

博士論文2021年度

教員の質の定量的測定および  
その形成要因に関する研究

慶應義塾大学大学院政策・メディア研究科

氏名 伊藤 寛武

# 主 論 文 要 旨

No.1

報告番号	甲 乙 第	号	氏 名	伊藤 寛武
主 論 文 題 目 :				
教員の質の定量的測定およびその形成要因に関する研究				
(内容の要旨)				
<p>教育政策を巡る議論では教員の資質向上が強く要請され、そのために多くの施策が実施されている。しかし、これらの教育施策が本当に教員・学校の質向上に結び付くのかについて十分に根拠が存在するわけではない。それどころか、子どもにとってどのような教員や学校が良いのかという問いについて、未だに研究間で結果が一致するような結果は得られていない。その一方で、教員が子どもの能力成長に強い影響力を持っているとする研究成果は強固かつ頑健である。</p> <p>そのような研究状況の中で、本稿では教員の有効性について主に日本の小学生・中学生のデータを用いながら多角的な検討を行った。</p> <p>第一に、教員の違いが子どもの成長に与える影響について日本のある市町村のデータを用いて分析した。その結果として学力や学習方略に対して個々の教員の違いが与える影響力が取りうる値の範囲は広がった一方で、非認知能力や学級での子ども・教員間の関係性に対して教員の違いは子ども全体の変動のうち少なくとも 5%から 8%以上を説明していたことを示した。さらに、教員付加価値間の相関係数から、学力を伸ばすことが得意な教員と非認知能力・学習方略を伸ばすことが得意な教員は異なるという結果を得た。以上の結果から、教員の違いは学力より非認知能力に対して重要である可能性が示唆された。</p> <p>第二に、日本のある自治体のデータを用いて効果的な指導方法について分析した。しかしどのような指導が授業で行われるかは必ずしも外生的に決定されるわけではなく、学校が置かれた状況によって内生的に決定される。本稿では時間変化しない観察不可能な生徒間異質性および時間変化する観察不可能な生徒間異質性をコントロールすることで、その内生性への対処を行った。結果として、子どもの自発性を促進する指導や授業手法を工夫すること・子どもの授業態度への指導は小学生の学力に対して正の効果を持った。</p> <p>第三に、指導方法の具体例としてコンピューター支援教育(CAI)をとりあげ、カンボジアで行った実証実験を通じてその効果を分析した。その結果、認知能力に対する平均処置効果は肯定的で統計的に有意である一方で、学習時間には変化がなかったことが示</p>				

唆された。これは、CAI が子どもの単位時間当たりの学習生産性を向上させることに成功していることを示している。さらに、CAI は子どもの大学進学に対する主観的な期待感を高めることがわかった。

キーワード： 教員、教員付加価値、指導方法、コンピューター支援教育

# 主 論 文 要 旨

No.1

報告番号	甲 乙 第	号	氏 名	ITO HIROTAKE
主 論 文 題 目 :				
Essays on the quality of the teacher and the factors				
(内容の要旨)				
<p>In the discussion over the education policy, there is a strong call for improving the quality of teachers, and many policies have been applied to it. However, there is not enough evidence to prove that these educational policies lead to the improvement of the quality of teachers and schools. On the contrary, there is still no consensus among studies on the question of what kind of teachers and schools are good for children. On the other hand, the research findings that teachers have a strong influence on children's ability to grow are strong and robust.</p> <p>In such a research context, this paper examines the effectiveness of teachers from multiple perspectives, mainly using data from Japanese elementary and junior high school students.</p> <p>First of all, we analyzed the impact of different teachers on children's growth using data from an anonymous municipality in Japan. Specifically, we estimated the teacher's value-added using individual data, and descriptively analyzed the impact of teacher's abilities on the growth of children. In this study, we estimated the value-added of teachers not only in terms of academic achievement but also in terms of learning strategies, non-cognitive abilities, and relationships between children and teachers in the classroom. The results showed that the range of possible values for the influence of individual teacher differences on academic achievement and learning strategies was wide, while teacher differences in non-cognitive abilities and classroom child-teacher relationships explained at least 5% to 8% of the total variation in children. Furthermore, the correlation coefficients among the teacher value-added indicated that teachers who were good at developing academic skills differed from those who were good at developing non-cognitive skills and learning strategies. These results suggest that differences among teachers may be more important for non-cognitive skills than academic skills.</p> <p>Second, we analyzed the effective teaching methods. Here, we estimate the impact of instructional methods on children's educational outcomes. However, the type of instruction that takes place in the classroom is not necessarily determined exogenously, but rather endogenously by the circumstances in which the school is located. In this paper, we address this endogeneity by controlling for time-invariant unobservable student heterogeneity and time-variant observable student heterogeneity. As a result, teaching that promotes children's spontaneity, devising teaching methods, and teaching children's classroom attitudes had a positive effect on the academic performance of elementary school students.</p> <p>Thirdly, as a concrete example of the teaching methods, we took up computer-assisted instruction (CAI) and analyzed its effects through an experiment carried out in Cambodia. A clustered randomized controlled trial was conducted in five elementary schools near Phnom Penh, Cambodia, with approximately 1,600 children. The results suggest that the treatment effect on cognitive ability was positive and statistically significant while there was no change</p>				

in learning time. This indicates that the CAI has been successful in improving children's learning productivity per unit of time. Besides, CAI was found to increase children's subjective expectations of going to college.

Keyword: teacher, teacher value-added, teaching method, computer-assisted instruction

# 目次

<b>第 1 章</b>	<b>問題意識と本稿の概要</b>	<b>3</b>
1.1	現代の教員施策とその背景	3
1.2	教員の質研究の現状と課題	6
1.3	教員の指導方法についての研究の現状と課題	7
1.4	コンピューター支援教育の効果	8
1.5	おわりに	9
<b>第 2 章</b>	<b>教員の有効性研究のこれまでの展開</b>	<b>11</b>
2.1	効果的な教員・学校の特徴	11
2.1.1	教員認証	12
2.1.2	教員の経験年数	13
2.1.3	教員の性別	14
2.1.4	教員の給与	15
2.1.5	学校の特徴	17
2.2	教員固定効果	18
2.2.1	教員固定効果研究が抱える限界	19
2.3	効果的な指導方法	20
<b>第 3 章</b>	<b>教員付加価値から見た教員の役割について - 日本の小学生を例にして -</b>	<b>30</b>
3.1	はじめに	30
3.2	先行研究	32
3.3	データ	34
3.3.1	セルフコントロール	37
3.3.2	自己効力感	37
3.3.3	学習方略	38
3.4	推定戦略と推定結果	44
3.4.1	教員付加価値の推定	44
3.4.2	教員付加価値の分布	49
3.4.3	教員付加価値推定の頑健性	52
3.4.4	教員付加価値間の相関	54
3.4.5	教員の属性と教員付加価値	56
3.5	結論	59

---

<b>第4章</b>	<b>子どもの教育成果に対する指導方法の効果の検証 -時間変化する要因を考慮した分析-</b>	<b>66</b>
4.1	はじめに . . . . .	66
4.2	データ . . . . .	69
4.2.1	学力 . . . . .	69
4.2.2	指導方法 . . . . .	69
4.3	推定戦略 . . . . .	72
4.3.1	教育生産関数の分解 . . . . .	72
4.3.2	推定戦略 . . . . .	74
4.4	分析結果 . . . . .	76
4.5	結論 . . . . .	78
<b>第5章</b>	<b>コンピューター支援教育は子どもの能力を向上させるか？</b>	<b>83</b>
5.1	はじめに . . . . .	83
5.2	先行研究 . . . . .	85
5.3	手法とデータ . . . . .	86
5.3.1	背景 . . . . .	86
5.3.2	ベースライン調査とフォローアップ調査 . . . . .	86
5.3.3	教育アプリケーション: Think!Think! . . . . .	89
5.3.4	Clustered RCT . . . . .	91
5.4	識別戦略および推定結果 . . . . .	92
5.4.1	推定に用いた定式化 . . . . .	92
5.4.2	変数の詳細 . . . . .	92
5.4.3	推定結果 . . . . .	93
5.5	結論 . . . . .	101
<b>第6章</b>	<b>おわりに</b>	<b>106</b>
	<b>謝辞</b>	<b>109</b>

# 第1章 問題意識と本稿の概要

## 概要

本章では本博士論文全体を概観する。教育政策を巡る議論では教員の資質向上が強く要請され、そのために多くの施策が実施されている。しかしこれまでの教員研究においては、「効果的な教員の具体的な特徴は何か」という問いに対して、研究によって異なる結論が導かれておりいまだに議論が続いている。その一方で、教員が子どもの成長に対して強い影響力を持っているとする研究は多い。そのような研究状況の中で、本稿は主に日本におけるデータを用いながら教員の有効性について多角的な検討を行った。本稿は3編の論文から構成される。第3章では、教員の違いが子どもの成長に与える影響について日本のある市町村のデータを用いて分析し、教員の違いが子どもの認知能力や非認知能力に対して与える影響を定量的に測定し、その記述的な分析を行った。そして、効果的な教員の具体的な特徴として教員の指導方法が重要であると指摘する研究が増えていることをうけ、第4章では日本のある都道府県のデータを用いて指導方法について分析し、幾つかの指導方法インデックスが子どもの学力に対して有意な影響を持っていることを指摘した。第5章では指導方法の具体例としてコンピューター支援教育をとりあげ、カンボジアで行った実証実験を通じてその効果を分析し、コンピューター支援教育は伝統的な教員による授業と比較してより高い効果を持ったことを指摘した。

### 1.1 現代の教員施策とその背景

近年、日本の教育政策を巡る議論では教員や学校教育の資質向上が強く要請されてきた。例えば、文科省中央教育審議会において2015年12月に「これからの学校教育を担う教員の資質能力の向上について」と題された答申がまとめられ(文部科学省, 2015)、そこでは「新たな知識や技術の活用により社会の進歩や変化のスピードが速まる中、教員の資質能力向上は我が国の最重要課題であり、世界の潮流でもある」と述べられている。この見解は文科省の教育施策の方針の一つでもあり、その下で教員免許制改革などの施策が検討されてきた。このように、子どもの能力を向上させるためには学校教育

の充実が必要であり、その中でも特に子どもと多大な時間を共にする教員の資質が重要な役割を持っているのではないかと考えられてきた。

このように教員や学校の質の向上が要請されてきた背景には少なくとも2つの要因が考えられる。第1の要因は、社会変化によって子どもに求められる能力が変化していることである。社会・経済のグローバル化や急速なIT化の進展によって、プログラミング能力などこれまでの学校教育では焦点があたってこなかった能力が今後の人的資本として重要になる可能性がある。また、変化の多い社会の下では、その変化に対応するためのコミュニケーション能力や課題解決能力といった学力に限らない（「生きる力」という言葉に代表される）多元的な能力が必要になるのではないかと考えられてきた。そして、このように子どもに必要とされる能力に変化があるならば、その子どもを教育する立場にある教員に求められる資質もまた変化があると考えられる。このような考えを反映して、平成24年8月28日に開かれた文科省中央教育審議会の答申では「社会の急激な変化に伴い、高度化・複雑化する諸課題への対応が必要となっており、学校教育において、求められる人材育成像の変化への対応が必要である」と述べられている（文部科学省, 2012）。

第2の要因は、資質の高い教員の継続雇用の可否に対する危機意識が高まっていることである。すなわち、今後労働市場において相対的に能力の高い人材を雇用することができなくなる恐れがある。その例の1つが教員の新卒採用である。図1.1では公立学校の教員採用試験の状況を年度別にプロットした。横軸は年度を表しており、縦軸は採用数（左軸）や競争倍率（右軸）を表している。図からは、採用試験における倍率が平成12年度の13.3倍をピークに下がり続けていることを確認することができる。また教員採用受験者の数も昭和54年以降下落傾向にあり、平成4年に最低値をつけたあと、その後その低い水準のまま推移している。すなわち、現代において教員採用の選抜は比較的緩やかになっており、以前に比べると容易に教員としての職を得ることができる。そしてそのように選抜が緩やかになれば、以前ならば教員として採用されなかったであろう人材も現代においては教員になりうる。他国と比較した時に日本では認知能力の意味で優れた人物が教員という職を得ていることが知られているが（Hanushek et al., 2019）、現在の教員採用状況が続けば、そのような優秀な教員は相対的に減ってしまう恐れがある。

加えて、教員の退職者数も今後増加傾向であることが見込まれる。図1.2は年度別に小学校における教員の年齢構成を示した表である。図では、各々の教員を「25歳未満」「25歳～30歳未満」「30歳～35歳未満」「35歳～40歳未満」「40歳～45歳未満」「45歳～50歳未満」「50歳～55歳未満」「55歳～60歳未満」「60歳以上」に分けて、それぞれのカテゴリの全体に対する構成割合を示している。図からは「55歳～60歳未満」の教員構成割合が増加傾向にあることを確認することができる。このカテゴリの教員は1995年には全体の7.8%を占めるに過ぎなかったのにも関わらず、2016年には全体の18.1%

# 1章 問題意識と本稿の概要

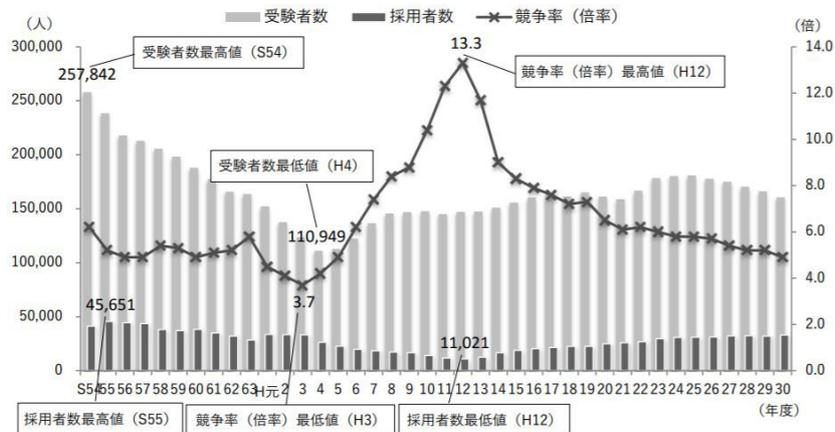


図 1.1: 教員採用試験の受験者数・採用者数・競争率（倍率）の推移

注: 本図は教員採用試験の受験者数・採用者数・競争率（倍率）の推移を示している。横軸は年度を表し、縦軸（左）は人数、縦軸（右）は倍率を表している。  
出典: 文部科学省 (b)

を占めるに至っている。そして、このカテゴリの教員は今後10年間で定年退職をしていくことになるが、その時には、その代わりとなる人材を雇用する必要があると考えられる。しかしその代替となる人材を確保しようにも、上述の通り教員という職業そのものの人気が低下しているため、今後優秀な教員の確保は重要な政策課題となりうる。

以上の理由から資質ある教員の雇用・育成が必要であるという認識の下で、これまで様々な教育施策が検討されてきた。例えば、必要な知識の維持や新しい知識の獲得を通じた教員の資質向上を意図して、教員免許更新制が2007年6月27日に教育職員免許法改正によって導入された<sup>1</sup>。また、高度な教員養成を行うことを目的とした平成19年に教職大学院設立に関する諸省令が交付されて以降、多くの大学が教職大学院の設置を行ってきた。

ただし、これらの教育施策が本当に教員・学校の質向上に結び付くのかについて十分に根拠が存在するわけではない。子どもにとってどのような教員や学校が良いのかという問いは教育経済学の世界では長く探求されてきたが、いまだに研究間で結果が一致するような結果は得られていない。例えば Hanushek, Eric A and Rivkin (2006) では論文発表時点での教員免許についての過去の研究をサーベイし、教員免許については研究間で一致した見解を得られていないとしている<sup>2</sup>。

その一方で教員が子どもの能力成長に強い影響力を持っているとする研究

<sup>1</sup>「求められる教員として必要な資質能力が保持されるよう、定期的に最新の知識技能を身に付けることで、教員が自信と誇りを持って教壇に立ち、社会の尊敬と信頼を得ることを目指すものです。」と説明されている。

<sup>2</sup>これらの論点については第2章で詳細な検討を行った。

# 1 章 問題意識と本稿の概要

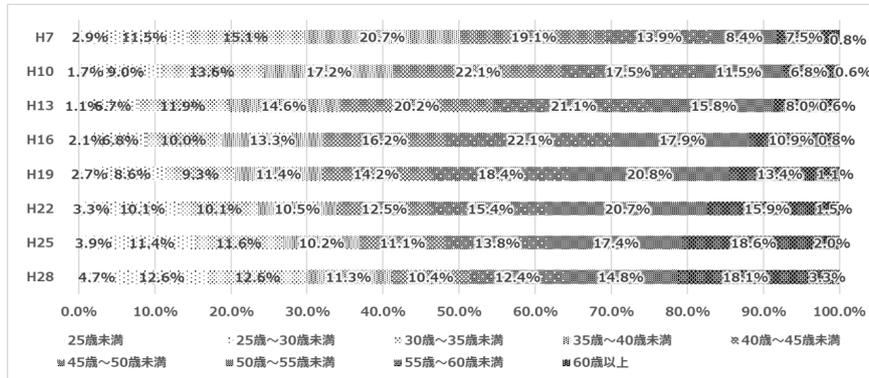


図 1.2: 教員の年齢構成割合の推移

注: 本図は平成 7 年から平成 28 年までの教員の年齢構成・平均年齢の推移を表している。教員の年齢を「25 歳未満」「25 歳～30 歳未満」「30 歳～35 歳未満」「35 歳～40 歳未満」「40 歳～45 歳未満」「45 歳～50 歳未満」「50 歳～55 歳未満」「55 歳～60 歳未満」「60 歳以上」の 9 つのカテゴリに分けて、それぞれの構成割合を年毎に表している。  
 出典: 学校教員統計調査平成 28 年度より筆者作成

成果は強固かつ頑健である。すなわち、より子どもの能力の促進させることができる「質の高い」教員がいる一方で、子どもの能力を成長させることができない「質の低い」教員も存在するということがわかりつつある<sup>3</sup>。このような教員による影響力の違いは教員付加価値研究として現在精力的に研究が進められている。

## 1.2 教員の質研究の現状と課題

しかしこの教員の質研究において、現在十分にわかっていない学問的もしくは政策的課題が存在する。それはまず第一に、日本において教員の違いは子どもの能力形成にどの程度影響を及ぼすかについてはほとんどわかっていないという点である。現在教員の質研究はほとんどアメリカにおける教員の質を対象としている。しかし教育制度は国によって大きく異なり、その制度によって教員に求められる役割は変わってくると考えられる。そのため、日本のようにアメリカと教育制度が大きく違う国で同様の議論を展開できる保証はない。場合によっては、日本のような全国的に非常に均質な教育を供給している国では、教員によるパフォーマンスの違いは小さい可能性が存在する。そのため、日本において教員の質を研究することには重要な学問的並びに政策的な意義が存在する。

第 2 に、学力以外の子どもの能力に与える教員の影響はほとんどわかって

<sup>3</sup>後述する教員付加価値の高低を指して「質の高い/低い」と呼称しているのであり、教員の質を「子どもの能力の多寡」で表すことができるというステイメントではない。

いないという点である。教員は必ずしも子どもの学力のみを伸ばすことが期待されているわけではなく、これまで学力以外の様々な観点に関心が集まってきた。その最たる例は非認知能力である。Heckman and Rubinstein (2001) などの研究が指摘するように、非認知能力は将来の労働市場で評価される重要な人的資本である<sup>4</sup>。それにもかかわらずこの非認知能力に対する教員の影響力はほとんど分かっていない<sup>5</sup>。この点は、教員が子どもの能力形成で果たしている役割を理解するためには非常な重要な論点となる。例えば、学力を伸ばすうえで「良い」教員は、果たして非認知能力を伸ばすという観点においても「良い」教員なのだろうか。場合によっては、学力を伸ばすために苛烈な指導を子どもに行った結果、学力は上昇しても非認知能力は下がってしまうかもしれない。

これらの点を検証することを目的として、第3章では日本のとある自治体における教員の質を推定し、その特徴を分析した。教員付加価値は学力だけでなく学習方略や非認知能力並びに学級での子ども・教員間の関係性を対象にして求めた。その結果、教員の影響の大きさは対象となる能力によって異なることがわかった。教員の影響力は非認知能力や子ども・教員間の関係性で大きく、学力では低いことがわかった。さらに学力の意味で「良い」教員と非認知能力の意味で「良い」教員は必ずしも同一ではないという示唆も得ている。すなわち、少なくとも当該自治体では「良い」教員のありかたというのは一様ではないことがわかった。「良い教員」や「良くない教員」とわかりやすい2つの区分に分けるのは難しく、学力の意味で「良い」教員が非認知能力の意味でも「良い」教員とは限らない。教員の質を考える時には、学力という単一の指標だけで評価をすることは、効率的ではないと考えられる。

### 1.3 教員の指導方法についての研究の現状と課題

しかし、上述の議論によって判明しているのは教員によって子どもへの影響力が異なるという点のみであり、教員付加価値研究として展開されている教員の質研究は「教員のどのような要素が子どもへの影響力の違いを生んでいるのか」という点について詳細な議論を展開しているわけではない。この点については、上述したようにはいまだに研究間で統一的な見解はなされておらず、さらなる研究の蓄積が求められている。

そしてこの「良い教員の特徴とは何か？」という問いに対して、「教員の指導方法が重要である」と答える研究が近年なされつつある<sup>6</sup>。例えば、授業中に大教室の前方で教壇に立って滔々と一方的に講義をする教員と、講義は一定程度にして子どもに理解度確認のためのテストを行ったり場合によ

---

<sup>4</sup>日本での事情などは李嬋娟 (2014) などに詳しい。

<sup>5</sup>教員の違いが子どもの非認知能力に与える影響を分析している研究として、Jackson (2018) などを挙げることができる。

<sup>6</sup>第2章で詳細な検討を行った。

ては視覚教材などを使って理解を促すような工夫を行っている教員では、後者に習う方が子どもはより授業内容を理解し身に付けることができるであろう。このように、指導方法についての研究では「教員が授業の中でどのように教えているか・振舞っているか」という要素が子どもの能力に影響を与えているのではないかという仮説を提示し検証している。この教員の指導方法という論点は、年齢や勤務経験などといった教員の履歴書的な特徴とは異なり、「教員がどのような実践を行っているか」という点に焦点をあてているという点で異なる。上述したように教員の履歴書的な特徴の子どもの学力への影響力の存否はいまだに議論の対象になっているが、教員の指導方法を重視する立場からすれば、どれほど教員としての経験を積み重ねたところでその経験が実際の実践に反映されなければ意味がないのではないかと考える。

この仮説を検証することを目的として、第4章では、教員の指導方法が子どもに与える影響を分析した。指導方法を分析する際に大きな課題になるのは、指導方法は学校や教員に割り当てられた子どもたちに応じて内生的に選択されるという点である。第4章では、そのような問題への対処方法を提案したうえで分析を行った。結果として、いくつかの指導方法についてはその有効性を認めることができた。さらに、その有効性は学校種別に応じて異なることを示唆した。ただし分析上の都合から、本章では指導方法の大まかな方向性を分析の対象としている。そのため、具体的にどのような指導がどの程度有効かそのものまでは踏み込むことができていない。

### 1.4 コンピューター支援教育の効果

そのため、最後に第3章では効果的でありうる指導方法を具体的に検討した。ここでは、教育アプリを用いた分析の効果について掘り下げて分析した。ICTを活用した指導は現在、その効果が大きく期待されている指導方法の1つである。2019年12月には文部科学省によって「ギガスクール構想」が発表されるなど(文部科学省, a)、ICTを教育に用いることで教育における技術革新をもたらすことができるのではないかという期待は大きい。特に、2020年3月に新型コロナウイルス感染拡大に端を発する全国的な学校閉鎖が行われたといった状況下では、オンライン上でどのような指導ができるか、またはどのような指導が効果的かという問いは強い社会性を帯びはじめている(Burgess and Sievertsen, 2020)。また学問的には、Comi et al. (2017) といった研究がICT教育の有効性について検証を行っている。このような動向をうけて、第5章ではカンボジアにおける実証実験を通じて、そのICTを活用した教育の効果进行分析した。

実験ではカンボジアのプノンペン近郊の5校において、通常授業の代わりにICTを活用した指導であるコンピューター支援教育(Computer-Assisted Instruction, CAI)を行った。分析の結果、CAIは子どもの学力を高める効

果を持ったことがわかった。しかし、その一方で CAI は非認知能力や勉強時間の増加には寄与しなかった。通常授業と代替的に行った CAI の実施によって勉強時間は増加することなく学力が上昇したことは、CAI によって子どもの学習生産性が上昇したことを意味する。すなわち、指導方法として CAI を採用することはその他の伝統的な教授方法と比較してより効果的であった。無論、カンボジアの子どもを対象にした分析になるため、その分析が日本の教育においてもそのまま当てはまるかという点は十分に検討される必要がある。特に、カンボジアにおいてここで比較対象となっている通常授業を行う教員の質は日本に比べると比較的低いと考えられ、日本のように比較的教員の質が高い国における CAI の活用の効果は本稿で示す結果よりも小さい可能性がある。しかしそれでもなお、教員の質が低いならば ICT などのサポートを得ながら指導を行うことが効果的である可能性を示したと言える。

### 1.5 おわりに

教員や学校の質の向上が求められる現代において、その「学校や教員の質とはどのようなものか？」についてのエビデンスはますますその重要性を増している。いかなる教員・学校施策も時間やコストがかかる以上、事前の十分な検討と事後の十分な検証が必要であり、そのためにはエビデンスを蓄積していくことが必要である。しかし、日本においてはそのエビデンスの蓄積は必ずしもなされておらず、本論文の各章で示した結果はそのエビデンスの重要な一つになる。第 3 章で示した教員の質の特徴は今後より注力すべき教員の資質についての示唆をもたらし、第 4 章や第 5 章で示した結果は、「どのような指導をするか」という点が子どもの学力の向上のうえでは重要でありうることを示した。もちろんこれらの結果は日本全体での結果ではなく、日本の一部の自治体や場合のよっては海外のデータを用いた研究であり、さらなる検証がなされねばならない。しかし、本稿の様なデータによるエビデンスの蓄積の継続によってはじめて、「教員や学校の質の向上のためには何が必要か？」という問いに対して十分な回答をすることができるだろう。

### 関連図書

Burgess, Simon and Hans Henrik Sievertsen (2020) “The impact of COVID-19 on education — VOX, CEPR Policy Portal”.

Comi, Simona Lorena, Gianluca Argentin, Marco Gui, Federica Origo, and Laura Pagani (2017) “Is it the way they use it? Teachers, ICT and student achievement”, *Economics of Education Review*, **56**, pp. 24–39.

- Hanushek, Eric A and Rivkin, Steven G (2006) “Teacher quality”, *Handbook of the Economics of Education*, **2** (5), pp. 579–599.
- Hanushek, Eric A, Marc Piopiunik, and Simon Wiederhold (2019) “The value of smarter teachers international evidence on teacher cognitive skills and student performance”, *Journal of Human Resources*, **54** (4), pp. 857–899.
- Heckman, James J and Yona Rubinstein (2001) “The importance of noncognitive skills: Lessons from the GED testing program”, *American Economic Review*, **91** (2), pp. 145–149.
- Jackson, C Kirabo (2018) “What do test scores miss? The importance of teacher effects on non-test score outcomes”, *Journal of Political Economy*, **126** (5), pp. 2072–2107.
- 文部科学省 (2012) 「教職生活の全体を通じた教員の資質能力の総合的な向上方策について (答申)」, [https://www.mext.go.jp/b\\_menu/shingi/chukyo/chukyo0/toushin/1325092.htm](https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1325092.htm), Accessed: 2020-07-31.
- (2015) 「これからの学校教育を担う教員の資質能力の向上について～学び合い、高め合う教員育成コミュニティの構築に向けて～ (答申)」, [https://www.mext.go.jp/b\\_menu/shingi/chukyo/chukyo0/toushin/1365665.htm](https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1365665.htm), Accessed: 2020-07-31.
- 「GIGA スクール構想の実現について」, [https://www.mext.go.jp/a\\_menu/other/index\\_00001.htm](https://www.mext.go.jp/a_menu/other/index_00001.htm), Accessed: 2020-07-31.
- 「平成 30 年度公立学校教員採用選考試験の実施状況について」, [https://www.mext.go.jp/a\\_menu/shotou/senkou/1416039.htm](https://www.mext.go.jp/a_menu/shotou/senkou/1416039.htm), Accessed: 2020-07-31.
- 李嬋娟 (2014) 「非認知能力が労働市場の成果に与える影響について」, 『日本労働研究雑誌』, 第 650 巻, 30–43 頁.

## 第2章 教員の有効性研究のこれまでの展開

### 概要

本章では教員の有効性を巡る研究の近年の動向を概観する。効果的な教員や学校の特徴を巡って、多くの研究が教員認証・経験年数・給与や性別といった様々な教員の特徴を取り上げ分析してきた。しかし、それらの分析の結果は研究によって異なり、統一的な見解は得られていない。しかしその一方で、教員付加価値研究は進展を見せている。子どもの成長にとって教員の違いは重要であるという示唆を導くこれらの研究は、その推定手法に一定の懸念が提出されながらも、実験的状況を通じてその有用性を示してきた。更に、近年多くの研究が効果的な教員の特徴として重要なのは教員の学級の中での指導方法や振る舞いではないかという仮説を検証している。今まさにこの指導方法についての研究が急速に進められており、さらなる成果の蓄積が求められる。

### 2.1 効果的な教員・学校の特徴

多くの研究が、学校や教員の存在が子どもの教育成果において重要な規定要因になっているのではないかと指摘してきた。Ammermüller et al. (2005) および Hojo and Oshio (2012) などの研究が学校要因がどの程度子どもの教育成果を説明できるかを考察しており、家庭背景および学校要因を加えた教育生産関数の決定係数は国によって異なるものの20%から40%であることを報告している。総じて、教員や学校、特に教員が子どもの教育成果の重要な規定要因であることは広く指摘されてされてきたことである。

それでは、どのような教員や学校が子どもの能力を向上させることができるのだろうか。高い学歴を持った教員が子どもを教えることが重要なのだろうか。それとも重要なのは教員としての経験の豊富さなのだろうか。もしくは、全く別の要因が決定的な役割を果たしているのだろうか。非常に多くの研究がこれまでに効果的な教員や学校の具体的な特徴の探求を行ってきた。以降では、これまで研究の対象となってきた教員の特徴とその研究の結果についてまとめていく。

### 2.1.1 教員認証

教員認証(教員免許)とは、教員候補をスクリーニングしその質の下限を保障しようとする施策である。この教員認証(教員免許)について、その制度が有効に機能しているかどうかはこれまで研究の対象になってきた。具体的には、教員認証(免許)を受けた教員と受けていない教員の間で子どもの担当した学力の成長に差があるかどうかは分析されてきた。Darling-Hammond et al. (2005)は通常の教員と Teach For America によって派遣された教員を比較し、後者によって指導された子どもの学力はより低いとした。Boyd et al. (2008)・Curran Neild et al. (2009)・Clotfelter et al. (2010)は米国におけるデータを用いて分析しており、教員の認証は子どもの学力に対して有意な影響を持つとしている。Andersson et al. (2011)はスウェーデンの義務教育の子どもの対象に分析し、教員が免許を持っていないことによる負の効果は大きいとしている。

一方で教員が認証を受けているかどうかと担当している子どもの学力の間には有意な関係はないとする研究も数多く存在する。Goldhaber and Brewer (2000)および Goldhaber and Brewer (2001)はOLSによる分析ではあるものの中学生の学力に対する教員認証の効果は非有意であるとしている。Jepsen (2005)は米国のデータを用いて教員認証の効果を推定している。その際に優位性についての議論はしていないが、教員年収の効果の推定値の標準誤差は大きいことを示している。その他、比較的近年の研究であると Croninger et al. (2007)・Kane et al. (2008)・Palardy and Rumberger (2008)・Sharkey and Goldhaber (2008)といった研究が教員の認証の効果が非有意であることを報告している。

ただし上述の教員認証に関する研究は、厳密には教員認証(免許)という制度そのものに対する分析ではない。これらの研究は、現行の教員認証制度の元における認証取得の有無によるパフォーマンスの違いを分析している。すなわち、それらの研究は「教員認証を取得していない教員」と「取得している教員」の比較を行っているのである。しかし、それらの研究が導く認証を受けた教員の有効性は必ずしも教員認証制度そのものの有効性とは同一ではない。認証を受けた教員の有効性とは別に、教員認証制度そのものの効果を考える必要がある。場合によっては、教員認証制度があることによって教員という職業に対する参入障壁が高くなってしまふことによる悪影響の可能性も存在する。

この教員認証制度そのものを評価した研究として、Angrist and Guryan (2008)がある。この研究では、教員認証においてテスト要件を課した州のデータから教員認証制度を厳しく設定することの効果の評価している。その結果から、教員認証制度は教員の賃金を上昇させるもののSATスコアなどで評価された教員の質を向上させたわけではないことが示されている。このように教員認証制度そのものが意図通り教員の質を向上させているかどうか

は、今のところ明らかではない。

### 2.1.2 教員の経験年数

訓練や経験を積むことで教員は指導に対して技能を積み上げ習熟するであろうという考えの下で、教員の経験年数は研究の対象になってきた。経験が浅い教員は授業や学級運営にも慣れていない一方で、経験を積むにつれてそれらの業務に習熟し高いパフォーマンスを発揮するのではないかという発想は非常に自然である。政策実務においては、平成24年08月28日に交付された「教職生活の全体を通じた教員の資質能力の総合的な向上方策について(答申)」(文部科学省, 2012)において「今後、このような資質能力を有する、新たな学びを支える教員を養成するとともに、『学び続ける教員像』の確立が必要である」と、教員にさらなる学習や訓練の機会を提供し高度な教員養成を目指す政策が謳われている。

このような想定の中で教員の経験年数を扱った研究のうちいくつかは、経験を積んだ教員は比較的高いパフォーマンスを発揮することを報告している。Jackson and Bruegmann (2009) は米国ノースカロライナにおける小学生のデータの分析から、教員がどのようにして教員として成長していくかを分析している。その結果、経験の浅い教員は同僚の教員から学ぶことで能力を蓄積をしていくことを示している。Carrell and West (2010) では高等教育において教員の経験年数が持つ効果をRCTを用いて分析している。その結果、講義の種類によって経験年数が持つ効果は全く異なることを示している。Harris and Sass (2011) では教員の経験年数は小学校・中学校教員の生産性を向上させる一方で、高校教員の実績性は向上させないことを示している。Wiswall (2013) では数学についてはキャリア後期での経験年数に高い効果があることを示している。Ost (2014) では教員の勤務経験が保つ効果について教職そのものの経験年数と学年別の経験年数の効果にわけ、前者をコントロールした上で後者について分析をしている。その結果、学年別の経験年数の効果はあるものの、急速に減耗していくことを示している。Papay and Kraft (2015) では米国のデータを用いて、キャリア職において経験年数の効果が大きいことを示している。

しかし、必ずしもすべての研究が教員の経験年数の有効性を示しているわけではない。Hanushek and Rivkin (2006) は先行研究をサーベイする中で、教員の経験年数が子どもの学力に与えた影響を調べた研究の結果をまとめている。その結果、その影響の統計的有意性が示された研究は半数程度であるとしている。Aaronson et al. (2007) はシカゴの公立学校のデータを用いた分析の中で、教員の経験年数の子どもの学力に対する影響は有意ではなかったことを示している。

まとめると、教員の経験年数が子どもの学力に正の影響を与えるとする研究は数多くあるものの、その効果は教員が置かれた状況によって異なりうる。

場合によっては教員の経験は効果的ではないことがあり、一概に「経験を積むことが教員にとって重要だ」と主張することは難しいことがわかる。

### 2.1.3 教員の性別

教員の性別、特に女性教員であることが担当した子どもに与える影響は頻繁に調査されてきたが、その際に中心的なトピックになるのは教える教員と教わる子どもの性別が同一であることが重要なのではという問いである。すなわち、同じ性別の教員に教わることで、学習者と教員の関係性は比較的良好になり授業もより効果的になる可能性がある。そのため実のところ、この問題意識は教員と子どものマッチングが重要なのではないかという問いと通底しており、教員の性別だけでなく同様の問題意識の下で教員の人種なども分析の対象となってきた。

そして多くの研究が教員の性別と子どもの性別が同一であることで子どもの学力が向上するという結果を報告している。Dee (2007) では、教員と子どもが同性であることが子どもの学力に対して正の効果があったことを示している。Muralidharan and Sheth (2016) ではインドにおける分析を示しており、その結果として、同性教員による効果は正であるとしている。この分析では単一学級における変動を用いており、後述の教員の学校内での内生的配置の影響の可能性は低いとしている。Carrell and West (2010) はアメリカの大学の生徒を対象にした分析をしており、女性教授は女性の生徒のSTEM 関連アウトカムを向上させるとしている。Lim and Meer (2017) は韓国においてRCT を実施し、その結果として女性教員は女子に正の影響を与えている。

一方で、子どもと教員の性別のマッチングが子どもの学力に与える影響は小さいとする研究も存在する。Neugebauer et al. (2011) はドイツの子どもを対象に分析を行っており、同性の教員を持つことが有益であるという証拠は男子でも女子でもほとんど見当たらないとしている。Winters et al. (2013) ではフロリダ州のデータを用いて分析を行い、教員と子どもが同性であると中学・高校では学力に正の影響があるものの、その効果量は小さいとしている。このような研究の中には Antecol et al. (2015, アメリカ) のように、子どもと教員のマッチングは子どもの能力成長において有害ですらあり得るとする研究もある。

このように研究によって結果が異なる理由として、幾つかの研究が教員が内生的に配置されている可能性を指摘している。ある学級にどの教員を割り当てるかという問題は、多くの場合学校や自治体による意思決定である。そのため、女性の子どもの多く所属する学級には女性の教員を配置するという意思決定が存在すれば、子どもの性別と教員のマッチングの間には内生性が存在することになる。例えば、Holmlund and Sund (2008) ではウェーデンに

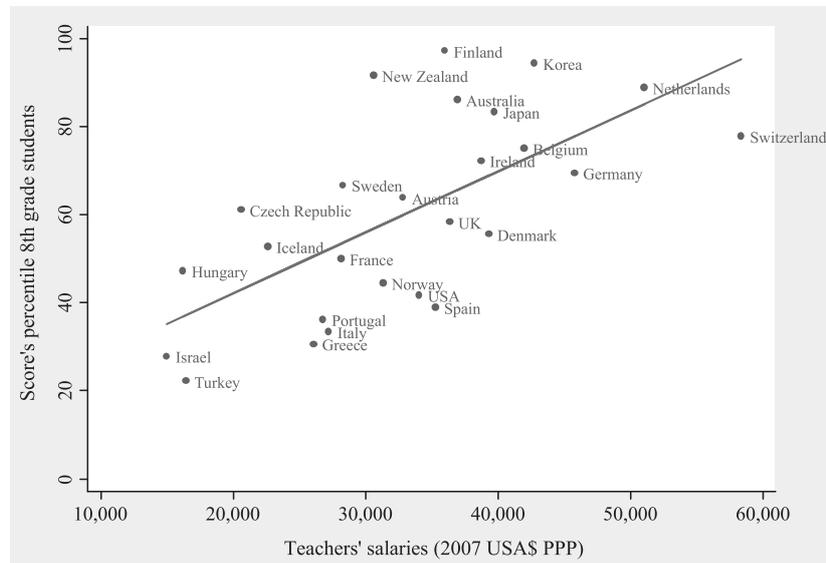


図 2.1: 教員の給与とテストスコア (中 2) の関係

注: 横軸は教員の給与を表し、縦軸はテストのスコアを表している。それぞれのマーカーは国を表し、プロットされたラインは回帰結果を表している。出典: Dolton and Marcenaro-Gutierrez (2011)

において、女性教員のクラスでは男女差が大きいものの、それは教員の内生的な配置によるものではないかと主張している。上述した研究の中には、Lim and Meer (2017) などのように RCT を用いてそのような内生性の影響を排除して行った分析もあるものの、教員と子どもの性別のマッチングが子どもの学力に与える影響について統一的な見解を主張するためにはさらなる研究蓄積が求められる。

#### 2.1.4 教員の給与

教員の給与は、子どもの能力を伸ばすことができるかどうかという意味での教員の能力を規定する要因として多くの研究の対象となってきた。図 2.1 は OECD が公表しているデータを用いて計算された教員の給与とテストスコア (中 2) の関係を国ごとに表した図である (Dolton and Marcenaro-Gutierrez, 2011)。横軸は教員の給与を表し、縦軸はテストのスコアを表している。図からは、教員の給与と子どもの学力の間に正の関係があることを確認することができる。ただし、背景に存在するメカニズムによって研究群を大別すると、2つの議論の潮流が存在する。

第 1 に、教員給与は労働市場における教員の相対的立ち位置を規定すると考える研究がある。すなわち、教員の平均的な給与水準をその他の職業の給

与と比較した相対賃金は、教員自身の労働市場における立ち位置を表しているとする。そこでは、教員の賃金が他の労働者と比べ平均的に高い時、より能力の高い労働者が教員になるという想定が存在している。Lakdawalla (2006) は米国の国勢調査のデータを用いて、教員と他職の給与比較に基づく相対賃金を教員の質を表す指標として分析を行っている。その結果、技術革新による労働者の賃金上昇や女性の就業機会の上昇が、米国の教員の質の低下を招いたとしている。Hendricks (2014) は米国のテキサス州のパネルデータを用いて、教員の給与を高く設定することで教員の離職率が低下することを示している。

このようにして能力の高い人材が教員になることで、その教員に教わる子どもの能力もまたより促進されるのではないかと教員給与についての研究では考えられてきた。Dolton and Marcenaro-Gutierrez (2011) では OECD が公表している 39 カ国のパネルデータを用いて教員給与のクロスカントリー分析を行い、能力の高い人材を教員に採用し給与の早期昇格を可能にすることは生徒の成績にプラスの効果をもたらすことを示した。Leigh (2012) ではオーストラリアの教員給与が教員志望者の学力を規定していることを示した。すなわち、教員給与が他の職業の給与と比較して1%上昇することは、教員課程に入学する生徒の相対順位を0.6%上昇させることがわかった Britton and Propper (2016) では英国の公立学校における教員給与規制を利用した DID(Difference in Difference) を用いて、教員とその他の職業の間にある賃金格差が10%広がると、平均学力が約2%低下することを示した。

ただし、教員の給与増加は必ずしも子どもの能力上昇に結びついたわけではない。Kingdon and Teal (2010) はインドの私立学校のデータから生徒の学力の科目間変動を用いて分析を行い、教員の組合加入(組合加入をした教員は給与が増加する)が生徒の学力低下と関連していることを示している。De Ree et al. (2017) ではインドネシアの公立学校で行われたRCTから、教員の給与を引き上げる介入は教員の勤務時間を短縮させ生産性を向上させたにも関わらず、子どもの学力は上昇には結びつかなかったとしている。Greaves and Sibieta (2019) ではイングランドにおいて一部の学校で高い給与水準を強制的に払う必要が生まれたという自然実験を利用して、高い給与水準は必ずしも子どもの学力上昇につながらなかったことを示した。

第2に、パフォーマンスに対するインセンティブとして教員給与が機能することを想定する研究がある。高いパフォーマンスを発揮することへの報酬として高い給与が支払われる時、教員は子どもへの指導に対してより動機づけられる。このような給与のインセンティブとしての側面に注目するのが、パフォーマンスペイ performance pay(もしくはメリットペイ merit pay) についての研究である。パフォーマンスペイとは、より高い成果を出した教員により高い給料を支払うような給与システムのことを指す。パフォーマンスペイの下では教員自身に担当する子どもの成長に対する動機付けが存在す

るため、教員はより子どもの能力促進に励むと考えられ、これまで研究がなされてきた。

しかし、それらの研究結果が示すのは、パフォーマンスペイが効果的であるかどうかはかならずしも自明ではないということである。まず、RCTを用いてパフォーマンス研究の結果は一貫していない。Glewwe et al. (2010, インド)・Muralidharan and Sundararaman (2011, ケニア)・Mbiti et al. (2019, タンザニア) はパフォーマンスペイの学力に対する正の効果を示す一方で、Springer et al. (2011, 米国ナッシュビル) 及び Fryer (2013, 米国ニューヨーク) はパフォーマンスペイは必ずしも子どもの学力向上に結びつかなかったことを示している。また準実験的な状況を用いた分析や固定効果をコントロールした分析も RCT を用いた研究と同様、一貫した結果を示していない。これらの研究は DID や RDD などを用いてパフォーマンスペイの効果を分析しているが、効果があるという研究 (Ladd, 1999; Lavy, 2002, 2009; Figlio and Kenny, 2007; Woessmann, 2011) や効果がなかったという研究 (Eberts et al., 2002) の両者が存在する。

### 2.1.5 学校の特徴

ここまで教員の特徴が子どもの能力に与える影響を分析した研究を概観してきたが、そのこれまでの研究動向は「一貫して効果があるとされる学校や教員の特徴は見つかっていない」と纏めることができる (Hanushek and Rivkin, 2006)。Aslam and Kingdon (2011) ではそのような状況を指して「履歴書的な教員の特徴は効果がない」と述べている。

この研究動向は効果的な学校の具体的な特徴についての研究においても同様である。効果的な学校の特徴の候補として、これまで頻繁に分析の対象となってきたのは学校による教育支出である。教育支出について量的に行われた研究として非常に大きな影響力を持ったのは、1950年代に米国の社会学者コールマンらによって行われた研究である (Coleman et al., 1966)。コールマンレポートと呼ばれることになったこの研究は、個々の人間の学力差は学校による影響をほとんど受けていないことを指摘した。その後この研究結果を受けて、学校への金銭的な投資は本当に子どもの学力に影響を与えていないのか、経済学の立場から検証する研究がハヌシェックらによって行われた (Hanushek, 1986, 2006)。これらの研究の中で問われていたのは、そもそも学校が子どもの教育成果に影響を与えているかどうかという点であり、結論としては、学校はあまり子どもの教育成果に影響を与えていないという結論

が現在のところ主流である。<sup>123</sup>。

## 2.2 教員固定効果

しかし、効果的な教員や学校の具体的な特徴が分からないことは、教員や学校の子どもに対する影響力そのものを否定するわけではない。そのような影響力の存在を裏付けるのが教員付加価値に関する研究群である。この教員付加価値とは、担当した子どもの能力の伸び具合を用いて評価した教員の質のことを指す。教員付加価値に関する研究では、教員付加価値を子どものデータから推定しその性質を調べる事を目的にする研究が多い。

多くの場合教員効果を推定するときは、次の様なモデル化のもとで推定を行う<sup>4</sup>。

$$A_{ijt} = \beta X_{ijt} + \nu_{ijt} \quad \text{where} \quad \nu_{ijt} = \theta_j + \mu_{jt} + \varepsilon_{ijt} \quad (2.1)$$

$i$  は生徒、 $j$  は教員、 $t$  は時間を指し、 $A$  は教育成果、 $X$  は生徒の個人特性、 $\varepsilon$  は観察不可能な要因を表す。 $\nu_{ijt}$  は教員効果  $\theta_j$ 、クラスルーム効果  $\mu_{jt}$ 、その他の要因  $\varepsilon_{ijt}$  で構成されるとする。上記式から  $A_{it} - \beta X_{it}$  は教員効果  $\theta_j$  についての情報を持つが、その中には今関心がないノイズである  $\mu_{jt}$  と  $\varepsilon_{ijt}$  を含んでいる。そのため、多くの研究は singular extraction 問題を解くことで教員効果を推定している。

この時、先の教育生産関数の式は教員効果  $\theta_j$  についての次の様なシグナルなのではないかと考えることができる。

$$s_{ijt} = A_{ijt} - \beta X_{ijt} = \theta_j + \mu_{jt} + \varepsilon_{ijt} \quad (2.2)$$

さらにこの式は、教員  $j$  が  $t$  期に担当している生徒を  $I_{jt}$  と書けば (集合  $I$  の要素数、すなわち生徒数を  $|I_{jt}|$  と書けば)、

$$s_{jt} = \frac{\sum_{i,i \in I_{jt}} s_{ijt}}{|I_{jt}|} \quad (2.3)$$

$$= \theta_j + \mu_{jt} + \frac{\sum_{i,i \in I_{jt}} \varepsilon_{ijt}}{|I_{jt}|} \quad (2.4)$$

$$= \theta_j + \nu_{jt} \quad (2.5)$$

という教員効果  $\theta_j$  についてのシグナルを每期受け取ると考えることができる。もし教員効果  $\theta_j$  に対する prior が期待値 0 及び分散  $V(\theta)$  であり、上記

<sup>1</sup>しかし、Hanushek (2006) も指摘する通り、これは学校が個人の教育成果に与える影響が学校資源の量を増やせば学力が上がる、といった単純なものではないだけで、学校そのものの価値や学校への投資が不要であることを示すものではない。

<sup>2</sup>Angrist and Lavy (1999) など研究されたクラスサイズ縮小政策は数少ない効果があるとされている学校施策である。ただし教育政策としての有効性を含めて議論は終わっていない。

<sup>3</sup>ただしより洗練された手法を用いて学校支出にはやはり一定の効果があるという Jackson et al. (2015) の様な研究も出てきている。

<sup>4</sup>Kane and Staiger (2008) に従う。

シグナルを  $t$  期分受け取ると (そのシグナルの集合を  $s_j = \{s_{jt} | t \in T\}$  と書くことにする)、

$$E(\theta_j | s_j) = \sum_{t, t \in T} s_{jt} \frac{\frac{1}{V(v_{jt})}}{\frac{1}{V(\theta)} + \sum_{t, t \in T} \frac{1}{V(v_{jt})}} \quad (2.6)$$

と教員  $j$  の効果  $\theta_j$  の期待値を書くことができる。多くの場合は得られたデータから、 $V(\theta_j)$  や  $V(v_{jt})$  を計算して妥当性の高い prior を構成した後、期待値を計算することが多い。

多くの教員付加価値研究はこのモデルを発展させていく中で分析を行っている。Lefgren and Sims (2012) は複数の科目に上述のモデルを拡張し、教科によって変わらない教員の能力を評価するモデルを提案している。Chetty et al. (2014a,b) は教員の能力は不変ではなく時間によって変動するようなモデルを考え分析を行っている。上述のモデルはあくまで基本的な形にすぎず、実際に知りたい情報などに応じた修正が今後ともなされながら分析されていくと考えられる。

### 2.2.1 教員固定効果研究が抱える限界

しかし、幾つかの研究は、上述 (2.2 節) の方法によって推定された教員付加価値は適切に教員の能力を反映していない可能性を指摘している。とくに、得られた教員付加価値が本当に教員の子どもに与える因果的な影響を表すことが出来ているのかという点において、推定上の成約などが頻繁に議論の対象になってきた。教員付加価値の推定では教員に割り当てられた子どもの情報を用いるが、その割り当ては必ずしも外生的に定まるとは限らない。例えばパフォーマンスが高い (とされる) 教員には能力の高い子どもを優先的に配置する可能性や、教員歴が浅い教員には比較的優秀で指導がしやすい学級を割り当てる可能性など様々な可能性を検討することができる。例えば、Rothstein (2010) は教員付加価値推定が教員のランダム割り付けを仮定していることを指摘し、さらにその仮定がノースカロライナのデータでは成立しないことを指摘した。また、Kinsler (2012) では Rothstein (2010) が提案したランダム割り付け検定の課題を指摘し、新しい手法を提案している。

しかしこれらの批判を考慮してもなお、上述の教員付加価値は教員のパフォーマンスの一部を十分に表現することができていると考えられる。すなわち、教員付加価値のうち少なくとも一部は、教員の能力を表していると考えることができる。例えば、上述の Kane and Staiger (2008) や Chetty et al. (2014a,b) といった研究は、得られた教員付加価値が本当に教員の能力を表しているのかを実験的環境を利用して検証をしている。そしてその結果として、教員付加価値が高かった教員に割り当てられた子どもは学力や非認知能力を表す行動変数を向上させることを報告している。教員の業務評価として教員付加価値を用いることには一定の議論が依然として必要ではあるものの、こ

これらの結果は教員の子どもに対する影響力そのものは存在していることを強く主張していると言える。

### 2.3 効果的な指導方法

教員付加価値研究から見えてくるのは（付加価値の意味で）効果的な教員もしくは効果的ではない教員そのものは存在するという事実である。しかし上述したように、その具体的な要因（「どのような教員が効果的か」）については未だに研究間で一貫性のある成果は存在しない。それでは良い教員とはなにかという問いに答えることは今後も難しく、教員政策については実証的な根拠に乏しい施策を検討するしかないのだろうか。あるいは、実証的にはサポートされるという理由で教員付加価値を用いるとして、教員付加価値を計算するためには各教員について少なくとも数年レベルのデータが必要になる。そのようなデータの蓄積を待たないと効果的な教員かどうかは分からず、教員施策を検討することもできないのだろうか。

以上の様な研究状況の中、教員の指導方法に関する実証研究が近年急速に蓄積されている。学級の中でどのように授業を行うかは、学校や教員によって大きく異なる。教壇の前にたって説明をすることに終始する授業もあれば、演習に重きをおいて多くの授業外課題を課すような授業もあるだろう。図 2.2 及び図 2.3 に示したのは、埼玉県で行われた学力調査の 2015 年から 2019 年のデータを用いて計算された、学年で用いられた指導方法の学校内での決定要因を示したものである。図 2.2 は、指導方法を学校要因・学年要因・科目要因・市町村要因に分解しており、指導方法によってその決定要因が大きく異なることを確認することが出来る。例えば「低学力向け少人数指導」では科目要因がその決定要因の多くを占めているのに対して、「対話による課題解決」では科目要因は決定要因に占める割合は比較的小さい。図 2.3 は、横軸に「学校内で年度ごとに指導方法を変えているか」とり縦軸に「学校内で科目ごとに指導方法を変えているか」を指導方法ごとに表したものである。図からは、学校の中でも年度による違いが大きい指導方法と、科目による違いが大きい指導方法に分かれていることを確認することができる。このように現在学級で用いられる指導方法は現場によって大きく異なっている。上述の教員の特徴についての研究と異なり、指導方法についての研究では学校や教員による授業の実践のあり方そのものに着目をする。

そのような指導方法についての研究では、授業における講義時間（いわゆる伝統的な立ち講義の時間）の影響というトピックが頻繁に議論の対象になってきた。教員が学級の前の黒板やホワイトボードの前に立ち子どもたちに対して一方的に説明をする授業スタイルに対して、一部の研究者や政策担当者は「主体的に子どもが学習に向かうことが重要である」として批判を行ってきた。このような論点を対象に、幾つかの研究が定量的な評価を試みている。

## 2章 教員の有効性研究のこれまでの展開

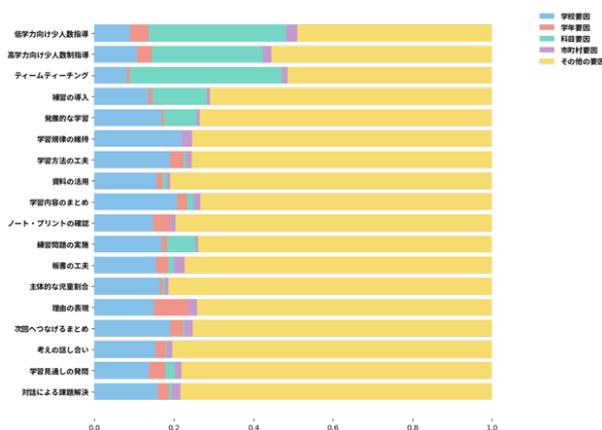


図 2.2: 指導方法の決定要因

注) 指導方法の規定要因として学校要因・科目要因・学年要因及び市町村要因の4つの要因を考え、その規定要因としての大きさを示した。手続きとしては、各要因をランダム効果として従う分布の推定を行ない、その分散を示した。分散の合計値が1になるように基準化してある。  
出典) 埼玉県学力調査から筆者作成。

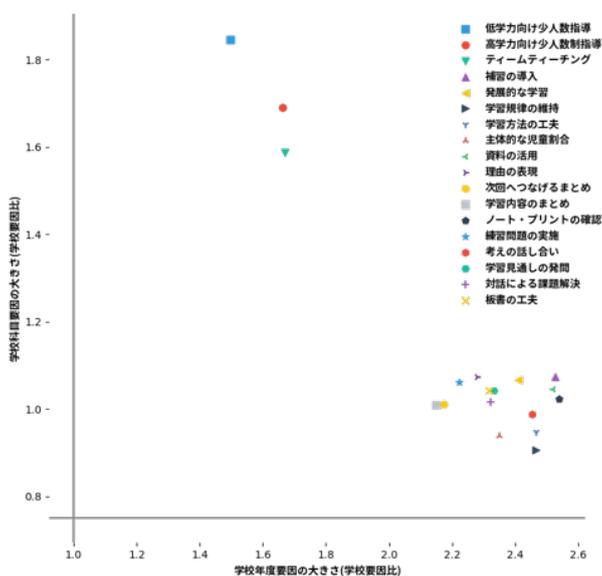


図 2.3: 指導方法の学校内決定要因

注) 各指導方法における規定要因として学校年度要因と学校科目要因の大きさを推定しプロットした。X軸は学校年度要因の大きさを表し、Y軸は学校効果要因の大きさを表す。基準化を行うため、学校年度要因・学校科目要因共に学校要因との比を取ることで基準化を行った。  
出典) 埼玉県学力調査から筆者作成。

Schwerdt and Wuppermann (2011) では講義時間の多寡が子どもの学力に影響を与えていることを示している。その一方で、Van Klaveren (2011) では授業時間の中で教員が教室の前で立って講義している間の割合が長くとも、子どもの学力には影響はないとしている。このように講義時間と子どもの学力の関係についてはまだ統一的な見解は形成されていない。

また、伝統的な指導方法と現代的な指導方法のどちらが子どもの学力を向上させる上で効果的かという点に着目する研究も多い。このような研究では、上述の講義時間の様に子どもに一方的に知識を伝授する伝統的指導方法と子どもの学習における主体性を重視する現代的指導方法を対比させ、後者が前者に対して優先するような仮説を検証の対象にしている。Bietenbeck (2014) では伝統的指導方法と現代的指導方法を対比し、前者が子どもの学力に対してポジティブな影響を与えているとしている。Lavy (2015) では伝統的授業方法は低いSESの子どもに高い正の影響を与えていることが示されている。Hidalgo-Cabrillana and Lopez-Mayan (2018) では授業で用いられた指導方法について教員と子どもではその認識が大きく異なることを示し、現代的指導方法は子どもの学力に正の影響があったとする。

またその他、個々の指導方法も検証の対象になってきた。Comi et al. (2017) ではICTを用いた指導方法を評価しその利用方法によって効果が異なりうるという結果を導いている。Tanaka and Ishizaki (2018) では日本において、リメディアル教育が子どもの学力に正の効果を与えうる可能性を指摘している。また、Bessho et al. (2019) も同様に日本のデータを利用した傾向スコアマッチングを用いて、指導方法が効果を持ちうる可能性を検討している。

以上のように、今まさに指導方法についての研究は急速に進められているところである。その研究結果は必ずしも研究間で整合的であるとは言えないものの、学術的な見解を下すほど研究は蓄積されてない。場合によっては教員の特徴についての研究のように、指導方法と子どもの学力の間には強い関係性はない可能性が高いという見解に今後落ち着く可能性さえある。しかしそのような議論を行うためにも、今後さらなる成果の蓄積が求められている。

### 関連図書

Aaronson, Daniel, Lisa Barrow, and William Sander (2007) “Teachers and student achievement in the Chicago public high schools”, *Journal of Labor Economics*, **25** (1), pp. 95–135.

Ammermüller, Andreas, Hans Heijke, and Ludger Wößmann (2005) “Schooling quality in Eastern Europe: Educational production during transition”, *Economics of Education Review*, **24** (5), pp. 579–599.

Andersson, Christian, Per Johansson, and Nina Waldenström (2011) “Do

- you want your child to have a certified teacher?”, *Economics of Education Review*, **30** (1), pp. 65–78.
- Angrist, Joshua D and Jonathan Guryan (2008) “Does teacher testing raise teacher quality? Evidence from state certification requirements”, *Economics of Education Review*, **27** (5), pp. 483–503.
- 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** (2), pp. 533–575, may.
- Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik (2015) “The effect of teacher gender on student achievement in primary school”, *Journal of Labor Economics*, **33** (1), pp. 63–89.
- Aslam, Monazza and Geeta Kingdon (2011) “What can teachers do to raise pupil achievement?”, *Economics of Education Review*, **30** (3), pp. 559–574.
- Bessho, Shunichiro, Haruko Noguchi, Akira Kawamura, Ryuichi Tanaka, and Koichi Ushijima (2019) “Evaluating remedial education in elementary schools: Administrative data from a municipality in Japan”, *Japan and the World Economy*, **50**, pp. 36–46.
- Bietenbeck, Jan (2014) “Teaching practices and cognitive skills”, *Labour Economics*, **30**, pp. 143–153.
- Boyd, Donald, Hamilton Lankford, Susanna Loeb, Jonah Rockoff, and James Wyckoff (2008) “The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools”, National Bureau of Economic Research.
- Britton, Jack and Carol Propper (2016) “Teacher pay and school productivity: Exploiting wage regulation”, *Journal of Public Economics*, **133**, pp. 75–89.
- Carrell, Scott E and James E West (2010) “Does professor quality matter? Evidence from random assignment of students to professors”, *Journal of Political Economy*, **118** (3), pp. 409–432.
- Chetty, Raj, John N Friedman, and Jonah E Rockoff (2014a) “Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates”, *American Economic Review*, **104** (9), pp. 2593–2632.

- , ———, and ——— (2014b) “Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood”, *American Economic Review*, **104** (9), pp. 2633–79.
- Clotfelter, Charles T, Helen F Ladd, and Jacob L Vigdor (2010) “Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects”, *Journal of Human Resources*, **45** (3), pp. 655–681.
- Coleman, James S, E Campbell, C Hobson, F McPartland, A Mood, F Weinfeld, and Others (1966) “Equality of educational opportunity study”, *Washington, DC: United States Department of Health, Education, and Welfare*.
- Comi, Simona Lorena, Gianluca Argentin, Marco Gui, Federica Origo, and Laura Pagani (2017) “Is it the way they use it? Teachers, ICT and student achievement”, *Economics of Education Review*, **56**, pp. 24–39.
- Croninger, Robert G, Jennifer King Rice, Amy Rathbun, and Masako Nishio (2007) “Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement”, *Economics of Education Review*, **26** (3), pp. 312–324.
- Neild, Ruth Curran, Elizabeth Nash Farley-Ripple, and Vaughan Byrnes (2009) “The effect of teacher certification on middle grades achievement in an urban district”, *Educational Policy*, **23** (5), pp. 732–760.
- Darling-Hammond, Linda, Deborah J Holtzman, Su Jin Gatlin, and Julian Vasquez Heilig (2005) “Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness”, *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, **13**, pp. 1–48.
- Ree, Joppe De, Karthik Muralidharan, Menno Pradhan, and Halsey Rogers (2017) *Double for nothing? experimental evidence on an unconditional teacher salary increase in indonesia*: The World Bank.
- Dee, Thomas S (2007) “Teachers and the gender gaps in student achievement”, *Journal of Human Resources*, **42** (3), pp. 528–554.
- Dolton, Peter and Oscar D Marcenaro-Gutierrez (2011) “If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance”, *Economic Policy*, **26** (65), pp. 5–55.

- Eberts, Randall, Kevin Hollenbeck, and Joe Stone (2002) “Teacher performance incentives and student outcomes”, *Journal of Human Resources*, pp. 913–927.
- Figlio, David N and Lawrence W Kenny (2007) “Individual teacher incentives and student performance”, *Journal of Public Economics*, **91** (5-6), pp. 901–914.
- Fryer, Roland G (2013) “Teacher incentives and student achievement: Evidence from New York City public schools”, *Journal of Labor Economics*, **31** (2), pp. 373–407.
- Glewwe, Paul, Nauman Ilias, and Michael Kremer (2010) “Teacher incentives”, *American Economic Journal: Applied Economics*, **2** (3), pp. 205–27.
- Goldhaber, Dan D and Dominic J Brewer (2000) “Does teacher certification matter? High school teacher certification status and student achievement”, *Educational Evaluation and Policy Analysis*, **22** (2), pp. 129–145.
- and ——— (2001) “Evaluating the evidence on teacher certification: A rejoinder”, *Educational Evaluation and Policy Analysis*, **23** (1), pp. 79–86.
- Greaves, Ellen and Luke Sibieta (2019) “Constrained optimisation? Teacher salaries, school resources and student achievement”, *Economics of Education Review*, **73**, p. 101924.
- Hanushek, Eric A (1986) “The economics of schooling: Production and efficiency in public schools”, *Journal of Economic Literature*, **24** (3), pp. 1141–1177.
- (2006) “School resources”, *Handbook of the Economics of Education*, **2**, pp. 865–908.
- and Steven G Rivkin (2006) “Teacher quality”, *Handbook of the Economics of Education*, **2**, pp. 1051–1078.
- Harris, Douglas N and Tim R Sass (2011) “Teacher training, teacher quality and student achievement”, *Journal of Public Economics*, **95** (7-8), pp. 798–812.
- Hendricks, Matthew D (2014) “Does it pay to pay teachers more? Evidence from Texas”, *Journal of Public Economics*, **109**, pp. 50–63.

- Hidalgo-Cabrillana, Ana and Cristina Lopez-Mayan (2018) “Teaching styles and achievement: Student and teacher perspectives”, *Economics of Education Review*, **67**, pp. 184–206.
- Hojo, Masakazu and Takashi Oshio (2012) “What Factors Determine Student Performance in East Asia? New Evidence from the 2007 Trends in International Mathematics and Science Study”, *Asian Economic Journal*, **26** (4), pp. 333–357.
- Holmlund, Helena and Krister Sund (2008) “Is the gender gap in school performance affected by the sex of the teacher?”, *Labour Economics*, **15** (1), pp. 37–53.
- Jackson, C. Kirabo and Elias Bruegmann (2009) “Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers”, *American Economic Journal: Applied Economics*, **1** (4), pp. 85–108, October.
- Jackson, C Kirabo, Rucker C Johnson, and Claudia Persico (2015) “The effects of school spending on educational and economic outcomes: Evidence from school finance reforms”, *The Quarterly Journal of Economics*, **131** (1), pp. 157–218.
- Jepsen, Christopher (2005) “Teacher characteristics and student achievement: Evidence from teacher surveys”, *Journal of Urban Economics*, **57** (2), pp. 302–319.
- Kane, Thomas J and Douglas O Staiger (2008) “Estimating teacher impacts on student achievement: An experimental evaluation”, National Bureau of Economic Research.
- , Jonah E Rockoff, and Douglas O Staiger (2008) “What does certification tell us about teacher effectiveness? Evidence from New York City”, *Economics of Education Review*, **27** (6), pp. 615–631.
- Kingdon, Geeta and Francis Teal (2010) “Teacher unions, teacher pay and student performance in India: A pupil fixed effects approach”, *Journal of Development Economics*, **91** (2), pp. 278–288.
- Kinsler, Josh (2012) “Beyond levels and growth estimating teacher value-added and its persistence”, *Journal of Human Resources*, **47** (3), pp. 722–753.

- Ladd, Helen F (1999) “The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes”, *Economics of Education Review*, **18** (1), pp. 1–16.
- Lakdawalla, Darius (2006) “The economics of teacher quality”, *The Journal of Law and Economics*, **49** (1), pp. 285–329.
- Lavy, Victor (2002) “Evaluating the effect of teachers group performance incentives on pupil achievement”, *Journal of Political Economy*, **110** (6), pp. 1286–1317.
- (2009) “Performance pay and teachers’ effort, productivity, and grading ethics”, *American Economic Review*, **99** (5), pp. 1979–2011.
- (2015) “Do differences in schools’ instruction time explain international achievement gaps? Evidence from developed and developing countries”, *The Economic Journal*, **125** (588), pp. F397—F424.
- Lefgren, Lars and David Sims (2012) “Using subject test scores efficiently to predict teacher value-added”, *Educational Evaluation and Policy Analysis*, **34** (1), pp. 109–121.
- Leigh, Andrew (2012) “Teacher pay and teacher aptitude”, *Economics of Education Review*, **31** (3), pp. 41–53.
- Lim, Jaegeum and Jonathan Meer (2017) “The impact of teacher–student gender matches random assignment evidence from South Korea”, *Journal of Human Resources*, **52** (4), pp. 979–997.
- Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani (2019) “Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania”, *The Quarterly Journal of Economics*, **134** (3), pp. 1627–1673.
- Muralidharan, Karthik and Ketki Sheth (2016) “Bridging education gender gaps in developing countries: The role of female teachers”, *Journal of Human Resources*, **51** (2), pp. 269–297.
- and Venkatesh Sundararaman (2011) “Teacher performance pay: Experimental evidence from India”, *Journal of Political Economy*, **119** (1), pp. 39–77.
- Neugebauer, Martin, Marcel Helbig, and Andreas Landmann (2011) “Unmasking the myth of the same-sex teacher advantage”, *European Sociological Review*, **27** (5), pp. 669–689.

- Ost, Ben (2014) “How do teachers improve? The relative importance of specific and general human capital”, *American Economic Journal: Applied Economics*, **6** (2), pp. 127–51.
- Palardy, Gregory J and Russell W Rumberger (2008) “Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning”, *Educational Evaluation and Policy Analysis*, **30** (2), pp. 111–140.
- Papay, John P and Matthew A Kraft (2015) “Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement”, *Journal of Public Economics*, **130**, pp. 105–119.
- Rothstein, Jesse (2010) “Teacher quality in educational production: Tracking, decay, and student achievement”, *The Quarterly Journal of Economics*, **125** (1), pp. 175–214.
- Schwerdt, Guido and Amelie C Wuppermann (2011) “Is traditional teaching really all that bad? A within-student between-subject approach”, *Economics of Education Review*, **30** (2), pp. 365–379.
- Sharkey, Nancy S and Dan Goldhaber (2008) “Teacher licensure status and student achievement: Lessons from private schools”, *Economics of Education Review*, **27** (5), pp. 504–516.
- Springer, Matthew G, Dale Ballou, Laura Hamilton, Vi-Nhuan Le, JR Lockwood, Daniel F McCaffrey, Matthew Pepper, and Brian M Stecher (2011) “Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT).”, *Society for Research on Educational Effectiveness*.
- Tanaka, Ryuichi and Kazumi Ishizaki (2018) “Do teaching practices matter for students’ academic achievement? A case of linguistic activity”, *Journal of the Japanese and International Economies*, **50**, pp. 26–36.
- Van Klaveren, Chris (2011) “Lecturing style teaching and student performance”, *Economics of Education Review*, **30** (4), pp. 729–739.
- Winters, Marcus A, Robert C Haight, Thomas T Swaim, and Katarzyna A Pickering (2013) “The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: Evidence from panel data”, *Economics of Education Review*, **34**, pp. 69–75.

Wiswall, Matthew (2013) “The dynamics of teacher quality”, *Journal of Public Economics*, **100**, pp. 61–78.

Woessmann, Ludger (2011) “Cross-country evidence on teacher performance pay”, *Economics of Education Review*, **30** (3), pp. 404–418.

文部科学省 (2012) 「教職生活の全体を通じた教員の資質能力の総合的な向上方策について (答申)」, [https://www.mext.go.jp/b\\_menu/shingi/chukyo/chukyo0/toushin/1325092.htm](https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1325092.htm), Accessed: 2020-07-31.

# 第3章 教員付加価値から見た教員の役割について – 日本の小学生を例にして –

## 概要

教員が子どもの能力の成長に与える影響について、多くの研究者や政策担当者がそれぞれの国のデータを用いて検証してきた。その一方で日本の教員を対象に行われた分析は非常に少なく、子どもの能力の成長に教員がどの程度影響を持っているのかはよく分かっていない。本稿では日本のある自治体における個票データを用いて教員付加価値を推定し、教員が子どもの能力の成長に与える影響について記述的に分析した。その際には、学力のみならず学習方略や非認知能力並びに学級での子ども・教員間の関係性を対象にして教員付加価値を求めた。その結果として学力や学習方略に対して個々の教員の違いが与える影響力が取りうる値の範囲は広がった一方で、非認知能力や学級での子ども・教員間の関係性に対して教員の違いは子ども全体の変動のうち少なくとも5%から8%以上を説明していた。更に、教員付加価値間の相関係数から、学力を伸ばすことが得意な教員と非認知能力・学習方略を伸ばすことが得意な教員は異なるという結果を得た。以上の結果から、教員の違いは学力より非認知能力に対して重要である可能性が示唆された。

## 3.1 はじめに

教員が子どもの能力の成長にどのように寄与しているのかという問いは、現代において教員付加価値に関する研究として深化している。教員付加価値とは教員が担当した子どものアウトプット（往々にして学力を指す）の平均的な増加分のことであり<sup>1</sup>、教員の質を計測した指標としてしばしば用いられる。多くの研究がこの教員付加価値をどのように推定するか、または推定された教員付加価値が子どもの能力の成長にどのように関係しているかを考察してきた (Kane and Staiger, 2008; Chetty et al., 2014a; Lefgren and Sims,

<sup>1</sup>Jackson et al. (2014) では “a label given to systematic variation in output across students assigned to the same teacher” と定義されている。

2012)<sup>2</sup>。そのみならず、実務面でも教員の評価に用いようとする動きもあり、研究・実務両面においては教員付加価値という指標は注目を集めている。

しかし、学力以外の教育成果に対して教員付加価値を分析した研究は必ずしも多くはない。教員付加価値についての研究の多くは学力を対象にしており、非認知能力などを対象に教員付加価値を推定し考察する研究は一部にとどまる (Jackson, 2018; Kraft, 2019)。教育成果について多くの人が最初に思い浮かべるのが学力ではあるものの、子どもにとっての教育成果は学力のみに限られるわけではない。特に、子どもの人的資本蓄積において非認知能力などの学力で表現できない能力の重要性が指摘されて久しい。そして教員が子どもに与える影響もまた、学力に対する影響に限られるわけではない。教員は子どもとのコミュニケーションを通じて非認知能力に対する影響を含めた多様な影響を子どもに与える。そのため、教員付加価値については様々な教育成果を対象とした研究が必要である。

更に、日本を対象にした教員の効果に対する定量的研究は非常に少ない。教育システムや学校を取り巻く状況は国によって大きな違いがあるため、海外での研究成果をどの程度日本においてそのまま適用可能かについては慎重な議論が必要である。そのため、その国独自のデータに基づく研究の蓄積は重要であるのにもかかわらず、日本の教員を対象にした研究は乏しいのが現状である。例えば、国際学力調査を用いた教育生産関数の文脈で教員の影響を分析する研究は存在する (Hojo and Oshio, 2012; Hojo, 2012) が、これらは教員の効果を分析することを目的とした研究ではなく教員の特徴を教育生産関数の1つの引数として捉えたものである。また二木 (2017) が日本における教員効果に関する数少ない量的研究であるが、あくまで都道府県レベルでの集計値に対する議論であり教員個人の教育効果を分析するものではない。教員の効果について知ることには政策的な意味も大きいため、日本を対象にして教員付加価値の研究を行う意義は大きい。

本稿では小学校の教員を対象にして教員付加価値の推定を行いその記述的な分析を行う。教員付加価値の推定には日本のある自治体における全ての子どもと教職員を対象に3年に渡って行われた調査のデータを用いた。その調査では、毎年4月に子どもに対する学力調査と質問紙調査及び教員に対する質問紙調査を行っている。特に、子どもに対する質問紙調査の中では非認知能力や学習方略に加えて学級の様子について尋ねており、本稿では学力に限らない多面的な教育成果に対する教員付加価値の推定をすることができた。そして、得られた教員付加価値の分布や相関を見ていくことで、教員付加価値が持っている性質を考察した。

結果として、学力や学習方略に対して個々の教員の違いが与える影響力が取りうる値の範囲は広がった一方で、非認知能力や学級での子ども・教員間

<sup>2</sup>元々、子どもは学校という場所を通して多くの時間を教員と共に過ごすため、教員が子どもに対して強い影響力を持ちうることを多くの研究が指摘してきた (Hanushek, 1986, 2006; Ammermüller et al., 2005; Hojo and Oshio, 2012)。

の関係性に対して教員の違いは子ども全体の変動のうち少なくとも5%から8%以上を説明していた。更に、学力の教員付加価値と非認知能力の教員付加価値の間の相関は小さく、統計的に有意ではなかった。これは、学力を伸ばすことが得意な教員と非認知能力を伸ばすことが得意な教員が異なることを示唆する。最後に、高い教員付加価値を持つ教員の観察可能な特徴を見つけることはできなかった。すなわち、年齢や性別といった教員の観察可能な特徴と推定された教員付加価値との間に統計的に有意な関係はなかった。

これらの結果は教員付加価値研究において次の様な意味を持つ。まず第一に、非認知能力や学習方略に対する教員付加価値のあり方を調べた数少ない研究であるということである。管見の限りセルフコントロールや自己効力感、学習方略に対する教員付加価値を現代的な手法で推定した研究は本研究が初めてである。特に学習方略については、認知能力が伸ばすことができる教員であっても学習方略を伸ばすことが出来る教員とは限らず、そのため子どもの学習のあり方を教員が変えると言うよりむしろ授業の方法など授業時間における教員のあり方の違いが教員付加価値の違いにもつながっているのではないかという示唆を得た。第二に、日本における現代的な手法による教員付加価値として管見の限り初出であるということである。その結果、非認知能力に対する教員付加価値について、その大きさの水準については小さい可能性があるものの、日本においてもアメリカにおける研究と同様の結論を得た。ただし、よりバイアスが少ない推定の必要性や子どもの長期的なアウトカムに対する教員の影響など、日本においては教員付加価値研究はさらなる進展を必要とする。

本稿の構成は以下の通りである。第3.2節では先行研究を概観する。第3.3節ではデータについて述べる。第3.4節では推定結果について述べる。第3.5節では本稿における結論を述べる。

## 3.2 先行研究

教員の子どもの教育成果への影響についての研究は古くから存在し<sup>3</sup>、Hanushek and Rivkin (2006)によればそれらの研究は3つの文脈に類型化できる。1つ目は労働市場一般における教員の立ち位置(賃金や志望倍率)を用いて教員の分析を行うものである。例えば日本を対象にした研究である二木(2017)は、労働市場における他職に比した教員の人気度を用いて分析を行なっている。これらの研究では総じて、他の産業と比較して優秀な人材が教員になっている事が子どもの成績を上げる可能性を検証している。

2つ目は教員の属性が子どもの教育成果に与える影響を分析するものである。ここで言う教員の属性とは、具体的には教員免許の有無や経験年数な

<sup>3</sup>教育経済学の分野で行われた初期の研究については、Hanushek and Rivkin (2006)にまとめられている。

どを指す。近年のこの種の研究例として Jackson and Bruegmann (2009) や Clotfelter et al. (2010) などを挙げるができる。しかし、これらの方向性での議論では研究によって結果が異なることが多く (Hanushek and Rivkin, 2006)、研究間で一致して「有効である」とされるような教員の属性は見つかっていない。

そして、3つ目が子どもの教育成果に対する教員付加価値を計算して分析する研究群である。教員付加価値とは子どもの教育成果の成長に対する教員の貢献を指した言葉であり<sup>4</sup>、もとより観察不可能な概念である。そのため、多くの場合は教員ごとの教員付加価値をデータから推定することになる。しかし、この推定をどの様に定式化するかについては必ずしも決まりきった手法があるわけではない。近年は Kane and Staiger (2008) に示された経験ベイズ的な手法を用いて教員付加価値を推定する研究が多く (Chetty et al., 2014a; Lefgren and Sims, 2012, など)、これらの研究は総じて教員付加価値が子どもの教育成果に対して一定の重要な役割を担っていることを報告している。その推定が含意している仮定の妥当性を巡った議論 (Rothstein, 2010; Kinsler, 2012) もなされているが、一方で実験的な状況を用いて教員付加価値が因果的な効果を持つ事を示す研究が存在する (Chetty et al., 2014a)。

しかし、学力以外の教育成果に対する教員付加価値の研究には蓄積が乏しい。教員が子どもに影響を与えるのは学力だけではなく、教室におけるコミュニケーションを通じて教員は様々な影響を子どもに与えうる。例えば Chetty et al. (2014b) では、良い教員は子どもの大学進学率や将来的な収入、貯蓄率や10代での妊娠率に影響を明らかにしていることを示している。しかし、上述の研究を始め多くの教員付加価値を扱った研究はテストスコアに対する付加価値を分析しており、学力以外への教育成果に対して教員が持っている影響力についてはよく分かっていない。数少ない例として、Jackson (2018) は子どもの出席率や停学といった社会行動に対する教員付加価値を推定し分析を行っている。しかしこの結果がどれほど一般的であるかなどはよく分かっておらず、更なる研究が必要である。

更に、非認知能力と教員付加価値に関する議論には教員の能力の多次元性という論点が存在する。すなわち、学力以外の教育成果を対象にして教員付加価値を計算できるということは、ある教員の教員付加価値は学力や非認知能力など教育成果の数だけ存在することになる。そのため、教員の能力は「良い/悪い」で表される1次元的な概念ではなく、「学力について良い/悪い」「非認知能力について良い/悪い」と続く多次元的なものになる。そして、この論点は、教員評価などの文脈では一層重要な問題になる。何故ならば、ある教員の付加価値への評価として「良い/悪い」を一概に定めることができなくなるからである。認知能力を伸ばすことに長けた教員が、非認知能力を伸ばすことに長けている保証はない。もしくは同じ認知能力であっても国語を伸ばすこと

<sup>4</sup>Jackson et al. (2014) は近年の教員の付加価値に関する研究動向をまとめている。

に長けた教員が算数を伸ばすことにも長けているとは限らない。このように、教員の能力の多次元性について言及した論文として、上述の Jackson (2018) では中学3年生という比較的高年齢の子どもを対象に、認知能力と非認知能力の教員付加価値の間の相関が弱いことを報告している。また、Goldhaber et al. (2013) では米国ノースカロライナ州のデータを用いて、小学校レベルでの教員の付加価値の科目間相関を算出し、国語科目と数学科目の付加価値相関係数が0.7程度であることを報告している。しかし、非認知能力についての研究がもとより少ないため、同様に上述の結果がどれほど一般的であるかなどはよく分かっていない。

最後に、日本を対象にした教員の付加価値についての研究は非常に少ないことについて詳述する必要がある。国際学力調査を用いた教育生産関数の推定の文脈で教員の影響を分析する場合は存在する (Hojo and Oshio, 2012; Hojo, 2012)。しかし、これらは教員の効果を分析することを目的とした研究ではない。上述の二木 (2017) は日本におけるある程度の規模のデータセットを用いた数少ない教員効果に関する研究であり、教員の質が子どもの欠席率を減少させようとしている。しかし、あくまで都道府県レベルでの集計値を用いた分析であり、個々の教員の違いを十分に反映できていない可能性がある (集計バイアス)。

### 3.3 データ

本稿では2016年 - 2018年 (調査は2015年から行われている) にかけて行われた日本のとある県Sにおける学力テストのデータを用いる。テストは毎年4月に行われ、テストと同時に非認知能力や普段の生活の様子を尋ねる質問紙調査を行った。テストを受けた生徒にはユニークなIDが振られ、経年で同一生徒を特定することが可能な設計になっている。調査の対象になったのは、S県における政令指定都市を除く全ての市町村の公立小学校及び公立中学校に所属する小学4年生から中学3年生までの生徒である。すなわち、小学4年生から中学3年生を対象にした悉皆調査として行われた。およそ生徒数では30万人程度、学校数では1064の学校 (小学校が708校、中学校が356校) の生徒が参加した。

上記の調査に加え、S県の市町村の一つであるT市に勤める小中学校の全ての教員を対象に質問紙調査を2016年から2018年にかけて行った。これは教員の授業での実践を尋ねるもので、教員一人一人にユニークなIDが振られ経年で同一教員を特定することが可能な設計になっている。ただし日本における公立学校の人事制度においてある教員は必ずしも同じ学校に所属し続けるわけではなく、T市以外の市町村への転出も多い。また全ての教員が毎年学級を一つ担当するわけでもなく、中には副担任などの形で自分の担当学級を持たない教員もいる。これらの理由から、複数年度で担当クラスを持つ

ている教員は少なくなってしまうことに注意をしなければいけない。本稿では教員のデータを用いるために、このT市のみのデータを取り扱った。

また今回分析に用いたのは小学校での教員データだけで、中学校での教員データは分析に用いなかった。その理由は、小学校と中学校では学級に対する教員の割り当て方として教科担任制か学級担任制かという違いがあるからである。日本の教育システムにおいては、多くの小学校の教員は一つの学級を受け持ち全ての科目を教えることになる<sup>5</sup>。一方で、多くの中学校では教員は専門科目を持ち、複数の学級で授業を行うことになる。その場合、中学生の能力の成長をどの教員の貢献と見做すことができるかは不明瞭である。例えば、子どもの国語の能力が成長していた場合は国語を担当した教員の貢献であると言えるだろう。しかし、非認知能力が成長していた場合は、どの科目の教員の貢献なのかを判断するのは困難である。以上の理由から、本稿では小学校の教員のみを分析の対象とした。まとめると、本稿で分析の対象となるのは、T市の小学校(その小学校数は12校)に通う4年生から6年生の子ども及び教員である。

S県T市の子どもが置かれている状況は必ずしも日本の教育状況そのものを表しているわけではないため、外的な妥当性には十分な検討が必要である。S県は東京近郊に所在する県であり、その中でもT市は人口13万人程度の中規模都市である。また、T市の公立小学校における2016年度の所属教員1人当たり児童数は20.5人であり<sup>6</sup>、日本の公立小学校平均の15.5人と比べると大きい。一方で、T市に住む人の平均的な課税対象所得は350万円<sup>7</sup>であり日本平均の321万円と比べ高く、日本の中では比較的所得の高い人が居住する都市であると言える。この様にT市をとりまく教育環境や家庭の状況は必ずしも日本のあらゆる地域と同一ではない。そのため、本校におけるT市を対象にした分析が他の市町村を対象にしたときにどの程度適用可能であるかについては十分に注意をする必要がある<sup>8</sup>。

表3.3および表3.4ではデータで用いる変数についての記述統計を示した。表3.3には子どもを単位とする変数について平均値、標準偏差及びサンプルサイズを年度・学年ごとに示した。「標準化の有無」と記された列に「○」がついている行の値は全て、標準化を行った後の値である<sup>9</sup>。また、調査が行われたのは毎年4月であることには注意が必要である。ある年の4月に行われた調査の結果は、その前年度の状況を反映していると考えられる。例えば、表

<sup>5</sup>ただし習熟度別学級の実施などにより、一部科目において複数の学級を担当するような状況は存在する。

<sup>6</sup>平成28年度学校基本調査より算出。なお本統計で言う教員とは学校に所属する教員全てを指し、その中には養護教諭など学級を受け持たない教員も含まれる。そのため、ここで示した値は平均的な学級規模とは異なることに注意が必要である。

<sup>7</sup>2013年度総務省『市町村税課税状況等の調』より、納税義務者一人当たり課税対象所得を算出した。

<sup>8</sup>一方で、学習指導要領などを通じて公立小学校では標準化されたカリキュラムが提供されており、T市と日本の他の地域の教育環境には共通する要素も多い。

<sup>9</sup>ただし、全ての子どもの対象に標準化を行った後に、前年度に紐づく教員がいないような子どもは取り除き計算したため厳密に平均0分散1になっているわけではない。

3.3において2016年5年生の値は、2015年4年生時の状況を反映した値と考えられる。そのため、教員もその前年の状況に紐付けられる（例えば、2016年5年生の子どものデータに紐付けられる教員は、2015年4年生時点の教員である）。表3.4では得られた教員についての情報をまとめている。表3.4上段では教員の数を年度及び学年ごとに分けて示し、表3.4下段では教員を単位とする変数の記述統計を示した。以降ではデータで用いる変数について記述する。

## 認知能力

認知能力を表す値として、国語と算数の学力の推定値を用いる。すなわち、S県学力調査では小学生に対して国語と算数のテストを実施しており、その結果からIRT(Item Response Theory)を用いて個人の科目ごとの学力の推定値を算出した。加えて本稿では分析にあたって全ての学力の推定値を年度・学年及び科目ごとに平均0・分散1に標準化を行なった。この操作によって標準化された学力は、実質的に学年内での相対的な高低を表すことになる。このIRTによって算出された標準化された学力推定値を以降では認知能力の値として分析の対象とする。

## 非認知能力及び学習方略

また、調査ではテストとは別に質問紙調査によって、子どもの非認知能力や学習方略を尋ねている。この「非認知能力」という言葉はある能力を具体的に指し示すものではなく、認知能力の尺度では捉えられないような性格的特徴などの個人的な属性を広く指す概念として用いられている(Heckman and Kautz, 2012)。特にその「能力」という用語法は、それらの属性のうち訓練などによって学習可能である部分に着目していること表している。この非認知能力という言葉が指し示す広範な概念のうち、本研究では「セルフコントロール」および「自己効力感」の2つの能力を対象に分析を行った。

このように非認知能力という言葉が指し示す範囲は非常に広範であるのにもかかわらず、本稿で分析しているのはそのうち「セルフコントロール」および「自己効力感」であることには注意が必要である。すなわち本稿で議論が可能であったのは、非認知能力そのものではなくその下位尺度でしかない。本稿ではこれらのセルフコントロールおよび自己効力感に対する教員付加価値の分析を通じて学力に限らない「非認知能力」の教員付加価値の特徴を考察していく。しかし、セルフコントロール・自己効力感以外にも子どもの非認知能力(社会情緒的能力)を構成するような能力は存在する。そしてその非認知能力に対する教員付加価値がセルフコントロール・自己効力感に対する教員付加価値と同様の性質を持つとは限らない。このような分析の射

程を巡っては注意深く結果を解釈する必要があり、また学術的な研究としてもこのような議論の射程を広げていくような研究が待たれる。本章では取得できたデータにおける都合から上記2つの非認知能力の下位尺度を対象にした分析を行うが、3.2で述べたように非認知能力を対象にした教員付加価値研究が少ない研究状況では本稿の価値は失われることはない。

以下ではそれらのセルフコントロールおよび自己効力感の定義の詳細について記述する。

### 3.3.1 セルフコントロール

セルフコントロールとは、「理想、価値観、道徳、社会的期待などの基準に合わせて自分自身の反応を変化させ、長期的な目標の追求をサポートする能力」と定義される (Baumeister et al., 2007)。このセルフコントロールの能力が高いことは、自らの行動の意識的な制御を通じて、学業や勉強時間 (Tsukayama et al., 2013)、また賃金などの労働条件などを予測することがこれまで示されてきている (李嬋娟, 2014)。

本研究では Tsukayama et al. (2013) で示されたセルフコントロール尺度を日本語に翻訳したものをを用いる。この尺度は学齢期の子どもを対象に開発されたセルフコントロール尺度であり、比較的簡易的にセルフコントロールを測定することができるという特徴をもっている。Tsukayama et al. (2013) では、この尺度で測定されたセルフコントロールの値によって、GPA や勉強時間・TV ゲームに費やす時間などを予測できることが示されている。尺度を構成する具体的な質問項目については表 3.1 に示した。各々の質問は5件法によって回答され、尺度を計算する際には点数が高いほどセルフコントロールが高いことを指し示すように変換を行った (例えば「授業で必要なものを忘れた」という項目に対して1(ほとんどない)と回答された場合は、指し示す値を5とした)。最終的にこれらの質問項目が指す値を合計して、「セルフコントロール」の値として用いた。

### 3.3.2 自己効力感

自己効力感とは「個人がある状況において必要な行動を効果的に遂行できる可能性の認知」と定義される (成田 健一他, 1995)。すなわち、自己効力感の低い人は特定のタスクを達成することができないと考え、そのタスクを避けようとするのに対し、自己効力感の高い人はタスクが困難であっても達成に向けて動こうとすると考えられる。自己効力感について心理学的研究は、高い自己効力感は高い学業成績 (Multon et al., 1991; Richardson et al., 2012; Behncke, 2012) と関係があることを示している。

本研究では Pintrich and de Groot (1990) で示された尺度のうち Self-Efficacy for Learning and Performance 下位尺度を日本語に翻訳したもの

を用いる。この尺度は子どもの学習動機づけを構成する尺度として開発され、子どもの学業成績と正の相関を持っていることが示されている。尺度を構成する具体的な質問項目については表 3.1 に示した。各々の質問は 5 件法によって回答された。最終的にこれらの質問項目が指す値を合計して、「自己効力感」の値として用いた。

### 3.3.3 学習方略

学習方略とは学習のやり方を表すメタ的な認知能力として主に教育心理学の分野で研究がなされてきている (佐藤, 1998; Pintrich et al., 1991)。学習方略が高い子どもは、学習のやり方を工夫する中でその学習の成果を高めることができると考えられる。Pintrich and de Groot (1990) といった心理学的研究が、高い学習方略が高い学業成績につながることをこれまで示している。

本研究では佐藤 (1998) で示された学習方略尺度を用いる。当該尺度は Pintrich and de Groot (1990) や Pintrich et al. (1991) をもとに、その下位尺度を整理しながら日本の小学生・中学生を対象に開発された心理尺度である。尺度を構成する質問項目については表 3.2 に示した。各々の質問は 5 件法によって回答され、尺度を計算する際には点数が高いほど学習方略が高いことを指し示すように変換を行っている (例えば「授業の内容が難しいときは、やらずにあきらめるか、簡単どころだけ勉強する」という項目に対して 1(ほとんどない) と回答された場合は、指し示す値を 5 とした)。最終的にこれらの質問項目が指す値を合計して、「学習方略」の値として用いた。

これらの値はそれぞれ複数の質問項目 (表 3.1・表 3.2) の回答 (5 件法) を足しあげることによって得られる。更に分析においては学年ごとに平均 0・分散 1 に標準化を行い、学年内での相対的な高低を表す値に直した。このうちセルフコントロールと自己効力感の 2 つの非認知能力は一部の学年でしか尋ねていないことに注意が必要である。すなわち、セルフコントロールは 2016 年に中学 1 年生であった学年と 2017 年に 5 年生であった学年にのみ尋ねており、自己効力感は 2016 年に 5 年生であった学年にのみ尋ねている。そのため非認知能力をアウトカムとして用いた分析では、他の変数をアウトカムとした分析に対して用いることができるサンプルが少なくなる。

データ

項目	逆転項目	質問
セルフコントロール	✓	授業で必要なものを忘れた
	✓	他の子たちが話をしているときに、その子たちのじゃまをした
	✓	何か乱暴なことを言った
	✓	机・ロッカー・部屋が散らかっていたので、必要なものを見つけることができなかつた
	✓	家や学校で頭にきて人やものにあたった
	✓	先生が、自分に対して言うことを思い出すことができなかつた
	✓	きちんと話を聞かないといけないときにぼんやりしていた
	✓	イライラしているときに、先生や家の人（兄弟姉妹を除きます）に口答えをした
自己効力感		授業ではよい評価をもらえるだろうと信じている
		教科書の中で一番難しい問題も理解できると思う
		授業で教えてもらった基本的なことは理解できたとと思う
		先生が出した一番難しい問題も理解できると思う
		学校の宿題や試験でよい成績をとることができると思う
		学校でよい成績をとることができると思う
		授業で教えてもらったことは使いこなせると思う
		授業の難しさ、先生のこと、自分の実力のことなどを考えれば、自分はこの授業でよくやっているほうだと思う

表 3.1: 質問リスト:非認知能力

注) 本表では変数「セルフコントロール」「自己効力感」及び「勤勉性」を構成する質問をまとめた。列「逆転項目」にチェックマークがついている場合、その質問は測定の対象が他の質問と逆になっている。また全ての質問に対して、回答は5件法で得られた。

項目	逆転項目	質問
学習方略すべて		勉強のやり方が、自分に合っているかどうかを考えながら勉強する
		勉強でわからないところがあったら、勉強のやり方をいろいろ変えてみる
		勉強しているときに、やった内容を覚えているかどうかを確かめる
		勉強する前に、これから何を勉強しなければならぬかについて考える
		勉強するときは、最初に計画を立ててから始める
		勉強をしているときに、やっていると正しくできているかどうかを確かめる
		勉強するときは、自分で決めた計画に沿って行う
		勉強しているときは、たまに止まって、一度やったところを見直す
		勉強するときは、参考書や事典などがすぐ使えるように準備しておく
		勉強する前に、勉強に必要な本などを用意してから勉強するようにしている
		勉強している大切だと思ったところは、言われなくてもノートにまとめる
		勉強で大切なところは、繰り返し書くなどして覚える
		勉強でわからないところがあったら、友達にその答えをきく
		勉強でわからないところがあったら、友達に勉強のやり方をきく
		勉強のできる友達と、同じやり方で勉強する
		勉強するときは、最後に友達と答え合わせをするようにする
		勉強するときは、内容を頭に思い浮かべながら考える
		勉強するときは、内容を自分の知っている言葉で理解するようにする
		新しいことを勉強するとき、今までの勉強したことと関係があるかどうかを考えながら勉強する
✓		学校で勉強しているとき、とてもめんどうでつまらないと思うことがよくあるので、やろうとしていたことを終える前にやめてしまう
		今やっていることが気に入らなかつたとしても、学校の勉強でよい成績をとるために一生懸命がんばる
✓		授業の内容が難しいときは、やらずにあきらめるか、簡単なところだけ勉強する
		問題が退屈でつまらないときでも、それが終わるまでなんとかやり続けられるように努力する
		学級での生活は楽しかったですか
		学級は落ち着いて学習する雰囲気がありましたか
		学校の友達とは自分のよいところを認めてくれましたか
		学校の先生たちは自分の悩みの相談にのってくれましたか

表 3.2: 質問リスト: 学習方略、学級での子ども・教員間の関係性

注) 本表では変数「学習方略」及び「学級での子ども・教員間の関係性」を構成する質問をまとめた。列「逆転項目」にチェックマークがついている場合、その質問は測定の見え方が他の質問と逆になっている。また「学習方略」を表す質問に対して回答は5件法で得られ、「学級での子ども・教員間の関係性」を表す質問に対しては回答は4件法で得られた。

## 子ども・教員間の関係性

更に学級風土 classroom climate、特にその中でも学級での子ども・教員間の関係性を対象にして教員付加価値を推定する。学級風土 (もしくは学校風土 school climate) とは子どもの学級生活の質や特徴を表し、主に教育心理学や教育社会学で研究されている概念である (Wang and Degol, 2016; Berkowitz et al., 2017)。その中では学級・学校における規範、価値観及び人間関係などが子どもの学級での生活に働きかけ、学力や行動にまで影響を与えるのではないかと考えられている (Hopson and Lee, 2011)。そして、その学級風土の要素の一つとして、学級における子ども・教員間の関係性は挙げられている (Wang and Degol, 2016)。

本稿では、この学級における子ども・教員間の関係性について分析を行う。調査では前年度の学級における教員及び友人との関係について、4つの質問 (全て4件法) によって子どもに尋ねた (表 3.2)。それらの項目全てを足し合わせて学級での子ども・教員間の関係性を表す変数とした。更に分析においては学年ごとに平均 0 分散 1 に標準化を行った。

## その他の変数

調査において子どもの家庭での状況などを子どもに尋ねて調査している。そのうち、本稿では (1) 週あたりの通塾時間 (8 件法) 及び (2) 家庭にある本の冊数 (5 件法) を用いる。これらは共に子どもの家庭に社会経済的地位 (Socio-economic Status、以降 SES と表記) を表す代理変数として用いる。前者は経済資本を表す変数であり、後者は文化資本を表す変数である。

## 教員

ある教員は必ずしも T 市に留まり小学校高学年を対象に教鞭を取り続けるとは限らない。表 3.4 上段では教員質問紙への回答を得ることができた教員の数を年度及び学年に分けて示し、同時に年度ごとに調査に新たに参加した教員の数及び調査への参加が認められなくなった教員の数を示した。調査に参加する教員数は学校内の配置転換や人事異動によって変わると考えられる<sup>10</sup>。例えば、T 市からの転出もしくは転入によって教員質問紙への調査の可否は変わってくる。そのようにして、ある年に調査に参加した教員のうち凡そ半分は次の年には調査に参加しなくなる。例えば、2017 年度において調査に参加した小学校教員の数は 80 人であったが、その内 37 人が 2018 年度の調査には参加していない。このように、調査に参加した教員ののべ総数に対して、継続して調査に参加し続ける教員の数は少なくなってしまう。本稿では 2016 年度 - 2018 年度に渡って一度でも調査に参加した教員を対象に分析を行う

<sup>10</sup>無論、あえて調査に参加をしないという教員も存在する。

が、そのうち単年度しか調査に参加していない教員も存在することには注意が必要である。

また小学4年から小学6年の学級を担当した全ての教員が調査に参加したわけではない。調査に参加していない教員の数を示すため、表3.4上段では調査の対象となった学級の数と同時に示している。例えば、2016年度において調査の対象となった学級は109学級あるのに対し、調査に参加した教員は82人であり全体の約80%である。これらの調査に参加していない教員が存在することには注意が必要である。次節の式3.1で求めるような教員付加価値の推定値が調査に参加していない教員の存在でバイアスされるわけではないが、その一方でやはり本稿の分析によって教員付加価値を得ることができる教員には偏りが出てしまう可能性は存在する。すなわち、教員という言葉で名指したい集団とは違う母集団を対象に分析をしてしまう可能性が存在する。平成28年度学校教員統計調査(文部科学省)によればT市の小学校の教員の平均年齢は40.2歳である。一方で、表3.4下段に示したようにT市教員調査上の教員の平均年齢は36.2歳であり、T市教員調査に参加した教員の年齢はT市の小学校に所属している教員の年齢と比べて4歳ほど低い。しかし、学校教員統計調査の対象となった教員には小学校低学年の教員や学級を持たない管理職の教員も含まれるため、この平均年齢の差がT市教員調査に参加していない教員によるものであるとすることは難しい。このように調査に参加していない教員の特徴をデータから分析することはできないが(そもそも調査に参加していないためデータ上に存在しない)、本稿の分析結果はT市教員が調査に参加している教員を母集団としたものであることには十分に留意して解釈を進める必要がある。

教員質問紙では担当した学級の他に、教員の特徴について尋ねている。表3.4下段にはその記述統計を示した。例えば平均的な年齢は36.2歳であり、全国の公立小学校教員の平均年齢である43.4歳<sup>11</sup>に比べると若いことがわかる。また、教員の出身大学については注意が必要である。すなわち、出身大学(大学院)について回答が得られたのは62人のみであり、データを得られた教員の総数の半分以下であった。これは、2016年に行った調査においてのみ教員に卒業した大学(大学院)を尋ねることが可能であり、2017年度-2018年度に新たに調査に参加した教員については相当する質問を行うことができなかったためである。

<sup>11</sup>平成28年度学校教員統計調査より。

年度 学年	2016			2017			2018			
	5	6	7	5	6	7	5	6	7	
	標準化の有無									
国語	○	-0.020 (849,1.009)	0.020 (837,1.000)	-0.003 (801,1.001)	0.049 (830,1.002)	0.013 (94,4,0.983)	0.025 (713,1.002)	0.022 (1212,0.984)	0.016 (1045,0.990)	0.020 (955,0.984)
算数	○	-0.054 (848,0.915)	0.024 (837,0.987)	0.003 (801,0.980)	0.046 (830,1.005)	-0.007 (94,4,0.989)	0.006 (713,0.998)	0.019 (1212,0.990)	0.012 (1045,0.964)	0.006 (955,0.976)
学習方略	○	-0.010 (785,0.987)	-0.026 (777,0.992)	-0.035 (756,1.010)	-0.028 (767,1.004)	-0.013 (876,0.992)	0.040 (674,0.991)	0.009 (1106,0.999)	-0.003 (968,1.009)	-0.006 (887,0.997)
セルフコントロール	○	-0.031 (839,0.998)		-0.028 (795,0.997)	-0.016 (817,1.009)				-0.019 (1030,1.016)	
自己効力感	○	-0.057 (837,1.015)	-0.045 (834,1.037)	0.070 (796,0.959)	-0.022 (828,1.006)	0.010 (942,1.003)	0.084 (713,0.963)	0.002 (1209,1.000)	0.022 (1045,0.999)	-0.008 (955,1.013)
子ども・教員間の関係性	○	3.018 (837,1.130)	3.220 (833,1.117)	2.955 (796,1.174)	3.048 (827,1.097)	3.116 (941,1.070)	3.179 (711,1.132)	3.145 (1205,1.143)	3.202 (1043,1.099)	3.184 (953,1.142)
家庭にある本の冊数		2.850 (834,2.227)	2.765 (831,2.160)	2.700 (797,1.886)	3.150 (800,2.431)	2.992 (920,2.245)	3.010 (696,1.982)	3.226 (1187,2.438)	3.179 (1026,2.325)	2.980 (947,1.990)
通塾時間										

表 3.3: 記述統計 (子ども)

注) 本表では、本稿で用いている変数のうち子どもを単位とする変数の記述統計を示している。各々のセルでは、対応する変数について「平均(サンプルサイズ、標準偏差)」という記述方法を用いている。データ単位は全て、子ども1人である。

(a) 調査に参加した教員 (学級) の数

	4年	5年	6年	ALL	IN	OUT	学級数
年度							
2015	28	26	28	82	82	0	109
2016	26	30	24	80	42	44	120
2017	37	32	34	103	60	37	122

(b) 教員の特徴

	年齢	教員歴	女性教員	出身大学偏差値
教員	36.210 (167,10.320)	12.416 (166,10.101)	0.509 (167,0.501)	49.338 (62,7.465)

表 3.4: 記述統計 (教員)

注) 本表では、本稿で用いている変数のうち教員を単位とする変数の記述統計を示している。上段では、本稿で教員付加価値の推定の対象とした教員の数を示した。小学校では1人の教員が1つの学級を担当するため、教員の数はそのまま学級の数を表す。行は年度を表す。1列目(「4年」)から3列目(「6年」)までは、学年ごとに教員が何人いるかを示し、4列目(「ALL」)はその合計を表す。5列目(「IN」)は新たに教員質問紙に参加した教員の数を表し、6列目(「OUT」)は教員質問紙に参加しなくなった教員の数を表す。7列目(「学級数」)は、調査の対象となった学級の数を表す。下段では、教員の特徴を表す変数の記述統計を示した。各々のセルでは、対応する変数について「平均(サンプルサイズ、標準偏差)」という記述方法を用いている。データ単位は全て、教員1人である。

### 3.4 推定戦略と推定結果

#### 3.4.1 教員付加価値の推定

教員付加価値を推定するために、ある子どもの教育成果を次のような教育生産関数で表すことができるとする<sup>12</sup>。

$$y_{it} = \alpha y_{it-1} + X_{it}\beta + \mu_j + \theta_c + \epsilon_{it} \quad (3.1)$$

$y_{it}$  は  $t$  年における子ども  $i$  の教育成果を表す。本稿では教育成果として、学力(国語、算数)の他に非認知能力(自己効力感、セルフコントロール)や学習方略及び学級での子ども・教員間の関係性を考える。 $X_{it}$  は観察可能な子ども  $i$  の特徴を表し、本稿ではSESの代理変数として家庭にある本の冊数及び通塾時間を用いる。 $\theta_c$  は子ども  $i$  が所属する学級  $c$  の効果を表し、 $\mu_j$  は学級  $c$  を担当する教員  $j$  の付加価値を表し、 $\epsilon_{it}$  は子ども  $i$  の  $t$  年におけるその他の要因全てを表す。 $\mu_j \cdot \theta_c \cdot \epsilon_{it}$  は全て観察不可能であり、互いに独立であることを仮定する。

<sup>12</sup>この定式化の妥当性については Appendix A.1 にて議論を行った

式 3.1 は付加価値モデルと広く呼ばれるモデルの一つである (Todd and Wolpin, 2003)。前期の能力を表す  $y_{it-1}$  をコントロールすることで、これまで蓄積された能力もしくは時間変動しない要因の影響を取り除くことを意図している。そのため、分析に用いる変動は子どもの能力の成長である。教員付加価値はこの子どもの能力の成長に対する教員の貢献  $\mu_j$  を評価することになる。

この教員の貢献の差  $\mu_j$  は教員の特定の特徴の有効性を表すものではなく、総体的な教員の貢献の差を表すことには注意が必要である。多くの教員の質的研究は「どのような教員が子どもの能力成長を促すことができるのか？」という教員の有効性の要因を探究してきた。本稿においても4章において、教員が学級で用いる指導方法に着目して分析を行なっている。しかし、教員付加価値研究はそのような有効性の要因研究とは異なる観点から分析を行う研究である。式 3.1 において教員の付加価値として表される  $\mu_j$  はデータにおいてなにか観察可能な値ではなく、教員が子どもに与える影響を表す潜在的な値として定義されている。その中にはあらゆる教員の要素が含まれている。例えば教員  $j$  の教員付加価値  $\mu_j$  には、教員  $j$  の性別や年齢からどのような指導方法を用いているか、これまでどのような研修を受けてきたのか、どのような家庭で育ったのかといった教員  $j$  を構成する全ての要素が含まれている。教員付加価値研究は「そもそも教員によってパフォーマンスの違いはあるのか?」「パフォーマンスの違いがあるのだとすると、それはどのような性質のものなのか?」という点に関心を寄せており、その意味では教員の有効性の要因研究よりも一つ上位の問いになる。もし教員によるパフォーマンスの違いが存在しないのならば、そのパフォーマンスの差の要因を探ったところで政策などに与える影響は小さい。いずれにせよ、本章で焦点をあててるのはそういった教員の相対的な貢献の差であることは強く強調しておく必要がある。

本稿では教員付加価値  $\mu_j$  を Kane and Staiger (2008) の手続きに従って求める。Kane らの手法では式 3.1 を推定し、その残差を用いて  $\mu_j$  の推定値を計算する。すなわち、式 3.1 を  $s_{it} = y_{it} - \alpha y_{it-1} - X_{it}\beta = \mu_j + \theta_c + \epsilon_{it}$  と書き直せば、 $s_{it}$  は  $\mu_j$  という求めたい値に  $\theta_c$  と  $\epsilon_{it}$  というノイズが加わった情報を表す変数と見做せる。この時、ある学級  $c$  はある教員  $j$  によって担任されるため、 $s_{it}$  を子ども  $i$  が所属するクラス (クラス  $c$  に所属する子どもの集合を  $I_c$  と表記する。またその集合の要素数すなわちクラスの人数を  $\|I_c\|$  と表記する。) で平均した値  $s_{jc}$  は

$$\begin{aligned} s_{jc} &= \frac{\sum_{i,i \in I_c} s_{it}}{\|I_c\|} \\ &= \mu_j + \theta_c + \frac{\sum_{i,i \in I_c} \epsilon_{it}}{\|I_c\|} \\ &= \mu_j + v_{jc} \end{aligned}$$

となる。ただし学級  $c$  に所属する子どもの集合を  $I_c$  で表している。この  $s_{jc}$  は教員付加価値  $\mu_j$  についてのノイズ付きの情報と見做せるため、教員  $j$  の教員付加価値の期待値  $\mu_j$  はある重み  $\alpha_{jc}$  を用いて

$$E(\mu_j | s_{jc}) = \sum_{c, c \in I_j} \alpha_{jc} s_{jc}$$

で表される<sup>13</sup>。本稿では以降特に断り無く「推定された教員付加価値」と記述されている場合、上述の手続きで求めた教員付加価値を指す。現代的な教員付加価値研究のほとんどが、このデータから得られる情報にその正確度に応じた重みを与える手法に準拠し教員付加価値を推定しており (Kane and Staiger, 2008; Lefgren and Sims, 2012; Chetty et al., 2014a; Jackson, 2018)、この手法で求めることはそれらの研究と比較しながら本研究を評価することが出来るため有用である。

教員付加価値を求める手法には幾つかの派生形がある。Kane and Staiger (2008) が示した教員付加価値を推定する方法は、ある教員が担当するクラスや生徒数は有限であることから必ずしも一致性などの統計的な性質が保証されているわけではない。また教員付加価値に期待される性質も状況によって異なる。例えば、上述の定式化では教員付加価値は時間を通じて一定であることを前提にしているが、新任の教員が業務経験を通じて徐々にその能力を高めていくことなどを考慮に入れようとすれば教員付加価値は時間を通じて変化しうるものとして定式化する必要がある。そのため、教員付加価値を推定する手法には複数の教育成果で共通する教員の付加価値を評価した Lefgren and Sims (2012) や教員の付加価値の時間による変化を許容する Chetty et al. (2014a) などの幾つかの派生系が存在する。更に、似た手法として式 (1) のうち  $\theta_c + \epsilon_{it}$  を誤差項を含む観察できない要因  $v_{it}$  などとして  $y_{it} = \alpha y_{it-1} + X_{it}\beta + \mu_j + v_{it}$  という式を推定し、そのうち教員の固定効果  $\mu_j$  を教員付加価値として用いる研究も存在する<sup>14</sup>。本稿では比較のため、この固定効果を用いた推定も報告する。

この式 3.1 において  $\mu_j$  を識別するための仮定は、 $\mu_j$  が  $\theta_c$  や  $\epsilon_{it}$  と独立である事である。すなわち、教員  $j$  が内生的に学校や学級に配置されているならず、そのような関係があれば推定された教員付加価値はその割り当て方に応じたバイアスを含むことになる。このバイアスの方向はその内生的割り当てがどのようにおこなわれているかによって異なる。例えば、経験の浅い教員の授業負担を軽くするために SES の高い子どもを優先的に割り当てているとする。そのような子どもが塾などの外部の教育リソースを活用して能力

<sup>13</sup>重み  $\alpha_{jc}$  は  $\mu_j$  に関する情報  $s_{jc}$  の正確度を表すパラメータである。具体的な計算方法は Kane and Staiger (2008) を参照せよ。

<sup>14</sup>Kane らの手法が教員固定効果の計算と大きく異なる点として、 $s_{it}$  の信頼度で教員付加価値を評価しているという点を挙げる事ができる。例えば 5 人しかいないクラスを担当する教員の固定効果と 30 人いるクラスを担当する教員の固定効果ではその信頼度に大きく差があり、Kane らの手法ではその点を考慮に入れて計算をしていると言う事ができる。

を伸ばしているといった場合、担当教員の教育効果を高く見積もってしまう可能性がある。本研究は実験的もしくは準実験的な状況を利用した研究ではないため、この仮定が成立しているかどうかは注意深く検討する必要がある。

その仮定の妥当性を考えるために、表 3.5 では教員と子どもの観察可能な特徴の間に相関が無いかを調べた。具体的には子どもの特徴を (学校固定効果・年度・学年の影響は取り除いた上で) 担当した教員の特徴に回帰し、その係数の推定値を示した。その結果、示した推定値は全て統計的に有意ではなかった。もちろん、観察されない子どもや教員の特徴を基に教員が内生的に配置されている可能性もありその検証は今後の研究において重要であるものの、子どもの特徴に合わせて教員を内生的に配置している傾向を見つけることはできなかった。

ただしこのような子どもの観察可能な特徴が教員の配置と相関がないことをもって内生的な配置が全くなかったとは言いきれない。本稿で用いるデータには含まれなかった要素と関係する形で内生的な配置が行われている可能性は存在する。しかしほとんど全ての場合において考えられるあらゆる要因と教員配置が関係ないことを示すことは非常に困難であると考えられる。そのような中で Kane and Staiger (2008) や Chetty et al. (2014a) の研究は、実験的な状況において推定された教員付加価値が教員のその後のパフォーマンスに対する予測能力を持つことをもって、そのような内生的配置などのバイアス要因だけではない教員付加価値の存在を実証していた。このように実験的状況を用いて推定された教員付加価値の検証を行うことが上述の仮定の検証としては最も妥当ではあるものの、本稿も含めて日本におけるそのような研究は管見の限り存在しない。このような検証の必要性は本稿の大きな限界点であり、更なる研究の蓄積が求められるところである。

		前年度学力 (国語)	前年度学力 (算数)	本の冊数	通塾	女性	就学援助
年齢	推定値	0.000 (0.001)	0.002 (0.001)	0.000 (0.001)	-0.001 (0.001)	0.000 (0.001)	0.000 (0.001)
	サンプルサイズ	8051	8048	8011	7904	8051	4904
経験年数	推定値	-0.003 (0.003)	0.001 (0.003)	-0.002 (0.003)	-0.002 (0.001)	0.000 (0.001)	0.001 (0.001)
	サンプルサイズ	6491	6488	6460	6381	6491	3811
女性教員	推定値	-0.003 (0.023)	-0.006 (0.023)	0.023 (0.026)	0.002 (0.011)	-0.001 (0.012)	0.000 (0.010)
	サンプルサイズ	8088	8085	8048	7940	8088	4904
出身学校の偏差値	推定値	0.000 (0.003)	-0.002 (0.003)	0.003 (0.003)	-0.001 (0.001)	0.000 (0.001)	0.000 (0.001)
	サンプルサイズ	3614	3613	3590	3553	3614	2812

表 3.5: 教員と子どもの観察可能な特徴間の相関

注) 本表では子どもの観察可能な属性 (前年度学力 (国語・算数)、本の冊数、通塾、女性、就学援助) を教員の観察可能な特徴 (年齢、経験年数、女性教員、出身学校の偏差値) に回帰した結果を表している。推定では学校固定効果、学年、年度の影響を取り除いて推定した。各々のセルでは推定値及び標準誤差・有意水準を示しており、「推定値 有意水準 (標準誤差)」という表記法を用いている。「\*\*\*」・「\*\*」・「\*」はそれぞれ 1%・5%・10%の水準で統計的に有意であることを表す。

また  $\mu_j$  が  $\theta_c$  や  $\epsilon_{it}$  に対して外生であったとしても、個々の  $\mu_j$  を正確に推

定するためにはある教員に対して複数の学級のデータが必要である (Jackson et al., 2014)。すなわち、教員が担当するクラス  $c$  固有の効果  $\theta_c$  は学級の生徒間では固有であり、学級内のサンプルを多く集めたところでその影響をとりぞくことはできない。そして、複数の学級のサンプルを用いることで、そのノイズの影響を軽減することができる。逆に言えばデータにおいて担当学級の数が少ない教員付加価値の推定値  $\hat{\mu}_j$  は  $\theta_c$  の影響を強く受けてしまう<sup>15</sup>。Kane and Staiger (2008) の様な経験ベイズ的手法が用いられるのは、まさにその様なノイズの大きさだけ教員の教員付加価値  $\hat{\mu}_j$  を割り引くためであった。本稿では小学校でのデータを取り扱うが、小学校ではある教員は年度ごとに1つの学級を担当するため、中には担当した学級が1つしかないような教員も存在する。そのため、 $\hat{\mu}_j$  の標準偏差  $\hat{\sigma}_\mu$  を報告する際には (表 3.6) そのような教員をデータから除いた場合の結果も報告する。

特にこの問題が重要になってくるのは、年度が変わってもある教員 ( $j'$  と表記する) は同じクラスを持ち上がって担当し、教員に割り当てられた子どもの構成が時間を通じて一定のケースの場合である (このような教員割り当てを「持ち上がり」とここでは表記する)。「持ち上がり」が教員  $j'$  に発生している時、教員効果  $\mu_{j'}$  とクラス効果  $\theta_c$  は識別することができない。ある教員  $j'$  の付加価値の推定値はある同じクラスの観察不可能な特徴による効果を含むことになり、内生性によるバイアスはより深刻化する。日本の教育システムにおいては、ある教員とあるクラスの割り当てがどの程度持続するかは自治体や学校に依存する。クラス替えが毎年発生し教員にクラスを毎年割り当てし直す学校もあれば、2年から3年程度同じクラスを教員に割り当て続ける学校もある。しかし、本稿におけるデータ内で「持ち上がり」の発生割合を計算すると、T市においては持ち上がりクラスは全体の1%と非常に稀にしか起きていなかった。そのため、クラスの持ち上がりの問題はほとんど発生していないと考えることができる。

更に、推定手法によって教員効果  $\mu_j$  の分散  $\sigma_\mu$  の推定値には異なるバイアスが存在する。教員固定効果を用いた推定では、測定誤差と真の教員効果  $\mu_j$  の識別ができない。そのため、その推定値の標準偏差  $\sigma_{\hat{\mu}_{FE}}$  は真の分散  $\sigma_\mu$  よりも大きくなるというバイアスが存在する。一方 Kane and Staiger (2008) による推定方法で求められた推定値の分散  $\sigma_{\hat{\mu}_{KS}}$  は真の分散  $\sigma_\mu$  よりも小さくなるというバイアスが存在する (Jacob and Lefgren (2005), Online Appendix C; Kraft (2019))。すなわち、真の教員付加価値の分散  $\sigma_\mu$  と  $\sigma_{\hat{\mu}_{FE}}$  および  $\sigma_{\hat{\mu}_{KS}}$  の間の関係は次の様にまとめることができる。

$$\sigma_{\hat{\mu}_{KS}} \leq \sigma_\mu \leq \sigma_{\hat{\mu}_{FE}} \quad (3.2)$$

<sup>15</sup>本稿では推定された個々の  $\mu_j$  そのものを直接分析の対象にはしなかったが、例えば教員付加価値を教員評価に用いることを考えた時にこの問題は強い制約となる。

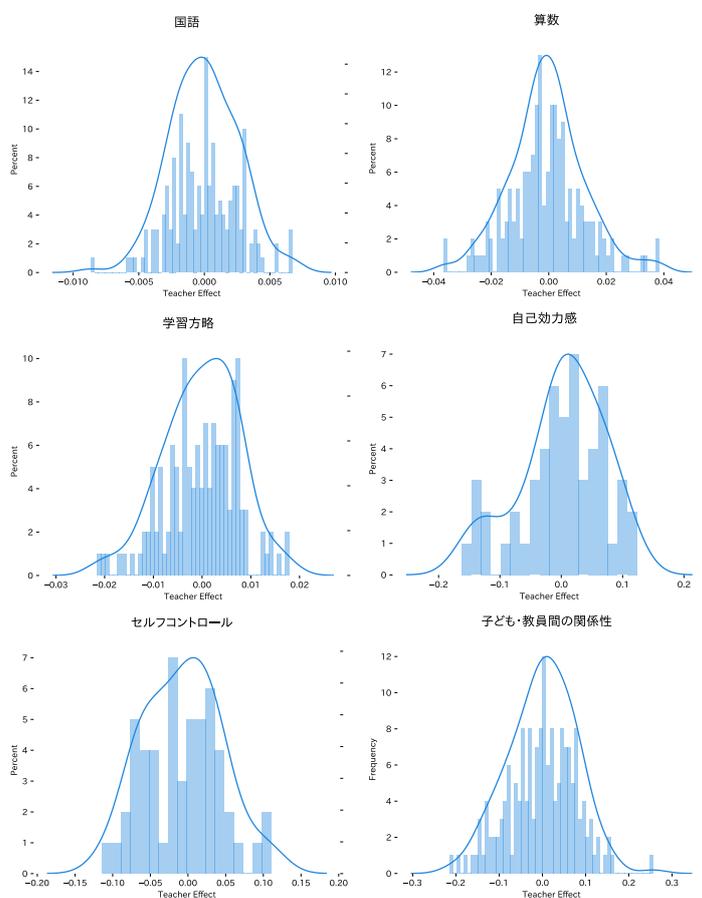


図 3.1: 推定された教員付加価値の分布

注) 本図では、推定された教員付加価値の分布を示している。教員付加価値のヒストグラムを棒グラフで示しており、推定された教員付加価値をカーネル密度推定することによって得られた密度関数を線グラフで示している。

### 3.4.2 教員付加価値の分布

推定された教員付加価値  $\mu_j$  について、その分布を図 3.1 に示した。図においては教員付加価値  $\mu_j$  のヒストグラムを棒グラフで示し、カーネル密度推定によって求められた密度関数を線グラフによって示した。これらの図からは、実際に推定された教員付加価値の分布を視覚的に確認することができる。どのような教育成果を対象にしても、教員付加価値は正規分布のような形の分布関数を描いており、極値が複数あるような複雑な形をしていない。

次に、教員の影響の大きさを評価するために、表 3.6 に教員付加価値の推

定値の標準偏差を示した<sup>16</sup>。(1)列には推定した教員付加価値の標準偏差を教育成果ごとに記している。比較のために以降の列では、他の手法で求めた教員付加価値や元々の記述統計を同時に示している。(2)列には固定効果法によって求めた教員付加価値の標準偏差を示した。(3)列には教員付加価値が推定された人数を記した。(4)列には教育成果の元々の標準偏差を示した。3.3節で示している通り各教育成果に対しては標準化を行っているため、ここでの標準偏差はほとんど1に近い。(5)列には教育成果のクラス平均の標準偏差を示した。

	(1) 教員付加価値 (標準偏差)	(2) 教員付加価値 (標準偏差)	(3) 教員の数	(4) $Std(y_{it})$	(5) $Std(\bar{y}_c)$
	KS	FE			
国語	0.003	0.140	170	0.996	0.330
算数	0.013	0.141	170	0.994	0.343
学習方略	0.008	0.194	140	0.999	0.305
セルフコントロール	0.051	0.165	54	1.000	0.292
自己効力感	0.069	0.159	54	0.997	0.242
子ども・教員間の関係性	0.078	0.323	170	0.999	0.397

表 3.6: 教員付加価値の標準偏差

注) 本表では求めた教員付加価値の標準偏差を示している。各々の行は教員付加価値を求める対象とした教育成果を表す。(1)列では本稿で主に用いている教員付加価値の標準偏差を示した。(2)列では、比較のために教員固定効果推定で求めた教員付加価値(教員固定効果)の標準偏差の値を示した。(3)列には教員付加価値が推定された人数を記した。(4)列では子どもの教育成果の標準偏差を示した。(5)列ではクラス平均の標準偏差を示した。

結果として、認知能力や学習方略に対する教員付加価値の標準偏差が取りうる値の範囲は広いことがわかった<sup>17</sup>。(1)列において認知能力(国語、算数)を対象にした教員付加価値の標準偏差はそれぞれ0.003・0.013であり、学習方略を対象にした教員付加価値の標準偏差は0.008である。これは全体の変動((4)列)やクラス平均の変動((5)列)のうちそれぞれ高々1%・3%程度しか推定された教員付加価値は説明していないことを表す。一方で(2)で示した固定効果によって推定された教員付加価値の標準偏差は、認知能力(国語、算数)を対象にした場合はそれぞれ0.140・0.141、学習方略を対象にした場合教員付加価値の標準偏差は0.194である。これは、全体の変動のうち15%から20%を教員付加価値が説明していることを表す。式3.2で示したように真の教員付加価値の標準偏差はKane and Staiger(2008)による方法で求められた標準偏差と教員固定効果による方法で求められた標準偏差の間にある

<sup>16</sup>またこれらの値が頑健であることを示すために、節3.4.3では推定の対象とする教員を変えた場合における推定値を示した。

<sup>17</sup>この「教員付加価値の標準偏差」の大きさは、教員の影響力そのものを意味しないことには注意する必要がある。ここで議論しているのは個々の教員の教員付加価値の違いであり、すなわち教員の違いの影響力である。

と考えられる。そのため、全体の変動に対して認知能力や学習方略に対する教員付加価値が説明するのは1%から20%の間のいずれかであるといえる。

この「少なくとも1%」という教員付加価値の大きさは、他の研究と比較した時に必ずしも大きいわけではない。Chetty et al. (2014a) では小学校において英語では11%、数学では15%ほどの教員付加価値の大きさを報告している。Kraft (2019) では認知能力に関する各種アウトカムにおいて13%から18%の教員付加価値の標準偏差を報告している。しかし、本研究で得られた分析では「認知能力や学習方略に対する教員付加価値は1%から20%の間にある」というところまでしか言えず、これらの海外における研究よりも教員の役割が小さい可能性もあれば大きい可能性もある。無論もしも教員付加価値の標準偏差の真の値が上限値の20%程度あれば海外の研究と比較しても同水準以上であるが、一方で教員付加価値の標準偏差の真の値が下限値の1%であれば1/10程度と非常に低い水準になる。結果として個々の教員の違いが子どもの学力に与える影響力の大小を評価することは難しいと言える。

Kane and Staiger (2008) の方法で導出した教員付加価値については図3.1の下段左図から視覚的に確認することができる。図3.1の下段左図では、算数への教員付加価値のヒストグラム（オレンジ色）と子どもの算数の学力のヒストグラム（青色）を同時に描いている。しかし、オレンジ色で描かれた算数への教員付加価値の分布は非常に小さく、視認をすることは難しい。この図からも、Kane and Staiger (2008) の方法で推定された場合認知能力に与える個々の教員の影響は小さいことを見て取ることができる。

次に、非認知能力や学級での子ども・教員間の関係性に対する個々の教員の影響力は全体の変動に対して少なくとも5%から8%以上あった。すなわち、(1)列において示されたKane and Staiger (2008) の方法によって求められた教員付加価値の標準偏差は非認知能力（セルフコントロール・自己効力感）を対象にした時にはそれぞれ0.051・0.069であり、学級での子ども・教員間の関係性を対象にした時には0.078である。これは元々の教育成果の標準偏差が1であることを踏まえると、個々の教員の付加価値の違いは全体の変動に対して5%から8%ほどを説明したことを意味する。そして式3.2からは、これらの値は真の値に対する下限として捉えることができる。

この「少なくとも5%から8%」という教員付加価値の大きさは、他の研究と比較した時にも同水準か少し小さい程度である。Jackson (2018) では中退率で測定した教員付加価値の標準偏差は10%、欠席率で測定した教員付加価値の標準偏差は3%であったとしている。Kraft (2019) では教員付加価値は成長志向やGRITといった非認知能力の変動を15%程度説明するとしている。これらの研究と比較すると今回得られた非認知についての教員付加価値の大きさは少し小さい程度である。無論この値は下限値であるため、場合によっては海外の研究と同水準の大きさである可能性もある。結果として、認知能力に対する教員付加価値と異なり、非認知能力や子ども・教員間の関係性に

対しては教員による一定の大きさの影響力を捉えることが出来たと言える。

このような関係性については図 3.1 の下段右図から視覚的に確認することができる。図 3.1 の下段右図では、学級での子ども・教員間の関係性への教員付加価値（オレンジ色）及び元々の子どもによる回答（青色）を同時にヒストグラムによって描いている。そしてこの図表では、図 3.1 の下段左図では算数への教員付加価値とは異なり、オレンジ色で描かれた教員付加価値の分布を視認することができる。

総じてこれらの結果は、小学校において個々の教員は非認知能力や子ども・教員間の関係性において重要な役割を担っていることを示している。子どもを担当する教員が変わることは、子どもとのコミュニケーションのあり方の違いなどを通じて、その子どもの非認知能力や教員との関係性を大きく変えうる。この非認知能力の変化は、労働条件などの長期的に成果に結実することを多くの研究が示している。教員付加価値研究では、Jackson (2018) は非認知についての高い教員付加価値は高校卒業や大学入学といった長期的なアウトカムにおいても正に有意な影響を持っていることを示している。また Hanushek (2011) は質の高い教員の雇用がもたらす社会的価値を試算し、平均よりも 1 標準偏差優秀な教員は 20 人の学級であれば現在価値で 40 万ドルの社会的価値を生むとしている。これらの研究はアメリカを対象に行われており日本では同様の研究がないことから、これらの議論をそのまま適用することが可能かどうかについてはさらなる議論が必要ではあるものの、教員の違いは非認知能力の改善を通じて子どもの長期的な達成や社会的便益において重要な役割をもちうると言えるだろう。一方で認知能力に対しては、推定結果が指し示す教員による違いの大きさの推定値の範囲は広く、外国の研究との比較などからその大小について議論することは難しかった。そのためその長期的影響においても、国による違いの影響などを除いてもなお、認知能力についての教員付加価値の長期的影響を研究した Chetty et al. (2014b) や上述の Hanushek (2011) が示すような高い効果がある可能性もあれば、それらの研究ほどの効果が無い可能性もある。

### 3.4.3 教員付加価値推定の頑健性

節 3.4.2 で得られた教員付加価値は頑健な推定値なのだろうか。ここでは教員付加価値を分析する教員の集団を変えながら推定値の頑健性を検討する。ここでは (1) 複数クラスを受け持った教員のみデータを絞った場合、および (2) 教員歴が 3 年以上ある教員にのみデータを絞った場合のそれぞれについて教員付加価値の推定値の標準偏差がどのような値をとるかを示す。前者は教員付加価値の推定においてクラス固有の効果の影響を一切軽減することができていない教員（節 3.4.3.4.1）を除いた場合を検討している。後者は、教員の質が教員歴と共に変わっていく可能性を考慮している。教員の経験年数

### 3章 教員付加価値から見た教員の役割について – 日本の小学生を例にして –

が子どもの能力に与える影響を分析する研究では、Harris and Sass (2011) の様に最初の数年の経験が教員の質にとって重要であるという知見が存在する。もしそうであれば、教員として勤務を始めて数年の教員の質は低くなりその結果、推定された教員付加価値の標準偏差が高くなってしまふ。そのような可能性を検討するために、教員歴が短い教員をデータから除いた上で推定した結果を報告する。

得られた推定値の標準偏差を表 3.7 に示した。(1) 列は本文と示した値と全く同一である。(2) 列では1クラスしか担当していない教員を除いて推定された教員付加価値の標準偏差を示した。(3) 列では教員歴が2年以下の教員を除いて推定された教員付加価値の標準偏差を示した。各々の行は教員付加価値を求める対象とした教育成果を表す。

結果として、推定された教員付加価値の標準偏差は推定の対象とする教員によらずほとんど同じような値をとった。例えば国語については考えると、本文で報告した標準偏差は0.003である((1)列)のに対して、複数学級を担当した教員のみで考えた場合の標準偏差は0.005であり((2)列)、経験年数3年以上の教員のみで考えた場合の標準偏差は0.003である((3)列)。これらの値の違いは本文における議論に修正を迫るような差ではない。この傾向は他のアウトカムにおいても同様である。

	(1) 教員付加価値 (標準偏差) KS	(2) 教員付加価値 (標準偏差) KS 複数クラス	(3) 教員付加価値 (標準偏差) KS 経験年数3年以上
国語	0.003	0.005	0.003
算数	0.013	0.017	0.012
学習方略	0.008	0.007	0.007
セルフコントロール	0.051	0.057	0.055
自己効力感	0.069	0.068	0.067
子ども・教員間の関係性	0.078	0.097	0.086

表 3.7: 教員付加価値推定の頑健性

注) 本表では求めた教員付加価値の標準偏差を示している。各々の行は教員付加価値を求める対象とした教育成果を表す。(1) 列では本稿で主に用いている教員付加価値の標準偏差を示した。(2) 列では1クラスしか担当していない教員を除いて推定された教員付加価値の標準偏差を示した。(3) 列では教員歴が2年以下の教員を除いて推定された教員付加価値の標準偏差を示した。

### 3.4.4 教員付加価値間の相関

次に教員付加価値間の関係性について考えるため、表 3.8 に教員付加価値の項目間相関を示した。表の行と列は教育成果の種類を表し、各々のセルは対応する教員付加価値間の相関を表す。例えば、[1] 行 (1) 列には、国語の教員付加価値と算数の教員付加価値の間の相関が示されている。

		(1)	(2)	(3)	(4)	(5)
		算数	学習方略	セルフコントロール	自己効力感	子ども・教員間の関係性
[1]	国語	0.488*** (0.067)	-0.017 (0.085)	0.213 (0.136)	0.087 (0.138)	0.093 (0.077)
[2]	算数		-0.143* (0.084)	0.101 (0.138)	0.211 (0.136)	0.237*** (0.075)
[3]	学習方略			0.345** (0.130)	0.510*** (0.119)	0.248*** (0.082)
[4]	セルフコントロール				0.254 (0.306)	0.363*** (0.129)
[5]	自己効力感					0.344** (0.130)

表 3.8: 教員付加価値の項目間相関係数

注) 本表では教員付加価値の項目間の相関を示している。各々のセルは相関係数を表し、括弧の中は標準誤差を表す。各々の行及び列は教員付加価値を求める対象とした教育成果を表す。「\*\*\*」・「\*\*」・「\*」はそれぞれ 1%・5%・10%の水準で統計的に有意であることを表す。

結果として第 1 に、子どもの国語の学力を向上させた教員は、算数の学力をも向上させた傾向があった。具体的には、国語と算数の教員付加価値の間の相関 ((1) 列 [1] 行) は統計的に有意であり、相関係数も 0.488 と正である。すなわち、国語を教える際に高いパフォーマンスを発揮している教員は算数でも高いパフォーマンスを発揮している。高いパフォーマンスを発揮している教員はある特定の科目を教えることが得意というよりむしろ、どのような科目であっても知識を伝達することを得意としているのかもしれない。この結果は英語と国語への教員付加価値の間に 0.7 の高い相関があることを示した Goldhaber et al. (2013) とも整合的な結果である。

第 2 に、子どもの非認知能力を向上させた教員は、学習方略や学級での子ども・教員間の関係性をも向上させた傾向がある。すなわち、学習方略と非認知能力 (セルフコントロール・自己効力感) 及び学級での子ども・教員間の関係性の教員付加価値はその間に統計的に有意な相関があり<sup>18</sup>、相関係数も正である。この結果は、子どもとの友好的コミュニケーションを通じて学級での子ども・教員間の関係性がよくなり、結果として子どもたちの非認知能力や学習方略が向上したと解釈することができるだろう。

第 3 に、学力 (国語・算数) の教員付加価値と学習方略・非認知能力の教員付加価値の間の相関は小さく統計的な有意性もなかった。例えば、算数を対象とした教員付加価値とセルフコントロールを対象にした教員付加価値の間の相関係数は 0.101 ([2] 行 (3) 列) と小さく統計的に有意ではない。学力を伸

<sup>18</sup>セルフコントロールと自己効力感の教員付加価値の間には有意な相関はないが、これはサンプルサイズが極端に少ないからであると考えられる。

ばすという意味で「良い」教員と学習習慣や非認知能力を伸ばすという意味で「良い」教員は必ずしも同一ではない。この結果は、学力に対する教員付加価値と授業への出席率や停学などの生徒の行動に対する教員付加価値の間には弱い相関しかないとした Jackson (2018) や Kraft (2019) とも整合的である。

またこの学習方略の教員付加価値の性質からは、教員はどのようにして子どもの学力を伸ばすかについて示唆を得ることができる。学習方略をあげることが出来た教員とは、子どもの「学習のやり方」をより効率的にすることが出来た教員である。しかし、そのように学習を効率的にしたからといって、子どもの学力そのものを上昇させているわけではない。ここから分かるのは、子どもの学習のあり方を教員が変えると言うよりむしろ、授業の方法を工夫することなど授業時間における教育のあり方の違いが学力にたいする教員付加価値の違いにもつながっている可能性である。この授業における教授方法の違いが子どもの学力に強い影響を与えているのではないかという論点は Bietenbeck (2014) や Hidalgo-Cabrillana and Lopez-Mayan (2018) といった研究が近年議論を行っている。教員は学習のありかたを教えるのではなく、子どもがより理解できてより授業内容を習得できるような授業を提供することで、子どもの学力を上昇させていると表 3.8 は解釈することができる。

まとめると、これらの結果が示唆するのは、教員の効果は多次元的であるということである。確かに、国語及び算数といった学力を成長させることができる教員は比較的共通である。同様に非認知能力・学習方略を伸ばすことができる教員もまた比較的似通っている。しかし学力を伸ばすことができる教員と非認知能力・学習方略を伸ばすことができる教員は必ずしも同一ではない。すなわち、一概に「良い」教員というものが存在するわけではなく、人的資本の種類に応じて良い教員が存在するのである。そのため、例えば、学力についての教員付加価値を用いて教員を評価することは、非認知能力という重要な人的資本に資する能力を評価できない可能性がある<sup>19</sup>。

一方で、学習方略への教員付加価値についての解釈には議論の余地がある。第1に学習方略と学力は、教員が異なることによる子どもの能力の成長への影響力という点で類似している。すなわち学力と学習方略は共に、推定された教員付加価値の標準偏差がとりうる値が広く、場合によっては非常にその影響力が非常に小さい可能性がある(表 3.6)。第2に、学習方略を伸ばすことができた教員は非認知能力を伸ばすことができた教員と共通する傾向がある。すなわち、学習方略についての教員付加価値は学力の教員付加価値との相関は高く有意であった(表 3.8)。更にその一方で、学習方略についての教員付加価値と学力の教員付加価値との間にある相関は小さかった。このように教員付加価値のどのような性質に着目するかによって、学習方略は学力とも非認知能力とも近い性質を持っている。教員付加価値と学習方略の関係

<sup>19</sup>ただし両者の関係性はトレードオフというわけではないと言う点には注意が必要である。

については管見の限り本稿が初出であり、以上の関係性についての議論には更なる慎重な検討が必要である。

### 3.4.5 教員の属性と教員付加価値

それでは具体的にどのような教員が、「良い」教員なのだろうか。多くの研究が観察可能な教員の特徴が必ずしも子どもの教育成果とは結びつかないことを報告しており (Hanushek and Rivkin, 2006)、研究間で一貫して効果的だとされる具体的な要因は見つかっていない。しかし、それらの研究はほとんど学力を対象にしたものであり、ここまで考えてきた非認知能力や学級での子ども・教員間の関係性などの学力以外の教育成果と教員の観察可能な特徴についての研究はほとんど存在しない。更に、そもそも日本の教員を対象にした場合、3.2節でも述べた通りほとんど研究が存在しないため海外での研究での結論を当てはめることが可能なかどうかは分からない。

そのため、教員質問紙で尋ねている教員の特徴と推定した教員付加価値の間の関係について分析した。また、高い教員付加価値を持った教員の観察可能な特徴を調べた。具体的には下記のような定式化で分析を行う。

$$\hat{\mu}_j = \mathbf{X}_{j,2018}\beta + \epsilon_j$$

$\hat{\mu}_j$  は推定された教員付加価値である。ただしここではアウトカム間で係数推定値の比較を行うために、平均0分散1に標準化した値を用いた。 $\mathbf{X}_{j,2018}$  は観察可能な教員の特徴のベクトルを表し、本稿では教員の年齢、教員歴、性別、出身大学の難易度を考える。ただし出身大学の難易度を考えるとサンプルサイズが非常に小さくなるため、出身大学の難易度をサンプルに加えた場合と加えていない場合でそれぞれ検討した。また、年齢などの年によって変動する値については、全て2018年度時点での値を用いた。加えて、非線形性を考慮するために年齢や教員歴については、その二乗項を推定式に加えて推定を行った。推定された $\hat{\beta}$ が本節で関心を寄せる値である。

ただし上記で得られる値 $\hat{\beta}$ が因果的な効果を表すわけではないことに注意が必要である。すなわち、 $\mathbf{X}_{j,2018}$ と $\hat{\mu}_j$ の間には交絡が存在しうる。 $\hat{\mu}_j$ の導出手続を考えると、上記式においてコントロールされているのは教員 $j$ が担当した子どもの前年度のアウトカムや社会経済的地位を表す代理変数である。しかし、 $\hat{\mu}_j$ の推定と同様に、(子どもの前年度のアウトカムで条件付けできない)教員の内生的配置などが存在した場合には、それらのコントロール変数では十分に交絡を制御することはできない。そのため、上記式で得ることができるのはあくまで $\mathbf{X}_{j,2018}$ と $\hat{\mu}_j$ の間の関係性であることに注意が必要である。ただし、後述するように統計的に強く有意な関係性は見られないことから、その因果的な効果は存在したとしても非常に小さいと考えられる。

推定結果を表3.9に示した。以下ではその結果について詳述していく。

被説明変数 (教員付加価値)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	国語	国語	算数	算数	学習方略	学習方略	セルフ コントロール	セルフ コントロール	自己効力感	自己効力感	子ども・教員間の 関係性	子ども・教員間の 関係性
年齢	0.180 (0.708)	1.385 (1.643)	-0.102 (0.704)	0.414 (1.473)	-0.258 (0.748)	4.410** (2.089)	1.277 (1.451)	-15.127 (12.740)	1.186 (1.245)	-1.175 (3.763)	1.186 (1.245)	-1.175 (3.763)
31歳 - 22歳												
教員歴	0.216 (0.621)	-1.127 (1.615)	0.572 (0.617)	0.638 (1.448)	0.118 (0.647)	-3.019 (1.965)	0.428 (1.115)	19.139 (12.518)	0.856 (0.966)	1.728 (3.359)	0.856 (0.966)	1.728 (3.359)
10年 - 1年												
女性教員	0.170 (0.182)	0.192 (0.416)	0.302* (0.170)	0.297 (0.360)	-0.088 (0.195)	-0.922** (0.407)	-0.455 (0.357)	-1.898 (1.515)	0.492 (0.329)	0.730* (0.367)	0.492 (0.329)	0.730* (0.367)
出身大学偏差値		0.007 (0.027)		-0.018 (0.023)		-0.088*** (0.022)		0.100 (0.057)		0.035 (0.039)		0.035 (0.039)
サンプルサイズ	133	48	133	48	113	30	44	10	47	18	47	18

表 3.9: 教員の特徴と教員付加価値の関係

注) 本表では年齢、教員歴、性別及び出身大学偏差値が教員付加価値に与える影響を示している。各々のセルでは推定値及び標準誤差・有意水準を示しており、「推定値 有意水準 (標準誤差)」という表記法を用いている。「\*\*\*」、「\*\*」、「\*」はそれぞれ1%・5%・10%の水準で統計的に有意であることを表す。推定の際には、被説明変数である教員付加価値は平均0分散1になるように標準化を行った。また標準誤差はロバスト標準誤差を求めた。また年齢および教員歴については二次の多項式を用いた定式化を行なっていているため、本表ではデルタメソッドを用いてそれぞれ「32歳と22歳の(推定された)教員付加価値の差」・「32歳と22歳の(推定された)教員付加価値の差」

### 年齢、教員歴

どのような教員が良い教員か考えた時に頻繁に挙げられる点として、年齢や教員歴といった仕事の経験の多寡を挙げることができる。教員という仕事は高度な専門職でありその業務経験が長ければ長いほど教員としてのスキルは向上する可能性がある。Hanushek and Rivkin (2006) では教員の年齢が子どもの教育成果に与える影響を分析した研究をサーベイし、研究によってその推定値は様々であり結論を未だに出せるような状況ではないとしている。

結果として、教員の年齢と教員付加価値の間に有意な関係は発見されなかった。表 3.9 において 32 歳と 22 歳の教員付加価値の差の推定値は、どの教員付加価値を被説明変数としても統計的に有意ではない。ただし学習方略を被説明変数とし出身大学偏差値を説明変数に加えた推定 ((6) 列) では、32 歳と 22 歳の教員付加価値の差の推定は 4.410 で 5% の水準で有意であった。しかしこれらの結果は、出身大学偏差値を説明変数に加えていない推定 ((5) 列) の結果とは統計的な有意水準やその効果の正負が異なり頑健な結果とは言い難い。

同様に、教員歴と教員付加価値の間に有意な関係は発見されなかった。すなわち、表 3.9 において教員歴 10 年の教員と教員歴 1 年の許員の教員付加価値の差の推定値は統計的に有意ではなかった。むしろサンプルサイズが小さいことを踏まえれば、教員の年齢や教員歴と教員付加価値の間に全く関係性がないとするのは早急かもしれない、その点については今後の研究の課題とする必要がある。しかし、本表のような単純な相関においても有意な関係性を見ることはできなかったことから、やはり年齢や教員歴が子どもの能力に与える影響については慎重に検討する必要がある。

### 性別

教員の性別も良い教員の特徴を考える時には頻繁に論点に挙げられる。例えば、労働市場における男女差は教員の能力にも影響しているかもしれない。日本の労働市場の中では教員という職業は性別による賃金差が比較的小さい(妹尾渉他, 2003)。そのため、相対的に労働市場で能力の高い女性が教員になっている可能性がある。

表 3.9 の「女性教員」行には教員の性別と教員付加価値の間にある関係性を示している。結果として、教員が女性であるか否かと教員付加価値との間の関係のその推定値の分散は大きい。ただし学習方略を被説明変数とし出身大学偏差値を説明変数に加えた推定 ((6) 列) では、「女性教員」に係る係数の推定値は -0.922 で 5% の水準で有意であった。しかし、出身大学偏差値を説明変数に加えるか否かで結果が変わっており、その頑健性には注意が必要である。

### 出身大学の偏差値

高い学力・認知能力を持つ教員は高いパフォーマンスを発揮するのではないかと言う仮説は頻りに検証されてきた。例えば二木 (2017) など教員の採用時の労働市場の状況が子どもの教育成果に与える影響を調べる研究は、暗に能力の高い労働者ならば教員としても能力が高い可能性があるという仮説を検証していると言えよう。Hanushek and Rivkin (2006) ではそのサーベイを行い、最終的には研究によってその推定値は様々であり結論を未だに出せるような状況ではないとしている。

本稿で用いるデータでは、2016 年度に T 市に所属した教員に対する質問紙調査の中で卒業大学及び卒業学科について尋ねている。それらの卒業大学の情報を 2018 年度に河合塾が算出した大学受験偏差値に紐付けることで、教員の学力を卒業大学の偏差値という形で取得した。無論入試難易度偏差値は時間によって変わるため、このひも付けは必ずしも適切ではないことに十分注意する必要がある。

表 3.9 の「出身大学偏差値」行には教員の卒業大学偏差値と教員付加価値の間にある関係性を示している。結果として、卒業大学偏差値が教員付加価値に与える影響は概ね正である一方でその推定値の分散は大きく統計的に有意ではなかった。ただし学習方略を被説明変数とした推定 ((6) 列) では、その推定値は  $-0.088$  と負で統計的に 5% で有意である。教員の認知能力が高いと担当した子どもの学習方略が下がってしまうことを示す結果であるが、そのサンプルサイズは 30 と非常に少ないこともあり、更なる研究と慎重な議論が必要である。

## 3.5 結論

本稿では子どもの教育成果に対する個々の教員の効果について考えた。教員効果を教育成果に対する付加価値として計測した上で、得られた推定値の性質を記述的に調べた。

まず教員が子どもの教育成果に与えている影響の大きさは、教育成果の種類に応じて異なることが分かった。学力や学習習慣に対する教員の影響力は小さい可能性もあれば海外の研究と同水準に大きい可能性もあった。一方で、非認知能力や学級での子ども・教員間の関係性に対する教員の影響は比較的大きかった。そのため、「教員は子どもに大きな影響を与える/与えない」といった言説について評価する時には、子どもの人的資本のどの部分を対象にしているかを考える必要がある。子どもの学力の向上を期待して能力の高い教員を雇用することは必ずしも効果的な施策ではない可能性があるが(一方で効果的である可能性もある)、非認知能力や学級への雰囲気への効果を期待して能力の高い教員を雇用することは高い効果を持ちうる。

次に教員付加価値間の相関を調べた結果、学力を伸ばすことが得意な教員と学習方略や非認知能力を伸ばすことが得意な教員は異なるのではないかという示唆を得た。そして、この結果を踏まえた時に、教員付加価値を用いて教員の評価をすることについては慎重に検討するべきであると考えられる。すなわち、教員を評価する時一概に「良い」か「悪い」の2つに分けることは難しく、学力において良いパフォーマンスを発揮した先生であっても、非認知能力では必ずしもそのようではない。

本稿の最大の限界はサンプルサイズの問題である。特に教員の数は少なく、解析の結果がどれほど頑健であるかという点に課題が残る。ただし、本稿では市町村に在籍する教員全ての情報を使っているため、もとより教員のパフォーマンス測定の間で教員固定効果を考える時にはより大きい行政区分単位で考えなければならない可能性が高い。

また本稿で得られた教員の付加価値は、教員が子どもの教育成果に対して持つ因果的な効果とも異なることに注意が必要である。Chetty et al. (2014a)らの議論のように、過去のデータから得られた教員付加価値が子どもの将来の教育成果に対して因果的な効果を持つことを示した研究は存在する。そして、本稿ではそれらの議論を踏まえて推定された教員付加価値をそのまま分析に用いている。しかし、日本においても教員付加価値が因果的な効果を持ちうるかどうかは必ずしも自明ではなく今後の検証を必要とする。

子どもの教育成果に対する教員の効果についての研究は日本のみならず全世界で未だ決まった結果の方向性が示されていない研究分野である。本稿で行った議論も多く存在する研究課題の一部を扱ったものであり、決して絶対的な知見であるとは言えないが、それでも本研究は日本における数少ない教員付加価値の研究として重要であると言える。

## Appendix 補論

### A.1. 式 3.1 の定式化の妥当性について

式 3.1 のような Kane and Staiger (2008) で用いられている定式化はアメリカにおける子どもの能力成長を描写するものであった。しかし、本稿では日本の子どもの対象にしているのであって、式 3.1 の定式化を用いることの妥当性は十分に検討される必要がある。それは Chetty et al. (2014a) などが検討したような教員付加価値がどのような構造をもっているかという点のみならず<sup>20</sup>、子どもの教育生産関数 (Hanushek, 2008) の定式化として式 3.1 が妥当なのかという問題である。式 3.1 では教育生産関数の変数として前年度の学力や非認知能力などのアウトカム、SES などといった子どもの特徴を表す

<sup>20</sup>この点については本稿でも多少言及・検討している。

値を考えていた。しかし、異なる定式化の方が子どもの発達の描写としてより適切である可能性がある。

そこで、式 3.1 とは異なる定式化の可能性をここでは検討する。例えば式 3.1 では AR(1) の形で定式化を行っているが、子どもの学力  $y_{it}$  を説明する上では 1 期前だけでなく 2 期前の子どもの学力も重要な要因になるかもしれない。そのような可能性を検証するために、ここでは

$$y_{it} = \alpha_1 y_{it-1} + \alpha_2 y_{it-2} + X_{it}\beta + \mu_j + \theta_c + \epsilon_{it} \quad (\text{A.1})$$

という定式化を検討する。この式 A.1 を推定した結果を表 10 に示した。

	国語	算数	学習方略	自己効力感	教員と子どもの関係性
$\hat{\alpha}_1$	0.470*** (0.033)	0.502*** (0.027)	0.650*** (0.036)	0.565*** (0.032)	0.331*** (0.033)
$\hat{\alpha}_2$	0.369*** (0.035)	0.370*** (0.029)	0.111*** (0.036)	0.191*** (0.033)	0.202*** (0.034)

表 10: 異なる定式化による教員付加価値推定の検討

注) 本表では各々のアウトカムごとに式 A.1 によって推定した結果を示している。各々のセルは推定値を表し、括弧の中は標準誤差を表す。「\*\*\*」・「\*\*」・「\*」はそれぞれ 1%・5%・10% の水準で統計的に有意であることを表す。

$y_{it-2}$  に係る係数  $\alpha_2$  の推定値はいずれも統計的に有意であった。すなわち、2 期前のアウトカムは現在の子どもの能力形成に影響を及ぼしているということである。この結果は様々に解釈することができる。ただ 1 期前の  $y_{it-1}$  の測定誤差などの精度を高める形で  $y_{it-2}$  は拾っている可能性もあれば、もしくは子どもの能力は長期的に形成されることを示唆しているのかもしれない。いずれにせよ、式 3.1 と比べた時に式 A.1 による定式化の方がより子どもの発達を適切に描写していると考えられ、教員付加価値を推定する際にもこの点を十分に検討する必要がある。

しかしそれにも関わらず、本稿では式 3.1 を用いて推定を行う。たしかに式 A.1 による定式化の方が子どもの能力形成をより適切に描写していると考えられる<sup>21</sup>。しかし、A.1 を教員付加価値の推定に用いると、ある年の教員付加価値を分析するためには過去 2 年分の情報が必要になってしまう。本稿で用いるデータでは 3 年分しか値が存在せず、そのため A.1 を用いてしまうとそのうち 2 年分しか推定に用いることはできない。その時、3.4.1 で記したように教員付加価値とクラス固有の効果との識別の問題が大きくなり、分析結果のうちクラス固有の効果もたらす影響が大きくなってしまふ。このような識別上の問題があるために本稿では式 3.1 を用いるものの、今後の研究においてより正確な教員付加価値の推定という文脈においては AR(2) の形やさらに他の定式化の可能性を十分に検討する必要がある。

<sup>21</sup>ただし、このような異なる定式化によって本稿で得られた結論が大きく変わるものではなかった。

## 関連図書

- Ammermüller, Andreas, Hans Heijke, and Ludger Wößmann (2005) “Schooling quality in Eastern Europe: Educational production during transition”, *Economics of Education Review*, **24** (5), pp. 579–599.
- Baumeister, Roy F, Kathleen D Vohs, and Dianne M Tice (2007) “The strength model of self-control”, *Current Directions in Psychological Science*, **16** (6), pp. 351–355.
- Behncke, Stefanie (2012) “How do shocks to non-cognitive skills affect test scores?”, *Annals of Economics and Statistics*, pp. 155–173.
- Berkowitz, Ruth, Hadass Moore, Ron Avi Astor, and Rami Benbenishty (2017) “A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement”, *Review of Educational Research*, **87** (2), pp. 425–469.
- Bietenbeck, Jan (2014) “Teaching practices and cognitive skills”, *Labour Economics*, **30**, pp. 143–153.
- Chetty, Raj, John N Friedman, and Jonah E Rockoff (2014a) “Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates”, *American Economic Review*, **104** (9), pp. 2593–2632.
- , ———, and ——— (2014b) “Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood”, *American Economic Review*, **104** (9), pp. 2633–79.
- Clotfelter, Charles T, Helen F Ladd, and Jacob L Vigdor (2010) “Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects”, *Journal of Human Resources*, **45** (3), pp. 655–681.
- Goldhaber, Dan, James Cowan, and Joe Walch (2013) “Is a good elementary teacher always good? Assessing teacher performance estimates across subjects”, *Economics of Education Review*, **36**, pp. 216–228.
- Hanushek, Eric A (1986) “The economics of schooling: Production and efficiency in public schools”, *Journal of Economic Literature*, **24** (3), pp. 1141–1177.
- (2006) “School resources”, *Handbook of the Economics of Education*, **2**, pp. 865–908.

- (2008) “Education production functions”, *The New Palgrave Dictionary of Economics: Volume 1–8*, pp. 1645–1648.
- (2011) “The economic value of higher teacher quality”, *Economics of Education Review*, **30** (3), pp. 466–479.
- and Steven G Rivkin (2006) “Teacher quality”, *Handbook of the Economics of Education*, **2**, pp. 1051–1078.
- Harris, Douglas N and Tim R Sass (2011) “Teacher training, teacher quality and student achievement”, *Journal of Public Economics*, **95** (7-8), pp. 798–812.
- Heckman, James J and Tim Kautz (2012) “Hard evidence on soft skills”, *Labour Economics*, **19** (4), pp. 451–464.
- Hidalgo-Cabrillana, Ana and Cristina Lopez-Mayan (2018) “Teaching styles and achievement: Student and teacher perspectives”, *Economics of Education Review*, **67**, pp. 184–206.
- Hojo, Masakazu (2012) “Determinants of academic performance in Japan: an economic perspective”, *Japanese Economy*, **39** (3), pp. 3–29.
- and Takashi Oshio (2012) “What Factors Determine Student Performance in East Asia? New Evidence from the 2007 Trends in International Mathematics and Science Study”, *Asian Economic Journal*, **26** (4), pp. 333–357.
- Hopson, Laura M and Eunju Lee (2011) “Mitigating the effect of family poverty on academic and behavioral outcomes: The role of school climate in middle and high school”, *Children and Youth Services Review*, **33** (11), pp. 2221–2229.
- Jackson, C Kirabo (2018) “What do test scores miss? The importance of teacher effects on non-test score outcomes”, *Journal of Political Economy*, **126** (5), pp. 2072–2107.
- Jackson, C. Kirabo and Elias Bruegmann (2009) “Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers”, *American Economic Journal: Applied Economics*, **1** (4), pp. 85–108, October.
- Jackson, C Kirabo, Jonah E Rockoff, and Douglas O Staiger (2014) “Teacher effects and teacher-related policies”, *Annual Review of Economics*, **6** (1), pp. 801–825.

- Jacob, Brian A and Lars Lefgren (2005) “Principals as agents: Subjective performance measurement in education”, National Bureau of Economic Research.
- Kane, Thomas J and Douglas O Staiger (2008) “Estimating teacher impacts on student achievement: An experimental evaluation”, National Bureau of Economic Research.
- Kinsler, Josh (2012) “Beyond levels and growth estimating teacher value-added and its persistence”, *Journal of Human Resources*, **47** (3), pp. 722–753.
- Kraft, Matthew A (2019) “Teacher effects on complex cognitive skills and social-emotional competencies”, *Journal of Human Resources*, **54** (1), pp. 1–36.
- Lefgren, Lars and David Sims (2012) “Using subject test scores efficiently to predict teacher value-added”, *Educational Evaluation and Policy Analysis*, **34** (1), pp. 109–121.
- Multon, Karen D., Steven D. Brown, and Robert W. Lent (1991) “Relation of Self-Efficacy Beliefs to Academic Outcomes: A Meta-Analytic Investigation”, *Journal of Counseling Psychology*.
- Pintrich, P. D. Smith, Teresa Duncan, and Wilbert Mckeachie (1991) “A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)”, National Center for Research to Improve Postsecondary Teaching and Learning.
- Pintrich, Paul R. and Elisabeth V. de Groot (1990) “Motivational and self-regulated learning components of classroom academic performance.”, *Journal of Educational Psychology*.
- Richardson, Michelle, Charles Abraham, and Rod Bond (2012) “Psychological correlates of university students’ academic performance: a systematic review and meta-analysis.”, *Psychological Bulletin*, **138** (2), p. 353.
- Rothstein, Jesse (2010) “Teacher quality in educational production: Tracking, decay, and student achievement”, *The Quarterly Journal of Economics*, **125** (1), pp. 175–214.
- Todd, Petra E and Kenneth I Wolpin (2003) “On the specification and estimation of the production function for cognitive achievement”, *The Economic Journal*, **113** (485), pp. F3–F33.

3章 教員付加価値から見た教員の役割について - 日本の小学生を例にして -

---

- Tsukayama, Eli, Angela Lee Duckworth, and Betty Kim (2013) “Domain-specific impulsivity in school-age children”, *Developmental Science*, **16** (6), pp. 879–893.
- Wang, Ming-Te and Jessica L Degol (2016) “School climate: A review of the construct, measurement, and impact on student outcomes”, *Educational Psychology Review*, **28** (2), pp. 315–352.
- 佐藤純 (1998) 「学習方略の使用と達成目標及び原因帰属との関係」, 『筑波大学心理学研究』, 第 20 巻, 115–124 頁.
- 成田 健一・下仲 順子・中里 克治・河合 千恵子・佐藤 眞一・長田由紀子 (1995) 「特性的自己効力感尺度の検討:生涯発達の利用の可能性を探る」, 『教育心理学研究』, 第 43 巻, 第 3 号, 306–314 頁.
- 二木美苗 (2017) 「子どもの学習に対する教員の質の効果: 都道府県パネルデータによる実証分析」, 『日本経済研究』, 第 74 号, 56–83 頁.
- 文部科学省「平成 28 年度学校教員統計調査」, [https://www.mext.go.jp/b\\_menu/toukei/chousa01/kyouin/kekka/k\\_detail/1395309.htm](https://www.mext.go.jp/b_menu/toukei/chousa01/kyouin/kekka/k_detail/1395309.htm), Accessed: 2021-07-25.
- 妹尾涉・松繁寿和・梅崎修 (2003) 「公務員および教員の男女間賃金格差-大卒者アンケート調査から」, 『大阪大学経済学』, 第 53 巻, 第 2 号, 96-108 頁, 9 月.
- 李嬋娟 (2014) 「非認知能力が労働市場の成果に与える影響について」, 『日本労働研究雑誌』, 第 650 巻, 30–43 頁.

## 第4章 子どもの教育成果に対する 指導方法の効果の検証 -時 間変化する要因を考慮した 分析-

### 概要<sup>1</sup>

どのような学校や教員が子どもの教育成果に対して良い効果を持つのか、その具体的な特徴については研究間で統一した見解がほとんど得られていない。そのような状況の中で、学校や教員によって用いられる指導方法に近年の研究では注目が集まっている。本研究では日本のある自治体における子どもの個票データを用いて、指導方法が子どもの教育成果に与える影響について推定を行う。しかしどのような指導が授業で行われるかは必ずしも外生的に決定されるわけではなく、学校が置かれた状況によって内生的に決定される。本稿では時間変化しない観察不可能な生徒間異質性および時間変化する観察不可能な生徒間異質性をコントロールをすることで、その内生性への対処を行った。結果として、子どもの自発性を促進する指導や授業手法を工夫すること・子どもの授業態度への指導は小学生の学力に対して正の効果を持った。

### 4.1 はじめに

どのような学校や教員が子どもの教育成果に対して良い効果を持つのだろうか。その具体的な特徴については研究間で統一した見解がほとんど得られていない。これまで多くの研究が教員や学校が子どもに対して強い影響を与えていることを指摘し (Hanushek, 1986, 2006; Ammermüller et al., 2005; Hojo and Oshio, 2012)<sup>2</sup>、影響力を規定する具体的な要因について分析を行ってきた。その分析の対象は、学級規模や学校の予算、教員の性別・学歴など広い範囲にわたる。しかし、その結果は研究によって大きく異なることが多

<sup>1</sup>本章は伊藤・田端 (forthcoming) として出版されたものを一部修正して掲載している。

<sup>2</sup>特に教員の重要性については、教員付加価値を推定する一連の研究が、教員が子どもの発達に強い影響を及ぼしていることを示してきた (Kane and Staiger, 2008; Lefgren and Sims, 2012)。

#### 4章 子どもの教育成果に対する指導方法の効果の検証 -時間変化する要因を考慮した分析-

く (Hanushek and Rivkin, 2006)、研究間で一貫して効果的だとされる要因は見つかっていない<sup>3</sup>。

そのような研究状況の中で、学校や教員による授業における指導方法に近年関心が集まっている (Schwerdt and Wuppermann, 2011; Van Klaveren, 2011; Aslam and Kingdon, 2011; Bietenbeck, 2014; Lavy, 2015; Algan et al., 2013; Comi et al., 2017; Hidalgo-Cabrillana and Lopez-Mayan, 2018; Bessho et al., 2019)。高い学力を持つ子どもを多く有する学校では授業で発展的な話題を中心に扱う指導を行う一方で、在籍する子どもの平均的な学力が低いため、基本的な話題を中心に学習から取り残される子どもを少なくすることを試みる指導を行う学校もあるだろう。そして、これらの指導方法の違いは学校や教員を特徴づける要素であり、子どもに与える影響力の重要な要因になりうる。このような関心から、Aslam and Kingdon (2011) や Bietenbeck (2014) などの研究が、指導方法が子どもの教育成果に与える影響力について分析を行ってきた。そしてそのような研究の多くが、何らかの指導方法が子どもの学力に対して正の効果を持ったことを報告している<sup>4</sup>。

しかし、指導方法の効果を推定することには困難が多い。何故ならば、実施される指導は必ずしも外生的に定まるとは限らないからである。どのような指導を行うかという問題は自治体や学校もしくは教員による意思決定を反映しており、内生性の問題を抱えやすい。例えば、学級に所属する子どもの学力が低ければ、授業において発展的な話題を扱うことを控えることで子どもたちの理解度を向上させようとする指導方法を採用する可能性がある。この場合、学級の学力が低いことは、授業における発展的な話題の取り扱い方の原因であって結果ではない。これまでの指導方法についての分析では、このような内生性が起きる原因として、①学校間のセレクションの問題や②学校内のセレクションの問題③教員の観察不可能な異質性の問題が大きく取り上げられてきた。これらの問題を乗り越えるために、RCT(Randomized Controlled Trial) の実施や擬似的な実験的環境を利用することが可能であれば良いが、研究にかかる費用や倫理的な側面もしくは擬似実験的な環境の発見の困難さなどからこれらの方法を用いることは必ずしも容易ではない。その際に頻繁に用いられるのは、子どもの固定効果をコントロールした上で指導方法の効果の推定することである。ただし、どのようにして固定効果を制御するか

<sup>3</sup>ただし、この指摘はそのまますべての教員の履歴書的な特徴が子どもに与える影響が存在しない、ということを示すわけではないことには十分な注意が必要である。ここで述べているのは、現在の研究水準においては頑健な関係性は見つかっていないということであり、今後の研究の進展によってはそのような研究状況が一変する可能性はある。また、教員の性別など子どもと教員のマッチングという意味で十分に研究が進展していない分野も存在する。

<sup>4</sup>Schwerdt and Wuppermann (2011) では授業において子どもに講義をしている時間が子どもの学力に影響を与えている報告されている一方で、Van Klaveren (2011) では教員が教室の前で講義をする時間が学力に影響を与えていなかったことが報告されている。また、Aslam and Kingdon (2011) では教員の属性ではなく、授業の準備や授業における発問といった授業プロセスが学力に影響を与えることを示している。さらに、指導方法の中でも効果がある指導方法と効果がない指導方法は異なりうる。Bietenbeck (2014) や Hidalgo-Cabrillana and Lopez-Mayan (2018) では指導方法を現代的な指導方法と伝統的な指導方法にわけ、その両方で効果が異なることを報告している。

については研究によっては特殊な方法が用いられる。一般的なパネルデータでは、個人の時間変化しない異質性をコントロールする。しかし、指導方法の効果进行分析をする際に頻繁に用いられるのは、複数の科目間での変動を考慮することである (Clotfelter et al., 2010; Schwerdt and Wuppermann, 2011; Van Klaveren, 2011; Aslam and Kingdon, 2011; Bietenbeck, 2014; Comi et al., 2017)。ある子どもの国語と算数の成績は当然異なるが、教員による指導も科目ごとにまた異なりうる。この科目間の変動を推定に利用することで、個人間の観察できない異質性をコントロールすることができる。この方法は複数年度のデータを必要としないこともあり、指導方法の効果进行分析するには広く用いられている。

しかし、推定の際にどの変動を用いるかは、どのような内生性の可能性を許容するかと密接に関わってくる。通常のパネルデータを用いた個人固定効果の制御による推定では、時間変動する内生要因は推定された値をバイアスさせる。一方で科目間の変動を用いた推定では、科目変動する内生要因が推定された値をバイアスさせる。詳しくは 4.3 節にて議論を行うが、このような要因によってこれまでの研究が報告してきた指導方法の効果は必ずしも頑健ではない可能性がある。

本稿の主たる問題意識は、固定効果推定では除くことのできないバイアスをコントロールした上で指導方法の効果进行分析することである。そのために、本稿では子どもの科目間異質性を明示的に推定式に加え、観察不能な時間変動しない生徒間異質性ならびに時間変動する生徒間異質性をコントロールした指導方法の効果推定を行った。そのため、指導方法の効果の推定が抱える内生性の問題に対して十分な対処を行うことができた。この点は過去の指導方法の効果の研究が抱えている課題を解決しているという点において、本稿の重要な貢献の一つとなっている。

結果として、一部の指導方法については子どもの教育成果に対する効果を見ることができた。すなわち、子どもの自発性を促進する指導や子どもの授業態度への指導および授業手法を工夫することは小学生に対してその効果は有意に正であった。一方で、学級編成を工夫することは統計的に有意ではなく、授業で扱う内容を工夫することはむしろ負の効果を持った。さらに、中学生に対しては指導方法と子どもの学力の間に強い関係性は観察されなかった。以上の結果を纏めれば、学校や教員によって様々に工夫され実践される指導の全てが子どもの学力向上に結びつくわけではないと言えよう。そのため、闇雲に様々な指導方法を試すのではなく、教育成果と結びつきうる指導方法において改善を志すことが重要であることが示唆された。

本稿の構成は以下の通りである。第 4.2 節ではデータについて述べる。第 4.3 節では推定戦略について議論をする。第 4.4 節では推定結果について述べる。第 4.5 節では本稿における結論を述べる。

## 4.2 データ

本稿では2017年－2018年<sup>5</sup>にかけて行われた日本のとある県Aにおいて実施された学力テスト(以降S県学力調査と記述)および質問紙調査のT市におけるデータを用いる。テストは毎年4月に行われ、テストと同時に学校に質問紙調査を行った。テストを受けた子どもにはIDが振られ、経年で同一生徒を特定することが可能な設計になっている。調査の対象になったのは、T市の全ての公立小学校及び公立中学校に所属する小学4年生から中学3年生までの子どもであった。ただし、中学1年生の子どもは学校回答と紐付けができなかったため、分析の対象から除外した。

表4.1にはデータで用いる変数についての記述統計を示した。以降ではデータで用いる変数について記述する。

### 4.2.1 学力

S県学力調査では国語と算数のテストを実施し、その結果から個人の科目ごとの学力の推定値が項目反応理論(Item Response Theory、以降IRTと表記)に従って算出されている。さらに、全ての学力の推定値を年度・学年及び科目ごとに平均0・分散1に標準化を行った。この操作によって、実質的に標準化された学力は学年内での相対的な高低を表すことになる。このIRTによって算出された学力推定値の標準化された値を、本稿では教育成果として分析の対象とする。

### 4.2.2 指導方法

S県学力調査では、子どもへの調査と同時に各々の学校に学年の状況や行なっている施策について尋ね回答を得ている。質問は全て、前年度用いた施策について尋ねている。子どもへの調査は上述の通り毎年4月に行われるため、そこで得られた学力はその年度の学校調査で回答された学校施策の効果を反映していると考えられる。ここで用いるのはその中でも、授業における指導方法についての質問項目である。質問は各学年の各科目(算数・数学、国語)ごとに尋ねており、そのため科目間で比較することが可能である。具体的な質問項目を表4.2に示した<sup>6</sup>。

<sup>5</sup>調査の実施は2015年から行われている。

<sup>6</sup>「授業ごとに学習の目標(めあて)を示した後に、児童自ら解決の方法を考えさせる発問をしていましたか」および「授業の課題解決の場面では、児童同士が互いの考えを比較検討しながら、他の児童の考えを取り入れて解決させるよう努めていましたか」の2つの質問に対しては「当てはまる」などの主観的な選択肢で回答された。どのような状況を指して「当てはまる」か否かは回答する教員によって異なると考えられる。4.3節で説明するように、本稿ではある個人の年度間変動もしくは科目間変動を利用しているため、そのような主観的な選択肢が小学校においてはバイアスを生むことは想定しにくい、中学校においては教員間の異質性を源泉としたバイアスを生じさせうる。そのため、中学校を対象にした分析ではこの点について十分注意をして解釈をする必要がある。

4章 子どもの教育成果に対する指導方法の効果の検証 -時間変化する要因を  
考慮した分析-

	2017年	2018年
子ども		
生徒数		
小学5年	1195	1272
小学6年	1165	1209
中学2年	1027	1050
中学3年	1024	1011
学力	0.007 (8820,0.997)	0.008 (9082,0.994)
学校		
学校数		
小学校	12	12
中学校	6	6
指導方法インデックス		
授業の編成	-0.100 (72,1.445)	-0.467 (72,0.990)
授業の内容の工夫	0.001 (72,1.185)	-0.335 (72,1.101)
生徒の授業態度への指導	0.233 (72,0.628)	0.246 (72,0.763)
授業手法の工夫	-0.094 (72,1.898)	0.109 (72,1.306)
生徒の自発性の促進	0.051 (72,1.462)	0.329 (72,1.274)

注) 本表では、本稿で用いている変数の記述統計を示している。表には、子どもや教員・学校のデータ数を表しているセルと変数の記述統計を表しているセルが存在する。記述統計を表すセルでは、対応する変数について「平均(サンプルサイズ、標準偏差)」という記述方法を用いている。

表 4.1: 記述統計

#### 4章 子どもの教育成果に対する指導方法の効果の検証 -時間変化する要因を考慮した分析-

インデックス名	項目	スケール	回答方法	記述統計	主成分得点
授業の編成	習熟の遅い児童に対して、学習内容が習得できるよう、少人数による指導を年間どのくらい計画的に取り入れられましたか	5	『年間の授業』	1.826 (144, 1.329)	0.706
授業の編成	習熟の早い児童に対して、少人数による指導を行い、発展的な内容を扱った指導を計画的に取り入れられましたか	5	『年間の授業』	1.729 (144, 1.154)	0.706
授業の編成	チーム・ティーチングによる指導を取り入れられましたか	5	『年間の授業』	2.451 (144, 1.581)	0.067
授業の内容の工夫	補充的な学習の指導を計画的に取り入れられましたか	5	『年間の授業』	2.708 (144, 1.121)	0.707
授業の内容の工夫	発展的な学習の指導を計画的に取り入れられましたか	5	『年間の授業』	2.458 (144, 0.953)	0.707
生徒の授業態度への指導	学習規律（私語をしない、話をしている人の方を向いて聞く、聞き手に向かって話をする、授業開始のチャイムを守るなど）を維持する指導を行っていましたか	5	『年間の授業』	4.757 (144, 0.661)	1.000
授業手法の工夫	学習方法（ノートの取り方、間違ったところを振り返って学習すること、目標・計画の立て方、自分をやる気にする方法など）に関する指導を行っていましたか	5	『年間の授業』	4.403 (144, 0.895)	0.415
授業手法の工夫	ワークシートや資料、模型などの具体物を使って、児童の様々な考えを引き出すための指導の工夫を行っていましたか	5	『年間の授業』	4.153 (144, 0.839)	0.448
授業手法の工夫	授業や単元（小単元）のまとめの場面では、学習した内容をどのように活用できるかを児童に書かせましたか	5	『年間の授業』	3.396 (144, 1.213)	0.423
授業手法の工夫	上記 No 6 2, 6 3 の場面で、ノートやプリントに書かせた児童の考えを教師が確認していましたか	5	『記述させた』	4.556 (144, 0.843)	0.389
授業手法の工夫	授業の中でドリルなどを使った練習問題を行っていましたか	5	『年間の授業』	3.722 (144, 1.137)	0.333
授業手法の工夫	思考を深める板書（授業の全体像と思考の流れが分かる、振り返りができるなど）を行っていましたか	5	『年間の授業』	4.340 (144, 0.795)	0.430
生徒の自発性の促進	児童に理由を付けて考えを発表させたり、書かせたりする指導を行っていましたか	5	『年間の授業』	4.319 (144, 0.799)	0.531
生徒の自発性の促進	授業のまとめの場面では、次の授業につながる新たな問いや疑問、次の授業で調べたいことを児童に書かせましたか	5	『年間の授業』	3.625 (144, 1.176)	0.464
生徒の自発性の促進	授業ごとに学習の目標（めあて）を示した後に、児童自ら解決の方法を考えさせる発問をしていましたか	5	『当てはまる』	4.368 (144, 0.736)	0.508
生徒の自発性の促進	授業の課題解決の場面では、児童同士が互いの考えを比較検討しながら、他の児童の考えを取り入れて解決させるよう努めていましたか	5	『当てはまる』	4.340 (144, 0.628)	0.495

表 4.2: 質問項目一覧

注) 本表では指導方法インデックスを構成する質問をまとめた。記述統計を表す列では、対応する変数について「平均 (サンプルサイズ、標準偏差)」という記述方法を用いている。

指導方法に関連する質問項目の数は非常に多く、そのまま分析に用いると議論が非常に煩雑になる恐れがある。そのため、ある程度似ていると判断される質問項目を主成分分析を行い1つのインデックスとしてまとめた。最終的には、「授業編成」「授業の内容の工夫」「生徒の授業態度への指導<sup>7</sup>」「授業手法の工夫」「生徒の自発性の促進」の5つのインデックスにまとめた。

インデックスを構成する質問項目、及びインデックスの組成に用いた主成分得点を表4.2に示した。このインデックスの作成の妥当性については慎重な議論が必要である。しかし、全ての質問項目における第1成分の符号は全て同一でありインデックス内での回答傾向は一貫している。そのため、本稿で用いるインデックスは学校の回答の傾向を表すことができていると考えられる。

### 4.3 推定戦略

#### 4.3.1 教育生産関数の分解

子どもの教育成果(学力)とその要因は、次の様な教育生産関数を用いて表現することができる。

$$y_{ikt} = \mu_i + \eta_{ik} + v_{it} + \xi_{ikt} \quad (4.1)$$

$y_{ikt}$  は  $t$  期における子ども  $i$  の科目  $k$  における学力を表す。 $y_{ikt}$  は時間や科目によって変わりうることを想定しており、そのため  $y_{ikt}$  は4つの要因( $\mu_i, \eta_{ik}, v_{it}, \xi_{ikt}$ )に分解することができる。それぞれについて詳記すれば下記の通りである。

- $\mu_i$  : 個人ごとに時間変化しない科目不変の要因 (time-fixed, subject-fixed)  
例) 生得的な能力や、親の社会経済的地位など
- $\eta_{ik}$  : 個人ごとに科目変化する要因 (time-fixed, subject-varying)  
例) 科目ごとの得意不得意など
- $v_{it}$  : 個人ごとに時系列変化する要因 (time-varying, subject-fixed)  
例) 勉強時間の増加や通塾の有無など
- $\xi_{ikt}$  : 個人ごとに時系列変化する更に科目変化する要因 (time-varying, subject varying)  
例) 特定科目での補習授業の受講など

---

<sup>7</sup>似たような項目がなかったため、対応する質問項目は1つである。

4章 子どもの教育成果に対する指導方法の効果の検証 -時間変化する要因を考慮した分析-

式 (4.1) は教育経済学において想定されてきた教育生産関数 (Hanushek, 2008) と基本的には相違はない。伝統的に教育生産関数では教育成果を規定する要因として生得的な能力や親の社会経済的地位、個人による学習時間投資、学校効果などを想定してきたが、それらは全て  $\mu_i$ 、 $\eta_{ik}$ 、 $v_{it}$ 、 $\xi_{ikt}$  のいずれかに分類することができる。また、科目要因を表す  $\eta_{ik} \cdot \xi_{ikt}$  は多くの研究において明示的には言及されてこなかったが、一方で多くの研究が科目別に推定を行ってきていることを踏まえれば暗黙の内にこれまでも想定されてきたと言える。

式 (4.1) を変量効果モデル及び固定効果モデルで推定<sup>8</sup>した結果を表 4.3 にまとめた。変量効果モデルで推定した結果については決定係数及び各要因の標準偏差を記し、固定効果モデルについては決定係数を記した<sup>9</sup>。また推定結果を視覚的に確認するために、図 4.1 では変量効果モデルで推定された  $\mu_i$ 、 $\eta_{ik}$ 、 $v_{it}$  の分布を図示した。図の横軸は推定された  $\mu_i$ 、 $\eta_{ik}$ 、 $v_{it}$  の値を表し、縦軸はその頻度を表している。

表 4.3: 教育生産関数の分解

	個人固定効果 ( $\sigma_\mu$ )	個人科目固定効果 ( $\sigma_\eta$ )	個人年度固定効果 ( $\sigma_v$ )	R2(adj)	Obs
固定効果モデル	✓	✓	✓	0.806	17902
ランダム効果モデル	0.841	0.361	0.049	0.783	17902

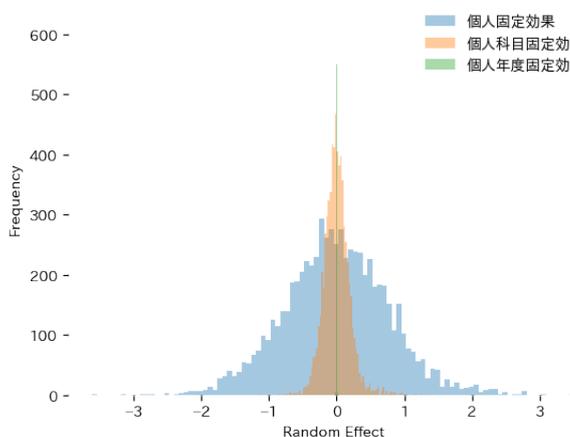
注) 本表では式 (4.1) に基づいて教育生産関数を推定した結果を示している。推定は式 (4.1) の各変数を固定効果として推定するモデル (固定効果モデル) もしくはランダム効果として推定するモデル (ランダム効果モデル) によって行なわれた。ランダム効果モデルについては、各ランダム効果の標準偏差の推定値を同時に示している。

子どもの教育成果はその約 80% を  $\mu_i$ 、 $\eta_{ik}$ 、 $v_{it}$  の 3 要因で説明することができる。すなわち、変量効果モデル及び固定効果モデルで推定した式 (4.1) の修正済み決定係数の値はそれぞれ 78.3%・80.6%であった。また、それぞれの要因の標準偏差はそれぞれ 84.1%・36.1%・4.9%(表 4.3) であり、それぞれ大なり小なり子どもの教育成果に対する一定の役割を果たしていた。これらの要因は様々な要因を内包しており、本稿で分析の対象とする指導方法や 4.1 節や 4.3.2 節で言及する内生要因も  $\mu_i$ 、 $\eta_{ik}$ 、 $v_{it}$  の中に含まれている。そのため、必ずしも十全に観察されるわけではない  $\mu_i$ 、 $\eta_{ik}$ 、 $v_{it}$  の 3 要因の影響をどのように考えるかが指導方法の分析においては重要な課題になる。次節の 4.3.2 節では、観察されない 3 要因の影響を如何に制御するかについて考察する。

<sup>8</sup>ただし  $\xi_{ikt}$  を除いた定式化で推定を行なった。 $\xi_{ikt}$  は被説明変数及び誤差項と線形従属であるために識別できない。

<sup>9</sup>ただし表 4.3 で報告される決定係数は、多くの研究で報告される固定効果を除いた変動に対する説明割合とは異なる。ここでは、固定効果が全体の変動をどれほど説明しているかを決定係数として示している。

図 4.1: 教育生産関数の分解



### 4.3.2 推定戦略

4.3.1 節における議論を踏まえ、指導方法の効果を推定するために本稿では次の様な推定式を考える。

$$y_{ikgt} = \beta tp_{kgt} + \mu_i + \eta_{ik} + v_{it} + \epsilon_{ikt} \quad (4.2)$$

$i$  は子ども個人、 $k$  は科目、 $g$  は子どもが在籍する学年、 $t$  は時点を表す。 $tp_{kgt}$  は  $t$  期での学年  $g$  における科目  $k$  の指導方法を表す。 $\epsilon_{ikt}$  は誤差項である。 $\mu_i$ ,  $\eta_{ik}$ ,  $v_{it}$  はそれぞれ生徒個人固定効果、生徒科目固定効果、生徒年度固定効果を表す<sup>10</sup>。本稿において関心を寄せている指導方法の効果は  $\beta$  で表される。

本稿の主要な貢献として、指導方法の効果  $\beta$  の推定値の内生要因の多くを制御して推定を行っている点を挙げることができる。 $\beta$  を一致推定量で求めるためには、指導方法  $tp_{kgt}$  が誤差項と無相関であることが必要になる。しかしその推定値をバイアスさせる要因として、これまで①子どもや教員が学校間で内生的に配置される可能性 (selection between schools)、②学校内で内生的に子どもや教員が配置される可能性 (selection within schools)、さらに③学校や教員による指導方法が内生的に選択される可能性などが挙げられてきた。これらの要因をコントロールすることを目的とするため、推定式には個人科目固定効果  $\eta_{ik}$  及び個人年度固定効果  $v_{it}$  が含まれている。このうち、4.3.1 節で記述したように個人年度固定効果  $v_{it}$  は個人ごとに年度によって変化する要因全てを観察可能・観察不能に関わらず含む。その中には小学校における上記②や③も含まれると考えられる。なぜならば、通常子どもの

<sup>10</sup> この式では教育を対象にした分析で多く用いられる付加価値モデルによる定式化と推定を行っていない。その理由は用いたデータは 2 期間分しか存在せず、付加価値モデルに直すと 1 期間の変動しか推定に用いることができなくなるためである。但し、データが 3 期間以上に渡って存在する場合においては、式 (4.1) における  $\epsilon_{ikt}$  の影響を少しでもコントロールすることを目的として付加価値モデルを用いることは有用な推定戦略になりうる。

配置は年ごとに行われ小学校においては教員の配置も年ごとに一定であると考えられるからである。そのため個人年度固定効果  $v_{it}$  をコントロールして推定を行うことで②及び③の問題を解決することが可能になる<sup>11</sup>。

また、上述の②及び③の問題を解決するためには過去の指導方法の効果の分析で用いられてきた手法は不十分であるという点は特に重要な論点である。過去の研究が用いたそのような手法の代表例として、子どもの科目間変動を推定に利用し個人固定効果  $\mu_i$  をコントロールする手法を挙げることができる (Clotfelter et al., 2010, など)。多くの場合、授業は科目ごとに行われ学力テストも科目ごとに行われる。そのため、指導方法や学力テストの点数は同一個人であっても科目ごとに異なりうる。その手法ではこの科目間変動を用いて推定を行う。例として式 (4.2) に基づいて考える。推定する対象の年を  $t'$  とし科目ごとの平均を引くことで (データの科目を表す集合を  $K$  と記す)、式 (4.2) は次のように変形される。

$$y_{ikt'} - \frac{1}{|K|} \sum_{k'' \in K} y_{ik''t'} = \beta \left[ tp_{kgt'} - \frac{1}{|K|} \sum_{k'' \in K} tp_{k''gt'} \right] + \left[ \eta_{ik} + \epsilon_{ikt'} - \frac{1}{|K|} \sum_{k'' \in K} (\eta_{ik''} + \epsilon_{ik''t'}) \right] \quad (4.3)$$

この操作を行うことで、結果として個人固定効果  $\mu_i$  及び個人年度効果  $v_{it'}$  の影響を取り除くことができる。そして、この手法はパネルデータ分析と異なり複数年度に渡ってテストをする必要がないという利点を持つ。単一年度のデータであっても、学力テストの値と分析を行う説明変数のとる値は科目によって変動するという状況があれば上述の定式化を行うことができる。そのため、指導方法の効果进行分析の研究の多くが科目間変動を用いて分析を行ってきた。

しかし、科目間変動を用いるこの手法は内生性の問題を解決する手段としては不十分であり、個人の科目間の異質性が内生要因である可能性を見逃しうる。すなわち、式 (4.3) において、(誤差項である  $\epsilon_{ikt}$  を除けば) 科目個人固定効果  $\eta_{ik}$  をコントロールすることはできていない。そのため  $\eta_{ik}$  が指導方法を表す  $tp_{kgt}$  と相関関係があれば、 $\beta$  の推定値はバイアスを含んだものになる<sup>12</sup>。そしてそのような状況は容易に想像することができる。例えば、算数を苦手とする子どもが多く所属するクラス A と算数を得意とする子どもが多く所属するクラス B があるとする。その場合教員はクラス A とクラス B で同じような指導を行うだろうか。むしろ、目の前の子どもの性質に合わせて、例えばクラス A では特別丁寧に問題解説を行うようにするなど、指導方法を変えるのではないだろうか。そのような状況があれば、科目個人固定効果  $\eta_{ik}$  をコントロールしないまま得られた指導方法の効果にはバイアスが含まれると言える。

<sup>11</sup>①の問題に対しては、その他の指導方法の効果进行分析の研究と同様に  $\mu_i$  をコントロールすることで対処している。

<sup>12</sup>これは科目間変動を用いる代わりに、年度間変動を用いた場合も同様の議論を展開することができる。その場合、観察不可能な時間変動要因  $v_{it}$  をコントロールすることはできなくなる。これはいわゆるパネルデータ分析における固定効果法とその特徴を書き直したに過ぎない。

これらの問題を乗り越えるためにも、本稿では個人固定効果  $\mu_i$  だけでなく個人科目固定効果  $\eta_{ik}$  および個人年度固定効果  $v_{it}$  を推定式に加えている。この様な定式化の下で推定を行うことで、通常の固定効果推定では取り除けない時間変動する異質性をもコントロールすることが可能になった<sup>13</sup>。

ただし、中学校においては個人科目固定効果  $\eta_{ik}$  及び個人年度固定効果  $v_{it}$  をコントロールしたとしてもなお内生要因は存在しうることには注意が必要である。中学校において、ある子どもを担当する教員は科目によって異なりその配置は年ごとに変わりうる。すなわち、中学校において担当する教員の効果は式 (4.1) においては  $\xi_{ikt}$  で表されるものである。そのため、個人科目固定効果  $\eta_{ik}$  や個人年度固定効果  $v_{it}$  をコントロールするだけでは、その影響を制御するためには必ずしも十分ではない。日本のデータを用いて中学校における指導方法の効果を式 (4.2) の下で分析する際には、その内生性の方向性などを注意深く検討する必要がある。

#### 4.4 分析結果

表 4.4 では式 (4.2) に基づき定式化を変えながら指導方法の効果  $\beta$  の推定値を示した。(1) 列では何もコントロールしていない定式化における推定値を示している。(2) 列では個人固定効果  $\mu_i$  のみをコントロールした定式化における推定値を示している。(3) 列では (2) に個人科目固定効果  $\eta_{ik}$  を加えた定式化における推定値を示している。(4) 列では (2) 列に個人年度固定効果  $v_{it}$  を加えてコントロールをした定式化における推定値を示した。(5) 列では個人科目固定効果  $\eta_{ik}$  及び個人年度固定効果  $v_{it}$  の両方をコントロールをした定式化における推定値を示している。この (5) 列では式 (4.2) に示した全ての固定効果を含めて推定を行っており、その推定値は最もバイアスの影響が小さいと考えられる。そのため以降では (5) 列で示した推定値を中心的に考察していく。小学校と中学校では教員配置のルールが異なることに鑑み、分析は学校種別ごとに行った<sup>14</sup>。

分析の結果、小学生の学力向上に効果を持った指導方法が存在した。具体的には、子どもの自発性を促進する指導・子どもの授業態度の規律の向上に努める指導及び子どもの学習の方法を工夫する指導<sup>15</sup>は小学校において子どもの教育成果の向上に効果を持ちうるということが分かった。すなわち、表 4.4 の

---

<sup>13</sup>観察できない固定効果を取り除く方法としては、層別化によるサブサンプルの分析といった手法も候補に上がる。しかし本稿ではそのような方法を用いなかった。なぜならば、個人固定効果を対象に層別化してしまうとその人数だけ層ができてしまい解釈をすることがとても困難であり、また個人固定効果や個人科目固定効果・個人年度固定効果全てを考慮するとすれば、各々の層のサンプルサイズは 1 つだけになってしまう。これらの限界を踏まえて本研究では層別化の手法は用いなかった。

<sup>14</sup>ただし、ロバストネスチェックとして全ての学年をプールし、学校種別のダミーを用いた推定も行なった。その結果は表 4.4 に示したものと大差がないことを確認している。

<sup>15</sup>それぞれ「生徒の自発性を促進」インデックス、「生徒の授業態度への指導」インデックス、「授業手法の工夫」インデックス。

4章 子どもの教育成果に対する指導方法の効果の検証 -時間変化する要因を考慮した分析-

表 4.4: 指導方法が子どもの学力に与える影響

小学校						
コントロール変数	(1)	(2)	(3)	(4)	(5)	サンプル数
		個人固定効果	個人固定効果 個人科目固定効果	個人固定効果 個人年度固定効果	個人固定効果 個人科目固定効果 個人年度固定効果	
[1] 授業編成	0.009 (0.010)	-0.004 (0.006)	0.004 (0.008)	-0.005 (0.007)	0.015 (0.014)	9682
[2] 授業の内容の工夫	0.002 (0.010)	0.005 (0.006)	0.003 (0.007)	-0.005 (0.010)	-0.042*** (0.016)	9682
[3] 生徒の授業態度への指導	0.003 (0.011)	0.013 (0.013)	0.058*** (0.019)	0.002 (0.015)	0.165*** (0.043)	9682
[4] 授業手法の工夫	-0.006 (0.010)	0.006 (0.006)	0.010* (0.006)	-0.009 (0.018)	0.142*** (0.041)	9682
[5] 生徒の自発性の促進	-0.036 (0.011)	0.022*** (0.007)	0.016** (0.007)	0.058*** (0.016)	0.120*** (0.038)	9682
中学校						
コントロール変数	(1)	(2)	(3)	(4)	(5)	サンプル数
		個人固定効果	個人固定効果 個人科目固定効果	個人固定効果 個人年度固定効果	個人固定効果 個人科目固定効果 個人年度固定効果	
[1] 授業編成	-0.018 (0.018)	-0.008 (0.009)	-0.046** (0.024)	-0.006 (0.010)	-0.035 (0.033)	8220
[2] 授業の内容の工夫	0.005 (0.011)	0.020* (0.011)	0.030** (0.012)	0.016 (0.014)	0.038* (0.023)	8220
[3] 生徒の授業態度への指導	-0.015 (0.013)	-0.002 (0.015)	0.001 (0.018)	0.002 (0.020)	0.039 (0.038)	8220
[4] 授業手法の工夫	0.007 (0.011)	-0.014* (0.008)	-0.001 (0.011)	-0.022** (0.011)	-0.012 (0.030)	8220
[5] 生徒の自発性の促進	-0.005 (0.011)	-0.010 (0.008)	-0.017 (0.011)	-0.009 (0.009)	-0.017 (0.012)	8220

注) 本表では指導方法が子どもの学力に与える影響を示している。各々のセルでは指導方法の効果  $\beta$ (式(4.2)) の推定値及び標準誤差・有意水準を示しており、「推定値 有意水準 (標準誤差)」という表記法を用いている。「\*\*\*」・「\*\*」・「\*」はそれぞれ 1%・5%・10%の水準で統計的に有意であることを表す。

小学校を対象にした回帰表の(5)列で、上記3つのインデックスの教育成果に対する効果の推定値は正で有意であった。特に「生徒の自発性を促進」インデックスについては定式化に関わらず(2)列から(5)列の全てで推定値は頑健に正で有意であった。一方で、小学生の学力向上に対して習熟度別少人数学級の導入など学級編成を工夫すること(「授業編成」インデックス)は統計的に有意ではなく、授業で扱う内容を工夫する指導(「授業の内容を工夫」インデックス)は負に有意であった((5)列)。総じて纏めれば、小学校を対象にした時に子どもの教育成果を向上させるためには、学級編成や学習内容を工夫することよりもむしろ学習の方法や学習動機付けを対象にして指導することこそが重要でありうる。

さらに、小学校と中学校では指導方法の有効性は大きく異なり、小学校では効果が認められた上述の指導方法は中学校では効果は認められなかった。すなわち、表4.4の(5)列において、小学校では「生徒の授業態度への指導」「授業手法の工夫」「生徒の自発性の促進」といった指導方法インデックスに係る係数の推定値は1%の水準で統計的に正に有意に出ているが、一方で中学校では5%の水準であっても統計的に有意なインデックスは存在しなかった。また「授業の内容の工夫」については中学校においては10%の有意水準で統計的に有意ではあるものの、小学校における係数の推定値にくらべ係数の推定値は絶対値は小さく標準誤差も大きい。この結果からは、中学校では子

ものの教育成果に対する指導方法の影響というのはそれほど強くない可能性が示唆される。無論、小学校と中学校では年齢が異なるうえ指導を行う教員を割り当てる方法も違うため、小学校と中学校での結果をそのまま比較することには困難が伴う。特に4.3.2節で記したように、中学生を対象にした分析では必ずしも教員の内生的な配置や観察できない異質性を影響をコントロールできているわけではない。しかしそれでもなお、小学校で有効でありうる指導方法は必ずしも中学校では有効でありえず、「どのような指導が有効か」という問いを考えた時には「どのような子どもを対象としているか」という点にも注意を向けなければならないことを上記の結果は示している。

加えて、クラスの子どもの状況に合わせて、授業で用いる指導方法は変化している可能性が示唆された。すなわち、指導方法の効果の推定値はコントロールする固定効果によって大きく異なった。具体的には、小学校においてその効果が正に有意であった指導方法のインデックス（「生徒の自発性を促進」インデックス、「生徒の授業態度への指導」インデックス、「授業手法の工夫」インデックス）の推定値は(5)列では0.1標準偏差を超える効果量であるのに対し、(1)-(4)列では効果量は小さくそのほとんどが有意ではない。例えば「生徒の授業態度への指導」インデックスの効果の推定値は(5)列では0.17(1%の水準で有意)であるのに対して、(3)列では0.06(1%の水準で有意)であった。この推定値の違いをもたらしているのは、個人年度固定効果をコントロールしているか否かである。同様に(4)列では0.00(有意ではない)であり、(5)列における推定値と比較すれば効果量は小さく統計的有意性も失われている。この推定値の違いをもたらしているのは、個人科目固定効果をコントロールしているか否かである。これらの結果から、個人科目固定効果や個人年度固定効果は指導方法と負の相関を持ち、観察できない異質性が推定値を負の方向にバイアスさせていたと考えることができる。このことは、学校や教員がどのような指導をするかを決めるにあたって、眼前の子どもたちの特徴に応じて決めている可能性を示唆する。例えばある学校では、算数を苦手とする子どもが多くいる学年において算数の授業では念入りに授業を行い工夫を凝らした指導を行うことで子どもたちの学習を促進しようとした可能性がある。もしくは同様に、学力水準が低い学年においては、念入りに授業を構成している可能性がある。この様な状況が存在していたため、(1)列-(4)列では効果量は小さく推定されてしまったと考えることができる。そのため、この結果は学校や教員による指導方法の内生的に選択されておりその影響を取り除くことの重要性を示唆している。

## 4.5 結論

本稿では授業で用いられた指導方法が子どもの教育成果に与える影響を推定した。その際には学校から回答を得た子どもの指導に関するアンケート結

果から似た項目を纏めた指導方法のインデックスを作成し、そのインデックスと子どもの教育成果の間にある関係を分析した。授業でどのような指導方法がなされるかは学級の子どもの特徴やどのような教員が学校や学級に配属されるかと相関しうるため、指導方法の効果の分析はその内生性の問題に対処をする必要がある。そのため、個人固定効果のみならず個人科目固定効果・個人時間固定効果をコントロールした上で指導方法の効果の推定を行なった。

その結果として小学校においては数種の指導方法には子どもの教育成果への正の効果が認められた。すなわち、子どもの自発性を促進する指導や授業手法を工夫すること・子どもの授業態度への指導は小学生の学力に対して正の効果を持った。一方で、学級編成を工夫することは統計的に有意ではなく、授業で扱う内容を工夫する指導することは負の効果を持った。中学校においては指導方法と子どもの教育成果の間には統計的に有意な関係性は観察されなかった。様々に工夫され実践される指導の全てが、子どもの学力向上に結びつくわけではない。本稿での結果は、闇雲に指導方法の改善を志すのではなく、教育成果と結びつきうる指導方法において改善を志すことの重要性を示唆する。

ただし、本稿における分析は幾つかの限界を抱えている。まず第一に子どもの教育成果に影響を与えうる要因のうち、時間および科目によって変動する要因のコントロールには至っていない点を挙げることができる。例えば、算数の学力が低く推移しているのを見た教員が補修授業を行なったといったような意思決定が広く存在する場合、本稿で得られた推定結果はバイアスを含む可能性がある。第二に、本稿で用いている授業で用いた指導方法とは、あくまで学校による回答を元に作成しているという点である。Hidalgo-Cabrillana and Lopez-Mayan (2018) が指摘しているように、学校による自己認知が客観的事実と必ずしも整合的である保証はない。

子どもの教育成果に対する学校や教員の効果についての研究は日本のみならず全世界で未だ決まった結果の方向性が示されていない研究分野である。本稿で行った議論も多く存在する研究課題の一部を扱ったものであり決して絶対的な知見であるとは言えないが、それでも本研究は日本における数少ない指導方法についての研究として重要であると言える。

## これからの教員の質研究

ここでひとまず、教員の質研究における本稿の立ち位置を改めて総括する。ある教員が担当した子どもの能力の成長の平均的増分として定義される教員付加価値を用いた研究は近年急速に進展を見せており、本稿第3章においても同様の問題を扱った。教員付加価値推定には多くの仮定が存在しその妥当性については慎重に判断する必要があるものの、本稿が示した学力や非認知能力に対する教員付加価値の性質は日本における教員の子どもに対する影響力とその性質の一端を明らかにすることができたと言えよう。

#### 4章 子どもの教育成果に対する指導方法の効果の検証 -時間変化する要因を考慮した分析-

---

一方で「どのような教員が子どもの能力を伸ばすのか？」という教員の有効性についての研究も相変わらず重要である。なぜならば「どのような教員が有効か」という問いに対する答えは教員政策をそのまま方向付けるからである。もし高い学力を有する教員がそのまま高い教員としての質を示すという研究成果が認められた時には、教員政策はそのような教員の採用や育成に向けてのものになるであろう。しかし、本章の冒頭で記したように、現在の研究水準では教員の履歴書的な特徴が教員の有効性と強い関係性があるとは考えられていない。

そのような中で本章が取り上げたのは教員の指導方法であった。クラスルームにおける教員や学校の実践そのものを表す指導方法はその概念的に子どもの能力の成長との強い関係性を期待されており、近年の教育経済学の研究はそのような期待を実証するようなものであった。本章の研究の結果もそのような近年の研究動向を反映したものになっている<sup>16</sup>。

そしてその中で、本稿はこれまでの指導方法研究の手法の妥当性について強調して言及を行った。すなわち、これまでの教育経済学による指導方法についての研究は一定のバイアスを含んだ結果である可能性があることを指摘した。その意味で、これまでの研究はそのバイアスの程度を割り引いてその結果を評価する必要がある、またこれからの研究はそういったバイアスに対してどのように対処をするかが求められる。

本稿で用いた定式化はそのようなバイアスへの対処方法として有効ではあるものの限界も多い。その中でも最も大きい限界の一つが、本章で提案した手法では学力以外のアウトカムを対象に分析をすることはできないという点である。本稿で用いた手法は科目間の変動を利用するものであった。しかし、それは科目という概念が存在する学力を対象にして行うことができる分析であり、非認知能力などを対象にして分析を行うときは本稿の結果を用いることはできない。「(伸ばすべき)非認知能力に応じた指導方法」というような非認知能力に応じた指導方法という高度な実践は現状調査の対象になっておらず、またそのような比較が妥当なのかも現状十分な知見の蓄積は存在しない。

そういった中、今後の指導方法についての研究を行うときには、RCTや擬似実験的な状況を上手く用いることが求められる。学力だけではなく様々なアウトカムに対して分析を実行可能でまた現代的な因果推論の基準を満たすためには、指導方法についての研究においてもそのような条件付けられた状況でランダムな変動をうまく用いる必要がある。そして、そのようなこれからの指導方法研究・分析の1つの例・ケーススタディでもあるのが、次章にて展開するカンボジアにおけるICT教育の実践についての分析である。当該研究においては、カンボジアの首都ペノムペンでクラスター RCT を行い、

---

<sup>16</sup>加えて、この結果は本稿第3章の結果とも整合性が高いと考えられる。第3章では教員の影響力は非認知能力に対してより高いことが示されているが、教員の指導によって変わるのは学力そのものではなく子どもの学習態度や学習への取り組み方といった子どもの生活の過ごし方であるのかもしれない。

ICT を用いた指導の効果について認知能力・非認知能力共に分析を行った。次章では ICT 教育にしばってそのような研究を行ったものの、今後の指導方法を対象にした分析ではこのようなエビデンスの蓄積が求められていると言えよう。

## 関連図書

Algan, Yann, Pierre Cahuc, and Andrei Shleifer (2013) “Teaching practices and social capital”, *American Economic Journal: Applied Economics*, **5** (3), pp. 189–210.

Ammermüller, Andreas, Hans Heijke, and Ludger Wößmann (2005) “Schooling quality in Eastern Europe: Educational production during transition”, *Economics of Education Review*, **24** (5), pp. 579–599.

Aslam, Monazza and Geeta Kingdon (2011) “What can teachers do to raise pupil achievement?”, *Economics of Education Review*, **30** (3), pp. 559–574.

Bessho, Shunichiro, Haruko Noguchi, Akira Kawamura, Ryuichi Tanaka, and Koichi Ushijima (2019) “Evaluating remedial education in elementary schools: Administrative data from a municipality in Japan”, *Japan and the World Economy*, **50**, pp. 36–46.

Bietenbeck, Jan (2014) “Teaching practices and cognitive skills”, *Labour Economics*, **30**, pp. 143–153.

Clotfelter, Charles T, Helen F Ladd, and Jacob L Vigdor (2010) “Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects”, *Journal of Human Resources*, **45** (3), pp. 655–681.

Comi, Simona Lorena, Gianluca Argentin, Marco Gui, Federica Origo, and Laura Pagani (2017) “Is it the way they use it? Teachers, ICT and student achievement”, *Economics of Education Review*, **56**, pp. 24–39.

Hanushek, Eric A (1986) “The economics of schooling: Production and efficiency in public schools”, *Journal of Economic Literature*, **24** (3), pp. 1141–1177.

——— (2006) “School resources”, *Handbook of the Economics of Education*, **2**, pp. 865–908.

4章 子どもの教育成果に対する指導方法の効果の検証 -時間変化する要因を考慮した分析-

---

- (2008) “Education production functions”, *The New Palgrave Dictionary of Economics: Volume 1–8*, pp. 1645–1648.
- and Steven G Rivkin (2006) “Teacher quality”, *Handbook of the Economics of Education*, **2**, pp. 1051–1078.
- Hidalgo-Cabrillana, Ana and Cristina Lopez-Mayan (2018) “Teaching styles and achievement: Student and teacher perspectives”, *Economics of Education Review*, **67**, pp. 184–206.
- Hojo, Masakazu and Takashi Oshio (2012) “What Factors Determine Student Performance in East Asia? New Evidence from the 2007 Trends in International Mathematics and Science Study”, *Asian Economic Journal*, **26** (4), pp. 333–357.
- Kane, Thomas J and Douglas O Staiger (2008) “Estimating teacher impacts on student achievement: An experimental evaluation”, National Bureau of Economic Research.
- Lavy, Victor (2015) “What makes an effective teacher? Quasi-experimental evidence”, *CEifo Economic Studies*, **62** (1), pp. 88–125.
- Lefgren, Lars and David Sims (2012) “Using subject test scores efficiently to predict teacher value-added”, *Educational Evaluation and Policy Analysis*, **34** (1), pp. 109–121.
- Schwerdt, Guido and Amelie C Wuppermann (2011) “Is traditional teaching really all that bad? A within-student between-subject approach”, *Economics of Education Review*, **30** (2), pp. 365–379.
- Van Klaveren, Chris (2011) “Lecturing style teaching and student performance”, *Economics of Education Review*, **30** (4), pp. 729–739.
- 伊藤寛武・田端紳 (forthcoming) 「子どもの教育成果に対する学校施策の有効性の検証-時間変化する要因を考慮した分析-」, 『国民経済雑誌』.

# 第5章 コンピューター支援教育は 子どもの能力を向上させるか？

## 概要<sup>1</sup>

本章では、コンピューター支援教育 (Computer-Assisted Instruction, CAI) が子どもの認知能力と非認知能力に与える効果を分析した。カンボジアのプノンベン近郊に所在する小学校5校、約1,600名の子どもを対象に、クラスター化無作為化比較試験を実施した。その結果、認知能力に対する平均処置効果は肯定的で統計的に有意である一方で、学習時間には変化がなかったことが示唆された。これは、CAIが子どもの単位時間当たりの学習生産性を向上させることに成功していることを示している。さらに、CAIは子どもの大学進学に対する主観的な期待感を高めることがわかった。

## 5.1 はじめに

世界銀行は近年「学習の危機 Learning Crisis」について言及している (World Bank, 2017)。例えば、市場での売買や家計のやりくり、銀行や他の金融機関との取引に必要な数学についての基礎的なスキルを学校で身につけることすらできていない生徒が、開発途上国の生徒の大きな割合を占めていることがわかっている (Hanushek and Woessmann, 2016)。

そして多くの低所得国ではここ数十年で小学校への就学者数が急速に増加しており、学習の危機を回避する上で大きな障害に直面している。第一に、初等教育への就学者数は増加してきたが、一人当たりの教育投入量の減少は初等教育の質を低下させる可能性が高い。第二に、多くの開発途上国では質の高い教師の雇用が困難である。なぜならば、教師の給与が他の同等の資格を持つ専門職よりも低いためである。第三に、低学力の生徒と高学力の生徒の能力に大きな差があると、教師が適切な指導レベルを設定することが困難になる。このような状況では、教師の指導レベルと生徒の習熟度の間にミスマッチが生じうる (Glewwe and Muralidharan, 2016)。

<sup>1</sup>本章は Ito et al. (2021) として出版されたものを一部修正して掲載している。

新技術は発展途上国におけるこのような問題を軽減する有望な方法を提供すると考えられる。教室でのコンピューターの利用そのものは必ずしも生徒の学習の質を向上させるものではないが (Barrera-Osorio and Linden, 2009)、よく設計されたコンピューター支援教育 (Computer-Assisted Instruction、以降 CAI と記す。コンピューター支援学習 (Computer-Assisted Learning、CAL) とも呼ばれる) の下では、生徒は教師が不足していても質の高い教材にアクセスでき自分のペースと習熟度に合わせて学習することが可能である。幾つかの研究がこの CAI の効果を検証している。Linden (2008) はインドにおける RCT を通じて特に成績の低い生徒の成績を大幅に向上させることを示している。また、ペルーとウルグアイの「子ども一人につき一台のラップトップ」プログラムでは生徒の読解力や数学の能力に CAI は影響を与えなかったことがわかっている (Cristia et al., 2017; De Melo et al., 2014)。

本稿は、カンボジア政府、国際協力機構 (JICA)、日本の民間企業であるはなまるラボと共同で行われた経済実験から、CAI が生徒の認知能力と非認知能力に与える因果関係を厳密に推定することを目的とする。実験では Think!Think! という教育アプリケーションが用いられた。この Think!Think! の主な目的は、小学生の算数の基礎力を養うことにある。

Think!Think! の効果を調べるために、2018 年 5 月から 8 月にかけて、プノンペン近郊の公立小学校 5 校で、1 年生 (G1) から 4 年生 (G4) までの 1,656 人の生徒を対象としたクラスター化無作為化比較試験 (RCT) を実施した。各学校は各学年に 2 つのクラスがあるため、生徒は 3 カ月間の介入期間中に Think!Think! を使用した 20 のトリートメントクラスのいずれか、または 20 のコントロールクラスのいずれかに無作為に割り付けられた。

結果として、数学のテストと IQ テストによって測定された認知能力に対する平均処置効果が正で統計的に有意であった。その効果の大きさは、特に発展途上国で実施された先行研究と比較して大きなものであり、ベースライン調査で調べられた性別・学年・出生月・親の教育・学校の時間変化しない異質性を考慮した場合でも、達成度に関する推定値は 0.68~0.77 標準偏差、IQ スコアに関する推定値は 0.66 標準偏差であった。さらに、CAI は学生の大学進学に対する主観的な期待感を高めることがわかった。しかし、モチベーションや自尊心などの非認知能力には有意な影響は見られなかった。

本研究の貢献は、主に以下のように要約される。(1) 第一に先行研究ではテストスコアの向上に焦点が当てられていたが、本研究ではテストスコアや IQ などの認知能力だけではなく、モチベーションや自尊心などの非認知能力や学習時間といった様々なアウトカムに対してコンピューター支援教育の効果を検証した。(2) 第二に、特に発展途上国では新たに開発された技術への期待が高まっているが、その外部妥当性を検証するためにより厳密な定式化による研究を行った。分析の結果、CAI が生徒の認知能力を向上させた要因は、授業時間数の増加ではなく単位時間当たりの学習生産性の向上によるもので

あることが明らかになった。

本章は下記のような構成になっている。5.2節では先行研究のレビューを行う。5.3節では、研究デザインとデータを説明する。5.4節では、推定結果を示す。5.5節では、政策的なインプリケーションを示す。

## 5.2 先行研究

過去の研究では、学校によるコンピューターへの投資は、(i) 情報通信技術 (ICT) または (ii) CAI のいずれかとして定義されている。特に近年では、インターネット接続を必ずしも必要としない CAI プログラムが公立学校でより広く使用されるようになった。そして、いくつかの研究では、適切に設計された CAI プログラムが、特に発展途上国の学力の低い学生の数学または科学の能力に強く正の影響を与えることが示された。Banerjee et al. (2007) は、インドの都市の G4 学生に対する CAI プログラムの効果を調査した。トリートメント群の学校に割り当てられた学生は、主に成績の悪い子どもたちの改善により、数学の成績を標準偏差 0.47 増加させた。プログラムが終了した後も、約 0.10 の標準偏差に減少はしたがこの肯定的な効果は残った。しかし他の研究では読書と言語テストのスコアにわずかな影響しか与えていなかった。例えば、Rouse and Krueger (2004) は、米国北西部の都市地区にあるコンピュータソフトウェアプログラム Fast For Word for G3~G6 の学生を使用して大規模な RCT を実行している。彼らの結果は、言語と読解力に対する CAI のプログラムの影響が小さいものの、統計的に有意であることを示した。

経済学では、コンピューター・インターネット・ソフトウェア、およびその他の技術への投資は通常教育生産関数の文脈で分析される。Bulman and Fairlie (2016) は、モデルを規定するもののうち一つが、指導に利用できる時間であることを指摘した。CAI においては、コンピューターを使った指導時間と従来の指導に費やされた時間との間のこのトレードオフにより、学校が CAI プログラムを使用すべきかは決定される。それにも関わらず、Rouse and Krueger (2004) および Banerjee et al. (2007) を含む多くの研究では、通常の授業外の CAI プログラムを用いる補習的な教育の効果を推定している。そのため、それらの研究からは、CAI プログラムが伝統的な指導を代替するものなのかについて判断することは難しい。

これらの問題に対処するために Barrow et al. (2009) では、ランダムに割り当てられたトリートメントクラスの中学生在が CAI を利用して教えられる一方、コントロールクラスの学生は伝統的にクラスで教えられる介入を考えている。これにより、新しく開発された CAI プログラムの効果と、限られた学校のリソースと時間の制約の下でのより伝統的な指導の効果の比較が可能になっている。2年間の実験の結果、トリートメントクラスの生徒はコントロールクラスの生徒よりも少なくとも 0.17 標準偏差だけ数学の能力が向上したこ

とがわかった。また、Carrillo et al. (2011) は、エクアドルで小学生を対象に同様の実験を行った。分析の結果、従来の授業の代わりにクラスで CAI を使用すると、数学のパフォーマンスは向上がなされた一方で、言語習得は改善されなかったことがわかっている。また、インドの都市部の中学校に関する最近の調査では、クラスで CAI を使用すると数学と言語の両方の能力に大きな影響があることが示されている (Muralidharan et al., 2019)。彼らの操作変数法を用いた推定は、トリートメントの学生が5か月の介入中に数学で0.37標準偏差、ヒンディー語で0.23標準偏差能力が向上したことを示唆している。彼らはまた、学力の低い学生ほど成績の向上が大きいことも発見している。本稿の分析はこの Muralidharan et al. (2019) の分析に基づいており、CAI プログラムが発展途上国の比較的恵まれない地域の低年齢の子どもたちに効果的かどうかを検証する。

## 5.3 手法とデータ

### 5.3.1 背景

本研究では、プノンペン周辺の半径約10km以内に位置する5つの公立小学校を対象としている。これらの学校は介入期間中に他の開発機関からの援助を受けていなかったため、他の外部介入による交絡要因は除外できる。また学校周辺の世帯の大部分は収入を得るために農業や漁業に従事しており、高等教育を受けている親の割合はわずかである。これら5つの学校の位置を図5.1に示した。

### 5.3.2 ベースライン調査とフォローアップ調査

介入に先立ち、2018年5月21日から5月25日まで、教職員の全面的な協力のもと、授業内でベースライン調査を実施した。ベースライン調査では、G3とG4の生徒を対象とした40分間の達成度調査を2種類、全生徒を対象とした40分間のIQテスト、全生徒と保護者を対象とした20分間のアンケート調査を実施した。

学生の認知能力を測定するために、カンボジア教育省が実施しているG3の学生を対象とした全国学力テスト(NAT)と、International Association for the Evaluation of Educational Achievement (IEA) が実施しているG4の学生を対象としたTrends in International Mathematics and Science Study(TIMSS)の2種類の学力テストを使用した。これらのテストでは介入対象の生徒が過去に受験したことのない問題が使われている。低学年の子どもの認知能力を測定するための標準化された調査がないため、G1とG2には達成度テストを実施しなかった。その代わりに、調査では年齢に応じた2つのIQテストを

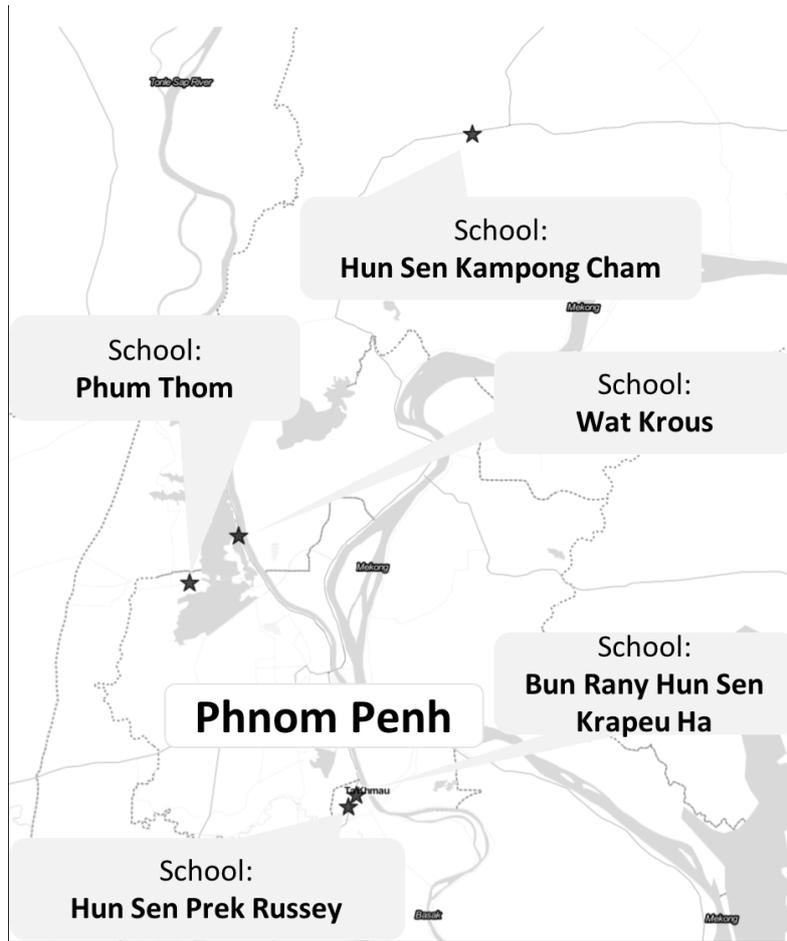


図 5.1: 介入の対象となった学校の位置

実施した。そのうちの1つが「新田中B型知能検査」である。田中B型知能検査 (Tanaka et al., 2003) は、日本をはじめとするアジアの国々で、子どもの年齢に応じた認知能力を測る尺度として古くから利用されてきた。この田中B型知能検査を現地の言語に翻訳し、現地の環境に合わせて適切に改変して用いた (紙幣や食べ物、人のイラストなど)。

ベースライン調査で実施されたもう一つの知能テストは、Goodenough Draw-a-Man (DAM) テスト (Goodenough, 1926) である。このテストでは、生徒は紙に人物全体の絵を10～15分かけて描くように求められる。収集された子どもたちの絵のいくつかの例を図5.2に示した。知能の尺度としてのこのテストの妥当性については批判もあるが、ある研究によるとDAMテストは低い知能の5-12歳児をスクリーニングするのに有効であることが示唆されている (Scott, 1981)。



図 5.2: Samples of Draw-a-Man Test

G1～G4の全生徒を対象とした調査では、性別・学年・出生月・自宅での学習時間・将来の大学進学の主観的な可能性などについて質問を行った。その調査には、ローゼンバーグ自尊感情尺度 (Rosenberg self-esteem scale (Rosenberg,

1965)) や学習意欲 (internal and external motivation for study)(Sakurai and Takano, 1985) などの非認知的能力を測定するための質問票も含まれている。また、保護者調査では、親の学歴や社会経済的地位を尋ねている。

3ヵ月間の介入の後、8月16日から8月25日にかけて追跡調査を実施した。大学への進学意欲や自宅での学習時間など、時間的に変化する変数のみに焦点を当てて再度達成度テスト、IQテスト、アンケート調査を実施した。

対象校において正式に登録されている1656名のうち、ベースライン調査とフォローアップ調査の両方に参加したのは77.2%であったが、ベースライン調査のみに参加したのは6.3%であった。このサンプルの減少は、トリートメントとコントロールの比較可能性を低下させる可能性がある。例えば、トリートメントに割り振られた低学力の生徒は介入が上手くいったことで介入期間中に退学しないかもしれないが、コントロールに割り振られた低学力の生徒はに退学するかもしれない。この場合、この介入の推定値は下方に偏る可能性がある。そこで、トリートメントとコントロールの両方について離脱率を計算し、その値が子どもの特徴と関連していないかを調べた。幸いなことに、トリートメントとコントロールでは、離脱率に差があったり離脱をした子どもの特徴が異なっているという証拠は見られなかった。しかし、ベースライン調査にもフォローアップ調査にも参加しなかった学生の9.2%についてはその詳細はよくわかっていない。世界銀行の最新の指標によると、2017年のカンボジア全国の中途退学率は9.4%である。私たちの介入は学期末の3ヶ月間に実施されたため、介入前や介入中に中退してしまう生徒もいるかもしれない。この問題に対処するために、ベースラインデータが欠落している場合は0とするダミー変数を作成し、ロバスト性チェックとしてANCOVAによる定式化で制御した。

### 5.3.3 教育アプリケーション: Think!Think!

今回の介入で使用したThink!Think!というアプリケーションは、株式会社はなまるラボが日本の児童向けの塾を数多く運営してきた経験を活かして開発したものである。このアプリケーションは、特に小学生の算数の基礎力を養うために開発された(図:例5.3)。具体的には、このアプリケーションでは独自のアルゴリズムによる適応学習を取り入れ、個々の生徒の習熟度に応じた算数の問題・教材・指導を提供している。多くの文献で示唆されているように、算数・理科の能力は国を超えて経済成長との関連性が高いとされており(e.g. Hanushek and Kimko, 2000; Jamison et al., 2007; Hanushek and Woessmann, 2016)、数学的習熟度は経済成長を牽引するだけでなく個人の所得を高める効果もあるとされる。例えば、デンマークでは高等学校の数学習得コストが制度的に低下することを利用し、数学を重視した高校の専門科目を選択することが将来の労働市場の収益に因果関係があることを Joensen and Nielsen (2009) が実証している。

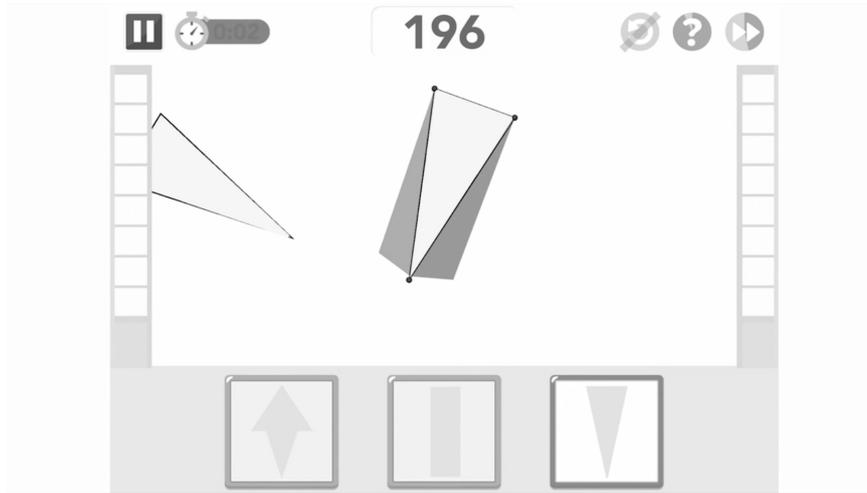


図 5.3: Sample Problem

Think!Think!はカンボジアの小学生向けに現地のカリキュラム基準に合わせて修正され、現地の言語であるクメール語に翻訳された。トリートメントに割り当てられた生徒には、授業中に追加の教員が必要になることがあった。そのため、介入では、技術的な問題や時間管理について生徒にアドバイスをするために教員経験のない3人のスタッフを追加で派遣した。

介入で使用した Think!Think!と、Muralidharan et al. (2019) が使用した Mindspark によるコンピュータ支援学習を注意深く比較すると、多くの特徴が非常に似ている。彼らは論文の中で、P1431 から P1432 まで、Mindspark を使うことの利点を強調している。それは (i) 質の高い教材、(ii) 個々の学生に合わせた「正しいレベルでの授業」を実現するためのアダプティブコンテンツ、(iii) 学生特有の概念的ボトルネックの緩和、(iv) インタラクティブなユーザーインターフェースといった要素で表すことが出来る。この類似性の結果として、本稿の結果は Muralidharan et al. (2019) と強く一致している。

一方で少し異なる点として、Mindspark は中学生 (G6~G9) 向けにヒンディー語 (言語) プログラムと数学のコンテンツを提供しているのに対し、Think!Think!は小学生向け (G1-G4) に特化している。さらに、Muralidharan et al. (2019) と本稿の結果の最も重要な違いの1つは、プログラムのあり方にある。Muralidharan et al. (2019) の介入は “blended learning” プログラムであり、すなわち “Mindspark コンピューター支援学習 (CAL) プログラムとグループごとの指導と追加の指導時間の組み合わせ” (p.1429) である。したがって、彼らの結果は CAL の純粋な効果を介入内容から分離することができない。しかし学校の中で教育に利用できる時間には上限があると仮定すると、学校が従来の授業での教育との比較は、開発途上国の政府にとってより重要な政策問題になる可能性がある。

### 5.3.4 Clustered RCT

もし子どもが自分の好みに基づいて CAI にアクセスできるようになったと考えた時、そのアプリケーションは成績の良い学生が利用する可能性が高い。新しい技術に触れることを含めより質の高い教育を受けようとしたことのある学生はそうでない学生に比べて、平均して勉強に熱心である。CAI ベースのソフトウェアへのアクセスを無作為に割り当てることで、この選択バイアスを回避することができる。

本研究において、トリートメントクラスの学生は毎日約 30 分間 Think! Think! を用いた。ピア効果はこの実験の内部妥当性に対する潜在的な脅威であり、学生間の相互作用により推定は SUTVA に違反する可能性がある。教育においてはクラスター化 RCT がより一般的であるというだけでなく、そのような状況を避けるために我々は教室内の個々の学生ではなく教室ごとに学生を無作為に介入群と非介入群に割り付けた<sup>2</sup>。

各学校は各学年に 2 つのクラスを持っているため、クラスターランダム化比較試験を使用した。すなわち、各学校において 1 学年に 1 つのトリートメントクラスを無作為に選んだ。この操作は、5 つの学校全体で 20 のトリートメントクラス (840 人の学生) と 20 のコントロールクラス (816 人の学生) を生んだ<sup>3</sup>。しかし、トリートメントクラスの生徒が、同じ学校のコントロールクラスの友人に、自分が学んだことを話してしまうのではないかという懸念は残る。そのような波及のリスクを減らすために、トリートメントクラスの生徒は授業外での Think! Think! を利用することは許可されなかった。さらに、タブレットやノートパソコンを家に持ち帰ることも許可されなかった。しかし、クラスレベルのランダム化比較試験では、トリートメントとコントロールの間のスピルオーバーを抑制するのに十分ではないという懸念は依然存在する。トリートメントを受けた仲間内での正の波及効果と、同じ学校でトリートメントを受けていない生徒への副次的な効果があれば、偏りのない推定値はより大きくなるかもしれない。

3ヶ月という比較的短い介入期間にもかかわらず、生徒たちは Think! Think! に非常に熱心に取り組んだ。この場合、ホーソン効果とジョン・ヘンリー効果と呼ばれるコントロールの評価による行動変化が介入効果に存在する可能性がある。トリートメント内での有意なモチベーションの変化は見られないものの、ホーソン効果はトリートメントの子どもの成績を向上させるため、コンピュータ支援教育の真の効果と比較して推定値は過大評価されている可能性がある。一方、ジョン・ヘンリー効果はコントロールの子どもの成績を向上させるため、コンピュータ支援教育の真の効果は過小評価されている可

<sup>2</sup>しかし、(Imbens and Wooldridge, 2009) が指摘しているように、個人への介入の直接的な影響と、その個人への仲間の間接的な影響とを区別することは技術的に困難である。

<sup>3</sup>同じ教室の学生の結果の間に観察されない相関関係がある可能性があり、クラスター標準誤差はそのような相関関係を補正するために使用することができる。しかし、このような相関関係を補正するためにはクラスター標準誤差の計算には少なくとも 42 クラスターが必要であることが示唆されているため (Angrist and Pischke, 2008) 我々の推定を適用することはできない。

能性がある。

## 5.4 識別戦略および推定結果

### 5.4.1 推定に用いた定式化

Think!Think!を使用した指導の因果効果を推定するために、以下の定式化に基づいて因果効果の推定を行った。

$$Y_{i,j,t} = \alpha + \beta T_{i,j,t} + \gamma Y_{i,j,t-1} + \delta \text{MissingBaseline}_{i,j,t} + X_{i,j,t} \sigma + \epsilon_{i,j,t} \quad (5.1)$$

ここで  $Y_{i,j,t}$  は、時間  $t$  における  $j$  校の学生  $i$  のアウトカムである。 $\text{MissingBaseline}_{i,j,t}$  は、生徒  $i$  がベースライン調査に参加したかどうかを示すダミー変数である。 $X_{i,j,t}$  はコントロール変数のベクトルで、 $\epsilon_{i,j,t}$  は誤差項である。 $Y_{i,j,t-1}$  をコントロールする ANCOVA を用いた。

この実証モデルの重要な仮定は、CAI ベースのソフトウェアを用いるという介入と学生の観察されない能力との関係が、コントロールを条件として誤差項に直交するというものである。この仮定のもとでは、(1) 式の  $\beta$  の推定値は、CAI ベースのソフトウェアが学生のアウトカムに与える因果効果と解釈することができる。

### 5.4.2 変数の詳細

表 5.1 では、ベースライン調査のバランスチェックを示した。Bruhn and McKenzie (2009) に従い学年ごとの固定効果を制御した時、トリートメントクラスに割り当てられた G3 の学生とコントロールクラスに割り当てられた学生との間の NAT の結果に統計的に有意な差はなく、トリートメントクラスの G4 学生はコントロールクラスの G4 の学生よりも TIMSS の結果がわずかに優れていた。

田中 B 型 IQ テストと DAM テストの結果を精神年齢 (MA) に換算し、MA を年代別年齢 (CA) で割ったものに 100 を乗じたものが IQ テストのスコアである。記述統計によると、田中 B 型 IQ テストの平均値は 78.612、標準偏差 13.451、DAM テストの平均値は 0.692、標準偏差 0.207 となっている。田中 B 型 IQ テストスコアと DAM 型 IQ スコアの間には統計的に有意な差はない。

非認知能力の値は、自尊心と動機づけそれぞれにおいて、関連する質問の平均値として計算されている。自尊心の尺度はトリートメントを受けた学生の方がわずかに高いが、動機づけの尺度は 2 つのグループの学生で同様であ

る。すべての認知的および非認知的能力の値は、回帰分析を実行する際に、平均を0標準偏差を1に正規化している。

大学への進学意欲は、学生の主観的な期待値をもとに3段階の尺度(1=「可能性がない」から3=「可能性が高い」まで)で表される。自宅で勉強した時間は、6段階の尺度(1=「全くない」から6=「4時間以上」まで)で測定される。その尺度から勉強時間の値の最小値を0、最大値を4とし、2(=「30分未満」)と5(=「2-3時間」)の間のカテゴリの回答はその中央値を用いて表した。 $T_{i,j,t}$ で示される介入を表す変数は、学生がトリートメントクラスに割り当てられている場合は1、そうでない場合は0としてコード化されたダミー変数として表される。

性別・年齢・親の学歴など、 $X_{i,j,t}$ で示される子どもの特徴は、トリートメントの子供とコントロールの子供の間で非常によく似ていた。親の学歴に関する値としては、どちらかの親の最高学歴を用いた。これらの値は学生調査と同時に実施された保護者調査から取得されていることに注意する必要がある。授業中に実施された学生調査の100%の回答率とは異なり親の調査の回答率は約85%であった。

観察可能な特性は2つのグループ間で類似しているが、G4の生徒の学力などのいくつかの値はトリートメント群とコントロール群の子供の間で有意な差があった。ランダム化が正しく行われていたとしても、トリートメント群とコントロール群の間で不均一性が発生する可能性がある。グループレベルによるランダム化においてグループ間での均一性が得られる可能性はサンプルサイズに依存するため、本研究で用いるデータにおける幾つかの変数で不均一であったことは必ずしもランダムな割り当てが機能していないことを意味するわけではない。ただし、学校のクラス構成は無作為に変更されているのにも関わらず、ベースライン調査の日における欠席または退学のために、トリートメントグループとコントロールグループの間に不均一性が依然として存在する可能性がある。そのため、施策効果を推定するには異質性のチェックに使用する特徴をコントロールした。

### 5.4.3 推定結果

#### 認知能力への効果

まず、生徒の成績に対するCAIの影響を推定した。推定値はロバスト標準誤差とともに表5.2にて示した。ここでの主な焦点は表に示したThink!Think!のG3のNATおよびG4のTIMSSへの推定された効果である。モデル1はコントロールがない場合の推定の結果を示している。モデル2は、ベースライン調査の成績とベースライン不参加ダミーをコントロールした結果を表す。モデル3では以前のテストのスコアとベースライン不参加ダミーに加えて、

5章 コンピューター支援教育は子どもの能力を向上させるか？

表 5.1: Descriptive statistics and balance test

	ALL	Treatment (A)	Control (B)	Difference (A)-(B)
Achievement Test (NAT, G3)	0.538 (0.207, 356)	0.554 (0.214, 179)	0.522 (0.198, 177)	0.031 (0.039)
Achievement Test (TIMSS, G4)	0.292 (0.203, 347)	0.252 (0.211, 173)	0.330 (0.187, 174)	-0.067* (0.035)
IQ Test (Tanaka-B)	78.612 (13.451, 1385)	78.795 (13.777, 685)	78.432 (13.131, 700)	0.401 (1.647)
IQ Test (Draw-a-man)	0.692 (0.207, 1217)	0.705 (0.207, 623)	0.678 (0.206, 594)	0.026 (0.033)
Self-esteem	2.762 (0.549, 1150)	2.794 (0.502, 615)	2.726 (0.596, 535)	0.039 (0.043)
Motivation	0.656 (0.142, 996)	0.660 (0.133, 525)	0.652 (0.150, 471)	0.010 (0.013)
Willingness to go to college	2.410 (0.771, 1051)	2.467 (0.734, 569)	2.342 (0.809, 482)	0.108 (0.109)
Minutes of studying at home week	168.667 (117.005, 949)	167.481 (123.173, 526)	170.142 (108.975, 423)	-3.111 (13.628)
Gender (male=1, woman=0)	0.525 (0.500, 1643)	0.519 (0.500, 830)	0.530 (0.499, 813)	-0.010 (0.016)
Age	8.485 (1.553, 1620)	8.470 (1.535, 817)	8.501 (1.573, 803)	-0.034 (0.048)
Highest parental education				
College or Graduate school	0.023 (0.149, 1236)	0.029 (0.167, 626)	0.016 (0.127, 610)	0.012* (0.005)
High school	0.457 (0.498, 1236)	0.441 (0.497, 626)	0.474 (0.500, 610)	-0.035 (0.038)
Junior high school	0.299 (0.458, 1236)	0.305 (0.461, 626)	0.292 (0.455, 610)	0.019 (0.022)
Elementary school	0.220 (0.414, 1236)	0.225 (0.418, 626)	0.215 (0.411, 610)	0.008 (0.030)
no education(ref)	0.002 (0.040, 1236)	0.000 (0.000, 626)	0.003 (0.057, 610)	-0.004 (0.003)
Birth of Month				
Jan. - Mar.	0.234 (0.423, 1620)	0.245 (0.430, 817)	0.223 (0.416, 803)	0.004 (0.026)
Apr.-Jun.	0.246 (0.431, 1620)	0.230 (0.421, 817)	0.263 (0.440, 803)	-0.037 (0.030)
Jur.-Sep.	0.249 (0.433, 1620)	0.244 (0.430, 817)	0.255 (0.436, 803)	-0.018 (0.024)
Oct.-Dec.	0.270 (0.444, 1620)	0.282 (0.450, 817)	0.259 (0.438, 803)	0.038 (0.031)

注) この表では分析に用いるデータの記述統計を表した。トリートメントとコントロールとは、子どもが CAI のクラスにランダムに割り当てられるかどうかを示している。そして、割り当てられたトリートメントに限らず、2018 年 5 月に学力テスト・質問紙調査・親への質問紙調査を行った。本表で表している値は全てその時の値になる。各セルでは、「平均値 (標準偏差, サンプルサイズ)」という表記方法を用いた。また、「Difference」列では子どもがトリートメントに割り当てられている場合 1、コントロールに割り当てられている場合 0 をとるようなダミー変数を、それぞれの変数に回帰させた場合を推定値を示した。「\*\*\*」「\*\*」「\*」という表記はそれぞれ 0.1%、1%、5% の水準で有意であることを表している。

性別、学年、誕生日、親の教育、学校学年固定効果などの基本的な人口統計学的特徴をコントロールした結果を表した。

結果は、NAT に対する推定値が 0.1 %レベルで正で統計的に有意であることを明確に示している (表 5.2、NAT)。特にモデル 3 における G3 の子どもにおける推定値は CAI の利用が平均的にテストスコアを約 0.77 標準偏差上昇させることを示している。

子どもの特徴と学校学年固定効果を追加しても、モデル間の係数の推定値は大きく異ならなかった。相互作用項を含め、性別、学年、親の教育に関する異質な効果を推定すると、ほぼすべての相互作用項で小さな点推定値が得られ、これらの係数間の差はテスト得点に対する有意な異質な効果の仮説を支持しなかった。さらに、学力の弱い生徒であっても達成度の向上は均質であった。

また、G4 の TIMSS の成績に対する推定値に対しても同様の結果を得ることができる (Table 5.2, TIMSS)。CAI の利用は、3ヵ月間のトリートメントで TIMSS のスコアを 0.68 標準偏差改善した (モデル 3)。コントロールを追加するとこの点推定値は増加し、これらの推定値の標準誤差は減少した。同時に、性別・学年・親の教育またはベースラインにおける学力がテストスコアに及ぼす異質な効果は有意には見出されなかった。

次に、生徒の IQ スコアに対する CAI の影響を推定した。Table 5.3 では、田中 B 型 IQ スコアに対する介入効果の推定値は正であり、0.1%レベルで統計的に有意であることが示されている。モデル 1 の IQ スコアに対する効果の推定値は 0.71 標準偏差である。その推定値はモデル 3 で子どもの特徴をコントロールした後も変化していない。しかし、DAM スコアに対する介入効果の推定値はモデルにかかわらず、統計的に有意ではない。

全体的に、我々の結果は、過去の研究で示唆された数値と比較して、認知スキルの大きさが非常に大きいように見えることを示している。Muralidharan et al. (2019) は非常によく似た CAL ソフトウェアをインドのデリーにある比較的貧しい生徒に用いており、その結果を本稿の結果と比較することは価値がある。そのうち比較可能な ITT の推定値は、介入の 4.5 か月後にトリートメントの学生はコントロールの学生と比較して 0.23 標準偏差スコアが高くなったことを示した。一方で本稿からはモデル 3 において 0.77 標準偏差 (G3、NAT)、0.68 標準偏差 (G4、TIMSS)、および 0.66 標準偏差 (G1-G4、Tanaka-B IQ) の効果量を得ることができた。Muralidharan et al. (2019) では、デリーの Mindspark センターと呼ばれる学校の生徒が採用され、その保護者は子どもたちがくじで授業料免除に選ばれるかどうかは決まると通告された (月額 200 インドルピー、3 米ドルに相当)。参加者は自身が参加することを選択するため (そしておそらく非常にやる気があり)、管理データは彼らが非参加者と比較してパフォーマンスが優れていることを示唆している。また、Muralidharan et al. (2019) は初期の学習レベルによって生徒の進歩にか

なりの不均一性があることを発見し、介入によるテストスコアの向上は初期の低学歴の生徒の方がはるかに大きいことを示した。そのためより成績の低い学生を含む代表サンプルから引き出された真の推定値は、彼らの論文で報告されている推定値よりもはるかに大きい可能性がある。本稿では公立学校のすべての生徒をカバーしたため、参加者は介入に自己選択されず介入の効果としてより正しい値を得ることができる。

3ヶ月間の介入後のスコア分布を視覚的に比較するために、カーネル密度推定を用いて達成度テストのスコアとIQスコアの両方について確率密度関数を求めた(図 5.4 - 5.6)。サンプル全体の DAM スコアの差や、学年との交互作用項でさえも統計的に有意ではないが、図からは低学年の生徒のスキルが向上しているように見える。

#### 非認知能力および学習習慣への効果

次に、非認知能力を用いて上記と同様に CAI の効果を推定した(表 5.4)。認知能力の結果とは異なり、モチベーションと自尊心で測定された非認知能力には有意な効果は見られなかった。この結果からは CAI の非認知スキルに対する正の効果は示されていないが、確率密度関数(Figure 5.7 - 5.8)は低学年ではわずかな変化の兆候を示唆している。

また、学習時間も教育生産関数の重要なインプットと考えられるため、自宅での学習時間への影響も推定した(表 5.5)。その結果、有意な推定値は得ることはできなかった。上記で述べたように、タブレット PC を自宅に持ち込むことは認められていない。そのため、このように自宅での学習時間を長くする効果が見られないことは納得できる。しかし結果として、トリートメントクラスの学生は自宅でも教室でも勉強時間に変化はなかったのに関わらず成績が大幅に向上した。このことは、CAI が学生の学習効率や生産性を向上させることに成功していることを示している。

大学進学意欲に関する推定値(表 5.5)は 0.1%レベルで正で統計的に有意であった。この結果は、授業中に CAI を利用した学生は将来より高度な教育を受けると考える可能性が高いことを示している。この推定値はモデル 2 で子どもの特徴をコントロールした後も一定であり、性別・学年・親の学歴による異質な効果は存在しないことを示唆している。

表 5.2: Effect of Treatment: Cognitive Skills

Dependent Variable	National Assessment			TIMSS		
	Model1	Model2	Model3	Model1	Model2	Model3
Treatment	0.814*** (0.291)	0.723*** (0.204)	0.767*** (0.223)	0.522*** (0.135)	0.630*** (0.091)	0.681*** (0.104)
Baseline Score		✓	✓		✓	✓
Control			✓			✓
Observations	369	369	298	350	350	303
Adjusted R2	0.131	0.619	0.695	0.051	0.096	0.213

注) 表では介入効果の推定値を示している。データの単位は各個人。「モデル1」列ではOLSによって得られた推定値を示している。「モデル2」列では介入前の学力およびベースライン不参加タミーをコントロールしている、「モデル3」列ではモデル1に加えて性別や学年、生まれ月、親の教育レベル、学校学年固定効果をコントロールして推定をしている。推定値の標準誤差は括弧の中で示しており、それぞれ学校レベルでクラスター化されている。「\*\*\*」「\*\*」「\*」という表記はそれぞれ0.1%、1%、5%の水準で有意であることを表している。

表 5.3: Effect of Treatment: IQ

Dependent Variable	IQ			Draw A Man		
	Model1	Model2	Model3	Model1	Model2	Model3
Treatment	0.705*** (0.117)	0.692*** (0.111)	0.664*** (0.111)	0.071 (0.080)	0.022 (0.097)	-0.003 (0.105)
Baseline Score		✓	✓	✓	✓	✓
Control			✓	✓		✓
Observations	1404	1404	1146	1390	1390	1133
Adjusted R2	0.076	0.4	0.51	0.001	0.182	0.287

注) 表では介入効果の推定値を示している。データの単位は各個人。「モデル1」列ではOLSによって得られた推定値を示している。「モデル2」列では介入前の学力およびベースライン不参加タミーをコントロールしている。「モデル3」列ではモデル1に加えて性別や学年、生まれ月、親の教育レベル、学校年固定効果をコントロールして推定をしている。推定値の標準誤差は括弧の中で示しており、それぞれ学校レベルでクラスター化されている。「\*\*\*」「\*\*」「\*」という表記はそれぞれ0.1%、1%、5%の水準で有意であることを表している。

表 5.4: Effect of Treatment: Non-cognitive Skills

Dependent Variable	Motivation		Self Esteem	
	Model1	Model2	Model1	Model2
Treatment	-0.023 (0.070)	-0.031 (0.069)	0.023 (0.052)	0.014 (0.059)
Baseline Score	✓	✓	✓	✓
Control		✓		✓
Observations	1402	1125	1396	1121
Adjusted R2	0.274	0.377	0.025	0.138

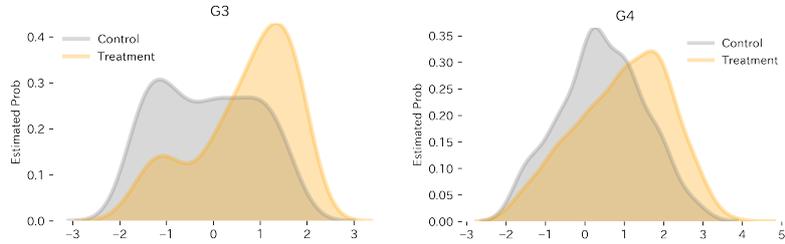
注) 表では介入効果の推定値を示している。データの単位は各個人。「モデル1」列では介入前の学力およびベースライン不参加ダミーをコントロールしている、「モデル2」列ではモデル1に加えて性別や学年、生まれ月、親の教育レベル、学校学年固定効果をコントロールして推定をしている。推定値の標準誤差は括弧の中で示しており、それぞれ学校レベルでクラスター化されている。「\*\*\*」「\*\*」「\*」という表記はそれぞれ0.1%、1%、5%の水準で有意であることを表している。

表 5.5: Effect of Treatment: Study Input

Dependent Variable	Study Time(minutes)		Willigness to go to college	
	Model1	Model2	Model1	Model2
Treatment	-0.032 (0.097)	-0.099 (0.101)	0.136* (0.073)	0.139* (0.083)
Baseline Score	✓	✓	✓	✓
Control		✓		✓
Observations	1299	1057	1367	1094
Adjusted R2	0.050	0.090	0.033	0.048

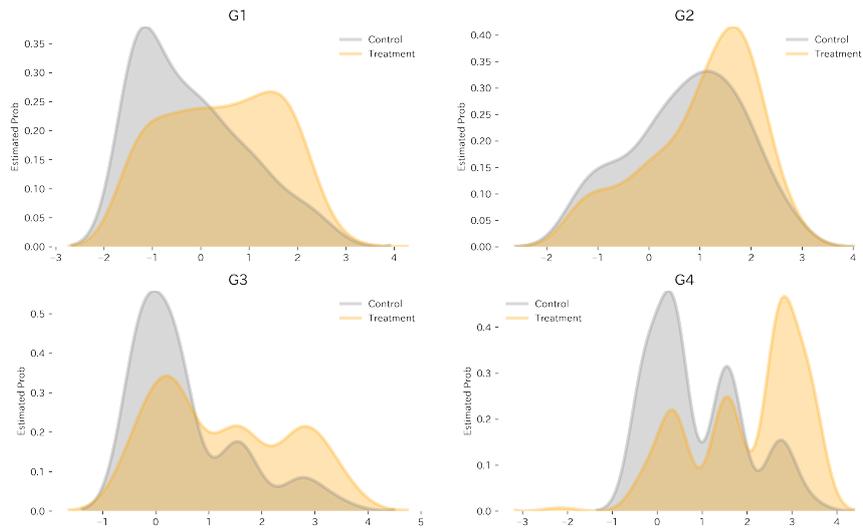
注) 表では介入効果の推定値を示している。データの単位は各個人。「モデル1」列では介入前の学力およびベースライン不参加ダミーをコントロールしている、「モデル2」列ではモデル1に加えて性別や学年、生まれ月、親の教育レベル、学校学年固定効果をコントロールして推定をしている。推定値の標準誤差は括弧の中で示しており、それぞれ学校レベルでクラスター化されている。「\*\*\*」「\*\*」「\*」という表記はそれぞれ0.1%、1%、5%の水準で有意であることを表している。

図 5.4: National Assessment Score and TIMSS



注) この図はフォローアップサーベイにおける標準学力テストおよび TIMSS のスコアの推定された確率密度関数を示している。オレンジ色で塗られた関数はトリートメントグループを表しており、灰色で塗られた関数はコントロールグループを示している。

図 5.5: IQ  
IQ(End-line)

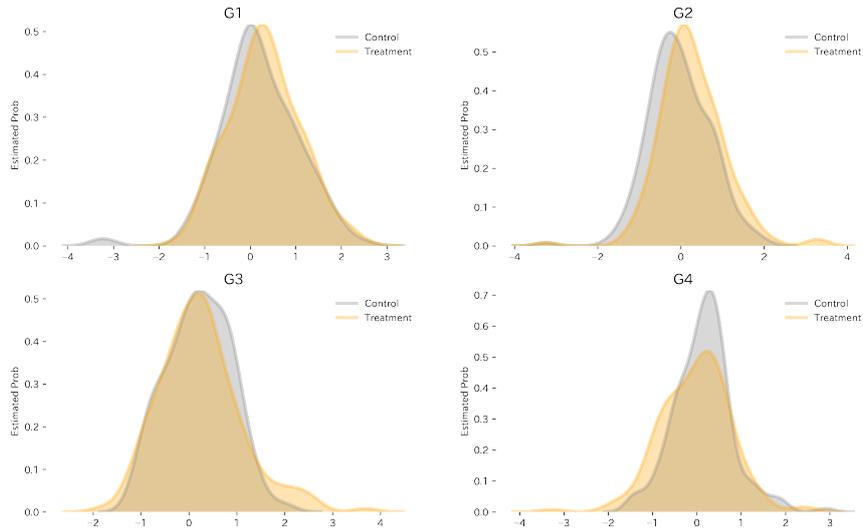


注) この図はフォローアップサーベイにおける IQ スコアの推定された確率密度関数を示している。オレンジ色で塗られた関数はトリートメントグループを表しており、灰色で塗られた関数はコントロールグループを示している。

## 5.5 結論

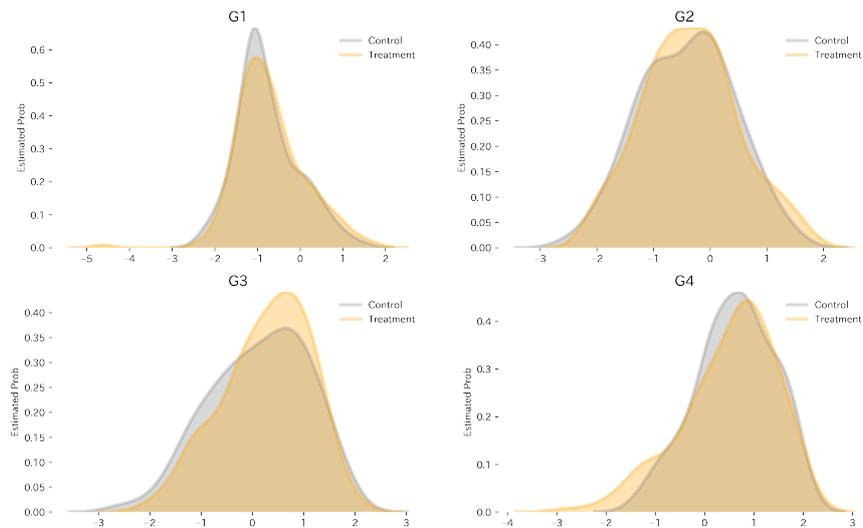
本稿では CAI が子どもの認知能力と非認知能力に与える因果的な効果を検討した。カンボジア政府と共同でプノンベン周辺の 5 つの小学校で 3 ヶ月間クラスター RCT を実施し、子どもは介入期間中に通常の算数の授業の代わりに CAI を使用することを許可された 20 のトリートメントクラスと、20 のコントロールクラスのいずれかに無作為に割り付けられた。我々の結果は、いく

図 5.6: Draw a Man Test  
Draw-a-Man(End-line)



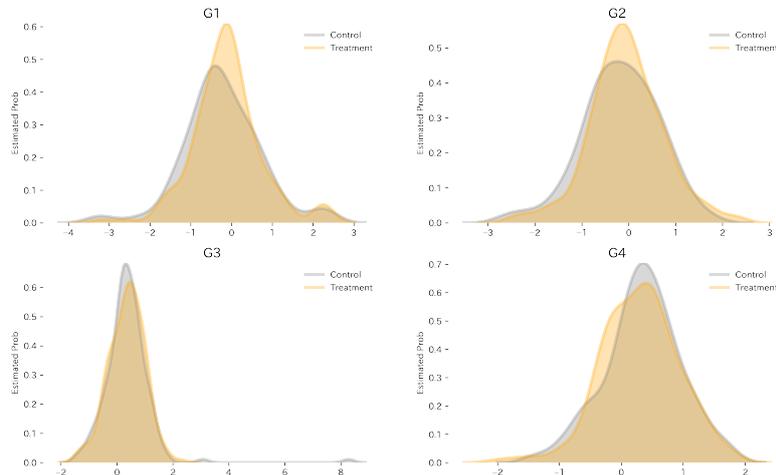
注) この図はフォローアップサーベイにおける Draw a man テストスコアの推定された確率密度関数を示している。オレンジ色で塗られた関数はトリートメントグループを表しており、灰色で塗られた関数はコントロールグループを示している。

図 5.7: Motivation  
Motivation(End-line)



注) この図はフォローアップサーベイにおけるモチベーションスコアの推定された確率密度関数を示している。オレンジ色で塗られた関数はトリートメントグループを表しており、灰色で塗られた関数はコントロールグループを示している。

図 5.8: Self-Esteem  
Self-esteem(End-line)



注) この図はフォローアップサーベイにおける自尊心の推定された確率密度関数を示している。オレンジ色で塗られた関数はトリートメントグループを表しており、灰色で塗られた関数はコントロールグループを示している。

つかのタイプの数学についての達成度テストとIQテストによって測定された認知能力に対する平均処置効果が正で統計的に有意であることを示している。効果の大きさは、特に発展途上国を対象とした先行研究と比較して大きい。生徒の学力に対する推定係数は0.68~0.77標準偏差、人口統計学的要因を考慮した場合でもIQスコアは0.66標準偏差であった。さらに、CAIは学生の将来の大学進学に対する主観的な期待感を高めることがわかった。しかし、モチベーションや自尊心などの非認知的能力には有意な効果は見られなかった。

このクラスター化RCTは3ヶ月間しか実施されていないため、これらの効果が長期的に持続するかどうかについてはさらなる調査が必要である。しかし、我々の結果はCAIが短期的にも長期的にも生徒の数学の成績を向上させる可能性があることを示唆している。

## 関連図書

Angrist, Joshua D and Jörn-Steffen Pischke (2008) *Mostly harmless econometrics: An empiricist's companion*: Princeton university press.

Banerjee, Abhijit V, Shawn Cole, Esther Duflo, and Leigh Linden (2007) "Remedying education: Evidence from two randomized experiments in India", *The Quarterly Journal of Economics*, **122** (3), pp. 1235-1264.

- Barrera-Osorio, Felipe and Leigh L Linden (2009) “The use and misuse of computers in education: Evidence from a randomized controlled trial of a language arts program”, *Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL)*. [www.leighlinden.com/Barrera-Linden](http://www.leighlinden.com/Barrera-Linden), **20**.
- Barrow, Lisa, Lisa Markman, and Cecilia Elena Rouse (2009) “Technology’s edge: The educational benefits of computer-aided instruction”, *American Economic Journal: Economic Policy*, **1** (1), pp. 52–74.
- Bruhn, Miriam and David McKenzie (2009) “In pursuit of balance: Randomization in practice in development field experiments”, *American Economic Journal: Applied Economics*, **1** (4), pp. 200–232.
- Bulman, George and Robert W Fairlie (2016) “Technology and education: Computers, software, and the internet”, in *Handbook of the Economics of Education*, **5**: Elsevier, pp. 239–280.
- Carrillo, Paul E, Mercedes Onofa, and Juan Ponce (2011) “Information technology and student achievement: Evidence from a randomized experiment in Ecuador”, IDB Working Paper No. 78.
- Cristia, Julian, Pablo Ibararán, Santiago Cueto, Ana Santiago, and Eugenio Severin (2017) “Technology and child development: Evidence from the one laptop per child program”, *American Economic Journal: Applied Economics*, **9** (3), pp. 295–320.
- Melo, Gioia De, Alina Machado, and Alfonso Miranda (2014) “The impact of a one laptop per child program on learning: Evidence from Uruguay”, *IZA Discussion paper*.
- Glewwe, Paul and Karthik Muralidharan (2016) “Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications”, in *Handbook of the Economics of Education*, **5**: Elsevier, pp. 653–743.
- Goodenough, Florence Laura (1926) *Measurement of intelligence by drawings.*: World Book Co.
- Hanushek, Eric A and Dennis D Kimko (2000) “Schooling, labor-force quality, and the growth of nations”, *American Economic Review*, **90** (5), pp. 1184–1208.
- and Ludger Woessmann (2016) “Knowledge capital, growth, and the East Asian miracle”, *Science*, **351** (6271), pp. 344–345.

- Imbens, Guido W and Jeffrey M Wooldridge (2009) “Recent developments in the econometrics of program evaluation”, *Journal of economic literature*, **47** (1), pp. 5–86.
- Ito, Hirotake, Keiko Kasai, Hiromu Nishiuchi, and Makiko Nakamuro (2021) “Does Computer-Aided Instruction Improve Children’s Cognitive and Noncognitive Skills?”, *Asian Development Review*, **38** (1), pp. 98–118.
- Jamison, Eliot A, Dean T Jamison, and Eric A Hanushek (2007) “The effects of education quality on income growth and mortality decline”, *Economics of Education Review*, **26** (6), pp. 771–788.
- Joensen, Juanna Schrøter and Helena Skyt Nielsen (2009) “Is there a causal effect of high school math on labor market outcomes?”, *Journal of Human Resources*, **44** (1), pp. 171–198.
- Linden, Leigh L (2008) “Complement or substitute?: The effect of technology on student achievement in India”, Working Paper, Columbia University: InfoDev.
- Muralidharan, Karthik, Abhijeet Singh, and Alejandro J Ganimian (2019) “Disrupting education? Experimental evidence on technology-aided instruction in India”, *American Economic Review*, **109** (4), pp. 1426–1460.
- Rosenberg, Morris (1965) *Society and the adolescent self-image*: Princeton university press.
- Rouse, Cecilia Elena and Alan B Krueger (2004) “Putting computerized instruction to the test: a randomized evaluation of a ”scientifically based” reading program”, *Economics of Education Review*, **23** (4), pp. 323–338.
- Sakurai, Shigeo and Seijun Takano (1985) “A new self-report scale of intrinsic versus extrinsic motivation toward: learning for children”, *Tsukuba Psychological Research*, **7**, pp. 43–54.
- Scott, Linda H (1981) “Measuring intelligence with the Goodenough-Harris drawing test”, *Psychological Bulletin*, **89**, p. 483.
- Tanaka, Kanichi, Kenroku Okamoto, and Hidehiko Tanaka (2003) *The New Tanaka B Intelligence Scale*: Kaneko shobo.
- World Bank (2017) *World Development Report 2018: Learning to Realize Education’s Promise*: The World Bank.

## 第6章 おわりに

子どもの能力を向上させることができる教員はどのような教員かという問いが本稿を通じて探求された問いであった。様々な教育政策が教員の資質向上を謳って施策を実行に移してきたが、第1章や第2章で指摘したように、学術的には必ずしもそれらの施策を裏付けるような一貫した研究成果が存在するわけではない。しかしそれは必ずしも、教員の子どもに対する影響力が存在しないことを意味しない。むしろ、いくつかの研究はそのような教員の影響力の存在を指摘しており、さらにその影響力は因果的な効果をも含んでいることが分かっている。そのような中で、本稿では日本のデータを用いながら改めて上述の問いを再検討し、日本における教員の有効性について定量的なエビデンスを提示した。

第3章で扱ったのは教員の違いが子どもの能力に与える影響についてであった。この問題は教員付加価値研究として定式化することができ、本稿では教員付加価値を測定しその記述的な分析を行った。その結果として次の3点を確認することができた。第1に、教員の違いは子どもの非認知能力に対して大きな影響を与えていた。第2に、認知能力を伸ばすことが得意な教員と非認知能力を伸ばすことが得意な教員は異なっていた。第3に、観察可能な教員の特徴と教員付加価値の間には相関はなかった。ただし、これらの分析で扱っている教員付加価値とは年齢や性別といった教員の特徴を全て合わせた個々のそれぞれの教員の総体的な影響力を示したものであり、具体的に「どの要因が効果的か」という問いに答えるものではなかった。

そこで第4章では、「どの要因が効果的か」という問いに対する一つの仮説として教員が学級で用いる指導方法が持つ可能性を提出した。指導方法についての研究は今の所世界的にも始まったばかりであり、その研究成果が十分に蓄積されているとは言い難い。そういった中で本稿では、先行研究が抱える課題を解決しながら、日本において指導方法の違いが子どもの学力と関係があるかどうかを検討したのであった。その結果、いくつかの指導方法インデックスが小学生の子どもの学力に正の影響を与えていることを示した。ただし、ここで分析をしたのは指導方法インデックであり、具体的な指導方法そのものではない。またその分析対象も学力(認知能力)に限られるものであり、非認知能力など学力以外のアウトカムに対する影響について議論をすることはできていなかった。これは本稿にとどまらず、観察データを用いた教員の指導方法研究が一般に抱える問題である。

第4章で立ち上がった指導方法の効果の分析の限界を乗り越えるためには、RCT や準実験的な状況を利用した分析が必要になる。そのような分析の例として、第5章では具体的な指導方法としてコンピューター支援教育を検討し、実験的な状況を利用して分析を行った<sup>1</sup>。ここでは、教員による授業に代わって教育アプリを用いて行う授業を考える。そのような教育アプリを用いて授業を行うという実証実験をカンボジアで実施をし、その結果を分析した。その結果として、コンピューター支援教育が子どもの学力に正の影響を与えていることを示した。この第5章で扱うことができたのはコンピューター支援教育という指導方法の1手法にとどまるが、今後の研究においては同様水準の分析が求められる。

これらの結果をまとめると次のようになる。教員の違いは子どもの非認知能力の成長に大きな影響を与えていた一方で、学力などの認知能力にたいしては必ずしもその大きさの程度は自明ではなかった。しかしその中でも、学級における指導方法の違いは子どもの学力にも違いをもたらしていた。そしてコンピューター支援教育はそういった効果的な指導方法の候補の1つになりうる。

本稿で明らかにすることができたのは教員の有効性研究のごく一部の分野であり、将来的に解決する必要がある課題を多く抱えている。例えば、教員付加価値研究として本稿では教員の因果的な効果を分析をすることはできていない。指導方法についての研究としては、本稿では指導方法のインデックスを取り扱ったのみで、個々具体的な指導方法の有効性を分析したわけではない。コンピューター支援教育の効果の分析としては、カンボジアにおける実証実験を通じた分析であり、その成果が日本においても適用可能かどうかについてはさらなる議論が必要である。

しかし、本稿は日本における教員の有効性研究として数少ない定量的なエビデンスを提示することができた。教員制度は国によって違うため、教員や学校の有効性のありかたもまた国によって違うと考えられる。それにも関わらず、第2章で指摘したように、日本のデータを用いて教員の有効性を定量的に研究した例は非常に少ない。そのような中で本稿が日本に教員についての教育施策にもたらす含意は決して少なくない。例えば第3章の教員付加価値研究からは、どのような教員を雇用すべきかについての指針を得ることが出来るだろう。無論、本稿で扱うことができなかつたトピックも数多く存在する。しかしいづれにしてもとりわけ大事なのは、これまでなされてきた膨大な教員研究を踏まえた上で、エビデンスを一つ一つ蓄積していくことである。そのような中で初めて、様々な教育施策・教育政策を根拠を持って実施する道が拓かれると言えよう。

---

<sup>1</sup>一方で筆者が関わった研究で準実験的な状況を利用した分析として、学級規模が子どもの能力に与える影響を分析した Ito et al. (2020) を挙げることができる。

## 関連図書

Ito, Hirotake, Makiko Nakamuro, and Shintaro Yamaguchi (2020) “Effects of class-size reduction on cognitive and non-cognitive skills”, *Japan and the World Economy*, **53**, p. 100977.

## 謝辞

本研究を進めるにあたってご指導と激励を賜りました指導教官である慶應義塾大学総合政策学部の中室牧子教授に心から感謝の意を表します。民間企業の社員であった私に研究指導のみならず研究者の心得等から一から指導をすることは並大抵のことではなかったと存じます。しかし、時に深夜にご連絡することになった時であっても、親身になって相談に乗って頂きご指導頂いたことで、なんとか博士課程を完走できたのではないかと確信しております。

また快く副査を引き受けてくださった秋山美紀教授・鈴木寛教授・和田龍磨教授にも深く御礼申し上げます。コロナウイルス流行に伴って対面でのコミュニケーションなどが大きく制限される中で、博士論文の審査頂くことには多大な困難があったことかと存じます。そういった中で、お忙しい時間の一部を割いて頂いたことに、心より感謝しております。

そして最後に、不安定なキャリアを凶らずも歩んでしまっている私をいつも見守ってくれている家族や友人に感謝の意を表します。博士課程を通じて（もしくはそれ以前から）順調とは言えないキャリアの中で、皆様が温かい目で応援してくださったことが大変励みになりました。

2021年3月  
伊藤寛武