

Working Paper Series No.107 (revised)
Faculty of Economics, Niigata University

国際学力調査結果を用いた教育生産関数の推計
—少人数指導・習熟度別指導・ピアグループの効果—

北條 雅一

新潟大学経済学部

2010年9月改訂

日本経済学会春季大会 2010（千葉大学）報告論文

本稿は「TIMSS 2007 を用いた教育生産関数の推定：少人数指導および習熟度別指導の効果」新潟大学経済学部ワーキングペーパー107（2009年6月）を改訂したものである。

連絡先：hojo@econ.niigata-u.ac.jp

国際学力調査結果を用いた教育生産関数の推計 —少人数指導・習熟度別指導・ピアグループの効果—

1. はじめに

本稿の目的は、最新の国際学力調査の国内結果を用いて、公立の小中学校における教育生産関数を推定することである。教育生産関数とは、学校や教師といった物的および人的資源をインプット（投入要素）、学力等の指標で示される生徒の達成水準をアウトプットと捉える考え方である。このような教育生産関数は、これまで欧米を中心として、実際の学力データを用いた実証分析が蓄積され、どういった投入要素が生徒の学力水準に影響を及ぼしているかが検証されてきた¹。しかしながら、日本国内における同様の分析は、近年まであまり行われてこなかったのが実情である。教育経済学の分野における本稿の位置づけは、利用可能な児童・生徒レベルの学力データを用いて、国内における教育生産関数の基礎的な推定結果を提示し、学力の規定要因を明らかにすることにある。

国内の義務教育に関しては、1990年代以降、子どもの学力低下を懸念する声が高まった。この背景には、1992年および2002年の学習指導要領の改訂により、授業内容や授業時間数の削減が実施されたことがある。また、2003年に公表された複数の国際学力調査結果において、日本の児童・生徒の得点順位が低下したことも影響したと考えられる。その後、文部科学省は授業内容や授業時間の削減方針を改め、主要教科の授業時間数を拡大するなどの対策を講ずるに至っている。また、学習効果を高める方策として、少人数指導や習熟度別指導、学校選択制などの新たな取り組みも浸透しつつある。

本稿の実証分析では、こうした新しい取り組みの効果についても検証をおこなう。本稿で検証するのは、少人数指導および習熟度別指導の効果である。前者は、欧米を中心とした研究において‘class-size effect’と呼ばれ、数多くの実証研究が蓄積されているが、その効果は必ずしも一様ではない²。少人数指導の実施には追加の物的および人的資源の投入が必要であり、その効果の検証は教育政策のみならず、財政的な観点からも重要な意味を持つ。習熟度別授業に関しては、とくに国内公立学校においては、児童・生徒に平等な教育を提供するという観点から導入に消極的であったと考えられる。しかしながら、近年になって急速に導入が進んでおり、その効果を検証することは教育の実践上、また教育政策の観点からも意義のあるものと考えられる。

教育生産関数の推定が日本国内で進まなかった理由は、Oshio and Senoh (2007) が指摘するように、実証分析に必要な個別学力データの利用が困難であったことが挙げられる。とはいえ、近年、そうしたデータが一部利用可能となり、国内においても学力データを用

¹ 米国を中心とした研究に関しては、Hanushek (2006) が詳細なサーベイを行っている。

² 米国における学級規模縮小の効果については、Pritchard (1999) がまとめている。

いた実証分析が蓄積されつつある³。本稿で使用するデータは、国際数学・理科教育動向調査 (Trends in International Mathematics and Science Study, 略称 TIMSS) の 2007 年調査の結果である。このデータには、小学 4 年生と中学 2 年生の算数・数学と理科の学力調査の結果に加え、児童・生徒の家庭環境に関する情報、授業担当教師の特徴、通学している学校の特徴に関するデータが含まれており、学力に影響を及ぼすさまざまな要因を考慮したうえで教育生産関数を推定することが可能となっている⁴。

国内における児童・生徒の学力の規定要因という観点については、経済学者による教育生産関数の推定とは別に、教育社会学の分野において多くの研究が蓄積されてきた。教育社会学において注目されてきたのは「学校の効果」と「学力の階層差」の 2 つの研究領域である。「学校の効果」とは、児童・生徒の学習成果に学校が及ぼす影響を意味しているが、中でも進展が著しいのは、Edmonds (1979) の提唱した「効果のある学校＝学力格差の小さい学校」という考え方に基づいた研究である⁵。また、「学力の階層差」に関する研究では、子どもの学力や学習意欲、学歴取得、人々の教育に対する意識などには明確な階層差が存在することが確認されている⁶。本稿で分析する教育生産関数には、学校や授業担当教師の特徴に関する説明変数が含まれている。これらは「学校の効果」の一側面を捉えるものである。また、家庭環境や親の学歴に関する説明変数は「学力の階層差」を代理するものと考えられる。

本稿の実証分析の結果、以下の点が明らかとなった。第 1 に、児童・生徒本人およびその家庭の要因が学力達成に強く影響していることである。具体的には、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、児童・生徒レベルでの試験の点数に強い影響をもたらしている。早生まれによる不利を示す効果（相対年齢効果と呼ばれる）および学力の階層差を追認する結果とも言える。第 2 に、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という点である。こうした結果は、米国においてこの分野の先駆けとなった「コールマン報

³ 「全国学力・学習状況調査」の千葉県公立校のデータを用いた篠崎 (2008)、首都圏・近畿圏の中高一貫校を対象に学校レベルの分析を行った小塩・佐野・末富 (2009)、都市部自治体の学力調査結果を使用した上野・三野・小塩・佐野 (2007) がある。

⁴ TIMSS を使用した教育生産関数の実証研究に Woessmann (2003)、Woessmann and West (2006) などがある。

⁵ 「効果のある学校」についての研究では、効果に関するいくつかの条件を満たす学校を「効果のある学校」として少数選び出し、その学校の特徴を詳細に調査するというアプローチが主流である。主な研究に鍋島 (2003)、志水 (2004) がある。他に、複数年度の学力調査結果に基づいて「効果のある学校」を抽出した川口・前馬 (2007)、地域の社会経済的条件に注目した舞田 (2008)、マルチレベルモデルを用いて学校の固定効果を定量的に算出した川口 (2009) がある。

⁶ 苅谷 (2001) によれば、階層とは「所得や職業の威信、学歴、権力などのさまざまな社会・経済・文化的資源と呼ばれるものを基準としてみた、社会的な地位やカテゴリーのこと」と定義されている。

告」(Coleman, et al. 1966) およびそれ以後の研究結果と似通ったものとなっている⁷。

3 点目として、少人数指導や習熟度別指導に一定の効果が確認されたことが挙げられる。具体的には、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が挙げられる。また、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も明らかとなった。この効果は、習熟度別授業の実施が「学力の階層差」を縮小する可能性を示唆するものと言える。

また、本稿の分析では「ピア・グループ効果」の存在についても検証をおこなっている。ピア・グループ効果とは、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である。例えば、優秀なクラスメイトと一緒に学校生活を送ることで本人の成績が向上するような効果である。ピア・グループ効果の検証は欧米諸国を中心に研究が蓄積されており、近年はピア・グループ効果と他の要因による効果を適切に識別することが研究上の重要な観点となっている。現状では、ピア・グループ効果を識別するための実証的な分析手法は多様であり、標準的な手法と呼ばれるようなものは存在しない⁸。本稿におけるピア・グループ効果の検証は、国内データを用いた分析の第一歩と位置づけられよう。本稿の分析結果によれば、ピア・グループ効果は両学年で存在が確認され、小学 4 年生では性別によって、中学 2 年生では習熟度別授業の実施有無によって効果が変わることが明らかとなっている。

本稿の分析結果には興味深いものもあるが、その解釈については慎重を要する。本稿の分析で使用するデータは調査時点の一時点のものであるため、その推定結果は変数間の因果関係を示すものではない場合がある⁹。ある生徒のある時点における学力を決定するのは、その生徒に対する教育資源の累積投入量（物的および人的資源の投入量）と、本人の（研究者には観測されない）能力であると考えられるが、累積的な教育資源の投入量に関するデータはほとんど存在せず、本人の観測されない能力を代理する指標も入手は難しい。別の方法としては、ある期間における同一生徒の学力の変化を、その期間に投入された教育資源の量によって説明する方法が考えられる。本稿の分析は、このいずれでもない点に注意が必要である。本稿の分析においては、児童・生徒の年度末の学力と、その年度に投入された教育資源およびその他の要因との関係を検証している。したがって、「ある説明変数

⁷ 「コールマン報告」は、1960 年代に米国で実施された大規模な学力調査をもとに、生徒の学力達成水準の決定要因を分析した結果をまとめたものである。この中で、学校や教師への資源投入を通じた生徒一人当たり教育支出額と学力の間には明確な相関がない一方で、家庭環境などの生徒の社会的・経済的背景と学力が強く相関していることが示された。

⁸ ピア・グループ効果に関する近年の研究成果として、西欧 6 カ国を対象に学校固定効果によって内生性に対処した Ammermueller and Pischke (2009)、テキサス州の学校を対象にクラス編成の変化を用いて識別している Hanushek, et al. (2003)、科目間階差をとることで観察されない能力によるバイアスに対処している Kang (2007) などがある。

⁹ 学力の決定要因を分析する方法と問題点については、Hanushek (2006) が詳しく論じている。

の増加／減少によって生徒の学力が高まる／低下する」といった因果関係を示す解釈が適切でない場合がある点には、十分な注意が必要である。

本稿の構成は以下のとおりである。次節では、TIMSS の国内データについて説明する。第 3 節は教育生産関数の推定モデルと分析アプローチについて詳述する。第 4 節は推定結果の報告、第 5 節は結論と今後の課題について述べる。

2. データ

本稿の分析で使用するデータは、国際数学・理科教育動向調査（TIMSS）の 2007 年調査（以後 TIMSS 2007 と表記）の調査結果である¹⁰。この国際学力調査は、国際教育到達度評価学会（IEA）によって行われてきた、第 1 回国際数学教育調査（1964 年）、第 1 回国際理科教育調査（1970 年）から続くものであり、1995 年実施の第 3 回国際数学・理科教育調査以降の調査を TIMSS の後ろに調査年を付加して呼ぶようになった。

本稿で使用する TIMSS 2007 には、59 カ国/地域が参加し、これまでの TIMSS 調査とほとんど同じ条件で、第 4 学年（小学校 4 年生）と第 8 学年（中学校 2 年生）を対象として実施された。国内の調査は国立教育政策研究所が担当し、2007 年 3 月に実施されている。調査の目的は、初等中等教育段階における児童・生徒の算数・数学および理科の教育到達度を国際的な尺度によって測定・研究すること、1995 年、1999 年、2003 年調査の同学年との比較を行うことである。

調査の種類は、児童・生徒を対象とした「問題」（算数・数学、理科の問題）および「児童質問紙」「生徒質問紙」、教師を対象とした「教師質問紙」、学校長を対象とした「学校質問紙」である。「問題」は 14 種類の問題冊子から児童・生徒ごとに 1 種類を指定し、解答時間は小学 4 年生が 72 分、中学 2 年生が 90 分である。1 人の児童・生徒が解く問題数は、小学校は約 50 題、中学校は約 60 題である。

「児童質問紙」および「生徒質問紙」は本人が約 30 分で回答することになっており、自身の性別や生年月に加え、家庭環境や家での時間の過ごし方、教科の理解や成績に関する自己評価、勉強の楽しさ、勉強が将来の役に立つと思うか、といった項目が質問されている。「教師質問紙」は、標本抽出された児童・生徒の算数・数学、理科の授業を担当している教師が調査の対象となっている。調査項目は、年齢、性別、教職経験年数、大学での専攻といった項目に加え、調査対象学級の学級規模、教科の指導に関する時間の使い方や問題点などについて詳細に尋ねられている。「学校質問紙」は、標本抽出された児童・生徒が通う学校の学校長が回答している。調査項目は、学校が立地する都市の規模などの学校の特徴や、学校全体として児童・生徒の問題行動の頻度や施設の不足感といったものである。後に詳述する習熟度別授業に関する質問も含まれている。

¹⁰ 1995 年以降に実施された調査のデータは、TIMSS & PIRLS International Study Center のウェブサイト (<http://timss.bc.edu/>) よりダウンロードすることができる。

標本抽出は、学校所在地の地域類型（層）を考慮した上で、学校基本調査をもとにまず学校を抽出し、そこから児童・生徒（学級）を抽出する層化二段階抽出によって実施されている。学校によって、1学級のみを抽出している場合と、2学級を抽出している場合がある。小学校・中学校の参加状況は以下のとおりである。小学校は148校の児童4487名と教師250名、中学校は146校の生徒4312名、数学教師216名、理科教師178名が調査の対象となった。表1は、TIMSS 1995以降の日本の平均得点の推定値の推移を示している。1999年の国内調査は中学2年生のみを調査対象としたため、小学4年生の結果は存在しない。2003年の調査結果では、前回調査に比べて小学4年の理科で10点、中学2年の数学で9点、国内平均点が低下し、学力低下を示唆するものとして注目を集めたことは記憶に新しい。なお、2007年調査における日本の得点順位は、算数・数学、理科ともに参加国中第5位以上となっている。

3. 推定モデルと分析アプローチ

3.1 推定モデル

本稿では、次式で示される教育生産関数を推定する。

$$A_i = \beta_0 + F_i\beta_1 + T_i\beta_2 + Z_i\beta_3 + \beta_4\bar{A}_i + \varepsilon_i$$

A_i は生徒*i*の点数、 F_i は生徒*i*の本人属性および家庭背景等の変数ベクトル、 T_i は生徒*i*を指導している教師や通っている学校に関する変数ベクトル、 Z_i は少人数指導や習熟度別授業といった教育上の取り組みに関する変数ベクトルである。 \bar{A}_i はピア・グループ効果を代理する変数であり、本稿の分析では、本人を除いたクラスメイトの平均点数である。 β はパラメータ、 ε は誤差項である。なお次小節に詳述するが、データの制約上、本稿では算数・数学に限って分析をおこなう。

点数*A*に関して、TIMSSのデータセットには素点のほかにもいくつかの種類が含まれている。本稿の分析で用いるのは、国内平均150点、標準偏差10点に変換された数値である。 F には、本人属性として女子ダミーと生まれ月ダミー、家庭に関するものとして家庭の蔵書数、学習机・コンピュータ・地球儀の有無、両親の学歴をあらわすダミー変数が含まれている¹¹。なお、これらの変数は児童・生徒への質問紙から得られたものであり、保護者から得た回答ではない点に注意が必要である。特に、両親の学歴については「わからない」と回答している生徒も少なくない。そのため、ダミー変数の作成に当たっては「わからない」を一つのカテゴリとした。

教師や学校に関する変数*T*には、授業担当教師の教職経験年数とその二乗項、性別、学

¹¹ 家庭の蔵書数には雑誌、新聞、教科書は含まれていない。また、コンピュータには家庭用ゲーム機は含まれていない。両親の学歴は小学4年生には質問していないため、中学2年生のサンプルでのみ利用可能である。

歴のダミー変数、学校における教材や教師の不足感、学校所在地の都市規模、学校全体における経済的に恵まれない生徒の比率、学年生徒数が含まれている。教師や教室施設の不足感は、学校質問紙内の「あなたの学校では、次のどれかが不足していたり、不十分であることが学習指導に影響を与えていますか」との質問に対し、「教材」「教室などの学習施設」および「教師」の項目において、それぞれ「影響がある」「たくさん影響がある」と回答している場合に1、そうでない場合に0をとるダミー変数としている。

教育上の取り組みを示すZには、教科授業時の学級規模（クラスサイズ）、習熟度別授業の実施有無に関するダミー変数（実施している場合に1をとる）が含まれている。学級規模の変数を説明変数とする場合には、内生性によるバイアスに注意する必要がある（Hoxby 2000）。例えば、学力の停滞している学校に対して、学力向上を目的としてより多くの教員が配置され、結果としてそのような学校において少人数学級が実現しているような場合には、学級規模と点数の同時決定によるバイアスが生じる可能性がある。また、調査者には観測されない変数と学級規模の間に相関があるような場合にもバイアスが発生する可能性が考えられる。こうしたバイアスを除去する方法として、操作変数を用いた推定が提案されてきた。本稿では、Angrist and Lavy (1999) で提案された操作変数法を適用する。

Angrist and Lavy (1999) は、イスラエルの小学校の学級編制において、「学級規模の上限を40人とすべきである」という古代の賢人が示したルール（Maimonides' s Rule）が厳格に運用されているという性質に着目し、実際に編制された学級規模がその学年の学年生徒数の増減によって決定されていることを示した。具体的には、例えば学年生徒数が40人以下の場合には1つのクラスが編制されるが、学年生徒数が41人となったところでルールが適用され、20人と21人の2クラスに編制される。つまり、学年生徒数が40の倍数を超える度に1クラス当たりの生徒数が小さくなるということであり、こうしたルールに基づいて導かれる学級規模の予測値を、実際の学級規模の操作変数として使用することによって学級規模の効果を識別している。予測値は次式によって導出できる。

$$\text{Predicted class size} = \text{enrolment} / \{\text{int}[(\text{enrolment}-1)/40]+1\}$$

図1は、本稿で使用するTIMSSデータを用いて、実際の学級規模と40人を上限とした場合のルールに基づく予測値を示している。やや乖離がみられるものの、1クラス40人上限ルールから導かれる予測値は実際の学級規模の変動をうまく説明しているといっていよう。言い換えれば、ルールから導かれる予測値は実際の学級規模との相関が高い。同時に、予測値の変動は学年生徒数にのみ依存するものであり、学校の平均学力や親による学校選択の影響は小さいであろう。したがって、ルールから導かれる予測値は、実際の学級規模の操作変数として望ましい性質をもつものと考えられる。分析においては、実際の学級規模と予測値の両方の対数をとったものを使用している。

習熟度別授業の実施有無に関するダミー変数についても注意が必要である。この変数は、

学校質問紙の「あなたの学校では、4年生の児童は算数の授業の中で能力別にグループ分けされますか」（小学4年の算数の場合）の質問に対して「はい」と回答している場合を1とするダミー変数である。したがって本稿では、TIMSSにおいて「能力別グループ分け」と表現されているものを「習熟度別授業」と読み替えていることになる。

ピア・グループ効果の代理変数として、本稿では本人を除いたクラスメイトの平均点数を使用している。同様の手法はTIMSS 2003を用いたKang (2007)においても適用されているが、ピア・グループ効果の検証には別の変数が使用されることもある。例えば、Ammermueller and Pischke (2009)は本人を除いたクラスメイトの家庭蔵書数の平均値を使用し、Hanushek et al. (2003)は本人を除いたクラスメイトの平均点数と経済的に恵まれないクラスメイトの比率を用いている。上述のように、本稿で使用するTIMSSデータには家庭の蔵書数に関する情報が含まれており、Ammermueller and Pischke (2009)と同様のピア・グループ変数の効果を検証することも可能であるが、本稿ではその結果は報告しない。なお、表2に本稿で使用するすべての変数の記述統計を報告している。

3.2 分析アプローチ

先に述べたように、本稿の分析の位置づけは、これまで蓄積の少なかった日本における教育生産関数を推定し、基礎的な分析結果を提示することにあると考えている。したがって、必要な説明変数を可能な範囲内で分析に含め、それらの変数と、点数であらわされる児童・生徒の学力達成との関係を統計的に検証することが目的の一つである。

本稿の実証分析では、調査対象となった小学4年生と中学2年生の算数・数学について、それぞれの教育生産関数を推定する。なお、それぞれの学年の理科得点を対象とした教育生産関数を推定することも可能であるが、本稿ではその結果は報告しないこととする。その理由の一つは、小学4年の理科の授業を習熟度別に行っている学校がデータに存在しないことである。また、中学2年のデータには理科の習熟度別授業を行っている学校が含まれているものの、その比率は約3.3%と小さいため、習熟度別授業の実施有無による影響を適切にコントロールすることが難しいという制約もある。以上の理由から、本稿の分析は算数・数学に限定することとした。また、中学2年生のデータセットには私立学校に通う生徒も含まれているが、本稿の分析では除外している。

小学4年生の算数、公立中学2年生の数学の全サンプルを対象とした推定に続き、次の2つのケースについてサンプルを分割し、サンプルの特徴による教育生産関数の構造の差異を確認する。

1つ目は、児童・生徒の男女別の推定である。これは、男子と女子でそれぞれの学力達成に及ぼす要因が異なる可能性を考察するためである。2つ目は、習熟度別授業の実施有無によるサンプル分けである。詳しくは後に報告するが、習熟度別授業の実施は中学2年生の数学得点に統計的に有意な影響を及ぼしている。習熟度別の授業を受けることで、その他の要因の得点に及ぼす影響が変化している可能性を考察する。

3.3 少人数指導・習熟度別指導・教師年齢構成の偏りの効果

本稿の分析では、近年導入が進んでいる教育上の取り組みのうち、少人数指導と習熟度別授業の効果に注目している。少人数指導について検定する仮説は、「少人数指導の実施と児童・生徒の点数の間には正の相関がある」というものである。学級規模の縮小は、教師一人当たりの生徒数を少なくすることを意味する。その期待される効果は、大人数学級に比べて生徒一人一人に教師の目が行き届くようになり、それが教師による生徒の状況把握を改善し、各生徒の状況に応じたきめ細かい指導を可能とすることで、生徒の学力向上をもたらすというものであろう。国内においては、文部科学省が2001年に発表した『第7次公立義務教育諸学校教職員定数改善計画』において、基礎学力の向上ときめ細かな指導を目指し、少人数指導のための教職員の加配を開始した。

しかしながら、学級規模縮小の効果が統計的に明らかに確認されているわけではない。1960年代から米国を中心に蓄積された実証研究をサーベイした Hanushek (2006) によれば、学級規模を説明変数とした教育生産関数の推定において、期待される符号で統計的に有意な効果が認められたケースは全体の14%にとどまることが報告されている。本稿ではまず、学級規模縮小の効果が存在するか否かを確認した上で、その効果が男女別・習熟度別授業の実施有無によって異なっているかを確認する。

習熟度別授業についての仮説は、「習熟度別授業の実施と児童・生徒の点数の間に正の相関関係がある」というものである。習熟度別授業は、同程度の理解度の児童・生徒を学習集団とすることによって理解度に応じた水準の授業を可能にし、児童・生徒の学力の向上を図るものである。文部科学省は、2002年に公表した『確かな学力の向上のための2002アピール（「学びのすすめ」）』において、「少人数授業・習熟度別指導など、個に応じたきめ細かな指導の実施を推進し、基礎・基本の確実な定着や自ら学び自ら考える力の育成を図る」としている。

児童・生徒の習熟度に応じた指導の実施が学力の向上に結び付くという考え方に対して、教育経済学においては「ピア・グループ効果」と呼ばれる効果の存在も指摘されてきた。同じクラスで勉強するクラスメイトの影響や刺激を受けることで、本人の理解度が変化するような効果である。習熟度別授業の実施によって同程度の理解度の児童・生徒が学習集団となるため、こうしたピア・グループ効果が変化する可能性が考えられる。

本稿では、まず習熟度別授業の実施が児童・生徒の点数とどのような関係にあるかを確認し、その上で、習熟度別授業の効果が児童・生徒の性別によって異なる可能性を確認する。また、習熟度別授業の有無によって他の要因の効果が変化している可能性を検証し、その中でピア・グループ効果の変化についても注目する。

教師の年齢構成について検定する仮説は、「授業担当教師の教職経験年数と児童・生徒の点数の間に正の相関がある」というものである。教師の経験年数の効果は先行研究でも数多く検証されてきたが、小中学校において教師の年齢構成に偏りが指摘される日本の現状

と照らし合わせると、この仮説の意味は重大である¹²。図1は、公立小中学校教師の年齢構成を示したものである。今後5～10年以内に60歳を迎える教師は公立小学校で約34%、公立中学校で約27%となっている。仮に、指導力の高いベテラン教師に教わることが児童・生徒の高い点数と結びついているとすれば、今後こうした教師の多くが教育現場を去る効果は無視できないものとなる可能性がある。

4. 推定結果

表3は、小学4年生の算数の点数、表4は中学2年生の数学の点数を被説明変数とした教育生産関数の推定結果である。すべての推定に際し、標本抽出の影響をコントロールするためのウェイトを使用し、学校レベルのClusteringによる影響をコントロールしたうえでOLSを適用している¹³。

4.1 小学4年生サンプルの推定結果

表3の第(1)列は、小学4年生の全サンプルに関する推定結果である。ここでまず確認されることとして、児童本人およびその家庭に関する説明変数の影響が大きいことが挙げられる。例えば、1～3月に生まれた早生まれの児童は、4～6月に生まれた児童に比べて2.6点(0.26 standard deviation)、統計的に有意に点数が低くなっている。これは、Bedard and Dhuey (2006) および Kawaguchi (2009) において確認されている同一学年内での年齢効果(相対年齢効果)を追認するものである。この効果は、10～12月生まれの子供についても確認されているが、その影響は1.5点とやや小さくなっている。また、家庭の蔵書数や所有物の効果も大きい。蔵書数が101～200冊の家庭に育った児童は、0～10冊の家庭の児童に比べて8.1点、得点が高いという結果となっている¹⁴。コンピュータや地球儀を所有している場合も同様に有意に得点が高くなっている。学習機の有無は有意な係数とはなっていないが、これは、学習機を保有していない児童の割合が非常に少ないことによるものと考えられる。

授業を担当する教師に関する説明変数の効果は、統計的には確認されていない。教師の学歴、性別、教職経験年数およびその二乗項はすべて統計的に有意ではないという結果になっている。次に、学校に関する変数の影響を確認する。学校が立地する都市規模につい

¹² 公立小中学校教師の年齢構成の偏りについては苅谷(2008)が詳細に論じている。

¹³ ウェイト変数としてデータセット中の *MATWGT* を用いている。本稿ではOLSの結果のみを報告しているが、分位点回帰(Quantile Regression)による推計でもOLSと同様の結果が得られることを確認している。なお、本稿のすべての推計にはStata version 10.1を使用した。

¹⁴ 家庭の蔵書数という変数の効果の解釈、言い換えれば何の代理変数なのかを判断することは容易ではない。家庭の所得水準、家庭の文化的環境、保護者の教育熱心さ、といった要因を代理していると考えられる。

ては、得点との間の有意な関係は確認されていない。しかしながら、経済的に恵まれない児童の割合が大きい学校では得点が低下していることが確認される。具体的には、経済的に恵まれない家庭から通う子供の比率が25%を超える学校に通う児童は、その比率が低い学校に通う生徒に比べて点数が有意に低くなっている。低所得の親が多い地域で得点が低くなっていることは、親の職業や所得水準が子どもの得点に影響を及ぼすことを示唆していると考えられる。なお、学校規模の代理変数である学年生徒数、教材や教師、教室設備の不足感を示す変数の効果は統計的に有意ではない。

近年導入が進んでいる教育上の取り組みをあらゆる変数については、学級規模も習熟度別授業の実施も児童の得点と統計的に有意な関係はほぼ確認されない結果となった。上述したが、この結果を「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。児童の授業理解度が停滞している学校においてこうした実践が積極的に導入されている可能性もあるからである。一方で、分析に使用しているTIMSS調査の実施時期が学年末の3月であったことを考慮すると、調査前のほぼ1年間の実践が反映された結果であるとも考えることも可能である。

ピア・グループ効果の代理変数である本人を除いたクラスメイトの平均得点の係数は、プラスで統計的に有意である。具体的には、クラスメイトの平均得点が10点上昇すると、本人の点数は約3.5点高くなっている。この結果はピア・グループ効果の存在を示唆するものではあるが、推定値の妥当性および解釈には注意が必要である。例えば、ある児童のクラスメイトの平均得点が高く、その影響によって本人の得点が上昇していると同時に、クラスメイトはその児童の得点上昇によって刺激を受け、得点が上昇している可能性がある。これはManski(1993)において指摘されている‘reflection problem’と呼ばれるものであり、こうした同時決定性を考慮すると、 β_4 はクラスメイトの学力から本人学力への因果関係を示すものではなく、児童相互間の影響の度合いと解釈するほうが適切である。

続いて、児童の性別によって上記の結果に変化が生じている可能性について確認する。第(2)列は男子児童のサンプル、第(3)列は女子児童のサンプルに関する推定結果を報告している。

説明変数のうち、男女でその効果が異なっているものとして本人や家庭の所有物の効果が挙げられる。まず、生まれ月に関しては、4~6月に生まれた児童に比べ、1~3月に生まれた早生まれの児童の点数が有意に低くなっている点は第(1)列と同様であるが、10~12月生まれを示すダミー変数は、女子においてのみ、その負の効果の統計的有意性が失われる結果となっている。次に、家庭の所有物についてであるが、男子児童の場合は地球儀の所有が、女子児童の場合は学習機の所有が、それぞれ統計的に有意な正の効果となっていることが確認される。

教師の効果に男女差は確認されないが、学校の立地に関する変数の効果は異なっている。大都市に立地する学校において得点が高くなる効果は女子児童のサンプルでのみ確認され、経済的に恵まれない児童の割合が多い学校における負の効果は男女の両サンプルにおいて

確認されている。居住地や親の経済力の影響が児童の性別によって異なることを示唆しており興味深い。

学級規模や習熟度別授業の実施有無の効果は男女間で違いは確認されていないが、ピア・グループ効果については男女でやや異なっている。具体的には、男子のほうが女子に比べて推定値が大きくなっており、この男女差は限界的にはあるが統計的に有意となっている。こうした結果は、ピア・グループ効果の大きさが児童の性別によって異なっている可能性を示唆しているが、その要因については明らかではない。

次に、習熟度別授業の実施有無によって、説明変数の効果が異なっている可能性について確認する。第(4)列は算数の習熟度別授業を実施していない学校に通う児童のサンプル、第(5)列は習熟度別授業を実施している学校に通う児童のサンプルに関する推定結果を報告している。なお、算数の習熟度別授業を実施している学校に通う児童の比率は38.1%である。

まず、習熟度別授業の実施有無によって都市規模の効果および経済的に恵まれない児童の割合の効果が異なっていることが確認される。具体的には、人口50万人以上および10万人から50万人の都市部に立地する学校で得点が高くなる効果、および経済的に恵まれない児童の割合が高い学校において得点が低くなる効果は、習熟度別授業の実施校では効果が縮小し、統計的有意性を失うという結果となっている。また、1~3月の早生まれの児童の得点が低くなる効果は、習熟度別授業の未実施校では約3点であるのに対して、実施校では約1.6点に縮小しており、この差は限界的にはあるが統計的に有意である(p-value=0.103)。習熟度別授業の重要な目的は、児童・生徒の理解度に応じた授業をおこなうことを通じた学力の向上である。この点については第(1)列で報告したように、習熟度別授業の実施が算数の得点を直接的に上昇させる効果は確認されていない。しかしながら、習熟度別授業の実施が早生まれの不利のみならず居住地や親の経済力の影響を小さくしているとすれば、世代を通じた格差の固定化を和らげるという重要な効果を示唆している可能性がある。

その他の変数では、授業担当教師の経験年数の効果が異なっている。教職経験年数およびその二乗項は、習熟度別授業を実施している学校のサンプルでのみ統計的に有意となっている。また、習熟度別授業の実施校では教室の不足感を示す変数が統計的に有意となっているが、これは習熟度別授業の実施によって多くの教室が必要になったことを示している可能性がある。

4.2 中学2年生サンプルの推定結果

次に、表4の第(1)列で報告されている中学2年生の数学の点数を被説明変数とした推定結果を確認する。生まれ月の効果は、1~3月生まれおよび10~12月生まれの場合に負で統計的に有意となっているが、小学4年生と比較するとその効果はやや小さくなっているといえる。同様に、家庭の蔵書数や所有物の効果もやや小さくなっているが、下に示すよ

うに、こうした変化は説明変数の追加によるものである。

中学 2 年生のサンプルでは、両親の学歴に関する情報が利用可能である。父親、母親について、それぞれ高卒未満を基準としたダミー変数を作成し説明変数として加えている。父親学歴の効果は明瞭である。父親の学歴が高い生徒ほど得点が高くなる傾向が確認され、特に父親が大卒以上である場合の効果が大きい。一方、母親学歴の効果はやや異なっており、母親が短大・高専卒の場合の効果が最も大きくなっている。いずれにしても、この結果は両親の学歴が生徒の点数と強く相関しており、学力の階層差を追認するものと解釈することは可能であるが、分析の精度に関しては以下の 2 点を踏まえるとやや慎重にならざるを得ない。第 1 に、前節で述べたとおり、両親の学歴に関する情報は生徒に対する質問紙から得られたものであり、両親本人から得られたものではないという点である。そのため、仮に生徒が両親の学歴を誤って認識していた場合には測定誤差が発生することになる。第 2 に、生徒が両親の学歴を知らない場合には「わからない」という選択肢が用意されているが、この選択肢を選んだ生徒が母親学歴で約 25%、父親学歴で約 30%存在している点である。これは無視できない大きさであるため、父母それぞれについて「わからない」を独立したダミー変数として加えているが、こうした方法が推定結果に影響している可能性も考えられる¹⁵。

次に、教師や学校の特徴に関する説明変数の効果を確認する。統計的に有意な推定値となっているのは、女性教師のダミー変数と、経済的に恵まれない生徒の比率（25%超）である。一方で、教師の経験年数、教材等の不足感、学年生徒数は統計的に有意な推定値にはなっていない。

興味深い結果となっているのは、近年導入が進んでいる教育上の取り組みに関する説明変数の効果である。学級規模の変数は統計的に有意ではないが、習熟度別授業の実施は正で有意な推定値となっている。具体的には、習熟度別授業を実施している学校に通う生徒は未実施の学校に通う生徒に比べて約 0.8 点、統計的に有意に得点が高くなっている。小学 4 年生の算数では有意な効果が確認されなかったことと比較すると、中学校段階の学習においては、習熟度別授業の実施が得点の上昇に直接結びついていることを示唆している。

次に、生徒の性別によってサンプルを分割した結果を確認する。第 (2) 列は男子、第 (3) 列は女子生徒に関する推定結果である。まず、生まれ月の効果が男女で異なっていることが読み取れる。具体的には女子生徒サンプルの推定において、1~3 月の早生まれおよび 10~12 月生まれによる負の効果が統計的有意性を失っている。両親学歴の効果については、女子生徒のほうが母親学歴の効果がやや強いようであるが、男女間に顕著な差は確認されていない。学校の効果では、学校所在地の都市規模に男女差が確認されている。

学級規模の効果は、男女ともに統計的に有意ではない。一方で、習熟度別授業の効果は

¹⁵ そもそも、両親の学歴を認識し回答していること自体が生徒の能力を反映している可能性がある。また、両親（生みの親）と調査時点の保護者（育ての親）が異なっている可能性もあるが、今回使用した TIMSS のデータでは分析不可能である。

男女で異なっている。具体的には、男子生徒では習熟度別授業の実施が正で統計的に有意な推定値となっているのに対し、女子生徒では推定値が小さくなり、有意性を失うという結果となっている。男子生徒の場合、習熟度別授業の実施によって約 1.3 点、得点が高くなっている。ピア・グループ効果には男女差は確認されていない。

続いて、習熟度別授業の実施有無によってサンプルを分割した結果を確認する。第（4）列は習熟度別授業を実施していない学校、第（5）列は習熟度別授業を実施している学校に通う生徒のサンプルによる推定結果である。なお、習熟度別授業を実施している学校に通う生徒の比率は約 40% である。ここで興味深い結果として、習熟度別授業の実施校に通う生徒では、生まれ月によるマイナスの効果が統計的な有意性を失っていることが挙げられる。習熟度別授業の実施が、早生まれの生徒が抱える不利を解消している可能性を示唆するものと言える。同様に、両親学歴の効果に関しても、母親学歴の効果が習熟度別授業の実施校では統計的な有意性を失っており、親の学歴が子どもの学力達成に及ぼす影響が小さくなっている（ただしこの差は統計的に有意ではない）。また、経済的に恵まれない生徒の割合については、その割合が 25% を超える場合の負の効果の有意性が習熟度別授業の実施校で弱くなっていることが確認される。こうした変化は、小学 4 年生のサンプルで確認されているものと同様のものである。習熟度別授業の実施が、児童・生徒の理解度の向上という本来の目的に加え、本人固有の要因や社会経済的要因の影響を薄めるという効果を持つとすれば、その影響は公平性の観点からも無視できないものとなる可能性がある。

教師や学校に関する変数では、学校としての教師不足感や教室等施設不足感の変数が習熟度別授業実施校で統計的に有意となっているが、これらの変数の解釈は慎重を要する。例えば、教師不足感を感じている学校に通う生徒ほど数学の点数が高いという結果になっているが、これは、習熟度別授業の実施によってより多くの教師が必要となったことを示しているのかもしれない。

また、ピア・グループ効果についても実施校と未実施校で効果に差が確認されている。具体的には、習熟度別授業の未実施校では統計的に有意なピア・グループ効果が確認されているのに対し、実施校では統計的な有意性を失っている。習熟度別に授業をおこなうことで均質な学習集団が形成され、生徒相互間の影響度合いが薄れているのかもしれないが、この点についてはより詳細な検証が必要であろう。

5. 結論および今後の課題

本稿では、日本の小学 4 年生と中学 2 年生を対象に、算数・数学の教育生産関数を推定し、児童・生徒の学力を規定する要因について分析をおこなった。日本の教育生産関数の推定はこれまでの研究蓄積が少ない。そのため本稿は、この分野における基礎的な分析結果を提供することを目的とし、欧米の先行研究を参考にしながら可能な限り必要な説明変数を分析に加えることとした。同時に、近年、日本の小中学校で導入が進んでいる少人数

指導や習熟度別授業の効果についても考察した。扱うデータが一時点のものであるという限界があるため結果の解釈に慎重を要することは、先に述べたとおりである。

本稿の実証分析から、以下の3点の結果が確認された。まず第1に、児童・生徒本人やその家庭の要因が、得点で示される学力に及ぼす影響が強いことである。具体的には、本人の生まれ月（早生まれ）による影響と、家庭の蔵書数や所有物、両親の学歴、居住地域の経済力による影響である。こうした結果は欧米の先行研究の結果を追認するものであり目新しいわけではないが、生まれ育った家庭や地域の環境が子どもの学力に強く影響するという実証結果は、格差の固定化が懸念されている日本の現状においては無視できないものと言える。

第2点目として、学校や教師への投入物に関する説明変数の影響が、全体としてそれほど大きいものではないという点が挙げられる。この結果も、学校に対する物的・人的資源投入の成績に対する効果が統計的に確認されない、という欧米の先行研究の結果を追認するものであるといえる。なお本稿では、日本の公立小中学校において教師の年齢構成に偏りがある点に注目し、近い将来、経験豊かなベテラン教師が教育現場を去ることによる影響を考察した。いくつかの推定において教職経験年数の変数が正で統計的に有意であることが確認されたが、その効果はそれほど頑健なものではなかった。したがって、ベテラン教師の大量退職による子どもの学力への影響は限定的であろうというのが本稿の結果が示唆するところである。

第3に、近年、日本の小中学校で導入が進んでいる少人数指導や習熟度別指導の効果を検証したところ、いくつかの興味深い結果が得られた。小学4年生に関しては、学級規模、習熟度別指導ともに児童の得点に有意な影響を与えていないことが確認された。一方で、中学2年生に関しては、男子生徒のサンプルにおいて習熟度別授業の効果が正で統計的に有意となっていることが確認された。中学校段階の学習内容になると、生徒間の理解度の差が大きくなることが影響しているのかもしれない。さらに、算数・数学に関して習熟度別授業を実施している小中学校では、早生まれによる不利や経済的に恵まれない児童生徒の比率が高いことによる負の効果が縮小していることが明らかとなった。習熟度別授業の実施によって、生まれ月による不利や生まれ育った家庭環境・社会経済環境の影響が薄まる効果を示唆しており、世代を通じた格差の是正や公平性の観点からも興味深い結果であると言える。また、生徒相互間のピア・グループ効果は習熟度別授業の未実施校でのみ統計的に有意であり、実施校では効果が確認されなかった。こうしたことを合わせて考慮すれば、将来的には、学級規模、ピア・グループ、習熟度別授業の相互的な影響についてのより詳細な分析が必要であると考えられる。

最後に、本稿の分析の限界と今後の課題について述べる。まず、繰り返し述べているように、本稿の分析に用いたデータは一時点のものであるため、本稿の結果が因果関係を示すものではない場合があることに注意が必要である。教育の成果は長年にわたって蓄積され、そうした積み重ねが学力を決定することは言うまでもない。そうした側面を分析する

ためには、長期間にわたって同一児童・生徒を追跡したデータを用いることが理想であり、少なくとも同一児童・生徒の 2 時点のデータを用いることが必要である。同時に、児童・生徒が通った学校に関する時系列のデータも必要である。そのようなデータの利用可能性が拡大することを期待するものである。

次に、本稿の分析で考慮していない 2 つの要因についてである。1 つは、学習塾や家庭教師といった学校外学習の効果を考慮していない点である。これは、TIMSS 2007 では児童・生徒の学校外での学習状況を詳しく調査していないためである。学習塾と成績の関係は先行研究でも分析されており、その重要性は否定できない。もう 1 つは、私立学校を分析に含めていない点である。私立学校への進学、さらには近年導入が進みつつある学校選択制によって、児童・生徒の入学時点の学力に何らかの変化が生じている可能性もあり、今後は教育生産関数の推定においてもそうした側面を考慮する必要があると考えられる¹⁶。

最後に、児童・生徒の家庭的背景に関するデータの信頼性である。本稿で用いた TIMSS の調査データは、児童・生徒本人に対する質問紙によるものであるため、例えば家庭の所得水準や親の職業についての情報は得られない。また、両親の学歴や家庭の蔵書数に測定誤差が発生している可能性も考えられる。こうした家庭的要因が子どもの学力達成に及ぼす影響が強いことを考えると、将来的な調査においては、保護者を対象とした質問紙調査が行われることが望ましいであろう。

¹⁶ 公立高等学校の学区変更による効果を分析した Akabayashi (2006) などがある。

参考文献

- 上野有子・三野孝一郎・小塩隆士・佐野晋平「学力調査結果からみた学校選択制、少人数指導、習熟度別指導の効果に関する実証分析」経済財政分析ディスカッション・ペーパー, 2007, DP/07-1.
- 岡田典子・櫻田裕美子・山崎博敏「教員から見た学級規模の教育的効果」『日本教育大学協会会報』2001, 第82号, pp. 49-63.
- 小塩隆士・佐野晋平・末富芳「教育の生産関数の推定—中高一貫校の場合」『経済分析』2009, 第182号, pp. 48-69.
- 苅谷剛彦『階層化日本と教育危機』2001, 有信堂高文社.
- 苅谷剛彦『学力と階層』朝日新聞出版, 2008.
- 川口俊明・前馬優策「学力格差を縮小する学校—「効果のある学校」の経年分析に向けて」『教育社会学研究』2007, 第80集, pp. 187-205.
- 川口俊明「マルチレベルモデルを用いた「学校の効果」の分析—「効果的な学校」に社会的不平等の救済はできるのか」『教育社会学研究』2009, 第84集, pp. 165-184.
- 国立教育政策研究所「学級規模に関する調査研究」『国立教育政策研究所紀要』2001, 第131集.
- 篠崎武久「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』2008, 第7章, pp. 73-97.
- 志水宏吉「低学力克服への戦略—「効果のある学校」論の視点から」苅谷剛彦・志水宏吉編『学力の社会学』岩波書店, 2004, pp. 217-235.
- 鍋島祥郎『効果のある学校』解放出版社, 2003.
- 舞田敏彦「地域の社会経済特性による子供の学力の推定—学力の社会的規定性を克服する教育条件の探求」『教育社会学研究』2008, 第82集, pp. 165-184.
- 文部科学省『確かな学力の向上のための2002アピール』2002.
- 文部科学省『学校教員統計調査』2007.
- Ammermueller, A. and J. Pischke “Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study,” *Journal of Labor Economics*, 2009, vol. 27, no. 3, pp. 315-348.
- Akabayashi, H. “Average Effects of School Choice on Educational Attainment: Evidence from Japanese High School Attendance Zones,” Working Paper, 2006.
- Bedard, K. and E. Dhuey “The Persistence of Early Childhood Maturity: International Evidence of Long-run Age Effects,” *Quarterly Journal of Economics*, 2006, 121 (4), pp. 1437-1472.
- Coleman, J. S. et al. *Equality of Educational Opportunity*. Washington, DC: U. S. Government Printing Office, 1966.
- Edmonds, R. R. “Effective Schools for the Urban Poor,” *Educational Leadership*, 1979, 37(1), pp. 15-24.

- Hanushek, E. A. "School Resources," in *Handbook of Economics of Education*, Vol. 2, eds. by Hanushek E. A. and F. Welch, North-Holland, 2006, pp. 865-908.
- Hanushek, E. A., J. F. Kain, J. M. Markman and S. G. Rivkin "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics*, 2003, vol. 18, pp. 527-544.
- Hoxby, C.M. "The Effects Of Class Size On Student Achievement: New Evidence From Population Variation," *Quarterly Journal of Economics*, 2000, vol. 115(4), pp.1239-1285
- Kawaguchi, D. "Actual Age at School Entry, Educational Outcomes, and Earnings," forthcoming in *Journal of the Japanese and International Economies*.
- Kang, C. "Academic interactions among classroom peers: a cross-country comparison using TIMSS," *Applied Economics*, 2007, vol. 39, pp. 1531-1544.
- Manski, C. F. "Identification of endogenous social effects: The reflection problem," *Review of Economic Studies*, 60, pp. 531-542.
- Oshio, T. and W. Senoh "The Economics of Education in Japan: A Survey of Empirical Studies and Unresolved Issues," *The Japanese Economy*, 2007, 34 (1), pp.46-81.
- Pritchard, I. "Reducing Class Size: What Do We Know?" Technical Report, Office of Educational Research and Improvement, U.S. Department of Education, available online at <http://www.ed.gov/pubs/ReducingClass/index.html>
- Woessmann, L. "Schooling Resources, Educational Institutions and Student Performance: the International Evidence," *Oxford Bulletin of Economics and Statistics*, 2003, vol.65, pp.117-170.
- Woessmann, L. and M. West "Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS," *European Economic Review*, 2006, vol.50, pp.695-736.

1

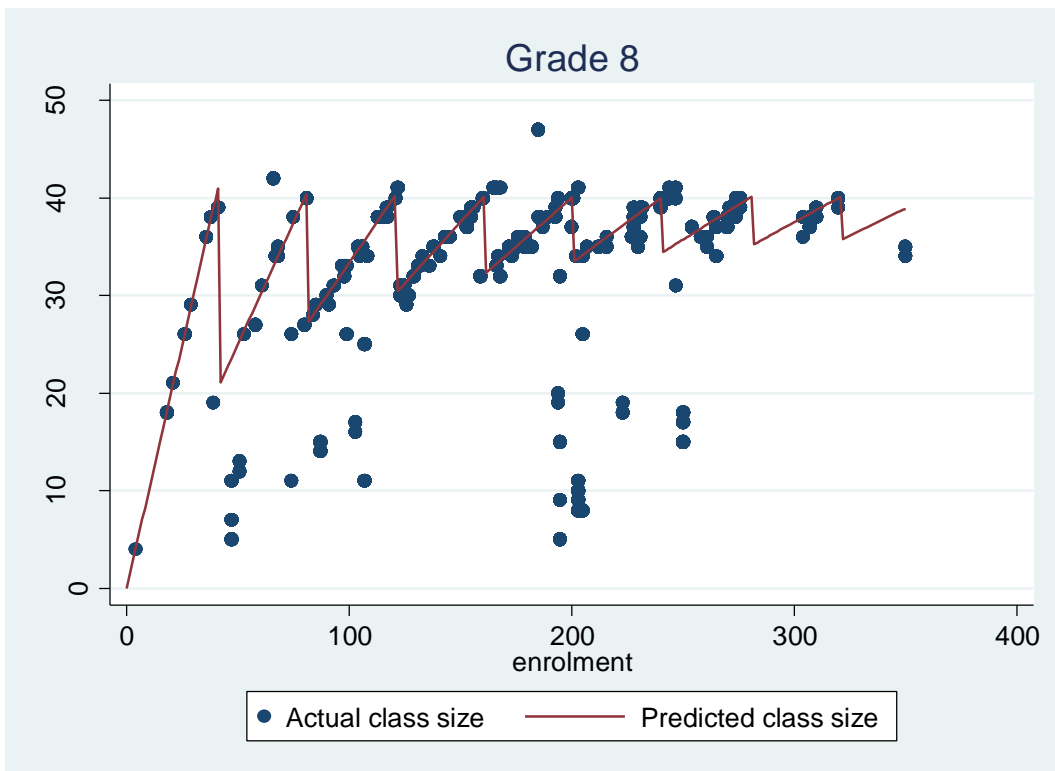
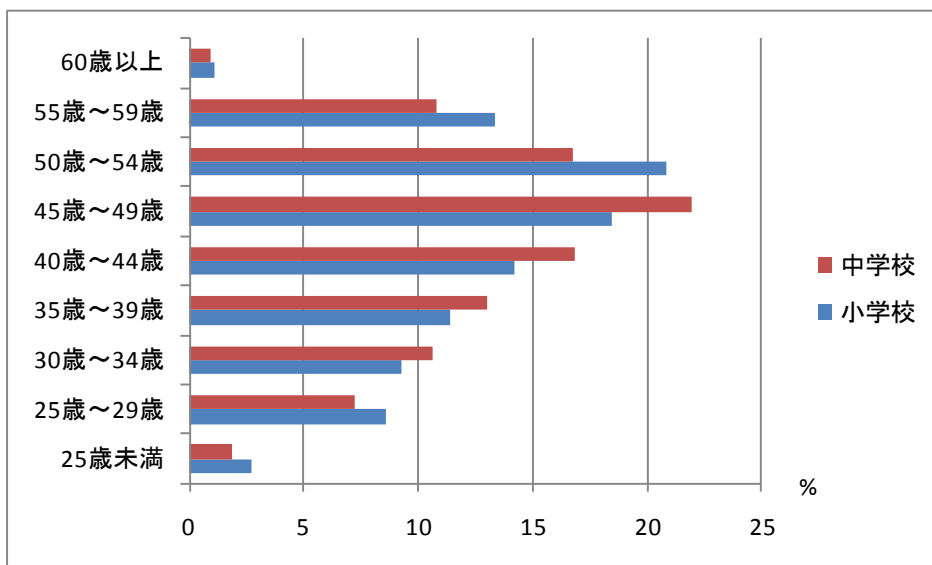


図2 小中学校における教師年齢構成



(出所) 文部科学省 (2007) 『学校教員統計調査』より筆者作成。

Table 1. Average Score of Japanese Students

	1995	1999	2003	2007
4th graders				
mathematics	567	–	565	568
science	553	–	543	548
8th graders				
mathematics	581	579	570	570
science	554	550	552	554

Table 2. Descriptive Statistics

	4th grade (N=4963)				8th grade (N=4602)			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Mathematics test score	150.15	9.88	108.02	179.91	149.13	9.37	106.96	186.77
girls	0.49	0.50	0	1	0.49	0.50	0	1
Month of birth								
April–June	0.26	0.44	0	1	0.25	0.43	0	1
July–September	0.27	0.44	0	1	0.27	0.44	0	1
October–December	0.24	0.43	0	1	0.24	0.43	0	1
January–March	0.24	0.43	0	1	0.24	0.43	0	1
Number of books at home								
0–10	0.14	0.35	0	1	0.16	0.36	0	1
11–25	0.28	0.45	0	1	0.22	0.41	0	1
26–100	0.37	0.48	0	1	0.33	0.47	0	1
101–200	0.13	0.34	0	1	0.16	0.36	0	1
more than 200	0.07	0.26	0	1	0.15	0.36	0	1
Study desk at home	0.94	0.24	0	1	0.95	0.22	0	1
Computer at home	0.82	0.39	0	1	0.87	0.33	0	1
Globe at home	0.41	0.49	0	1	0.38	0.49	0	1
Teacher's education: master degree	0.03	0.17	0	1	0.05	0.22	0	1
Teacher: female	0.61	0.49	0	1	0.41	0.49	0	1
Years of teaching	18.50	10.78	1	39	15.47	9.33	1	35
School location								
more than 500,000 people	0.22	0.42	0	1	0.21	0.41	0	1
100,001–500,000 people	0.53	0.50	0	1	0.41	0.49	0	1
50,001–100,000 people	0.13	0.34	0	1	0.18	0.38	0	1
15,001–50,000 people	0.10	0.30	0	1	0.17	0.37	0	1
less than 15,000 people	0.02	0.13	0	1	0.03	0.18	0	1
Percentage of students from economically disadvantaged homes								
0–10%	0.66	0.47	0	1	0.52	0.50	0	1
11–25%	0.24	0.43	0	1	0.38	0.48	0	1
more than 25%	0.10	0.30	0	1	0.10	0.30	0	1
enrollment in the 4th or 8th grade	101.12	40.82	5	216	171.77	69.19	4	350
shortage: instructional materials	0.08	0.28	0	1	0.13	0.34	0	1
shortage: classrooms	0.17	0.38	0	1	0.34	0.47	0	1
shortage: teachers	0.08	0.27	0	1	0.11	0.31	0	1
class size	31.88	6.37	2	40	32.65	8.49	4	41
peer mean score	149.95	3.01	137.88	159.72	149.02	2.89	138.14	155.69
grouped by ability	0.38	0.49	0	1	0.40	0.49	0	1
Mother's education								
junior high school	–	–	–	–	0.03	0.16	0	1
high school	–	–	–	–	0.35	0.48	0	1
junior college	–	–	–	–	0.21	0.41	0	1
university or more	–	–	–	–	0.16	0.37	0	1
unknown	–	–	–	–	0.25	0.44	0	1
Father's education								
junior high school	–	–	–	–	0.05	0.21	0	1
high school	–	–	–	–	0.28	0.45	0	1
junior college	–	–	–	–	0.08	0.27	0	1
university or more	–	–	–	–	0.28	0.45	0	1
unknown	–	–	–	–	0.32	0.47	0	1

Table 3. Regressions for Mathematics Test Score for 4th Graders										
	(1) All sample		(2) Boys		(3) Girls		(4) Mixing		(5) Grouping	
girl	0.255	(0.335)					0.117	(0.413)	0.496	(0.601)
July–September	-0.211	(0.387)	-0.397	(0.599)	0.065	(0.576)	-0.032	(0.512)	-0.551	(0.517)
October–December	-1.458	(0.391)***	-1.872	(0.592)***	-0.944	(0.637)	-1.355	(0.519)***	-1.765	(0.490)***
January–March	-2.561	(0.376)***	-2.151	(0.620)***	-2.875	(0.597)***	-2.949	(0.447)***	-1.648	(0.675)**
number of books at home (11–25)	3.402	(0.491)***	3.963	(0.762)***	2.513	(0.647)***	3.118	(0.614)***	4.249	(0.814)***
number books at home (26–100)	5.583	(0.530)***	5.777	(0.794)***	5.054	(0.622)***	4.901	(0.624)***	7.296	(0.944)***
number books at home (101–200)	8.126	(0.625)***	8.926	(0.846)***	6.948	(0.800)***	7.585	(0.733)***	9.876	(1.089)***
number books at home (more than 200)	7.450	(0.824)***	8.016	(1.163)***	6.573	(1.172)***	7.227	(0.999)***	8.466	(1.320)***
study desk at home	1.007	(0.682)	0.278	(0.948)	2.338	(0.914)**	1.393	(0.927)	0.575	(1.009)
computer at home	2.704	(0.444)***	3.092	(0.651)***	2.395	(0.560)***	2.234	(0.573)***	3.394	(0.755)***
globe at home	0.713	(0.321)**	1.064	(0.459)**	0.371	(0.428)	0.679	(0.396)*	0.719	(0.534)
teacher: master degree	0.459	(0.911)	-0.112	(1.178)	1.108	(1.354)	0.604	(1.188)	0.662	(0.839)
teacher: female	0.369	(0.272)	0.361	(0.385)	0.379	(0.435)	0.526	(0.340)	0.034	(0.582)
years of teaching	-0.041	(0.045)	-0.044	(0.071)	-0.049	(0.068)	0.026	(0.053)	-0.278	(0.105)***
years of teaching squared	0.161	(0.126)	0.211	(0.200)	0.139	(0.187)	0.034	(0.147)	0.743	(0.313)**
school location: more than 500,000 people	0.703	(0.492)	-0.290	(0.753)	1.841	(0.841)**	1.145	(0.584)*	-0.214	(1.163)
school location: 100,001–500,000 people	0.700	(0.483)	0.636	(0.683)	0.933	(0.758)	1.010	(0.550)*	0.222	(1.132)
school location: 50,001–100,000 people	0.326	(0.501)	-0.742	(0.683)	1.563	(0.804)*	0.442	(0.601)	0.390	(1.202)
economically disadvantaged students: 11–25%	-0.499	(0.381)	0.055	(0.496)	-0.939	(0.527)*	-0.365	(0.432)	-1.015	(0.939)
economically disadvantaged students: more than 25%	-1.124	(0.482)**	-1.046	(0.613)*	-0.969	(0.827)	-1.544	(0.631)**	-0.799	(0.984)
number of enrollment in the 4th grade	0.239	(0.493)	0.145	(0.621)	0.301	(0.664)	0.096	(0.582)	1.366	(1.119)
shortage: instructional materials	-0.479	(0.717)	-0.606	(1.178)	-0.358	(1.061)	-0.012	(0.934)	-0.410	(1.721)
shortage: classrooms	0.402	(0.356)	1.074	(0.557)*	-0.174	(0.608)	-0.126	(0.466)	1.533	(0.616)**
shortage: teachers	0.989	(0.669)	1.476	(0.985)	0.581	(0.800)	0.803	(0.776)	1.261	(1.412)
class size (instrumented by predicted value)	-0.668	(1.511)	0.823	(2.203)	-2.424	(2.200)	-0.672	(1.821)	-1.146	(4.525)
peer mean	0.350	(0.061)***	0.429	(0.079)***	0.297	(0.086)***	0.280	(0.097)***	0.304	(0.125)**
grouped by ability	0.053	(0.323)	-0.083	(0.471)	0.187	(0.518)				
constant	90.967	(10.320)***	74.375	(13.239)***	104.320	(13.504)***	101.615	(16.704)***	95.215	(14.719)***
R-squared	0.128		0.143		0.124		0.117		0.171	
Observations	4963		2518		2445		3073		1890	

Note: Weighted least squares regressions using students' sampling probability as weight. Standard errors are reported in parentheses and robust to clustering at the school level.

* p<0.10, ** p<0.05, *** p<0.01

Table 4. Regressions for Mathematics Test Score for 8th Graders										
	(1) All sample		(2) Boys		(3) Girls		(4) Mixing		(5) Grouping	
girl	-0.344	(0.293)					0.013	(0.355)	-0.992	(0.514)*
July–September	-0.215	(0.345)	-0.485	(0.582)	-0.043	(0.507)	-0.058	(0.345)	-1.016	(1.171)
October–December	-1.088	(0.365)***	-1.577	(0.603)***	-0.673	(0.563)	-1.096	(0.420)***	-1.385	(0.852)
January–March	-1.350	(0.369)***	-1.979	(0.689)***	-0.864	(0.538)	-1.871	(0.468)***	-0.653	(0.927)
number of books at home (11–25)	1.933	(0.544)***	2.742	(0.787)***	1.352	(0.672)**	2.064	(0.683)***	1.963	(0.948)**
number books at home (26–100)	3.320	(0.519)***	3.665	(0.789)***	3.105	(0.617)***	3.622	(0.595)***	3.060	(1.008)***
number books at home (101–200)	4.098	(0.548)***	4.479	(0.852)***	3.833	(0.673)***	4.345	(0.662)***	3.572	(1.086)***
number books at home (more than 200)	4.290	(0.618)***	4.616	(0.846)***	4.109	(0.812)***	4.890	(0.764)***	3.107	(1.070)***
study desk at home	2.363	(0.849)***	2.373	(1.373)*	2.318	(0.857)***	2.393	(1.155)**	2.297	(1.289)*
computer at home	1.890	(0.454)***	1.662	(0.642)***	2.037	(0.626)***	2.063	(0.570)***	1.456	(1.313)
globe at home	0.747	(0.316)**	0.848	(0.468)*	0.669	(0.411)	0.731	(0.359)**	0.538	(0.805)
mother’s education: high school	2.411	(0.979)**	2.048	(1.657)	2.632	(1.428)*	3.287	(1.169)***	0.257	(2.136)
mother’s education: junior college	3.867	(0.982)***	3.656	(1.758)**	3.955	(1.373)***	4.661	(1.159)***	2.092	(2.097)
mother’s education: university or more	2.965	(1.031)***	2.362	(1.698)	3.471	(1.469)**	3.743	(1.225)***	0.982	(2.343)
mother’s education: unknown	2.137	(0.969)**	2.278	(1.763)	1.590	(1.463)	2.691	(1.188)**	0.731	(1.949)
father’s education: high school	1.694	(0.891)*	1.815	(1.233)	1.707	(1.138)	1.853	(0.909)**	1.664	(2.172)
father’s education: junior college	2.379	(1.029)**	3.357	(1.580)**	1.658	(1.245)	2.220	(1.153)*	2.773	(2.424)
father’s education: university or more	4.953	(0.902)***	5.339	(1.257)***	4.780	(1.179)***	4.927	(0.934)***	5.791	(2.097)***
father’s education: unknown	2.248	(0.914)**	2.446	(1.420)*	2.300	(1.143)**	2.315	(1.035)**	2.554	(1.843)
teacher: master degree	0.710	(0.549)	0.714	(0.652)	0.707	(0.837)	1.506	(0.621)**	1.122	(2.949)
teacher: female	0.350	(0.218)	0.374	(0.329)	0.381	(0.352)	0.307	(0.254)	1.831	(2.829)
years of teaching	0.030	(0.038)	-0.011	(0.079)	0.072	(0.076)	0.037	(0.047)	0.079	(0.152)
years of teaching squared	-0.084	(0.115)	0.070	(0.257)	-0.240	(0.243)	-0.109	(0.141)	0.044	(0.733)
school location: more than 500,000 people	-0.400	(0.311)	0.065	(0.620)	-0.834	(0.470)*	-0.209	(0.409)	-0.721	(4.042)
school location: 100,001–500,000 people	-0.109	(0.296)	0.156	(0.539)	-0.400	(0.462)	0.488	(0.359)	-1.282	(1.295)
school location: 50,001–100,000 people	-0.488	(0.376)	-0.259	(0.580)	-0.647	(0.559)	-0.679	(0.380)*	2.079	(4.450)
economically disadvantaged students: 11–25%	-0.264	(0.240)	-0.700	(0.426)	0.226	(0.340)	-0.268	(0.268)	-2.291	(3.244)
economically disadvantaged students: more than 25%	-0.099	(0.433)	-0.212	(0.752)	0.164	(0.584)	0.191	(0.605)	2.006	(7.058)
number of enrollment in the 8th grade	0.075	(0.326)	-0.007	(0.630)	0.307	(0.537)	-0.377	(0.350)	1.373	(1.259)
shortage: instructional materials	0.395	(0.257)	1.052	(0.686)	-0.111	(0.586)	0.495	(0.225)**	2.072	(6.639)
shortage: classrooms	-0.546	(0.325)*	-0.325	(0.557)	-0.911	(0.475)*	-0.223	(0.337)	-3.834	(3.474)
shortage: teachers	0.245	(0.328)	0.475	(0.778)	0.066	(0.608)	-0.417	(0.284)	4.797	(5.116)
class size (instrumented by predicted value)	-0.979	(1.334)	-1.766	(2.818)	-1.078	(2.035)	0.353	(1.152)	-7.912	(19.547)
peer mean	0.584	(0.056)***	0.571	(0.069)***	0.603	(0.066)***	0.583	(0.055)***	0.289	(0.432)
grouped by ability	0.113	(0.301)	0.447	(0.577)	-0.419	(0.427)				
constant	53.665	(8.008)***	58.568	(11.029)***	49.960	(9.766)***	49.730	(8.524)***	116.263	(28.422)***
R-squared	0.215		0.178		0.263		0.254		0.055	
Observations	4602		2337		2265		2744		1858	

Note: Weighted least squares regressions using students’ sampling probability as weight. Standard errors are reported in parentheses and robust to clustering at the school level.

* p<0.10, ** p<0.05, *** p<0.01