという結果が出て、ストックが十分な分野、国、時代に着目して調査すれば、教育投資には効果がないという結果が出る。

教育投資に収穫逓減性があるとすると、各分野への教育投資の限界的な効果（教育投資一単位の増加がもたらす効果の増加分）が等しくなるように、教育投資を調整することが最も効率的な教育投資であることになる。

## <引用・参考文献>

赤林英夫・中村亮介『学級規模縮小が学力に与えた効果の分析－横浜市公開データにもとづく実証分析－』KEIO/KYOTO GLOBAL COE DISCUSSION PAPER SERIES DP2011-005、2011年

<http://www.gcoe-econbus.keio.ac.jp/2011/07/dp2011-005.html>

浅野哲・中村二郎『計量経済学〔第2版〕』有斐閣、2009年

荒井一博『学歴社会の法則 教育を経済学から見直す』光文社新書、2007年

アンソニー・E・ボードマン、デヴィッド・H・グリーンバーグ、アイダン・R・ヴァイニング、デヴィッド・L・ワイマー『費用・便益分析 公共プロジェクトの評価手法の理論と実践』岸本光永監訳、ピアソン・エデュケーション、2004年

小川啓一・野村真作『教育統計学【基礎編】 「万人のための教育」に向けた理論と実践的ツール』学文社、2009年

小塩隆士『教育の経済分析』日本評論社、2002年

小塩隆士・妹尾渉『日本の教育経済学：実証分析の展望と課題』ESRI Discussion Paper No.69、2003年

[http://www.esri.go.jp/jp/archive/e\\_dis/e\\_discus.html](http://www.esri.go.jp/jp/archive/e_dis/e_discus.html)

大竹文雄「就学前教育の投資効果から見た幼児教育の意義－就学前教育が貧困の連鎖を断つ鍵となる－」『BERD』No. 16、Benesse 教育研究開発センター、2008年

<http://benesse.jp/berd/center/open/berd/backnumber/b2008.html>

株式会社三菱総合研究所『文部科学省委託調査研究 平成20年度 教育改革推進のための総合的調査研究～教育投資の費用対効果に関する基本的な考え方及び文献の収集・整理～報告書』、2009年

- [http://www.mext.go.jp/a\\_menu/kyoikuhi/chousa/1307884.htm](http://www.mext.go.jp/a_menu/kyoikuhi/chousa/1307884.htm)  
亀井慶二「小中学校の教育生産関数の推計」『プール学院大学研究紀要』第 49 号、  
pp.181-192、2009 年  
<http://ci.nii.ac.jp/naid/110007478311>
- 国立教育政策研究所「特集 学級規模に関する調査研究」『国立教育政策研究所紀要』第 131  
集、2002 年  
[http://www.nier.go.jp/03\\_laboratory/kankou\\_kiyou.html](http://www.nier.go.jp/03_laboratory/kankou_kiyou.html)
- 国立教育政策研究所『学級編制と少人数指導形態が児童の学力に与える影響についての調  
査 報告書』、2012 年  
[http://www.nier.go.jp/04\\_kenkyu\\_annai/div04-shochu.html](http://www.nier.go.jp/04_kenkyu_annai/div04-shochu.html)
- 須田康之・水野孝・藤井宣彰・西本裕輝・高旗浩志「学級規模が授業と学力に与える影響  
ー全国 4 県児童生徒調査からー」『北海道教育大学紀要 教育科学編』第 58 巻第 1 号、  
pp.249-264、2007 年  
<http://s-ir.sap.hokkyodai.ac.jp/dspace/handle/123456789/478>
- 杉江修治「学級規模と教育効果」『中京大学教養論叢』第 37 巻第 1 号、pp.147-190、1996  
年  
<http://ci.nii.ac.jp/naid/110004653198>
- 橋本俊詔・八木匡『教育と格差 なぜ人はブランド校を目指すのか』日本評論社、2009 年  
田中廣滋『費用便益の経済学的分析 環境と公共分野の理論』中央大学出版部、2003 年  
千葉県検証改善委員会（本部：東京大学教育学部、代表：苅谷剛彦）『平成 19 年度「全国  
学力・学習状況調査」分析報告書』2008 年  
<http://www.p.u-tokyo.ac.jp/kikou/chiba.htm>
- 野崎祐子・平木耕平・篠崎武久・妹尾涉『学力の生産関数の推定：底上げをどう図るか』  
広島大学経済学部 Discussion Paper Series no.2011-03、2011 年  
<http://ir.lib.hiroshima-u.ac.jp/00031454>
- 橋野晶寛「教育支出の政策分析における論点」『東京大学大学院教育学研究科教育行政学論  
叢』第 27 号、pp.121-132、2008 年  
<http://hdl.handle.net/2261/36856>
- 橋野晶寛「準実験的手法による学級規模効果研究のメタアナリシス」『東京大学大学院教育  
学研究科教育行政学論叢』第 30 号、pp.49-60、2011 年  
<http://hdl.handle.net/2261/43580>
- 樋口美雄・太田清・新保一成『入門 パネルデータによる経済分析』日本評論社、2006 年  
ヘンリー・レヴィン、パトリック・マキューアン『教育の費用効果分析 学校・生徒の教  
育データを使った政策の評価と立案』赤林英夫監訳、日本評論社、2009 年  
北條雅一『国際学力調査結果を用いた教育生産関数の推計 ー少人数指導・習熟度別指導・  
ピアグループの効果ー』新潟大学経済学部ワーキングペーパー No.107、2010 年 9 月改

訂版

<http://www.econ.niigata-u.ac.jp/~hojo/research.html>

北條雅一「大学在学中の学習成績・学習への取り組みと卒業後の所得」『新潟大学経済論集』  
第 89 号、pp. 111-120、2010 年

<http://hdl.handle.net/10191/13276>

三好皓一編『評価論を学ぶために』世界思想社、2008 年

文部科学省「公立義務教育諸学校の学級規模及び教職員配置の適正化に関する検討会議（第  
11 回）」議事録

[http://www.mext.go.jp/b\\_menu/shingi/chousa/shotou/084/gijiroku/1322062.htm](http://www.mext.go.jp/b_menu/shingi/chousa/shotou/084/gijiroku/1322062.htm)

文部省教育職員養成審議会『養成と採用・研修との連携の円滑化について（第 3 次答申）』  
1999 年

[http://www.mext.go.jp/b\\_menu/shingi/12/yousei/toushin/991201.htm](http://www.mext.go.jp/b_menu/shingi/12/yousei/toushin/991201.htm)

山崎博敏「学級規模の教育効果に関する諸理論」『日本教育社会学会大会発表要旨集録』（60）、  
pp.321-322、2008 年

<http://ci.nii.ac.jp/naid/110007011712>

山田昌弘『希望格差社会 「負け組」の絶望感が日本を引き裂く』筑摩書房、2004 年

八尾坂修「アメリカにおける学級編制研究の推移と州政策の方向」『教育學研究』第 66 卷  
第 1 号、pp.78-82、1999 年

<http://ci.nii.ac.jp/naid/110001175943>

山下絢「米国における学級規模縮小の効果に関する研究動向」『教育學研究』第 75 卷第 1  
号、pp.13-23、2008 年

<http://ci.nii.ac.jp/naid/110006680578>

リチャード・E・ニスベット『頭のでき 決めるのは遺伝か、環境か』水谷淳訳、ダイヤモンド社、2010 年

Barnett W. S. and Leonard N. Masse, 2007. “Comparative benefit-cost analysis of the  
Abecedarian program and its policy implications.” *Economics of Education Review*,  
Vol.26, pp.113-125

Betts Julian R., 1996. “Is There a Link between School Inputs and Earnings? Fresh  
Scrutiny of an Old Literature.” Discussion Paper 96-09, University of California, San  
Diego (Department of Economics)

Bohrnstedt, G. W. and B. M. Stecher eds., 2002. “*What We Have Learned About Class  
Size Reduction in California.*” California Department of Education

<http://www.classize.org/techreport/index-02.htm>

Card, David and Alan B. Krueger, 1996. “*Labor Market Effects of School Quality:  
Theory and Evidence.*” Working Paper #357, Princeton University (Industrial  
Relations Section)

<http://arks.princeton.edu/ark:/88435/dsp013f462543x>

- Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov, 2006. "Interpreting the Evidence on Life Cycle Skill Formation." *Handbook of the Economics of Education*, Vol.1, pp.698-812
- Currie, Janet, and Duncan Thomas. 1999. "Early Test Scores, Socioeconomic Status, and Future Outcomes." Working Paper No. 6943. Cambridge, Mass.: National Bureau of Economic Research
- Finn, Jeremy D. 2002. "Small Classes in American Schools: Research, Practice, And Politics" *Phi Delta Kappan*, March 2002, pp.551-560
- Glass, Gene V., and Mary Lee Smith, 1978. "*Meta-Analysis of Research on the Relationship of Class-Size and Achievement.*" Far West Laboratory for Educational Research and Development, San Francisco, CA.
- Goldstein, Harvey, Min Yang, Rumann Omar, Rebecca Turner, and Simon Thompson, 2000, "Meta-Analysis Using Multilevel Models with Application to the Study of Class Size Effects." *Journal of the Royal Statistical Society : Series C*, V.9, I.3, pp.399-412
- Hanushek, Eric A. 1996. "Measuring Investment in Education." *Journal of Economic Perspectives*, Vol.10, No.4, pp.9-30
- <http://hanushek.stanford.edu/publications/measuring-investment-education>
- Hanushek E.A., 2006, "School Resources." in Hanushek E.A. and Welch F. eds. ,*Handbook of Economics of Education* Vol.2, Ch. 14, pp.865-908
- <http://hanushek.stanford.edu/publications/school-resources>
- Hanushek, Eric A. 1999. "Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects." *Educational Evaluation and Policy Analysis*, Vol. 21, No. 2, pp.143-163
- <http://hanushek.stanford.edu/publications/some-findings-independent-investigation-tennessee-star-experiment-and-other>
- Hanushek, Eric A. 2011. "The economic value of higher teacher quality." *Economics of Education Review* Vol.30 No.2, pp.466–479
- <http://hanushek.stanford.edu/publications/economic-value-higher-teacher-quality>
- Hanushek, Eric A., John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin. 2005. "*The Market for Teacher Quality.*" NBER Working Paper No.11154
- <http://hanushek.stanford.edu/publications/market-teacher-quality>
- Hanushek, Eric A., and Steven G. Rivkin. 2010. "*Constrained job matching: does teacher job search harm disadvantaged urban schools?*" NBER Working Paper No.15816
- <http://hanushek.stanford.edu/publications/constrained-job-matching-does-teacher-job>

[-search-harm-disadvantaged-urban-schools](#)

Hanushek, Eric A., and Steven G. Rivkin. 2010. "Generalizations about Using Value-Added Measures of Teacher Quality." *American Economic Review*, Vol.100 No.2, pp.267-271

<http://hanushek.stanford.edu/publications/generalizations-about-using-value-added-measures-teacher-quality>

Hanushek, Eric A., and Ludger Woessmann, 2009. "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation." NBER Working Paper No.14633

<http://hanushek.stanford.edu/publications/do-better-schools-lead-more-growth-cognitive-skills-economic-outcomes-and-causation>

Hanushek, Eric A., and Lei Zhang, 2009. "Quality-Consistent Estimates of International Schooling and Skill Gradients." *Journal of Human Capital*, Vol.3, No.2, pp.107–143

<http://hanushek.stanford.edu/publications/quality-consistent-estimates-international-schooling-and-skill-gradients>

Heckman, James J., 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science*, Vol.312, pp.1900-1902

Krueger, Alan B., 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics*, 114(2), pp.494-532

Krueger, Alan B. and Diane M. Whitmore. 1999. "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR." Working Paper #427, Princeton University (Industrial Relations Section)

<http://arks.princeton.edu/ark:/88435/dsp010v838058q>

Mishel, Lawrence, and Richard Rothstein, eds., 2002. "THE CLASS SIZE DEBATE." ECONOMIC POLICY INSTITUTE, Washington, D.C.

Molnar, Alex, Philip Smith, John Zahorik, et al., 1998. "1997-98 Evaluation Results of the Student Achievement Guarantee in Education (SAGE) Program"

<http://nepc.colorado.edu/publication/1997-98-evaluation-results-student-achievement-guarantee-education-sage-program>

Molnar, Alex, Philip Smith, John Zahorik, et al., 2000. "1999-2000 Evaluation Results of the Student Achievement Guarantee in Education (SAGE) Program"

<http://nepc.colorado.edu/publication/1999-2000-evaluation-results-student-achievement-guarantee-education-sage-program>

Molnar, Alex, Philip Smith, John Zahorik, et al., 2001. "2000-2001 Evaluation Results of

- the Student Achievement Guarantee in Education (SAGE) Program”  
<http://nepc.colorado.edu/publication/2000-2001-evaluation-results-student-achievement-guarantee-education-sage-program>
- Murnane, Richard J., John B. Willett, and Frank Levy, 1995. “The Growing Importance of Cognitive Skills in Wage Determination.” *The Review of Economics and Statistics*, Vol.77, No.2, pp.251-266
- Neal, Derek, and William Johnson. 1996. “The Role of Premarket Factors in Black-White Wage Differentials.” *Journal of Political Economy* 104 (October), pp.869-95
- Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos. 1999. “The Long-Term Effects of Small Classes: A Five-Year Follow-up of the Tennessee Class Size Experiment.” *Educational Evaluation and Policy Analysis*, Vol.21, No.2, pp.127-142
- OECD 教育研究革新センター編著『教育とエビデンス 研究と政策の協同に向けて』岩崎久美子・菊澤佐江子・藤江陽子・豊浩子訳、明石書店、2009年
- Pritchard, Ivor, 1999. “*Reducing Class Size: What Do We Know?*” Technical Report, Office of Educational Research and Improvement, U.S. Department of Education  
<http://www2.ed.gov/pubs/ReducingClass/index.html>
- Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. “Teachers, Schools, and Academic Achievement.” *Econometrica*, Vol.73, No.2, pp.417-58  
<http://hanushek.stanford.edu/publications/teachers-schools-and-academic-achievement>
- Schweinhart, L. J., J. Montie, Z. Xiang, W. S. Barnett, C. R. Belfield, & M. Nores, 2005. “*The High/Scope Perry Preschool study through age 40: Summary, Conclusion, and Frequently Asked Question.*” High/Scope Educational Research Foundation.  
<http://www.highscope.org/Content.asp?ContentId=219>
- Smith, Phil, Alex Molnar, and John Zahorik, 2003. “*Class Size Reduction in Wisconsin: A Fresh Look at the Data.*” Education Policy Research Unit, Arizona State University
- Smith, Phil, Alex Molnar, and John Zahorik, 2003. “Class Size Reduction in Wisconsin: A Fresh Look at the Data.” *EDUCATIONAL LEADERSHIP*, Vo.61, No.1, pp.72-74  
<http://nepc.colorado.edu/publication/class-size-reduction-a-fresh-look-data>
- Unlu, Fatih, 2005. “California Class Size Reduction Reform: New Findings from the NAEP.” Working Paper, Princeton University

(2012年3月6日初出、2012年7月13日改訂)

<福田光宏のホームページ> <http://fukuda.mond.jp/>