

# クラスサイズと学業成績および情緒的・行動的問題の因果関係

——自然実験デザインとマルチレベルモデルによる検証——

伊藤大幸\* 浜田 恵\*\* 村山恭朗\*\*\* 高柳伸哉\*\*\*\*  
野村和代\*\*\*\*\* 明翫光宜\*\*\*\*\* 辻井正次\*\*\*\*\*

クラスサイズ（学級の人数）が学業成績および情緒的・行動的問題に及ぼす影響について、要因の交絡とデータの階層性という2つの方法論的問題に対処した上で検証した。第1に、学年ごとの人数によってのみクラスサイズが決定されている学校を調査対象とする自然実験デザインにより、学校の裁量に起因する要因の交絡や逆方向の影響の発生を防いだ。第2に、マルチレベルモデルの一種である交差分類（cross-classified）モデルを用いて、データの特殊な階層性を適切にモデル化した。第3に、学校内中心化によって学校間変動を除外することで、クラスサイズの純粋な学校内効果を検証するとともに、学校規模との交絡を回避した。9回の縦断調査で得られた小学4年生から中学3年生のデータ（11,702名、のべ45,694名、1,308クラス）に基づく分析の結果、クラスサイズの拡大は、(a) 学業成績を低下させること、(b) 教師からのサポートを減少させること、(c) 友人からのサポートや向社会的行動の減少をもたらすこと、(d) 抑うつを高めることが示された。こうした影響の広さから、クラスサイズは学級運営上、重大な意味を持つ変数であることが示された。

キーワード：クラスサイズ、学業成績、情緒的・行動的問題、マルチレベルモデル、マルコフ連鎖モンテカルロ法

## 問 題

教育政策や学校運営の在り方を論じる上で、クラスサイズ（学級の人数）が児童生徒の学業成績や情緒的・行動的問題に与える影響は中心的な論点の1つである。20世紀半ば以降、多くの先進諸国では、クラスの少人数化が進められてきている。わが国も例外ではなく、義務教育標準法が定めるクラスサイズの上限は、制定当初の1959年度には50人であったが、64年度には45人、80年度には40人、2011年度には小1のみ35人に引き下げられた。しかし、全国の小中学校におけるクラスサイズの平均は、小学校で27人、中学校で33人と、経済協力開発機構（Organization for Economic Co-operation and Development: OECD）に加盟する34か国における平均（初等教育で21人、前期中等教育で24人）を大きく上回っており（OECD, 2015）、国際的に見て、

わが国の小中学校におけるクラスサイズは大きいという現状にある。

### クラスサイズの影響

**学業成績への影響** クラスサイズの影響については、教育経済学の分野を中心に多くの研究が行われてきた。代表的な研究として、1980年代に米国のテネシー州で行われたSTAR（Student/Teacher Achievement Ratio）プロジェクトがある（Finn, Fulton, Zaharias, & Nye, 1989）。このプロジェクトでは、79校329クラスの児童（約6,000名）を、13-17人の少人数学級と22-26人の普通学級にランダムに割り当て、幼稚園から小学3年生までの4年間にわたり学業成績や学習態度への影響を検証した。その結果、(a) 少人数学級に割り当てられた児童では数学と読字のテスト得点が普通学級の児童よりも平均して約0.2SD高いこと（Krueger, 1999）、(b) この差は小学3年生まで維持されること（Krueger, 1999）、(c) 小学4年生で普通学級に戻された後、8年次（わが国の中学2年生にあたる）における追跡調査や大学の入学試験でも一定の差が残存すること（Krueger & Whitmore, 2001）などが示された。

国内では、クラスサイズの影響に関する実証的知見は少ない。特にSTARプロジェクトのような実験的手法を用いた研究は見られず、いずれの研究も観察的

\* 浜松医科大学  
〒431-3192 静岡県浜松市東区半田山1-20-1  
ito\_hiroyuki@pd5.so-net.ne.jp

\*\* 名古屋学芸大学

\*\*\* 神戸学院大学

\*\*\*\* 愛知東邦大学

\*\*\*\*\* 常葉大学

\*\*\*\*\* 中京大学

手法を用いている。結果は一貫しておらず、国立教育政策研究所(2002)、篠崎(2008)、Hojo & Oshio(2010)ではクラスサイズの有意な効果が見られず、Akabayashi & Nakamura(2014)では小6、中3の国語と数学のうち、小6の国語のみで有意な効果(10人の増加につき0.15 SDの低下)が示された。一方、妹尾・篠崎・北條(2013)では、小6、中3の国語と算数・数学の全てでクラスサイズの有意な効果が見られた(10人の増加につき0.02 SD-0.06 SDの低下)。こうした知見の交錯には、後述する方法論上の問題が影響していると考えられる。

**情緒・行動面への影響** STARプロジェクトをはじめ、クラスサイズに関する先行研究では、主に学業面への影響に焦点が当てられてきた。しかし、クラスサイズの拡大によって担任教員の目が個々の児童生徒に行き届かなくなるとすれば、その影響は必然的に学業面だけでなく、情緒・行動面にも及ぶと考えられる。クラスサイズが情緒・行動面に与える影響については、実証的知見が不足しているが、Finn, Pannozzo, & Achilles(2003)は、クラスサイズと情緒・行動面の関連に関する10の実証研究(いずれもアメリカで実施)をレビューし、クラスサイズの小ささは、反社会的行動の少なさと関連するが、向社会的行動との関連は不明瞭であると結論づけている。一方、情緒・行動面への影響について、国内での体系的な検討は見られない。

クラスサイズの情緒・行動面への影響に関する研究では、標準化テストの開発が進んでいる学業成績とは様相が異なり、研究によって扱う問題の内容や測定の方法が共通しておらず、エビデンスが比較可能な形で十分に集積されていない。実際、多くの研究では、信頼性・妥当性の確認された尺度ではなく、個別の質問項目によって情緒・行動的問題を測定しており、心理測定学的な観点からは望ましくない。

**本研究の目的** 以上を踏まえ、本研究では、学業成績(国語および算数・数学)に加え、情緒・行動的問題に対するクラスサイズの影響を検討する。これまでのクラスサイズ研究において学業以外の側面として取り上げられてきた概念を整理すると、(a)クラス内の対人関係の問題(友人・教師との関係)、(b)クラス内での援助行動(向社会的行動、ソーシャルサポート)、(c)メンタルヘルスの問題(内在化・外在化問題)の3つに大別できる。本研究では、(a)の指標として友人関係および教師との関係上の問題、(b)の指標として向社会的行動と友人および教師からのソーシャルサポート、(c)の指標として抑うつ(内在化問題)および攻撃性

(外在化問題)を設定するとともに、いずれも信頼性・妥当性の確認された尺度によって測定を行い、クラスサイズの影響を包括的・体系的に検証する。

### 方法論上の問題

クラスサイズの因果的効果<sup>1</sup>を検証する上では、少なくとも2つの重要な方法論的問題に対処する必要がある。1つは外部変数との交絡の問題(経済学分野では内生性バイアスと呼ばれている)、もう1つはデータの階層性の問題である。

**要因の交絡** 交絡とは、統計モデル内の独立変数と相関し、かつ、従属変数に影響を及ぼす外部変数が存在することによって、独立変数から従属変数への効果が過大または過小に推定されてしまうことである。わが国では、義務教育標準法によってクラスサイズの上限が規定されているが、その上限を上回らない範囲では各校の裁量によって柔軟なクラス編成を行うことが可能になっている。したがって、例えば、社会経済的背景にめぐまれた児童生徒が多い学区では、学級運営上の問題が生じにくいいため、大きいクラスサイズのままでクラス編成を行うといった状況や、教育熱意が高い学校では、教育効果を高めるために小さいクラスサイズでクラス編成を行うといった状況が起こりうる(妹尾・篠崎・北條, 2013)。社会経済的背景や教育熱意は、従属変数である学業成績や情緒的・行動的問題にも影響を与えうるため、これらの外部変数とクラスサイズの間で交絡が生じる可能性がある。

こうした外部変数との交絡の問題に関して、先行研究では様々な対処が行われてきた。最もエビデンスレベルの高い方法は、STARプロジェクトのように、児童生徒をランダムに大小それぞれのクラスサイズに割り当て、その効果を検証するという実験的手法である。この方法は、交絡変数の影響をほぼ完全に統制し、最も信頼度の高いエビデンスを得ることができるが、人為的に学業成績や情緒的・行動的問題に格差を生み出す可能性があるという倫理的な問題をはらんでおり、現代の研究倫理規範のもとでは実施が難しくなっている。

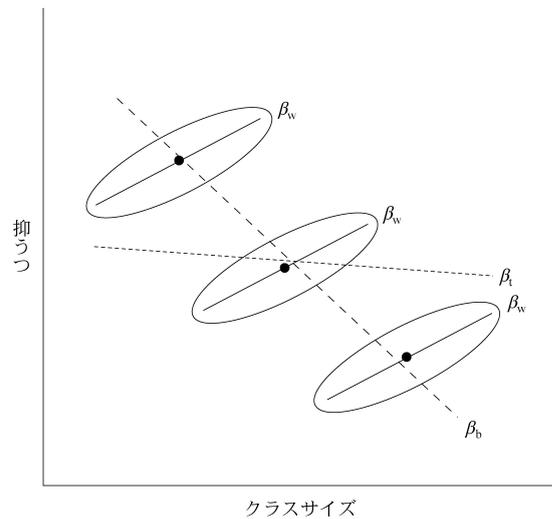
<sup>1</sup> 因果関係を示すための基準として、Lazarsfeld(1955)は、(a)原因が時間的に先行すること、(b)相関関係が存在すること、(c) bが第三の変数による疑似相関でないこと、という3つの基準を提唱している。これらの基準の他に、結果の外的妥当性(研究の結果をどの程度一般化しうるか)に関する点を基準として挙げるものもあるが、本研究では、内的妥当性(観察されたデータ内での因果関係の有無)に関する上記の3基準に基づいて「因果関係」や「因果的効果」の用語を使用する。

このような倫理的問題を避けるため、クラスサイズに関する近年の研究では、実験的介入以外の方法で外部変数の交絡を防ぐ工夫が用いられている。代表的な方法としては、法的または慣習的なクラスサイズの上限を利用した Angrist & Lavy (1999) の方法がある。例えば、法的なクラスサイズの上限が 40 人である場合、その上限にのみ基づいてクラス編成が行われれば、学年の人数が 40 人ならクラス数は 1 でクラスサイズは 40、41 人ならクラス数は 2 でクラスサイズの平均は 20.5 になることが予測される。この予測値を道具的変数として利用することで、上記のような学校の裁量によって生じる交絡を回避することができる。この方法は、原理的に、学校規模（学年ごとの人数）以外の外部変数との交絡を完全に統制することができるため、一種の自然実験（人為的なランダム割付を行わないものの、独立変数がほぼランダムに決定されているデザイン）と見なすことができる (Angrist & Pischke, 2010)。日本では、Akabayashi & Nakamura (2014) が、この方法を利用している。

本研究でも、学年ごとの人数によってのみクラスサイズが決定されている学校を調査対象とすることで、これらの研究と同様、学校の裁量に起因する交絡の問題を回避する。これにより、社会経済的背景や教育熱意などの外部変数との交絡だけでなく、学業成績や情緒的・行動的問題からクラスサイズへの逆方向の影響の発生（例：成績が低い学年だから、小さいクラスサイズで編成する）も防ぐことができる。

ただし、この自然実験デザインのもとでクラスサイズが学校規模（学年ごとの人数）によってのみ決まるとしても、学校規模そのものとの交絡の問題は残る。学校規模は、クラスサイズを規定する一方、学区内の教育産業の発展度合や世帯の社会経済的背景など、学業成績や情緒的・行動的問題に影響しうる多くの外部変数と関連を持つ可能性があり、クラスサイズに関連する中心的な交絡要因である。その上、クラスサイズは学校規模の複雑な非線形の関数であるため、通常重回帰モデル (Ordinary Least Squares regression; OLS) の共変量として学校規模を投入したとしても、学校規模との交絡を完全に防ぐことが難しいという問題がある。この問題への対処については次項の後半に述べる。

**データの階層性** もう 1 つの方法論的問題として、データの階層性の問題がある。一般に、学校で収集されるデータは、複数の個人が 1 つのクラスにネストされ（組み込まれ）、複数のクラスが 1 つの学校にネストされるという階層構造を成している。ここで、学業成



**Figure 1** 学校間効果および学校内効果に関する模式図（仮想データ；Raudenbush & Bryk, 2002 を参考に作成）

績や情緒的・行動的問題は、各個人間で変動を持つという意味で個人レベルの変数である。一方、クラスサイズは、クラス内の個人間では変動を持たず、クラス間で変動を持つという意味でクラスレベルの変数である。また、学校間でのみ変動を持つ学校レベルの変数としては、学校規模や学区の平均地価などがある。

こうした階層データでは、変数間の関連性をレベルごとに分けて扱う必要がある。Figure 1 に示した模式図の横軸はクラスサイズ、縦軸は抑うつを表し、楕円は各学校における 2 変数の散布図を表している（仮想データ）。この例では、それぞれの学校内の散布図における回帰直線の傾き（図の  $\beta_w$ ：添字の w は within を意味する）は正であるが、各学校の散布図の中心（図の黒丸）を結んだ回帰直線の傾き（図の  $\beta_b$ ：添字の b は between を意味する）は負となっている。前者は学校内でのクラスサイズの効果、後者は学校間でのクラスサイズの効果を表している。クラスサイズ研究におけるリサーチクエスチョンは、クラスサイズが平均的に大きい（小さい）学校の特徴を知ることではなく、クラスサイズが大きい（小さい）クラスの特徴を知ることにあるため、学校間効果ではなく学校内効果が本来の関心の対象となる。

しかし、データの階層性を適切に考慮しない分析では学校内効果を正しく検証することができない。例えば、学校単位の学力検査の平均値を使用した先行研究

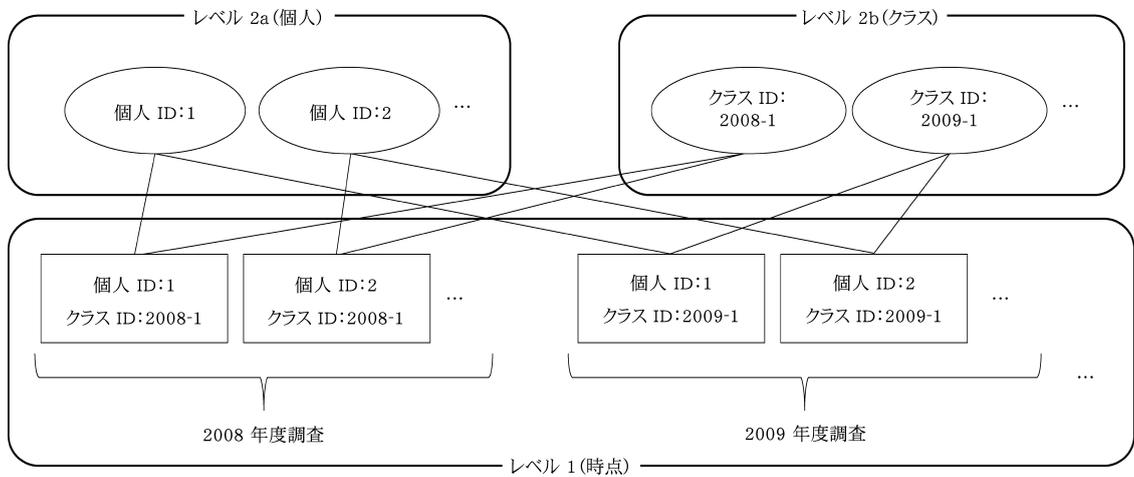


Figure 2 本研究のデータの階層構造

(Akabayashi & Nakamura, 2014 など)で行われているように、クラスサイズを学校内で平均するなどして学校レベルの変数として扱うと、学校間効果は推定できるが、学校内効果については検討できない。また、クラスが学校にネストされているという事実を考慮せず、同一学校内のデータと異なる学校のデータを均質なデータと見なして分析を行うと、標準誤差が過小推定されるだけでなく、得られる点推定値は学校間効果と学校内効果が混合した不明瞭な値 (図の  $\beta_t$ : 添字の  $t$  は total を意味する) となる (Enders & Tofighi, 2007)。特に Figure 1 のように学校間効果と学校内効果が著しく異なるケースでは、 $\beta_w$  の代替として  $\beta_o$  や  $\beta_i$  を用いることは重大な結論の誤りにつながりうる。したがって、クラスサイズの効果を正確に検証するには、データの階層性を適切に考慮した分析によって、学校間効果と学校内効果を分離して推定する必要がある。これまでのクラスサイズ研究では、専ら階層性を考慮しない OLS が用いられており、こうした問題について適切な対処が行われていない。

学校間効果と学校内効果を分離することは、前項に述べた要因の交絡の問題を回避する上でも有効な方法である。前述のように、クラスサイズは学校規模の複雑な関数であるため、OLS で学校規模との交絡を完全に防ぐことは難しい。しかし、学校規模は学校レベルの変数であり、学校内では変動を持たないため、学校内効果との交絡は生じえない。したがって、学校内効果を学校間効果と分離して推定することで、学校規模との交絡を回避し、クラスサイズの効果についてより正確な検証を行うことが可能になる。

そこで本研究では、階層構造を成すデータについてレベルごとにモデリングを行うマルチレベルモデルの手法を用いてクラスサイズの効果を検証する。ところが、本研究のデータは Figure 2 に示したような特殊な階層構造を成している。本研究では、ある中規模都市の全小中学校における 9 年分のパネル調査データを用いて、各年度のクラスサイズと学業成績および情緒的・行動的問題の横断的な関連を検証する。縦断データを使用するが、先行研究との比較のため、時点間の縦断的な関連ではなく時点内の横断的な関連のみを検証する。通常、学校データは「個人」レベル、「クラス」レベル、「学校」レベルの 3 レベル構造を成しているが、本研究では、同一の個人に複数回の測定を行っているため、個人レベルの下位に「時点」レベルを想定する必要がある。また、調査年度ごとに、クラス替え、転校、卒業・入学などによってクラスの構成が変化するため、「個人」レベルの上位に「クラス」レベルを位置づけることができない。

このような特殊な構造のデータに適用できるモデルとして交差分類 (cross-classified) モデル (Goldstein, 1994) がある。マルチレベルモデルの一種である交差分類モデルは、完全な階層関係が保たれない 2 つの属性 (本研究では個人とクラス) を並列的に扱うモデルであり、本研究のようなデータ構造にも対応できる。

交差分類モデルを用いることで、階層関係が保たれない「個人」レベルと「クラス」レベルを適切にモデル化することができるが、現在のところ、交差分類モデルで 3 レベル構造を扱う方法は考案されていないため、「クラス」レベルの上位に「学校」レベルを設定

することができない。したがって、このモデルではクラスサイズの学校内効果と学校間効果を分離して推定することができない。しかし、上記のように、クラスサイズ研究における関心の対象は学校内効果であり、学校間効果は本来検討が必要な対象ではない。そこで本研究では、独立変数であるクラスサイズと従属変数である学業成績および情緒的・行動的問題の双方について、学校内で中心化（観測値から学校内の平均値を減算する処理）を行うことで学校間変動を取り除くこととした。模式的には、Figure 1の各学校の散布図を全て中央に集めた上で学校内効果 ( $\beta_w$ ) のみを推定するという方法を取るようになる。こうして学校間変動を除外することで、本来の関心の対象である学校内効果を正確に推定できるだけでなく、「学校」レベルの変数である学校規模との交絡の問題も回避できる。

**本研究のアプローチ** 以上を要約すると、本研究では、第1に、学年ごとの人数によってのみクラスサイズが決定されている学校を調査対象とする自然実験デザインによって、学校の裁量に起因する要因の交絡や逆方向の因果関係の発生を防ぐ。第2に、マルチレベルモデルの一種である交差分類モデルを用いることで、反復測定データの特殊な階層性を適切にモデル化する。第3に、学校内中心化によって学校間変動を除外することで、クラスサイズの純粋な学校内効果を検証するとともに、学校規模との交絡を回避する。上記の第1の方法で学校規模以外の要因との交絡や逆方向の因果関係を防ぎ、第3の方法で学校規模との交絡も回避することで、本研究では事実上、STARプロジェクトなどの実験的研究と同様に、クラスサイズと学業成績および情緒的・行動的問題の因果関係を直接的に検証することが可能となる。また、第2の方法により複数年度の調査データをプールして分析することで、十分なサンプルサイズのもとで因果関係の強度を正確に推定することができる。

## 方 法

### 参加者

本研究は継続中のコホート研究プロジェクトの一部である。このプロジェクトでは、児童青年のメンタルヘルスと問題行動のメカニズムを明らかにすることを目的として、中部地域に位置する1つの中規模都市の全ての保育所と公立小中学校を対象とした悉皆調査を2007年度から縦断的に実施している。子どもの個人特性（発達障害特性、知能、気質など）および環境要因（家庭、友人、教師など）の相互作用が、思春期以降に深刻

**Table 1** 個人レベルのサンプルの内訳

入学年度	男子	女子	合計
2000	376	351	727
2001	381	373	754
2002	418	403	821
2003	449	402	851
2004	455	400	855
2005	486	488	974
2006	462	439	901
2007	546	493	1039
2008	509	502	1011
2009	529	450	979
2010	443	494	937
2011	480	435	915
2012	461	477	938
合計	5995	5707	11702

化するメンタルヘルス上の問題（抑うつ、攻撃性、不登校、自傷行為、いじめ、非行など）に及ぼす影響について検証している。調査対象市は大都市への通勤可能圏内であると同時に、工業、農業が盛んであり、都市で勤務する家庭や、地方型の勤務家庭など、多様な社会経済的状態の家庭が含まれている。人口は約9万人であり、政令市と東京23区を除く全国の市区平均（約10万人）にほぼ等しい。

本研究では、2007年度から2015年度までの計9回の調査における小学4年生から中学3年生（計11,702名、のべ45,694名、1,308クラス）のデータを分析に使用した（個人レベルの内訳をTable 1、時点レベルの内訳をTable 2、クラスレベルの内訳をTable 3に示した）。ただし、変数によって調査年度が異なった（各変数の調査年度は「尺度」の項を参照）。調査対象校は、小学校が9校、中学校が4校であった。調査時期は各年度の9月（学業成績のみ5月）であり、各年度の在籍児童生徒における有効回答率は96.8—97.5%であった。

クラスサイズの総平均は小学校で33.99、中学校で35.17であった。本研究の調査対象校では、義務教育標準法と県が独自に定める基準に基づいて、学年ごとの人数によってのみ自動的にクラスサイズが決定されていた。2009年度に定められた県の基準では、中1のクラスサイズの上限が35人と定められたため、2009年度以降、中1のクラスサイズがやや減少している。なお、全ての調査対象校では、年度ごとにクラス替えを行っているため、各クラスのメンバー構成は年度間で異なっていた。各校の学校規模（全学年の児童生徒数）は、小学校で358—1,149名（ $M=609.56$ ,  $SD=227.06$ ）、中学校で503—952名（ $M=693.50$ ,  $SD=171.57$ ）

Table 2 時点レベルのサンプルの内訳

入学 年度	2007年度		2008年度		2009年度		2010年度		2011年度		2012年度		2013年度		2014年度		2015年度		合計
	学年	n																	
2000	中2	702	中3	721															1423
2001	中1	737	中2	745	中3	746													2228
2002	小6	780	中1	775	中2	776	中3	818											3149
2003	小5	811	小6	801	中1	799	中2	847	中3	795									4053
2004	小4	804	小5	800	小6	795	中1	841	中2	787	中3	801							4828
2005			小4	897	小5	897	小6	939	中1	879	中2	881	中3	885					5378
2006					小4	830	小5	867	小6	835	中1	841	中2	850	中3	854			5077
2007							小4	987	小5	943	小6	943	中1	943	中2	950	中3	924	5690
2008									小4	932	小5	933	小6	933	中1	1010	中2	916	4724
2009											小4	929	小5	939	小6	972	中1	908	3748
2010													小4	898	小5	934	小6	882	2714
2011															小4	913	小5	881	1794
2012																	小4	888	888
合計		3834		4739		4843		5299		5171		5328		5448		5633		5399	45694

であった（2015年度時点；他年度もおおむね同様）。

毎年の調査に先立って、児童生徒および保護者に対して、調査への回答は任意であり、回答しないことによる不利益は生じないことを説明文書によって教示し、研究協力の同意を得た。本研究の手続きは、第1著者の所属機関の倫理委員会の審査と承認を受けた（2007年3月承認，第18-166号）。

#### 尺度

**学業成績** 学業成績の評価には、教研式標準学力検査 NRT（集団基準標準検査；辰野・石田・服部・筑波大学附属小各科教官，2002）を使用した。集団式の学力検査である NRT は、信頼性・妥当性が確認されており、日本全国の教育現場で広く使用されている。本研究では国語および算数の偏差値を使用した。偏差値は NRT の標準化サンプルにおける学年ごとの得点分布をもとに算出された。学業成績については 2009 年度以降の調査（7 回）のみデータを収集した。

**対人関係** 対人関係のうち、友人関係上の問題（以後、友人関係問題）の評価には Strengths and Difficulties Questionnaire (SDQ) 日本語版の自己評定フォーム (Goodman, 1997) の「友人問題」の下位尺度（5 項目）を使用し（項目例：他の子から、いじめられたり、からかわれたりする）、教師との関係における問題（教師関係問題）の評価には小中学生用社会的不適応尺度（伊藤他，2015）の「教師との関係」の下位尺度（4 項目）を使用した（項目例：先生に嫌われていると感じる）。いずれも自己評定式の尺度であり、国内で信頼性・妥当性が確認されている（原田他，2014；伊藤他，2015）。友人関係問題については 2011 年度以降（5 回）、教師関係問題については 2012 年度以降（4 回）の調査のみデータを収集し

た。

**向社会的行動** 向社会的行動の評価には SDQ 日本語版の自己評定フォーム (Goodman, 1997) の「向社会的行動」の下位尺度（5 項目）を使用した（項目例：誰かが心を痛めていたり、落ち込んでいたり、嫌な思いをしているときなど、すすんで助ける）。向社会的行動については 2011 年度以降の調査（5 回）のみデータを収集した。向社会的行動については 2011 年度以降（5 回）の調査のみデータを収集した。

**ソーシャルサポート** 村山他（2016）の小中学生用ソーシャルサポート尺度を用いた。自己評定式の尺度であり、国内で信頼性・妥当性が確認されている（村山他，2016）。周囲の友人からのサポート（以後、友人サポート）を測定する「友人からのサポート」（項目例：あなたの友達は、あなたが助けてほしいときに力になってくれますか）と、親、教師などの周りの大人からのサポート（大人サポート）を測定する「大人からのサポート」（項目例：あなたの周りの大人の人は、あなたが助けてほしいときに力になってくれますか）の 2 下位尺度（各 6 項目）から各 3 項目ずつ計 6 項目（項目 1, 2, 5, 7, 8, 11）を選択して使用した。ソーシャルサポートについては 2011 年度以降の調査（5 回）のみデータを収集した。

**メンタルヘルス** 抑うつ の尺度として Birleson Depression Self-Rating Scale for Children (DSRS-C) 日本語版（村田・清水・森・大島，1996）の短縮版（並川他，2011）、攻撃性の尺度として Buss-Perry Aggression Questionnaire の日本語版である Hostility-Aggression Questionnaire for Children (HAQ-C; 坂井他，2000) の短縮版を用いた。いずれも自己評定式の尺度であり、国内で信頼性・妥当性が確認されている（浜田他，2016；並

Table 3 クラスレベルのサンプルの内訳

調査年度	学年	クラス数	クラスサイズ			調査年度	学年	クラス数	クラスサイズ		
			M	SD	range				M	SD	range
2007	小4	25	32.12	5.53	23-40	2012	小4	28	33.21	2.85	27-37
	小5	24	33.63	4.43	22-40		小5	28	33.21	3.95	27-38
	小6	24	32.42	3.89	27-40		小6	27	34.89	4.24	25-39
	中1	21	34.95	1.86	32-38		中1	27	31.19	1.30	29-33
	中2	20	34.75	1.55	32-38		中2	24	36.79	0.98	35-39
	中3	—	—	—	—		中3	22	36.55	2.79	32-41
	全体	114	33.48	3.99	22-40		全体	156	34.17	3.53	25-41
2008	小4	26	34.27	3.75	25-39	2013	小4	26	34.04	4.31	28-41
	小5	25	32.04	5.62	23-40		小5	28	33.04	3.55	25-38
	小6	24	33.21	4.44	21-40		小6	28	32.82	3.75	27-37
	中1	21	36.90	2.10	34-40		中1	29	32.07	2.07	29-35
	中2	21	35.52	1.89	33-39		中2	23	36.48	0.51	36-37
	中3	20	36.80	1.32	35-39		中3	24	36.25	1.29	34-38
	全体	137	34.64	4.02	21-40		全体	158	33.97	3.37	25-41
2009	小4	23	36.17	3.43	28-41	2014	小4	26	33.81	3.93	27-39
	小5	26	34.62	4.12	25-42		小5	26	33.69	4.45	28-40
	小6	25	31.88	5.59	22-39		小6	28	33.04	3.63	24-38
	中1	24	33.33	1.49	31-37		中1	28	32.68	1.61	31-35
	中2	21	37.05	2.31	33-41		中2	25	36.84	1.60	34-39
	中3	21	35.62	1.83	32-38		中3	23	36.43	0.59	36-38
	全体	140	34.68	3.87	22-42		全体	156	34.32	3.38	24-40
2010	小4	26	36.81	3.93	25-40	2015	小4	26	34.12	2.83	29-40
	小5	24	35.25	5.07	22-40		小5	26	33.62	4.06	26-40
	小6	25	36.20	3.79	26-41		小6	26	33.77	4.74	27-40
	中1	25	31.52	1.42	29-33		中1	27	33.74	1.46	31-36
	中2	23	34.61	1.12	33-36		中2	24	37.75	1.67	35-40
	中3	21	36.71	2.24	33-40		中3	25	36.48	1.81	33-39
	全体	144	35.16	3.72	22-41		全体	154	34.86	3.39	26-40
2011	小4	27	34.59	4.64	27-40	全体	小4	233	34.33	4.14	23-41
	小5	26	36.46	3.56	25-39		小5	233	33.94	4.43	22-42
	小6	24	34.92	5.13	22-39		小6	231	33.68	4.50	21-41
	中1	28	31.43	1.23	29-33		中1	230	32.93	2.33	29-40
	中2	21	37.48	1.91	35-41		中2	202	36.39	1.88	32-41
	中3	23	34.57	1.20	33-36		中3	179	36.17	1.85	32-41
	全体	149	34.78	3.83	22-41		全体	1308	34.47	3.69	21-42

川他, 2011)。DSRS-C 短縮版は「抑うつ気分」, 「活動性および楽しみの減退」の2つの下位尺度から抽出した9項目から構成される(項目例: とても悲しい気がする)。HAQ-C 短縮版は, HAQ-C の「身体的攻撃」, 「短気」, 「敵意」の3下位尺度から抽出した8項目によって構成される(項目例: すぐにケンカをしてしまう)。抑うつ・攻撃性については全年度の調査(9回)でデータを収集した。

#### 分析

本研究では, 調査年度ごとのクラスサイズと学業成績および情緒的・行動的問題の横断的な関連を検討した。ただし, 学業成績についてはクラス編成直後の5月に測定が行われたため, 前年度のクラスサイズとの関連を検証した。情緒的・行動的問題については9月

の測定値と当該年度のクラスサイズの関連を検討した。

9年分の調査データについて, 個人(レベル2a)およびクラス(レベル2b)をクラスター変数とする交差分類モデルに基づいて分析を行った。交差分類モデルでは, 以下のように従属変数が3つのレベルの成分に分解される。

$$Y_{jk} = Y_{1jk} + Y_{2aj} + Y_{2bk}$$

ここで,  $Y_{jk}$  はクラス  $k$  の個人  $j$  における従属変数の観測値である。また,  $Y_{2aj}$  は個人レベルのランダム切片(個人  $j$  における平均的な従属変数の水準),  $Y_{2bk}$  はクラスレベルのランダム切片(クラス  $k$  における平均的な従属変数の水準),  $Y_{1jk}$  は観測値から個人およびクラスレベルのランダム効果を除いた時点レベルの成分を表す

潜在変数である。これら3つのレベルの成分について、個別に回帰式が設定される。本研究では、以下の回帰式を設定した。

レベル1 (時点)

$$Y_{1jk} = \gamma_{00} + \varepsilon_{1jk}$$

レベル2 a (個人)

$$Y_{2aj} = \gamma_{2a1} \text{性別\_女子}_j + \gamma_{2a2} \text{入学年度\_2000}_j + \dots \\ + \gamma_{2a13} \text{入学年度\_2012}_j + \varepsilon_{2aj}$$

レベル2 b (クラス)

$$Y_{2bk} = \gamma_{2b1} \text{クラスサイズ}_k + \gamma_{2b2} \text{学年\_小5}_k + \dots \\ + \gamma_{2b6} \text{学年\_中3}_k + \varepsilon_{2bk}$$

ここで  $\gamma_{00}$  は切片、 $\gamma_{2a1} - \gamma_{2a13}$  はレベル2 a (個人) の各独立変数の偏回帰係数 (傾き)、 $\gamma_{2b1} - \gamma_{2b6}$  はレベル2 b (クラス) の各独立変数の偏回帰係数 (傾き)、 $\varepsilon_{1jk}$ 、 $\varepsilon_{2aj}$ 、 $\varepsilon_{2bk}$  は各レベルの残差を表す。

本研究の調査対象校では、学年ごとの人数によってのみクラスサイズが決定されており、レベル1 (時点) の変数がクラスサイズと交絡する可能性が排除されているため、レベル1の独立変数は設定しなかった。レベル2 a (個人) の独立変数 (共変量) としては、性別および学年コホート (入学年度) を設定した。性別は男子を0、女子を1とコーディングしたダミー変数を分析に使用した。学年コホート (入学年度) は、最も人数の多い2007年度を基準カテゴリとし、その他の入学年度のダミー変数を作成し、分析に使用した。レベル2 b (クラス) の独立変数には、クラスサイズに加え、共変量として学年を設定した。学年については、小4を基準カテゴリとし、その他の学年のダミー変数を作成し、分析に使用した。学校間変動を除外するため、クラスサイズと各従属変数はあらかじめ学校内で中心化した<sup>2</sup>。具体的には、個々のデータから各年度の学校内の平均値を減算した。ただし、比較のために、学校内中心化をしていないデータでも同様に分析を行った。推定値の解釈を容易にするため、各従属変数はさらに個人レベルの平均値と標準偏差によって標準化 (z得点化) した。

<sup>2</sup> 通常のマルチレベル分析において、学校をクラスター変数として設定するのであれば、学校ごとのランダム切片 (各学校の平均的な従属変数の水準) が考慮されるため、従属変数について中心化を行う必要はない。しかし、今回の分析では、交差分類モデルの制約のため、学校をクラスター変数として設定できず、学校ごとのランダム切片がモデルに含まれないため、あらかじめ学校内中心化によって学校間の従属変数の変動を取り除いた。

分析には Mplus 7.31 (Muthén & Muthén) の交差分類モデルオプション (ANALYSIS: TYPE=CROSSCLASSIFIED) を使用し、マルコフ連鎖モンテカルロ法 (MCMC) に基づくベイズ推定 (ESTIMATOR=BAYES) によりパラメータを推定した (Asparouhov & Muthén, 2010, 2014)。交差分類モデルは、従来の最尤法などのアプローチでは計算が困難であり (Asparouhov & Muthén, 2010)、代替手段として MCMC に基づくベイズ推定を用いた方法が提案されている (Browne, Goldstein, & Rasbash, 2001)。MCMC のアルゴリズムは Gelman, Carlin, Stern & Rubin (2003) の Gibbs サンプラーを用い、生成チェーン数は2とした。収束指標には Potential Scale Reduction を用い、収束基準は1.05とした。ただし、最小反復回数を1,000とした。事前分布は無情報分布 (平均や固定効果は正規分布、分散・共分散は逆 Wishart 分布) を用いた。各チェーンの前半の半数の反復はバーンイン期間として除外され、後半の半数の反復に基づいて事後分布が推定された。

なお、小学校と中学校でクラスサイズの効果が異なるかを検討するため、初めに小学校と中学校のデータを分けて分析を行ったが、クラスサイズの効果の推定値はおおむね同様の数値を示した。また、小・中学生の別を表すダミー変数とクラスサイズの交互作用を検討したが、いずれの従属変数についても有意な交互作用は見られなかったため、小・中学生のデータをプールして分析を行った。

## 結 果

### 記述統計

Table 4 に各変数の記述統計量と級内相関係数を示した。ただし、標準偏差は交差分類モデルにおける各レベルの分散の推定値をもとに算出したため、レベル1 (時点) 変数の標準偏差は、3つのレベルに分解されている。級内相関係数は、各変数の全分散に占める各レベルの分散の割合を意味する。学校内中心化の有無で比較すると、レベル2 b (クラス) の標準偏差と級内相関係数が学校内中心化によって縮小していることが見て取れる。これはクラス間変動に混在していた学校間変動が学校内中心化によって取り除かれたことを意味している。

各変数の級内相関係数を見ると、全体にレベル2 a (個人) の級内相関係数が多いが、とりわけ学業成績の2変数 (国語, 算数・数学) が大きい値を示しており、他の変数よりも個人内の経時的な安定性が高いことがわかる。一方、レベル2 b (クラス) の級内相関係

Table 4 各変数の記述統計量

M	SD						級内相関係数				範囲	
	ローデータ			学校内中心化			ローデータ		学校内中心化			
	レベル1 (時点)	レベル2a (個人)	レベル2b (クラス)	レベル1 (時点)	レベル2a (個人)	レベル2b (クラス)	レベル2a (個人)	レベル2b (クラス)	レベル2a (個人)	レベル2b (クラス)		
レベル1 (時点) 変数												
国語	52.09	4.19	8.54	1.05	4.20	8.52	0.76	.796	.012	.799	.006	15-79
算数・数学	52.48	4.51	9.37	1.64	4.53	9.35	1.15	.792	.024	.800	.012	5-78
友人関係問題	1.35	1.16	1.13	0.32	1.16	1.13	0.26	.470	.037	.473	.026	0-10
教師関係問題	0.99	1.18	0.89	0.39	1.18	0.89	0.38	.341	.067	.342	.061	0-8
向社会的行動	6.13	1.36	1.42	0.31	1.36	1.42	0.30	.509	.025	.510	.023	0-10
友人サポート	9.21	1.44	1.44	0.30	1.44	1.43	0.26	.488	.021	.489	.016	3-12
大人サポート	9.07	1.57	1.58	0.42	1.58	1.57	0.34	.486	.035	.486	.022	3-12
抑うつ	4.52	2.08	2.19	0.52	2.08	2.18	0.37	.510	.029	.515	.015	0-18
攻撃性	16.67	3.32	4.05	0.79	3.32	4.04	0.75	.585	.022	.585	.020	8-32
レベル2b (クラス) 変数												
クラスサイズ	34.47		3.69			2.64						21-42

数は全体的に小さいが、教師関係問題は比較的大きな値を示しており、他の変数よりもクラス間の変動が大きい(クラス内の相関が高い)ことが見て取れる。これはクラスによって担任教師が異なる(クラス内では同じ)ことを考えれば自然な結果である。

#### 交差分類モデル

交差分類モデルに基づくマルチレベル分析によるパラメータ推定値をTable 5(学校内中心化なし)とTable 6(学校内中心化あり)に示した<sup>3</sup>。学校内中心化なしの分析では、全般的にクラスサイズの係数の絶対値が学校内中心化ありの分析より小さく、有意な効果が見られたのは算数・数学のみであった。

一方、学校内中心化によって学校間変動を除外し、純粋な学校内効果を推定した場合、国語、算数・数学、

向社会的行動、友人サポート、大人サポートの5変数でクラスサイズの効果が有意であり、ポジティブな変数(国語、算数・数学、向社会的行動、友人サポート、大人サポート)には負の効果、ネガティブな変数(抑うつ)には正の効果を示した。非標準化係数の値から、クラスサイズが10人増加するごとに、これらの従属変数が0.07-0.17 SD、望ましくない方向に変化することが示された。

## 考 察

#### 方法論的考察

学校内中心化なしの分析では、学校内中心化を行った場合よりも全般的にクラスサイズの効果の推定値が小さかった。この背景には学校規模との交絡の問題がある。学校規模が大きくなるほど少人数学級は生じにくく、クラスサイズは大きくなりやすい。したがって、もしクラスサイズが学業成績や情緒的・行動的問題に対して望ましくない効果を持つ一方で、学校規模がこれらの変数に望ましい方向の効果を持つとすれば、クラスサイズの効果は学校規模との交絡によって相殺され、実際よりも小さく推定されることになる。クラスサイズは学校規模の複雑な非線形の関数であるため、従来の研究(Akabayashi & Nakamura, 2014; Angrist & Lavy, 1999)のように学校規模を共変量として投入しても、交絡を完全に防ぐことはできない。

本研究では学校内中心化によって学校間変動を除外し、クラスサイズの純粋な学校内効果を推定することで、このような学校規模との交絡の問題を回避した。また、本研究の調査対象校では、学年ごとの人数にのみ基づいてクラスサイズが決定されていたため、学校

<sup>3</sup> 本研究のモデルを含む線形回帰モデルでは、従属変数が正規分布に従うことが仮定されている。この仮定を検証するため、各変数のクラスレベルにおける歪度および尖度を算出した。歪度・尖度に関して統一的な基準はないが、シミュレーションに基づいて、歪度は絶対値で2以下、尖度は絶対値で7以下であれば、推定結果に大きな影響を及ぼさないという知見が報告されている(West, Finch, & Curran, 1995)。各変数の学年ごとの歪度の範囲は、国語で-0.63-0.28, 算数・数学で-1.00-0.40, 友人関係問題で-0.07-0.56, 教師関係問題で0.08-1.33, 向社会的行動で-0.31-0.15, 友人サポートで-0.56-0.09, 大人サポートで-0.56-0.09, 抑うつで-0.04-0.39, 攻撃性で0.04-0.60であった。また、学年ごとの尖度の範囲は、国語で-0.51-0.83, 算数・数学で-0.04-2.76, 友人関係で-0.48-0.42, 教師関係問題で-0.71-3.43, 向社会的行動で-0.68-0.77, 友人サポートで-0.48-0.42, 大人サポートで-0.62-0.23, 抑うつで-0.32-0.57, 攻撃性で-0.28-1.10であった。いずれの変数も、上記の基準内の歪度・尖度に留まっており、推定結果に顕著な影響を及ぼすほどの非正規性はないと判断した。

**Table 5** 交差分類モデルにおけるパラメータ推定値 (学校内中心化なし)

	国語				算数・数学				友人関係問題			
	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>
レベル 2a (個人)												
性別	.362	.023	.182	<.001	-.026	.022	-.012	.250	-.217	.027	-.108	<.001
レベル 2b (クラス)												
学年 (基準: 小4)												
小5	-.041	.017	-.133	.020	-.002	.022	-.004	.910	-.249	.030	-.330	<.001
小6	-.002	.016	-.005	.874	.066	.019	.140	<.001	-.441	.031	-.586	<.001
中1	.143	.019	.465	<.001	.261	.019	.558	<.001	-.548	.034	-.744	<.001
中2	.142	.016	.426	<.001	.267	.023	.524	<.001	-.588	.036	-.742	<.001
中3									-.490	.039	-.615	<.001
クラスサイズ	-.002	.002	-.062	.334	-.007	.002	-.141	<.001	.004	.004	.053	.266
	教師関係問題				向社会的行動				友人サポート			
	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>
レベル 2a (個人)												
性別	-.388	.028	-.192	<.001	.618	.022	.309	<.001	.728	.027	.363	<.001
レベル 2b (クラス)												
学年 (基準: 小4)												
小5	-.058	.066	-.050	.380	-.126	.034	-.188	<.001	.117	.032	.211	<.001
小6	-.093	.065	-.078	.150	-.118	.034	-.173	<.001	.245	.033	.435	<.001
中1	-.140	.066	-.123	.040	-.260	.034	-.395	<.001	.316	.029	.573	<.001
中2	.370	.073	.300	<.001	-.392	.036	-.557	<.001	.343	.032	.591	<.001
中3	.137	.078	.113	.060	-.324	.037	-.459	<.001	.296	.035	.510	<.001
クラスサイズ	-.002	.007	-.018	.680	-.004	.004	-.048	.200	-.006	.003	-.099	.080
	大人サポート				抑うつ				攻撃性			
	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>
レベル 2a (個人)												
性別	.401	.025	.199	<.001	.235	.022	.117	<.001	-.348	.021	-.173	<.001
レベル 2b (クラス)												
学年 (基準: 小4)												
小5	-.018	.029	-.023	.600	.027	.022	.042	.300	.058	.021	.116	<.001
小6	-.088	.030	-.114	<.001	.022	.022	.036	.260	.056	.022	.112	.014
中1	-.215	.032	-.280	<.001	.189	.023	.296	<.001	.011	.022	.022	.614
中2	-.514	.036	-.632	<.001	.390	.023	.584	<.001	-.022	.023	-.041	.414
中3	-.574	.034	-.704	<.001	.571	.023	.810	<.001	-.114	.025	-.206	<.001
クラスサイズ	-.006	.004	-.074	.060	-.002	.002	-.033	.350	-.002	.002	-.048	.286

注) *B* および *SD* は、事後分布における非標準化推定値の平均値と標準偏差、 $\beta$  は標準化推定値の平均値を表す。*p* は事後分布における片側 *p* 値 (正の推定値については事後分布の 0 以下の面積の割合、負の推定値については事後分布の 0 以上の面積の割合) を 2 倍した値を示しており、この値が .05 を下回る場合、95%信用区間が 0 を含まないことを意味する。推定値の解釈を容易にするため、各従属変数は個人レベルで標準化した。性別は男子を 0、女子を 1 としてコーディングした。レベル 2a には学年コホートをダミー変数として投入したが、スペースの節約のためパラメータ推定値を表示していない。

規模以外の要因との交絡や従属変数からの逆方向の因果関係も生じえない。したがって、本研究では実験的研究と同様に、独立変数であるクラスサイズの外生性を確保し、その因果的効果を直接検証することができた。その結果、クラスサイズは学業成績や情緒的・行動的問題に全般的に望ましくない効果を持つことが示された。従来の研究では、クラスサイズの有意な効果が見られていない研究 (Hojo & Oshio, 2010; 国立教育政策研究所, 2002; 篠崎, 2008) や、限定的な効果しか見られて

いない研究 (Akabayashi & Nakamura, 2014) も多数存在するが、その主たる原因はこうした方法論上の問題にあると考えられる。

**学業成績への効果**

クラスサイズが 10 人増加するごとに、平均して国語の得点が 0.07 *SD*、算数・数学の得点が 0.10 *SD* 低下することが示された。STAR プロジェクトでは、普通学級より平均して 7 人少ない少人数学級で読字と数学の合成得点 (小1-小3) が 0.22 *SD* 高く (Krueger,

Table 6 交差分類モデルにおけるパラメータ推定値 (学校内中心化あり)

	国語				算数・数学				友人関係問題			
	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>
レベル 2a (個人)												
性別	.365	.021	.185	<.001	-.023	.022	-.011	.280	-.214	.027	-.107	<.001
レベル 2b (クラス)												
学年 (基準: 小 4)												
小 5	-.038	.016	-.172	.020	.007	.020	.020	.740	-.230	.028	-.479	<.001
小 6	.015	.016	.068	.430	.095	.017	.280	<.001	-.394	.026	-.823	<.001
中 1	-.015	.017	-.066	.380	.014	.019	.040	.510	-.228	.030	-.484	<.001
中 2	.015	.015	.064	.330	.051	.022	.145	.010	-.242	.033	-.480	<.001
中 3									-.138	.033	-.268	<.001
クラスサイズ	-.007	.002	-.218	<.001	-.010	.003	-.197	<.001	.003	.005	.049	.426
	教師関係問題				向社会的行動				友人サポート			
	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>
レベル 2a (個人)												
性別	-.388	.028	-.193	<.001	.614	.023	.310	<.001	.726	.027	.364	<.001
レベル 2b (クラス)												
学年 (基準: 小 4)												
小 5	-.038	.059	-.037	.520	-.125	.026	-.296	<.001	.116	.028	.297	<.001
小 6	-.078	.058	-.076	.180	-.143	.028	-.341	<.001	.230	.028	.596	<.001
中 1	-.263	.061	-.263	<.001	.013	.029	.030	.700	.137	.032	.361	<.001
中 2	.179	.065	.167	.020	-.151	.031	-.341	<.001	.178	.025	.442	<.001
中 3	-.048	.070	-.044	.550	-.109	.033	-.240	<.001	.094	.029	.242	<.001
クラスサイズ	.015	.008	.099	.070	-.007	.004	-.119	.015	-.015	.005	-.277	<.001
	大人サポート				抑うつ				攻撃性			
	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>	<i>B</i>	<i>SD</i>	$\beta$	<i>p</i>
レベル 2a (個人)												
性別	.402	.025	.201	<.001	.229	.008	.307	<.001	-.345	.020	-.174	<.001
レベル 2b (クラス)												
学年 (基準: 小 4)												
小 5	-.019	.025	-.036	.550	.021	.021	.058	.340	.074	.017	.214	<.001
小 6	-.096	.026	-.184	<.001	.017	.022	.048	.480	.101	.018	.289	<.001
中 1	.174	.030	.341	<.001	-.138	.024	-.369	<.001	.076	.020	.217	<.001
中 2	-.109	.033	-.201	<.001	.030	.023	.077	.154	.070	.018	.188	<.001
中 3	-.200	.029	-.362	<.001	.194	.023	.464	<.001	.025	.020	.066	.234
クラスサイズ	-.017	.004	-.227	<.001	.007	.003	.123	.006	-.002	.003	-.048	.454

注) *B* および *SD* は、事後分布における非標準化推定値の平均値と標準偏差、 $\beta$  は標準化推定値の平均値を表す。*p* は事後分布における片側 *p* 値 (正の推定値については事後分布の 0 以下の面積の割合、負の推定値については事後分布の 0 以上の面積の割合) を 2 倍した値を示しており、この値が .05 を下回る場合、95%信用区間が 0 を含まないことを意味する。分析に際し、学校間変動を除外するため、各従属変数とクラスサイズは学校内で中心化した。推定値の解釈を容易にするため、各従属変数はさらに個人レベルで標準化した。性別は男子を 0、女子を 1 としてコーディングした。レベル 2a には学年コホートをダミー変数として投入したが、スペースの節約のためパラメータ推定値を表示していない。

1999)、イスラエルにおける Angrist & Lavy (1999) の研究では、10 人の増加につき読字と数学の得点 (小 4・小 5) が 0.13—0.25 *SD* 低下することが示されている。一方、国内の先行研究では、妹尾・篠崎・北條 (2013) が 10 人の増加につき国語と算数・数学の得点 (小 6・中 3) が 0.02—0.06 *SD* 低下すること、Akabayashi & Nakamura (2014) が 10 人の増加につき国語の得点 (小 6) が 0.15 *SD* 低下することを報告

しており、本研究の結果とおおむね類似している。

これらの研究間には、前節で述べたような研究デザインや解析の問題を始め、研究期間、対象者の学年、学業成績の測定方法など、多くの方法論的な差異があるため、単純な比較は難しいが、本研究を含め、国内の研究におけるクラスサイズの学業成績への効果量は全体的に海外の研究よりも小さい傾向にある。この原因は定かでないが、補助教員の配置などの学校内での

工夫や学校以外の教育産業の発展度合などの違いが関係している可能性があり、今後の検証が待たれる。

しかし、長期的な視点で見た場合、0.1 SD 程度の学業成績の変化も無視できない意味を持っている。Krueger (2003) は、学業成績と後の収入の関連を検討した複数の研究のレビューから、1 SD の学業成績の上昇が8%の収入増加につながると結論づけている。わが国における平均給与は415万円であり(国税庁, 2015)、仮に学業成績が0.1 SD 上昇し、0.8%の収入増加が生じた場合、22歳から65歳まで勤めた場合の生涯年収は約146万円増加する計算になる。120人からなる学年で、クラス数を3から4に増やし、クラスサイズを40から30に縮小した場合、1名の教員(平均年収665万円; 総務省, 2015)の追加で120人の生涯年収の総額は単純計算で1億8千万円程度も増加することになる。また、本研究や国内の先行研究で検討されているのは、1年間のクラスサイズの効果であり、義務教育の9年間を通してその効果が蓄積した場合には、効果がさらに増大する可能性も考えられる。この点については、今後、さらなる検証が必要である。

#### 情緒的・行動的問題への効果

教師との関係性においては、クラスサイズの拡大は教師との関係上のストレスには有意に影響しないが、教師からのソーシャルサポートの減少をもたらすことが示された。海外の先行研究でも、クラスサイズが小さいほど教師が個々の児童生徒を深く理解し、児童生徒の行動に対する寛容性が高まることが示されている(Bourke, 1986; Finn, Forden, Verdinelli, & Panno, 2001)。クラスサイズの拡大により教師の時間的・精神的な余裕が減少することで、個々の児童生徒に寄り添った共感的な指導スタイルを取ることが難しくなると考えられる。こうしたメカニズムについては、今後より詳細な検証が求められる。

友人との関係性においても、クラスサイズが増加することで、いじめ被害などのネガティブな問題の増加は見られないものの、向社会的行動やソーシャルサポートといったポジティブな行動が減少することが示された。クラスサイズの拡大にともない、個々の児童生徒間の関係が希薄化することで、関係上のトラブルは増えないが、助け合いの行動が取りにくくなると考えられる。この結果は、クラスサイズが攻撃性には有意な効果を持たず、抑うつにのみ有意に影響するという結果とも符合する。つまり、クラスサイズの拡大は、いじめやケンカのような外在化問題よりも、クラス内の友人・教師との関係の希薄化を介して、ストレスを

内側に溜め込む内在化問題につながりやすいと考えられる。

クラスサイズの影響に関する海外の研究では、主に外在化問題(反社会的行動)に焦点が当てられており、一貫して正の影響が見出されている(CSR Research Consortium, 2000; Finn et al., 2001)。援助行動(向社会的行動)については、一部でクラスサイズの負の影響が報告されているが(Finn et al., 2001)、研究が不足している。また、抑うつを含む内在化問題への効果を示す研究は見られない(Jakobsson, Persson, & Svensson, 2013)。一方、本研究の結果は、上記のようにクラスサイズの拡大が外在化問題よりも援助行動の減少や内在化問題につながりやすいことを示唆しており、わが国の学校現場に固有の特徴を示している可能性がある。しかし、こうした差異について明確な結論を下すには実証的知見が不足しており、今後、内在化問題を含めた情緒・行動面の問題に対するクラスサイズの効果について、国内外でさらなるデータの蓄積が必要である。

#### 結論

本研究では、クラスサイズに関する観察的研究の障壁となってきたデータの階層性と要因の交絡という2つの方法論的問題を解決し、クラスサイズと学業成績および情緒的・行動的問題の因果関係を検証した。その結果、クラスサイズの拡大は、(a) 学業成績を低下させること、(b) 教師との関係上のストレスには有意に影響しないが、教師からのサポートを減少させること、(c) 友人関係におけるトラブル(いじめ、ケンカなど)には有意に影響しないが、相互の援助行動の減少をもたらすこと、(d) 攻撃性には有意に影響しないが、抑うつを高めることが示された。こうした影響の広さを鑑みると、クラスサイズは学級運営上、重大な意味を持つ変数であると結論づけられる。

本研究は単一都市における検証であったが、平均的な中規模都市の全ての小中学校を対象としたものであり、サンプルの選択バイアスは小さいと考えられる。ただし、本研究のサンプルにおけるクラスサイズは、21-42人の範囲であり、20人以下の少人数学級に知見を一般化することは難しい。今後、他の地域でもデータを集積し、知見の一般化可能性を確認する必要がある。

また、本研究では縦断研究のデータを分析に用いたが、先行研究との比較のため、各年度内での横断的な関連の検証のみを行った。今後、長期にわたっての効果の蓄積や残存についても検証を行うことが求められる。

る。それによって、少人数学級の費用対効果についても、より明確な結論を得ることができると考えられる。ただし、そうした縦断的な効果を十分な検出力のもとで検証するには、本研究の数倍の規模のサンプルが必要になると考えられ、多機関連携などによる全国規模の調査が求められる。

### 引用文献

- Akabayashi, H., & Nakamura, R. (2014). Can small class policy close the gap?: An empirical analysis of class size effects in Japan. *Japanese Economic Review*, 65, 253-281. doi:10.1111/jere.12017
- Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114, 533-575. doi:10.1162/003355399556061
- Angrist, J. D., & Pischke, J. S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24, 3-30. doi:10.3386/w15794
- Asparouhov, T., & Muthén, B. (2010). Bayesian analysis using Mplus: Technical implementation. Muthén & Muthén. September 29, 2010. Retrieved from <https://www.statmodel.com/download/Bayes3.pdf> (January 13, 2017)
- Asparouhov, T., & Muthén, B. (2014). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. Muthén & Muthén. November 19, 2014. Retrieved from [https://www.statmodel.com/download/NCME\\_revision2.pdf](https://www.statmodel.com/download/NCME_revision2.pdf) (January 13, 2017)
- Bourke, S. (1986). How smaller is better: Some relationships between class size, teaching practices, and student achievement. *American Educational Research Journal*, 23, 558-571. doi:10.3102/00028312023004558
- Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling*, 1, 103-124. doi:10.1177/1471082X0100100202
- CSR Research Consortium. (2000). *Class size reduction in California: The 1998-99 evaluation findings*. Sacramento, CA: California Department of Education.
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12, 121-138. doi:10.1037/1082-989X.12.2.121
- Finn, J. D., Forden, M. A., Verdinelli, S., & Pannozzo, G. M. (2001). *Evaluation of the class size reduction initiative: Buffalo Public Schools (2000-2001)*. Buffalo, NY: State University of New York at Buffalo, Graduate School of Education.
- Finn, J. D., Fulton, D., Zaharias, J., & Nye, B. A. (1989). Carry-over effects of small classes. *Peabody Journal of Education*, 67, 75-84.
- Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The "Why's" of class size: Student behavior in small classes. *Review of Educational Research*, 73, 321-368. doi:10.3102/00346543073003321
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis* (2nd ed.). New York: Chapman and Hall.
- Goldstein, H. (1994). Multilevel cross-classified models. *Sociological Methods & Research*, 22, 364-375. doi:10.1177/0049124194022003005
- Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581-586. doi:10.1111/j.1469-7610.1997.tb01545.x
- 浜田 恵・伊藤大幸・片桐正敏・上宮 愛・中島俊思・高柳伸哉…辻井正次 (2016). 小中学生における性別違和感と抑うつ・攻撃性の関連 発達心理学研究, 27, 137-147.
- 原田 新・伊藤大幸・望月直人・田中善大・大嶽さと子・高柳伸哉…辻井正次 (2014). 日本語版 Strengths and Difficulties Questionnaire 自己評定フォームの構成概念的妥当性—抑うつ, 攻撃性, 親評定フォームとの関連から 小児の精神と神経, 53, 343-351.
- Hojo, M., & Oshio, T. (2010). What factors determine student performance in East Asia?: New evidence from TIMSS 2007. PIE/CIS Discussion Paper, No. 494.
- 伊藤大幸・田中善大・村山恭朗・中島俊思・高柳伸哉・野田 航…辻井正次 (2015). 小中学生用社会的

- 不適応尺度の開発と構成概念妥当性の検証 精神医学, 56, 699-708.
- Jakobsson, N., Persson, M., & Svensson, M. (2013). Class-size effects on adolescents' mental health and well-being in Swedish schools. *Education Economics*, 21, 248-263. doi:10.1080/09645292.2013.789826
- 国立教育政策研究所 (2002). 学級規模に関する調査研究 国立教育政策研究所
- 国税庁 (2015). 平成 26 年分民間給与実態統計調査結果について 2015 年 9 月 Retrieved from <http://www.nta.go.jp/kohyo/press/press/2015/minkan/> (2017 年 1 月 13 日)
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, 114, 497-532. doi:10.1162/003355399556052
- Krueger, A. B. (2003). Economic considerations and class size. *Economic Journal*, 113, F34-F63. doi:10.1111/1468-0297.00098
- Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *Economic Journal*, 111, 1-28. doi:10.1111/1468-0297.00586
- Lazarsfeld, P. F. (1955). Interpretation of statistical relations as a research operation. In P. F. Lazarsfeld & M. Rosenberg (Eds.), *Language of social research* (pp. 115-124). New York: Free Press.
- 村田豊久・清水亜紀・森陽二郎・大島祥子 (1996). 学校における子どものうつ病—Birlersonの小児期うつ病スケールからの検討 最新精神医学, 1, 131-138.
- 村山恭朗・伊藤大幸・大嶽さと子・片桐正敏・浜田恵・中島俊思…辻井正次 (2016). 小中学生におけるメンタルヘルスに対するソーシャルサポートの横断的効果 発達心理学研究, 27, 395-407.
- 並川 努・谷 伊織・脇田貴文・熊谷龍一・中根 愛・野口裕之・辻井正次 (2011). Birlerson 自己記入式抑うつ評価尺度 (DSRS-C) 短縮版の作成 精神医学, 53, 489-496.
- Organization for Economic Co-operation and Development (OECD) (2015). *Education at a glance 2015: OECD indicators*. Paris: OECD Publishing.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publishing.
- 坂井明子・山崎勝之・曾我祥子・大芦 治・島井哲志・大竹恵子 (2000). 小学生用攻撃性質問紙の作成と信頼性, 妥当性の検討 学校保健研究, 42, 423-433.
- 妹尾 渉・篠崎武久・北條雅一 (2013). 単学級サンプルを利用した学級規模効果の推定 国立教育政策研究所紀要, 142, 161-173.
- 篠崎武久 (2008). 教育資源と学力の関係 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 (pp. 73-97) 千葉県検証改善委員会
- 総務省 (2015). 給与・定員等の調査結果等 Retrieved from [http://www.soumu.go.jp/main\\_sosiki/jichi\\_gyousei/c-gyousei/teiin-kyuuyo02.html](http://www.soumu.go.jp/main_sosiki/jichi_gyousei/c-gyousei/teiin-kyuuyo02.html) (2017 年 1 月 13 日)
- 辰野千壽・石田恒好・服部 環・筑波大学附属小各科教官 (2002). 標準学力検査 NRT (集団基準準拠検査) 図書文化
- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Newbury Park, CA: Sage.

(2016 年 5 月 10 日受稿, 2017 年 7 月 15 日受理)

*Effects of Class Size on Academic Achievement and Emotional and Behavioral Problems: Evaluation With a Natural Experimental Design and Multilevel Model*

HIROYUKI ITO (HAMAMATSU UNIVERSITY SCHOOL OF MEDICINE),

MEGUMI HAMADA (NAGOYA UNIVERSITY OF ARTS AND SCIENCES), YASUO MURAYAMA (KOBE GAKUIN UNIVERSITY),

NOBUYA TAKAYANAGI (AICHI TOHO UNIVERSITY), KAZUYO NOMURA (TOKOHA UNIVERSITY),

MITSunobu MYOGAN (CHUKYO UNIVERSITY) AND MASATSUGU TSUJII (CHUKYO UNIVERSITY)

*JAPANESE JOURNAL OF EDUCATIONAL PSYCHOLOGY*, 2017, 65, 451–465

The present study examined effects of class size in elementary and middle school on students' academic achievement and emotional and behavioral problems, while addressing some important methodological issues. Adoption of a natural experimental design in which class size was determined only by enrollment by grade prevented confounding bias and reverse causation due to the schools' discretion. The hierarchical nature of the data was modeled using a cross-classified model, a kind of multilevel model. The elimination of between-school variation by school-mean centering enabled a test of the pure within-school effects of class size, and avoided confounding by school enrollment. Data were collected from 4th- to 9th-graders in a 9-wave longitudinal investigation ( $N = 45,694$ ). Analysis of these data revealed that an increase in class size resulted in (a) a reduction in students' academic achievement, (b) a decrease in support from teachers, (c) a decrease in support from peers and in prosocial behavior, and (d) an elevation in the students' depression. These results suggest that class size may be a critical variable in class management, due to its extensive influence on the functioning and psychosocial adjustment of students.

Key Words: class size, academic achievement, students' emotional and behavioral problems, multilevel model, Markov chain Monte Carlo methods