

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

橋野 晶寛

A Meta-Analysis of the Class Size Effect Researches by the Quasi-Experimental Designs

Akihiro HASHINO

From 1990's the empirical methods by quasi-experimental designs have grown, many papers on class size effects have been published. However these estimated magnitudes of the class size effects are varying and mixed. This paper aims to review the literatures systematically by the meta-analysis.

First, we synthesize the previous researches by quasi-experimental design such as instrumental variable regression, regression discontinuity design and difference-in-differences. The estimated mean class size effect is +0.01 SD per student reduction. Second, we analyze the factors of heterogeneities of the class size effect. The nonlinearity of the class size effects and the factor about the policy on teacher salary are examined.

目次

- I. 課題設定
- II. 本研究の視点
- III. 対象と方法
 - A. 対象の選定
 - B. 分析モデル
- IV. メタアナリシス
 - A. 既存研究の統合
 - B. 異質性の考察
 - 1. 方法に関する要因
 - 2. データの特質に関する要因
 - 3. 政策的要因
- V. まとめ

I. 課題設定

この論稿は、準実験デザインによる学級規模効果研究のメタアナリシスを通じて、知見の統合と外的妥当性の考察を行うことを目的とするものである。

我が国において、1958年の「公立義務教育諸学校の学級編制及び教職員定数の標準に関する法律」以降 7

次に渡る教職員定数改善が行われ、特に教育条件・労働条件の改善としての学級規模の縮小は継続的に教育政策のアジェンダとして存在してきた。義務教育段階における県費負担教職員給与費国庫負担分の算出の基準となる学級規模の標準は第5次改善以降40人に据え置かれているが、近年では、政権党となった民主党が少人数学級の促進を掲げ、中央教育審議会答申においても35人への縮減を提言するなど教員定数改善策は政策的アジェンダとして再論されている。また第177通常国会では小学校1年生を対象とした35人学級実施のための改正標準法案および予算案が提出される見込みとなっている。このような背景のもとで研究の文脈においても、山崎他(2009)などにおいて、新しい計量的手法によって学級規模の効果に関する実証分析が報告されている。

海外に目を転じると、学級規模の効果に関する実証分析の歴史自体は非常に古く、アメリカでは100年以上もの蓄積が存在する⁽¹⁾。実証分析に際して一貫して問題とされてきたのは、因果的推論(Causal inference)であり、近年の施策評価(program evaluation)の文脈においても因果的推論を重視した手法が求められている。因果的推論を行う上での王道は、処置(treatment)を被験者に対して無作為に割り当てる実験的手法であるが、社会科学分野においては政策実験が行われるこ

と自体稀であり、行われるとしてもその厳密な実施は難しい⁽²⁾。実行が難しい実験的手法の代替として社会科学分野の政策分析・施策評価において主流を占めているのは準実験的方法による観察データの分析であり、1990年代以降の学級規模効果の分析においても方法の確立・適用が進み、多くの研究が生み出されている。

もともと、数々の準実験的方法による学級規模効果の推定結果には、研究間で一定の幅が存在しており、また決定的となりうる研究が有るわけではない。このような幅のある既存研究の結果をどのように統合するのか、あるいは研究間での異質性は何に由来するかという点を考察することは重要な研究課題であると言える。本稿は、既存研究論文に記載されている効果の推定値自体をデータとして扱うメタアナリシスと呼ばれる手法を用いて、これらの課題について答えようとするものである。メタアナリシス自体は教育学分野では古くから用いられている手法であり、学級規模効果研究を対象としたものとしては、頻繁に引用されるGlass&Smith(1979)のほか、Goldstein et al.(2000)などがある。ただし、これらのメタアナリシスで対象とされているのは、比較的古い時期の実験およびマッチングによる研究であり、90年代以降に行われた研究を対象としていない。その点で本稿はGoldstein et al.(2000)の知見について対象となる研究を広げつつ、アップデートするものであると言える。

メタアナリシスに基づく本稿の主な結論は以下の2点にある。まず第1は、学業成績向上という点での学級規模縮小の平均的効果は、1人あたり+0.01標準偏差程度であり、モダレートであるということである。第2は、学級規模効果の推定値の相違は、分析手法の選択と少人数学級が実施される環境に依存している可能性があるということである。本稿の構成は以下の通りである。2節では既存研究における準実験的方法について見た上で、分析のための視点を提示する。3節ではメタアナリシスの対象となる実証分析について概観する。4節ではメタアナリシスを行い、5節では分析結果をふまえた考察と含意を述べる。

II. 本研究の視点

この節では、施策評価における論点と本稿で行うメ

タアナリシスとの関係を明確にしておきたい。

まず、Leamer(2010)の議論に沿って既存の施策評価に関する論点を概観しておく。施策効果について線形回帰分析の枠組みの中で分析を行う場合、式(1)のような形となる。 y は結果に関する変数、 t は処置(treatment)の有無・程度に関する変数であり、また w は結果に影響を与える加法的交絡要因の行ベクトルである。式のうちパラメータ β が関心の対象となる処置効果である。

$$y_i = \alpha + \beta t_i + w_i \delta + \epsilon_i \quad (1)$$

処置効果の推定において問題となるのが、交絡要因 w に関する完全なデータがない場合である。パラメータ δ が非ゼロでなおかつ加法的交絡要因 w と処置変数 t とに相関があるならば、処置効果 β はバイアスを持つことになる。この問題は「内生性」の問題として知られている。無作為に処置を割り当てる実験的手法では w と t の相関をなくし、バイアスを除去した推定値を得ることができる。しかし、先にも述べたように社会科学において無作為割り当てを施したデータを得るのは難しく、観察データのみが利用可能なケースがほとんどである。準実験的方法は観察データという制約の中で内生性の問題に対処する手法であり、主なものとしては、操作変数法、regression discontinuity法(以下RD法)、パネルデータを用いた方法——固定効果モデルおよびdifference-in-differences法(以下DID法)——などがある。以下でこれらの手法について簡単な説明をしておく。

まず操作変数法は、内生的説明変数 t と相関し、かつ ϵ と相関を持たない外生変数(=操作変数)の分散によってパラメータ β の一致推定を行うという方法である。操作変数を z とするならば、実際の実証分析では t を z およびその他の外生変数に回帰させ、それによる t の予測値を用いて式(1)の推定を行うという手続きをとる。

RDデザインは、図1のようにある閾値 c を境に断絶をもたらす外生変数 x によって推定を行う方法である。実際の分析では、閾値付近のサンプルで、結果変数を処置変数に回帰させるという手続きを採る。処置の割当について個人が閾値付近を操作的に正確に選択できるのでない限り、閾値付近では局所的に処置の割当は無作為化される(Lee&Lemieux, 2010)。学級規

図1 Regression Discontinuity デザイン

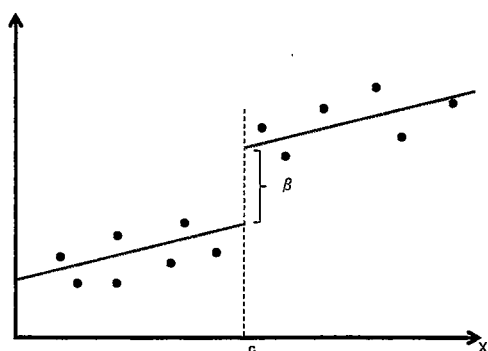


表1 Difference in Differences 法

	処置群 (y^t)	統制群 (y^c)	処置群と統制群の差
処置前 (t_0)	\bar{y}_0^t	\bar{y}_0^c	$\bar{y}_0^t - \bar{y}_0^c$
処置後 (t_1)	\bar{y}_1^t	\bar{y}_1^c	$\bar{y}_1^t - \bar{y}_1^c$
処置前と処置後の差	$\Delta \bar{y}^t$	$\Delta \bar{y}^c$	$\Delta \bar{y}^t - \Delta \bar{y}^c$

模効果に関する研究では、多くの国で Maimonides ルール（学年在籍者数が学級規模上限の倍数を超えると、学級数が1つ増え、児童・生徒が均等に分割される）が採られているため、学年在籍者数が断絶をもたらす x に該当することになる。

パネルデータ（多数の個体について複数時点で記録されたデータ）を用いた方法としては固定効果モデルおよび DID 法がある。式(3)は線形のパネルモデルの典型的な例である。固定効果モデルは時間的に変動しない個体レベルの固有の効果 α_i を各個体のダミー変数で推定するものであり、時間的に変動しない加法的交絡要因による内生性に対処することができる。

$$y_{ij} = \beta t_{ij} + \alpha_i + \epsilon_{ij} \quad (2)$$

線形固定効果モデルの推定は通常、被説明変数から各個体の時間を通じた被説明変数の平均値を減じた値を説明変数から各個体の時間を通じた説明変数の平均値を減じた値に回帰させることで推定が行われ、この推定値は、漸近的には、被説明変数の階差を説明変数の階差に回帰させることと同じものとなる。

DID 法もパネルデータにおける方法である。DID 法は表1のように、処置を受けた個体の処置前と処置後の結果の差 (Δy^t) の平均と処置を受けない個体の前後の結果の差 (Δy^c) の平均の差分を処置効果とするもの

である。

実際の推定においては式(3)のモデルを推定することになる。j は時点を第1期（処置前）と第2期（処置後）を示す添え字であり、t は第2期で処置を受けたか否かを示すダミー変数、d は第2期に属するか否かを示すダミー変数である。

$$y_{ij} = \alpha_0 + \alpha_1 t_i + \alpha_2 d_j + \beta t_i \cdot d_j + \epsilon_{ij} \quad (3)$$

この式のパラメータと表1との対応関係は、以下のようになり、関心のある処置効果に関するパラメータは β となる。

$$\bar{y}_0^t = \alpha_0$$

$$\bar{y}_0^c = \alpha_0 + \alpha_1$$

$$\bar{y}_1^t = \alpha_0 + \alpha_2$$

$$\bar{y}_1^c = \alpha_0 + \alpha_1 + \alpha_2 + \beta$$

$$\therefore \Delta \bar{y}^t - \Delta \bar{y}^c = \beta$$

また、データが2時点の場合、DID 法の推定は式(4)のような時間効果をもつ固定効果モデルと等しくなる。式(4)は式(5)のように階差をとれば、個体固有の効果について除外することができ、最小二乗法によって処置効果 β の推定値を得ることができる。

$$y_{ij} = \beta t_{ij} + \alpha_i + \delta_j + \epsilon_{ij} \quad (4)$$

$$\therefore \Delta y_{ij} = \beta \Delta t_{ij} + (\delta_j - \delta_{j-1}) + \Delta \epsilon_{ij} \quad (5)$$

学級規模効果に関する実証分析ではこれらの準実験的方法によって多くの研究がなされているが、複数の研究の推定値の情報を統合することが、本稿のメタアナリシスの第1の作業課題となる。

以上の議論は処置効果が一定であることを前提としたものである。施策評価において内生性の他の重要な問題として異質性の問題があり、これは式(6)のように交互作用的交絡要因の問題として捉える事ができる。

$$y_i = \alpha + (\beta + z_i \gamma) t_i + w_i \delta + \epsilon_i \quad (6)$$

変数 z は個人レベル変数ないし処置が行われた文脈に関する変数である。パラメータ γ が非ゼロであれば、処置効果には異質性があることになる。変数 z について社会経済的背景や性別など観察可能な個人レベルの要因を考えるのであれば、異質性を考慮することは難しくない。しかし、データ内で変動しない文脈変数の影響を考慮することはできない。例えば、データがある年の1つの自治体で採られたデータであるならば、国・自治体レベルの教育政策の方針が処置効果にどれだけ作用しているかを明らかにすることはできない。

複数の研究のパラメータ推定値自体をデータとして扱うメタアナリシスは、このような単一の研究では考慮することができない要因の影響を明らかにする際に有効な方法である。効果の推定値の相違をもたらす文脈変数の作用の検出が、本稿のメタアナリシスの第2の作業課題である。

III. 対象と方法

A. 対象の選定

以下の3つの基準に基づいて、メタアナリシスの対象となる研究を選定した。

- (1)分析手法において、学級規模変数の内生性の問題を考慮した、準実験的方法を用いていること。具体的には、操作変数法、RD デザイン、DID 法、固定効果モデルを用いた分析を対象とする。
- (2)学級規模の効果にあたる被説明変数について、標準

化されたテストのスコアを用いていること。

- (3)査読付きの学術雑誌に掲載された研究であること。施策評価のレポートやワーキングペーパーとしてみ公表されたものは対象としない。

論文の検索にあたっては、2つの方法を用いた。まず第1に、教育学文献データベースERICにおいて、“Class Size”と“Effect”または“Achievement”のキーワードで検索した上で、上記の3つの条件に該当する研究をピックアップした。第2に経済学関係の英文ジャーナルのウェブサイトにおいて同様に論文の検索し、選定を行った。対象とした雑誌は総合誌の *Econometrica*、*American Economic Review*、*Journal of Political Economy*、*Quarterly Journal of Economics*、*Review of Economics and Studies*、*Economic Journal* およびフィールドジャーナルの *Economics of Education Review*、*Education Economics*、*Journal of Human Resources* である。また、これらの過程において選定された掲載論文に引用された実証分析についても精査し、上記の3つの基準を満たす論文についても加えた。その結果、選定された研究は22論文となった。

メタアナリシスの対象となる22論文の概要は表1に示されている。対象となる研究の傾向について指摘しておくべき点は次の3点である。まず第1に先進国の初等教育段階を対象とした分析が多く、数学と言語関係の教科の標準化テストへの効果を分析したものが多ということである。各研究における学級規模の範囲は概ね15~40人の範囲にある。

第2に用いられた方法としては操作変数法が圧倒的に多数を占めるということである。操作変数法において操作変数として用いられているのは、学級規模の学年平均 (Akerhielm, 1995) もしくは“Maimonides”ルール⁽³⁾によって学年児童生徒総数から予測される学級規模である (Angrist&Lavy, 1999)。その他の手法については、RD デザインが3研究、DID 法が1研究、固定効果モデルにのみ基づいた研究が2研究となっている。

第3に、定式化には $y = \beta_0 + \beta_1$ (学級規模) + ... という線形モデルが多く用いられているということである。この定式化は学級規模の限界効果は規模によらず一定であることを意味している。線形モデルによって推定された効果は、学級規模1人増加につき-0.07標準偏

確実的手法による学校規模効果研究のメタアナリシス

表2 メタアナリシスの対象となる先行研究

No	研究	国	学年	教科	推定法	備考
1	Averchinn(1995)	アメリカ(NELS1988)	8	数学、英語、理科、歴史	操作変数法	
2	Angst&Lavy(1999)	イスラエル	3,4,5	言語	操作変数法・Regression Discontinuity	
3	Asadikh(2005)	インドネシア	中等教育		操作変数法	
4	Bonesonmg(2005)	ルウェー	8-9,10	数学	操作変数法	
5	Boonen&Kouset(2001)	アメリカ(NELS1988)	10,12		操作変数法	
6	Barband et al.(2005)	ケンタッキー州	3	数学、スペル、読解、国語、理科	操作変数法(同時方程式)	学校規模の「乗項を含むモデル
7	Bressoux et al.(2009)	フランス	3	数学、読解	操作変数法	
8	Datar&Masou(2008)	イギリス	1	数学、読解	操作変数法	学校規模の「乗項を含むモデル
9	Dobrichsen et al.(2002)	オランダ(PRIIMA)	4,6,8	数学、国語	操作変数法	
10	Heinesen(2010)	デンマーク	7,8,9	フランス語	操作変数法	
11	Hoxby(2000)	コネチカット州	4,6	数学、読解、ライティング	操作変数法・Regression Discontinuity	説明変数=学校規模の対数
12	Iacovou(2002)	イギリス	1	数学、読解	操作変数法	
13	Jakobovsk&Sakowski(2006)	ポーランド	6	数学、読解	操作変数法	説明変数=学校規模の対数
14	Jepsen&Rivkin(2009)	カリフォルニア州	2,3,4	数学、読解	固定効果モデル	
15	Kueger(1999)	テキサス州	K,1,2,3	SAT(読解、単語、認知、数学)	操作変数法	
16	Larven et al.(2008)	ルウェー	9	数学	操作変数法	
17	Lavy(2001)	オランダ	4,6,8	数学	操作変数法	
18	Lindahl(2005)	スウェーデン	5	数学	Difference in Differences	
19	Rivkin et al.(2005)	テキサス州	4,5,6	数学、読解	固定効果モデル	
20	Sims(2009)	カリフォルニア州	5	数学、国語	操作変数法	
21	Urbida(2006)	ポリビア	3	数学、国語	操作変数法・Regression Discontinuity	
22	Weissmann&West(2006)	オーストラリア、カナダ、チエコ、フランス、ギリシャ、イスラエル、ポルトガル、ルーマニア、シンガポール、スロベニア、スペイン	8	数学、理科	操作変数法・固定効果モデル	

差 (Boozer&Rouse、1999) $\sim +0.01$ 標準偏差 (Dobbelsteen at al、2002) と幅広い。線形でない効果を想定した研究としては、説明変数に学級規模の2乗項を加え、最適規模があることを前提としたものが2研究 (Borland et al、2005; Datar&Mason、2008)、説明変数に学級規模の対数を用い、限界効果が学級規模に反比例することを前提としたものが2研究ある (Hoxby、2000; Jakubowski&Sakowski2005)。

B. 分析モデル

変量効果モデルによるメタアナリシスは、下式 (7) のように表現される。 β_i は i 番目 ($i=1, 2, \dots, n$) の研究における推定値、 σ_i はその標準誤差、 τ^2 は研究間分散である。

$$\hat{\beta}_i = \mu + v_i + \epsilon_i \quad (7)$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$$

$$v \sim \mathcal{N}(0, \tau^2)$$

これは階層線形モデルの1種であり、第1段階の誤差 ϵ の分散が既知のモデルである (Raudenbush&Bryk、2002:Ch7)。各研究を統合した μ は平均的效果を示すパラメータである。 μ の事前分布を $\mathcal{N}(\mu_0, \sigma_0^2)$ とすると、 τ^2 が与えられた下での μ の推定値は式 (8) のような β と μ_0 の加重平均として表現される。

$$\hat{\mu} = w_0 \mu_0 + \sum_{i=1}^n w_i \hat{\beta}_i \quad (8)$$

$$w_0 = \frac{\sigma_0^{-2}}{\sigma_0^{-2} + \sum_{i=1}^n (\sigma_i^2 + \tau^2)^{-1}}$$

$$w_i = \frac{(\sigma_i^2 + \tau^2)^{-1}}{\sigma_0^{-2} + \sum_{i=1}^n (\sigma_i^2 + \tau^2)^{-1}}$$

この式から、 μ の事前分布の分散 σ_0^2 が小さいほど、事前情報が μ の値に反映されやすく、推定値の標準誤差が小さい研究のウェイトがより大きくなるのが分かる。つまり、メタアナリシスによる統合は、事前情報と各研究の推定値の精度の大小を考慮して行われるのである。

本稿のメタアナリシスに用いる分析モデルは、式(9)のような各研究の学級規模の効果の推定値を被説明変数とした変量効果モデルである。

$$\Delta_{ij} = \mu + v_i + \epsilon_{ij} \quad (9)$$

$$v_i \sim \mathcal{N}(0, \tau^2)$$

$$\Delta_{ij} = \hat{\beta}_{ij}$$

$$\text{Var}(\epsilon_{ij}) = \widehat{\text{Var}}(\hat{\beta}_{ij})$$

式(9)における Δ_{ij} は研究 i 、モデル j における学級規模の効果であり、学級規模1人増加によるスコアの変動分 (単位=標準偏差) である⁽⁴⁾。各研究では、推定法や被説明変数、統制変数の相違によって複数のモデルの推定が行われており、定式化の相違を反映させるために、1つの研究の中の複数の推定値をデータとして用いる。ただし、各研究で推定したモデルの数に応じてウェイトを加重する。例えば5つのモデルによる5つの推定値を含む研究では各モデルに1/5のウェイトを掛け、2つの推定値を含む研究では各モデルに1/2のウェイトを掛ける。 μ は全ての研究・定式化から推定された平均効果である。 ϵ は推定値固有の誤差であり、その分散は各研究に記載されている係数の標準誤差を2乗した値である。 v は研究レベルの特性であり、平均0、分散 τ^2 の正規分布に従うと仮定する。この式(9)のモデルを用いた分析の目的は既存研究を統合することによって、平均的な効果を推定することにある。

式(9)については、さらに説明変数を加えた下式 (10) のようなモデルを考えることができる。 z は説明変数の行ベクトルである。

$$\Delta_{ij} = \mu + z_{ij} \gamma + v_i + \epsilon_{ij} \quad (10)$$

説明変数として考えられるのは、データの取られた国・地域における特質や制度、あるいは研究に関するメタレベルの変数 (サンプルの性質や方法の相違) などのコンテキスト変数が考えられる。式(10)のモデルを用いた分析の目的は、学級規模効果の推定値の研究ごとの相違を考察することにある。

IV. メタアナリシス

A. 既存研究の統合

まず、式 (9) のモデルによって既存研究の統合を行

う。分析は、 μ について 2 つの事前分布を設定したベイズ推定によって行う。第 1 の事前分布は、平均 0、分散 100^2 の正規分布である。この分散の大きな事前分布は実質的に事前情報が分析に影響を与えないことを意味している。第 2 の事前分布は、平均 -0.022 、分散 0.007^2 の正規分布であり、この事前分布は実験およびマッチングによる研究を対象にメタアナリシスを行った Goldstein et al.(2000)の推定値に基づいている。メタアナリシスにおけるサンプルサイズも小さいことから、この小さな分散の事前分布は、事前情報がデータの情報と統合され、分析に反映されることを意味する。パラメータ推定はギブズサンプラーによって行う。

表 3 はパラメータの推定結果を示したものである。上段は線形性 (限界的効果は一定) を想定した 18 研究を対象としたメタアナリシスの結果である。中段は非先進国における推定値を除外した結果である⁽⁶⁾。パラメータ μ は各研究を統合した平均効果であり、数値は学級規模が 1 人増加した時に学力テストのスコアが何標準偏差変動するかを示したものである。事前分布の設定によって結果は異なっており、第 2 の事前分布を用いた分析では若干効果が大きく推定されているが、4 つの推定値から総合すると、学級規模 1 人減少につき、スコアが 0.013 標準偏差増加する程度の効果があるということが出来る。90%区間推定値では 0.007~0.019 標準偏差となっている。効果の線形性を前提とすれば、これは、学級規模を 40 人から 30 人へと縮小した時に偏差値 (平均 50、標準偏差 10) にして 1~2 上昇させる程度の効果となる。

上記の推定値は線形モデルを用いた研究のみを対象とした分析である。3 節でも述べたように、非線形効果を想定したものが 4 研究あり、定式化が異なるためメタアナリシスに反映させることができない。具体的には、Borland et al.(2005)と Datar&Mason(2008) は説明変数に学級規模の 2 乗項が含まれており、Hoxby(2000)と Jakubowski&Sakowski(2005)は学級規模の対数を説明変数として用いている。これらをメタアナリシスに含めるために、以下のような形で線形近似する。学級規模の取りうる範囲が \min 人から \max 人であるとする、線形近似のアイデアは、式 (11) のように被説明変数を \min ~ \max 人の範囲を 1 人単位で分節化し、限界効果について考えるというものである。

$$\Delta_{ij} = 1\mu + 1v_i + \epsilon_{ij} \quad (11)$$

$$\Delta_{ij} = \begin{bmatrix} \Delta_{ij,\min} \\ \Delta_{ij,\min+1} \\ \vdots \\ \Delta_{ij,\max-1} \end{bmatrix}$$

$$Var(\epsilon_{ij}) = \begin{bmatrix} Var(\epsilon_{ij,\min}) & 0 & \cdots & 0 \\ 0 & Var(\epsilon_{ij,\min+1}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Var(\epsilon_{ij,\max-1}) \end{bmatrix}$$

(1) 線形モデルの時 ($y = \alpha + \beta CS + \dots$)

$$\Delta_{ijk} = \hat{\beta}_{ij}$$

$$Var(\epsilon_{ijk}) = Var(\hat{\beta}_{ij})$$

$$k = \min_{ij} \cdots \max_{ij} - 1$$

(2) 説明変数に学級規模の 2 乗項を含む時 ($y = \alpha + \beta_1 CS + \beta_2 CS^2 + \dots$)

$$\Delta_{ijk} = \hat{\beta}_{1ij} + (2k+1)\hat{\beta}_{2ij}$$

$$Var(\epsilon_{ijk}) = [1, 2k+1] Var(\hat{\beta}_{ij}) [1, 2k+1]'$$

(3) 説明変数が学級規模の対数である時 ($y = \alpha + \beta \ln CS + \dots$)

$$\Delta_{ijk} = \frac{\hat{\beta}_{ij}}{(k+0.5)}$$

$$Var(\epsilon_{ijk}) = \frac{Var(\hat{\beta}_{ij})}{(k+0.5)^2}$$

被説明変数である Δ_{ij} は $(\max-\min) \times 1$ のベクトルとなる。その要素 $\Delta_{ij,k}$ は、学級規模が k 人から $k+1$ 人に増加した時の限界的効果である。 Δ の要素については上記の通り、線形モデルの場合、限界効果は一定であるため全て同一であり、一方で非線形モデルの場合は区間によって限界効果は異なるため、異なる値をとる。

表 3 下段は式 (11) について分析を行った結果を示している。式 (9) と同様にパラメータ μ について 2 種類の事前分布を設定し、ギブズサンプラーによってパラメータを推定した。表から学級規模 1 人減少につきスコアが 0.01 標準偏差前後増加する程度の効果があるということが読み取れる。線形モデルに基づいた研究を対象とした場合に比べて、非線形モデルに基づいた研究を含めた場合には平均効果はやや小さくなっている。

表3 線形モデルのメタアナリシスの結果

1) 線形モデル

	事前分布 $N(0,100^2)$		
	事後平均	90%CI下限	90%CI上限
μ	-0.01186	-0.01707	-0.00697
τ	0.00941	0.00510	0.01492

研究数18、モデル数83

	事前分布 $N(-0.022,0.007^2)$		
	事後平均	90%CI下限	90%CI上限
μ	-0.01353	-0.01707	-0.01707
τ	0.00973	0.00510	0.00510

研究数18、モデル数83

2) 線形モデル(途上国を除外)

	事前分布 $N(0,100^2)$		
	事後平均	90%CI下限	90%CI上限
μ	-0.01181	-0.01715	-0.00683
τ	0.00939	0.00484	0.01517

研究数16、モデル数75

	事前分布 $N(-0.022,0.007^2)$		
	事後平均	90%CI下限	90%CI上限
μ	-0.01352	-0.01707	-0.01707
τ	0.00975	0.00510	0.00510

研究数16、モデル数75

3) 線形モデル+非線形モデル

	事前分布 $N(0,100^2)$		
	事後平均	90%CI下限	90%CI上限
μ	-0.00852	-0.01325	-0.00414
τ	0.00951	0.00587	0.01427

研究数22、モデル数100

	事前分布 $N(-0.022,0.007^2)$		
	事後平均	90%CI下限	90%CI上限
μ	-0.01042	-0.01534	-0.00608
τ	0.00998	0.00622	0.01493

研究数22、モデル数100

B. 異質性の考察

前項では複数の研究における推定値を統合し、平均的効果を明らかにするという目的でメタアナリシスを行ったが、この項では各研究の推定値の異質性に着目し、その要因について明らかにすることを試みる。

異質性を説明する要因としては大きく分けて、方法、データの特質、制度・政策の3つの要因が考えられる。以下では試論的にではあるが、これらの3つの要因について分析することとしたい。分析は式(10)のモデルに従って行う。

1. 方法に関する要因

2節で観察データを用いた施策評価の分析方法を紹

介したが、特定の手法の採用自体が系統的に処置効果の推定に影響を与えている可能性がある。既存研究における実際の分析方法は大まかに、クロスセクションデータを用いた操作変数法・RD デザイン、パネルデータを用いた方法(固定効果モデル、DID法)に分けられる。パネルデータでも操作変数法、RD法を用いることは考えうるが、パネルデータを用いた分析では基本的には観測されない非時変的要因の統制に主眼があるため、メタアナリシスではパネルデータの使用の有無に着目して分析する。

表4の(1)は、パネルデータを用いた方法か否かという2つのダミー変数を説明変数として投入した分析の結果である。表ではダミー変数の係数の点推定値が正

の値となっており、標準誤差が大きいため断言は控えるべきではあるが、RD デザイン、パネルデータを用いた分析は操作変数法による分析と比較して、学級規模の効果は小さめに推定されるということを示している。

2. データの特質に関する要因

第2に考えられる要因は、分析対象となるサンプルの特質に関する要因である。学級規模の効果に関する研究について言えば、データはどの学年について収集されたのか、データにおいて学級規模の取りうる値はどの範囲にあるのか、という点などが具体的に挙げられるであろう。後者は言いかえれば、学級規模の限界効果は規模に対して一定なのかという線形性の問題である。

表4の(2)では、対象学年に関する2つのダミー変数——第3学年以下か否か、第7学年以上か否か——と、データにおける学級規模の平均の値、その2乗、学級規模の標準偏差を説明変数とした分析の結果を示している。表より次の2点が指摘できる。まず第1に第7学年以上のダミー変数の値が正の値となっており、学級規模の影響は中等教育段階で小さくなる。そして

第2に学級規模の増減による限界効果は一定ではなく、それゆえ学校規模の効果に非線形性がある。Glass&Smith(1978)のメタアナリシスでは規模が小さくなるほど限界効果が大きくなるという点を示したが、この分析では異なる結果を得ている。学級規模が小さくなるほどその限界効果が大きくなるならば、学級規模平均の係数は正になり、その2乗項の係数は0または正になるはずであるが、本分析では2乗項の係数の点推定値は負の値となっており、学級規模が27人以上の区間では学級規模が小さくなるほど限界効果は小さくなるという関係がみられる。

3. 政策的・制度的要因

第3は地域・国レベルの要因である。通常のデータ収集は特定の地域・国で行われるため、地域・国レベルで変動する要因については考慮することができない。地域・国レベルの要因としては最も重要なのが政策的・制度的要因である。学級規模の効果に関して言えば、労働条件などの教員の質や需給関係に関わる政策的要因、教員の職能成長に関わる制度が影響を与えうる候補として考えられる。本稿の分析ではデータの制約から教員の給与水準のみを対象とする。

表4 説明変数を伴うメタアナリシスの結果

	事後平均	標準誤差	90%CI下限	90%CI上限	収束判定統計量
定数項	-0.01354	0.00410	-0.02027	-0.00687	-0.30822
RD	0.00712	0.00789	-0.00585	0.02007	-0.45886
パネルデータ	0.00230	0.00680	-0.00902	0.01315	0.95168
τ	0.00995	0.00334	0.00516	0.01598	1.27175

研究数18、モデル数84

	事後平均	標準誤差	90%CI下限	90%CI上限	収束判定統計量
定数項	-0.09627	0.04060	-0.16271	-0.03017	-0.97563
第3学年以下	0.00340	0.00424	-0.00339	0.01053	0.29141
第7学年以上	0.01008	0.00501	0.00196	0.01847	1.22031
学級規模平均	0.00660	0.00343	0.00096	0.01226	0.54316
学級規模平均2乗/100	-0.01234	0.00692	-0.02383	-0.00102	-0.56698
学級規模標準偏差	-0.00089	0.00118	-0.00284	0.00104	0.08695
τ	0.00798	0.00282	0.00415	0.01315	-0.34560

研究数18、モデル数84

	事後平均	標準誤差	90%CI下限	90%CI上限	収束判定統計量
定数項	-0.02115	0.01327	-0.04350	-0.00030	-0.18561
15年目教員給与/1人あたりGDP	0.00761	0.01081	-0.00974	0.02563	0.01771
τ	0.01080	0.00359	0.00593	0.01735	0.65510

研究数14、モデル数65

表4の(3)は給与水準を説明変数として加えた分析である。給与水準のデータは国レベルのデータであり、OECDのEducation at a Glanceにおける15年勤務教員給与の対1人当たりGDP比である。表から、係数の点推定値は正の値となっており、給与水準が高い国では学級規模の影響は小さくなるという関係が見出せる。ただしパラメータ推定における精度が小さく、区間推定値が正負に跨っているため、あくまでも参考にとどめるべき結果であるが、給与水準が学級規模の効果に影響を与えるという可能性が見出せる。海外の学級規模効果に関する実証研究として参照される機会が多いのは教員給与水準の低いアメリカのものであるが、その知見(特に学級規模縮小の効果があるという知見)が日本などの比較的教員給与水準が高い国にも当てはまるかは留意する必要がある。

V. まとめ

学級規模に関する因果的推論を重視した実証分析は我が国においては本格化していないが、数千億円のオーダーで生じる教員給与費負担に照らせば、効果に関する政策評価は不可欠である。本稿では、準実験的方法による既存の学級規模効果研究について、メタアナリシスによって、統合と異質性の要因分析の2つの作業を行った。

変量効果モデルに基づいた既存研究の統合において、学校規模の平均的改善効果は1人減少につき、0.013標準偏差程度であることが明らかにされた。これは、あくまでも線形性を前提とした既存研究のメタアナリシスによる結果であるが⁶⁾、その大きさについて、実質的に効果があると言えるかは微妙なところである。その判断は個々人の主観的判断によるとも言えるが、代替手段の存在の有無に依存しており、実証的かつ多角的に議論されるべきものでもある。また、既存研究における推定値の異質性の要因分析では、学級規模縮小による改善効果は初等教育の場面に限定されること、学級規模の限界効果は一定ではなく、効果に非線形性があること、必ずしも学級規模が小さいほどより効果(限界的効果)が大きくなるとは言えないことが明らかにされた。また、教員給与水準の相違も学級規模の効果に影響を与える可能性が見出された。

今後の課題としては、異質性の要因分析の発展が挙げられよう。特に政策的・制度的要因の考察については本稿での分析は試論的なものであり、非常に限定的なものであったが、その含意から考えれば最も必要なものであると考えられる。仮に定数改善による学級規模縮減施策を行うとすれば、どのような環境の下でより効果的となるのか、総額裁量制と財政的制約の下でどのような資源配分をすればよいのかという点についての分析は教育政策・教育財政研究における喫緊の課題である。またそのためにも基礎的な作業として、研究が行われたコンテキストに関する比較可能な情報の収集も求められるのである。

注

¹⁾ 学校規模の効果のメタアナリシスとしてよく知られているGlass&Smith(1979)によれば、1970年代までの学級規模効果研究は、実験的手法導入以前の第1期(1895~1920年)実験的手法が導入された第2期(1920~1940年)、大規模調査データの分析による第3期(1950~1970年)、個別教授法を対象とした第4期(1970年~)に分けられているという。

²⁾ 例えば、学級規模効果研究で最も引用されるテネシー州のSTAR(Student Teacher Achievement Ratio)プロジェクトにおける政策実験に関して、処置群と統制群間の児童の移動や児童のattritionによって、無作為割り当てが不十分であったことは既に指摘されることである(Hanushek, 1999)。

³⁾ “Maimonides”ルールとは、学年在籍者数が学級規模の上限の倍数を超えた場合、学級数を1つ増やし、均等に児童・生徒数を配分するルールである。日本も基本的にこのルールによって運用されている。このルールが操作変数法及びRD法に適すのは、このルールによる決まる学級規模が外生変数となるからである。

⁴⁾ β は各論文の係数の数値を記述統計に記載されている被説明変数の標準偏差で除した数値である。標準誤差についても同様である。

⁵⁾ 除外したのは、バングラデシュの中等学校を対象としたAsadullah(2005)、ボリビアの中等学校を対象としたUrquiola(2006)および、Woessmann&West(2005)におけるルーマニアとスロベニアの推定値である。

⑥例えば、学級規模が40人から30人に変化した場合、0.13標準偏差程度の改善ということになる。PISAやTIMSSといった国際学力調査で標準偏差は100ポイントであり、これらのスコアでいえば、0.13標準偏差の改善は13ポイントの上昇ということになる。偏差値でいえば50から51.3になる程度の効果である。

文献

- Akerhielm, Karen, 1995, "Does Class Size Matter?", *Economics of Education Review*, 14(3), pp.229-241
- Angrist, Joshua D. and Victor Lavy, 1999, "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", *Quarterly Journal of Economics*, 114, pp.533-75
- Bonesronning, Hans, 2003, "Class Size Effects on Student Achievement in Norway: Patterns and Explanations", *Southern Economic Journal*, 69, (4), pp. 952-96
- Borland, Melvin V. and Roy M. Howsen and Michelle W. Trawick, 2005, "An Investigation of the Effect of Class Size on Student Academic Achievement", *Education Economics*, 13(1), pp.73-83
- Boozer, Michael and Cecilia Rouse, 2001, "Intraschool Variation in Class Size: Patterns and Implications", *Journal of Urban Economics*, 50(1), pp.163-189
- Bressoux, Pascal and Francis Kremer and Corinne Prost, 2009, "Teacher's Training, Class Size and Students' Outcomes: Learning from Administrative Forecasting Mistakes", *Economic Journal*, 119, pp.540-561
- Datar, Ashlesha and Bryce Mason, 2008, "Do Reductions in Class Size "Crowd Out" Parental Investment in Education?", *Economics of Education Review*, 27, pp.712-723
- Dobbelsteen, Simone and Jesse Levin and Hessel Oosterbeek, 2002, "The Causal Effect of Class Size on Scholastic Achievement: Distinguishing the Pure Class Size Effect from the Effect of Changes in Class Composition", *Oxford Bulletin of Economics and Statistics*, 64(1), pp.17-38
- Finn, J. D. and C. M. Achilles, 1990, "Answers and Questions about Class Size: A Statewide Experiment", *American Educational Research Journal*, 27(3), pp.557-577
- Glass, Gene and M. L. Smith, 1979, "Meta-Analysis of Research on Class Size and Achievement", *Educational Evaluation and Policy Analysis*, 1(1), pp.2-16
- Goldstein, Harvey and Min Yang and Rumana Omar and 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*, 49, pp.1-14
- Heinesen, Eskil, 2010, "Estimating Class-Size Effects Using within School Variation in Subject-Specific Classes", *Quarterly Journal of Economics*, 120, pp.737-760
- Hoxby, Caroline, Minter, 2000, "The Effects of Class Size on Student Achievement: New Evidence from Population Variation", *Quarterly Journal of Economics*, 115(4), pp.1239-1285
- Iacovou, Maria, 2002, "Class Size in the Early Years: Is Smaller Really Better?", *Education Economics*, 10(3), pp.261-290
- Jakubowski, Maciej and Pawel Sakowski, 2006, "Quasi-Experimental Estimates of Class Size Effect in Primary School in Poland", *International Journal of Education Research*, 45, pp.202-215
- Jepsen, Christopher and Steven Rivkin, 2009, "Class Size Reduction and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size", *Journal of Human Resources*, 44(1), pp.223-250
- Krueger, Alan. B., 1999, "Experimental Estimates of Education Production Functions", *The Quarterly Journal of Economics*, 114(2), pp.497-532
- Leamer, Edward E., "Tantalus to the Road to Asymptopia", *Journal of Economic Perspectives*, 24(2), pp.31-46
- Lee, David S. and Thomas Lemieux, 2010, "Regression Discontinuity Designs in Economics", *Journal of Economic Literatures*, 48, pp.281-355
- Leuven, Edwin and Hessel Oosterbeek and Marte Ronning, 2008, "Quasi-Experimental Estimates of the Effects of Class Size on Achievement in Norway", *Scandinavian Journal of Economics*, 110(4), pp.663-693

- Levin, Jesse, 2001, "For Whom the Reductions Count: A Quantile Regression Analysis of Class Size and Peer Effects on Scholastic Achievement", *Empirical Economics*, 26(1), pp.221-246
- Lindahl, Mikael, 2005, "Home versus School Learning: A New Approach to Estimating the Effect of Class Size on Achievement", *Scandinavian Journal of Economics*, 107 (2), pp.375-394
- Rivkin, S.G and Eric A. Hanushek and J.F. Kain, 2005, "Teachers, Schools and Academic Achievement", *Econometrica*, 73(2), pp.826-852
- Sims, Christopher A., "But Economics Is Not an Experimental Science", *Journal of Economic Perspectives*, 24(2), pp.59-68
- Sims, David P., 2008, "Strategic Response to Class Size Reduction: Combination Classes and Achievement in California", *Journal of Policy Analysis and Management*, 27(3), pp.457-478
- , 2009, "Crowding Peter to Educate Paul: Lessons from a Class Size Reduction Externality", *Economics of Education Review*, 28(4), pp.465-473
- Urquiola, Miguel, 2006, "Identifying Class Size Effects in Developing Countries: Evidence from Rural Bolivia", *Review of Economics and Statistics*, 88, pp.171-177
- Woessmann, L. and Martin R. West, 2006, "Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS", *European Economic Review*, 50(3), pp.695-736
- 山崎博敏・藤井宣彰・水野考「学級規模と指導方法が小学生の学力に及ぼす影響：共分散構造分析とマルチレベルモデル分析の適用」『広島大学大学院教育学研究科紀要』第三部 58号, 2009, 9-16 頁