RIETI Discussion Paper Series 25-E-111

The Effect of Subject-area Specialization on Student Achievement: Evidence from a cluster-randomized experiment in elementary schools

IGEI, Kengo
Keio University

NAKAMURO, Makiko
RIETI

MURAKAWA, Tomoya
Harvard University

LE, Chien Quang
Embassy of Japan in Vietnam

Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 25-E-111

December 2025

The Effect of Subject-area Specialization on Student Achievement: Evidence from a cluster-randomized experiment in elementary schools*

Kengo IGEI (Keio University)

Makiko NAKAMURO (Keio University/RIETI)

Tomoya MURAKAWA (Harvard University)

Chien Quang LE (Embassy of Japan in Vietnam)

Abstract

This study estimates the causal impact of deploying part-time subject-specialist teachers in elementary schools on students' academic outcomes, drawing on a cluster-randomized controlled trial conducted in Chiba Prefecture, Japan. In schools randomly assigned to receive part-time science specialists, students' science achievement increased by 0.153–0.162 standard deviations (SD), which is a relatively large effect compared to many other educational interventions such as class size reduction. Moreover, mathematics achievement improved by 0.101–0.108 SD, while Japanese language achievement remained unaffected. In contrast, the deployment of part-time mathematics specialists had no statistically significant effect on student performance. There is no evidence that the introduction of part-time subject-specialists altered teachers' classroom preparation time for other subjects. Science is a subject that demands a high level of content expertise, and prior studies indicate that as teachers gain more experience, their anxiety in teaching science decreases, while their self-efficacy increases. Given that the assigned specialists were relatively older and experienced part-time teachers, the results suggest that leveraging their expertise and confidence may have contributed to the observed academic gains. The findings highlight the potential of strategically utilizing experienced specialist teachers to improve science education in the upper grades of elementary school.

Keywords: Cluster Randomized Controlled Trial, Subject-Specialist Teaching System, Teacher Working Hours

JEL Classification: I0, I2

* This study is conducted as a part of the Project "Comprehensive study to promote Evidence-Based Policy Making (EBPM) in Japan" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This paper is part of the outcomes from the commissioned research project by the Chiba Prefectural Board of Education, titled "Study on the Analysis of Effectiveness Verification Data for the Deployment of Part-Time Subject-Specialist Teachers in Elementary Schools." An earlier draft of this paper was presented at the RIETI DP seminar for the paper. We are grateful for the valuable comments received from Fumio Ohtake and other participants of the meeting. We also wish to express our sincere gratitude to the Chiba Prefectural Board of Education for their generous support in the preparation of this paper. This research was conducted with the approval of the SFC Research Ethics Committee (Approval No. 475, approved on March 6, 2023).

1 Introduction

In recent years, the shortage of teachers and the problem of excessive working hours in Japanese primary schools have become urgent policy concerns. In response, the government has implemented a range of reforms aimed at improving teachers'working conditions and securing a sustainable teaching workforce. Among these reforms, the introduction of subject-specialist teachers has attracted particular attention.

Traditionally, Japanese primary schools have operated under a *homeroom-based system*, in which a single teacher is responsible for teaching all subjects to one class. Since the 2022 academic year, however, subject specialization has been introduced in the upper grades (Grades 5 and 6) of public primary schools. Because subject specialization has long been the norm in middle schools, this reform was also expected to mitigate the so-called "seventh-grade gap"—a phenomenon in which students experience adjustment difficulties and increased absenteeism due to the sharp pedagogical discontinuity between primary and middle school.

Although the initial implementation of subject specialization targeted upper grades, policymakers have begun to discuss its possible expansion to Grades 3 and 4. In its August 2024 report, Central Council for Education, MEXT proposed extending subject specialization to Grade 3 and 4 classrooms [1].

Grade 4 students in Japan receive approximately 1,015 hours of instruction annually—roughly equivalent to the instructional time in upper elementary grades and even in middle school. To alleviate the substantial workload placed on teachers under the traditional homeroom-based model, some schools have begun piloting departmentalized teaching in Grade 4. Consistent with the CCE's recommendations, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) announced on December 24, 2024, its plan to expand the use of subject-specialist teachers at this level. MEXT subsequently reached an agreement with the Ministry of Finance on the fiscal year 2025 budget, and corresponding updates to relevant regulations and curriculum guidelines are ex-

[1] Foreign language, science, math, and physical education are identified as priority subjects for subject-specialist instruction.

pected to follow.

While the government has strongly promoted subject specialization, its impact on students'academic performance and non-cognitive skills remains underexplored. Previous studies have reported mixed findings, suggesting the need for a rigorous causal evaluation in the Japanese context.

Proponents of subject specialization argue that reducing the number of subjects allocated to each teacher allows for deeper preparation, repeated teaching opportunities, and ultimately higher instructional quality (Chan & Jarman, 2004). Critics, however, contend that subject specialization increases the number of students per teacher, making it more difficult to monitor individual progress (Anderson, 1962). Others have pointed to potential losses in instructional coherence (McGrath & Rust, 2002) and to difficulties integrating concepts across subjects, which may hinder deeper understanding (Bastian & Fortner, 2020).

Empirical evidence is similarly mixed. For instance, Fryer Jr. (2018) conducted a randomized controlled trial (RCT) in Texas elementary schools and found that departmentalized instruction lowered student achievement and increased disciplinary problems and absenteeism. In contrast, more recent quasi-experimental studies in the United States suggest that the effects may depend on subject area, teacher experience, and the implementation model (Hwang and Kisida, 2022; Backes et al., 2024). Yet, to our knowledge, no prior research has rigorously examined this issue in Japan, a context where teaching practices, teacher assignment, and professional norms differ substantially from those in Western settings.

This study addresses that gap by utilizing field experiment data from public primary schools in Chiba Prefecture, Japan, to evaluate the impact of introducing subject-specialist teachers on both academic performance and non-cognitive skills. Unlike a complete shift to a departmentalized teaching model, the intervention in this study retained the homeroom system while assigning part-time specialist teachers specifically for mathematics and science classes. We conducted an RCT across 60 public elementary schools identified as relatively low-performing within the prefecture. Of these, 20 schools were randomly selected to receive part-time mathematics specialists, another 20 to receive part-time science specialists, while the remaining 20 schools served as the control group.

Introducing science specialist teachers led to an increase of 0.178 standard deviations (SD) in students' science test scores with statistical significance. Moreover, mathematics achievement improved by 0.127 SD with statistical significance, which suggests possible spillover effects from enhanced science instruction. Positive effects were observed in students' learning strategies and non-cognitive skills as well. In contrast, no significant changes were found in either subject in schools where mathematics specialists were introduced, and no spillover effects on Japanese language achievement was observed in either intervention. Placebo tests using data from fifth grade students, who were not part of the intervention, further confirmed that the observed effects were unlikely to be coincidental.

The intervention appears to be highly cost-effective as well. An analysis of fifth grade data shows that implementing subject-specialist teaching only through sorting teachers within a school and without adding part-time specialists did not improve student achievement. Also, consistent with existing literature, a negative impact on performance in mathematics was observed. A cost-effectiveness comparison showed that raising science achievement by 1 SD per student through introducing part-time science specialists cost 79,413 JPY (539 USD). Conversely, it cost 966,446 JPY (6,555 USD) to introduce a smaller class size policy (see Appendix B.2 for calculation details).

One possible explanation for strong effects only appearing in the science subject-specialist intervention is the subject's higher degree of specialization, and the fact that many teachers experience lower self-efficacy and higher anxiety when teaching science. Prior research indicates that such anxiety and low efficacy tend to be mitigated with experience (Nakajima and Kusaka, 2020), suggesting that the intervention's reliance on older, more experienced part-time teachers may have been advantageous. Notably, homeroom teachers did not increase their lesson preparation time in other subjects. However, teachers'growth mindset improved following the intervention.

This paper proceed as the follows. Chapter 2 reviews domestic and international literature on subject-specialist teaching and related economic literature. Chapter 3 describes the details of the intervention. Chapters 4 and 5 explains the data and empirical strategy. Chapter 6 reports the effects of the intervention on student achievement and teacher work practices. Chapter 7 concludes with policy implications. This study provides insights for refining education policy for Chiba Prefecture, and broader education strategies. Chiba Prefecture's approach underscores the importance of evidence-based policymaking and offers critical insights into how education systems can be updated.

2 Literature Review

Many studies conducted overseas have argued that, in theory, departmentalization and the implementation of subject-specialist teachers could enhance instructional productivity by exploiting teachers' comparative advantages. Empirical findings, however, reach an opposite conclusion, particularly at the primary education level. Improvements in student achievement are rarely observed, and in some cases, negative effects have been documented (Fryer Jr., 2018; Bastian and Fortner, 2020; Hwang and Kisida, 2022). Among the existing studies, Fryer Jr. (2018) is particularly relevant to the present research. The author conducts a cluster RCT in low-performing public elementary schools in Houston, Texas. In 23 randomly selected schools, a departmentalized system was introduced, while the remaining 23 schools continued with the conventional homeroom-based model. After one year, students in the treatment schools scored 0.12 (SD) lower than their peers in the control group, and this negative effect persisted into the succeeding year. The impact was especially large among students in special education programs and those taught by inexperienced teachers. Furthermore, the incidence of major behavioral infractions increased by a factor of 1.13, and students attended 0.36 fewer days per year on average. Despite teaching fewer subjects, teachers were responsible for a larger number of students, making it more difficult to tailor instruction to individual learning needs. Teacher satisfaction in the treatment group was also significantly lower than in the control group.

Similarly, Hwang and Kisida (2022) showed that the increase in the number of students per teacher under departmentalization led to weaker teacher–student relationships, resulting in negative effects on student achievement. Once again, the most severe impacts were observed among students from disadvantaged backgrounds, such as those with low prior achievement or from low-income households. However, the study found no significant effects on overall school level academic performance, absenteeism, or disciplinary incidents, suggesting that any positive effects for some groups may have been offset by negative effects for vulnerable students, leading to negligible net outcomes.

In contrast, Backes et al. (2024) reported positive effects of departmentalization on student achievement. Using eight years of administrative data from elementary schools in Massachusetts, they found a gain of 0.03–0.07 SD in English language arts (ELA), 0.04–0.06 SD in science, and 0.00–0.04 SD in mathematics. The authors attributed these improvements to a key feature of their reform. Teachers were assigned to subjects that aligned with their strengths, enabling gains through better congruence with teachers' comparative advantages. In other words, the benefits of departmentalization depend less on reducing the number of subjects taught by each teacher, and more on creating environments where each teacher can leverage their own strengths.[2] Nevertheless, negative effects on student adjustment and teacher-student attachment were observed. Survey data indicated that students taught under the departmentalized model reported lower ratings of teacher–student relationships and learning environments.

The effects of departmentalization also have been found to vary by subjects and grades. Bastian and Fortner (2020) analyzed the impact of departmentalization on test scores from approximately 660,000 fourth and fifth grade students, and data from 50,000 teachers in North Carolina. They found negative effects on mathematics and reading achievement but positive effects of 0.04–0.06 SD in science. Moreover, the negative impacts were stronger among younger students. Additionally, Backes et al. (2025) found that departmentalization in elementary school had beneficial effects on academic performance and adaptation to middle school using data from Massachusetts. By exploiting variations in the timing of departmentalization across grades within the same schools, the

[2] A key difference between Fryer Jr. (2018) and Backes et al. (2024) lies in their institutional design. In the former, departmentalization was implemented within individual grade levels, whereas the latter study reassigned teachers across the entire school level, which may account for the difference in outcomes.

study found that each additional year of exposure to departmentalization in elementary school was associated with increases of 0.01 SD in mathematics and ELA and 0.02 SD in science in middle school. The effects were largest immediately after transition to middle school but persisted through grade 8 in science, even as they faded in mathematics and ELA.

Departmentalization also has significant impact on teachers. Fryer Jr. (2018) reports that teachers in the departmentalized schools taught 23% more students than those in homeroom systems. The increased teaching load across multiple classes made it more difficult to engage deeply with individual students, leading to declines in both teacher satisfaction and instructional quality. On the other hand, Bastian et al. (2023) reported that departmentalization reduced teacher attrition, particularly among teachers working in urban or suburban schools (by 2.7 percentage points) and among Black teachers (by 5.3 percentage points).

Overall, the literature in overseas settings offers a mixed set of findings on the impact of departmentalization on student achievement, the persistence of its effects, subject and grade level heterogeneity, and teacher outcomes. A plausible explanation is that potential productivity gains from leveraging teachers'comparative advantages are offset by negative consequences, such as lower student adjustment and weaker teacher–student attachment. These trade-offs underscore the importance of institutional design for the successful implementation of departmentalization, particularly in the early years of schooling.

How has departmentalization been evaluated in Japan? As noted above, MEXT announced plans to implement departmentalization nationwide for upper elementary grades beginning in the 2022 academic year, and several municipalities have adopted the system (see Appendix B.1 for details). A number of domestic surveys have examined the effects of departmentalization on student achievement and teacher workload.

Some findings are positive. Departmentalization allows teachers more time to plan lessons and develop pedagogically, which potentially enhances instructional quality and leads to improvements in student comprehension and motivation. For example, a survey conducted in Oita Prefecture reported that after introducing a subject-exchange model of departmentalization, a large portion of students agreed with statements such as “I like this subject”and “I understand the content better” (Oita Prefectural Board of Education, Outcomes and Challenges of Departmentalization, 2022). Similarly, Tokoro et al. (2022), conducted surveys for students and teachers in Iwate Prefecture, and found that both students and teachers believed departmentalization contributed to higher academic achievement and increased motivation due to the novelty of being taught by different teachers for each subject. Asada and Nakanishi (2018) surveyed elementary schools in Yamaguchi Prefecture, and reported that students taught under departmentalization expressed significantly lower levels of anxiety about learning and teachers upon entering middle school, suggesting that departmentalization may help mitigate the “seventh-grade gap.”

On the other hand, some studies report negative results. Kajita and Doho (2022) conducted a survey of principals and teachers at schools selected for departmentalization and received additional staffing support. Respondents generally viewed departmentalization positively in terms of increasing student motivation, providing emotional security through exposure to multiple teachers, and enhancing instruction through specialization. On the other hand, many were skeptical about its potential to reduce teacher’s workload. Similarly, Tokoro et al. (2022) found that over half of surveyed teachers expressed anxiety about teaching in other classrooms and managing students, and that subjects requiring substantial preparation, such as science, could actually increase workload. Importantly, most of these evaluations are based on subjective perceptions from students and teachers, without comparison to the counterfactual of schools that did not implement departmentalization. Above all, the lack of quantitative evidence remains a critical gap in Japanese literature.

3 Setting and Research Design

3.1 Overview of Departmentalization

In this study, we conducted a cluster-randomized controlled trial and linked the intervention to Chiba Prefecture’s academic assessment data to rigorously estimate the causal effects of departmentalization on students' educational outcomes.

We begin by describing the institutional design of departmentalization in Chiba Prefecture. The model implemented in Chiba is not a simple arrangement in which a teacher within a school or grade teaches only one subject. Because Japan faces demographic decline and a shortage of prospective teachers. Under these conditions, it is crucial to adopt a form of departmentalization that is both feasible and is expected to be effective. Drawing on insights from studies conducted overseas, the Japanese system was designed with two features intended to raise student achievement.

First, we retained the conventional homeroom-based system and assigned part-time subject-specialists for selected subjects. Studies such as Fryer Jr. (2018) suggest that a straightforward form of departmentalization may adversely affect individual students'achievement. Especially for younger pupils, the homeroom model plays an important role in promoting school adjustment, sustaining teacher–student attachment, facilitating instruction tailored to individual proficiency, and maintaining teacher satisfaction. Therefore, our design therefore sought to combine the strengths of both approaches: keeping the homeroom structure as the default while deploying additional part-time specialists in subjects that demand higher levels of content expertise.

Second, we employ individuals with prior teaching experience, primarily retired teachers. Previous research indicates that departmentalization staffed by inexperienced teachers can have larger negative effects on achievement (Fryer Jr., 2018) and that the model is more likely to be adopted when teacher quality is comparatively low (Hwang and Kisida, 2022). Hence, we consider that ensuring teacher quality is essential when introducing departmentalization. However, the applicant-to-position ratio for elementary school teacher recruitment in Chiba was only 1.9:1 in fiscal year 2024, making it difficult to hire teachers with both expertise and experience.

Therefore, we employed former teachers as part-time teachers. Recruitment explicitly targeted part-time subject-specialists in science and mathematics, and hiring prioritized candidates with demonstrated expertise in these subjects. The part-time teachers hired through this intervention had an average age of 55.76 years, and they taught an average of 9.68 classes per week. A single specialist could be assigned to multiple schools, multiple grades, and multiple classes. In total, 38 specialists accepted posts for fourth-grade classes, which form the basis of our analysis. Beyond potential gains in student learning, this assignment of additional teachers also was expected to reduce homeroom teachers'workloads and contribute to work-style reforms.

A related reference point is the United Kingdom's supply teacher system, which helps cover sudden staffing shortages due to absences or parental leave. Contracts range from single-day arrangements to term or year-long appointments and are typically brokered through private agencies. In Japan, MEXT is considering the establishment of a similar system at a regional level to employ retired teachers as part-time teachers to fill temporary vacancies (Yomiuri Shimbun, 2025).

In 2024, the UK Department for Education published a report summarizing the current state and challenges of the supply teacher system (Department for Education, 2024). According to the report, more than 98% of supply teachers hold a bachelor's degree or higher, over half are in their fifties, and many have prior experience teaching as full-time teachers. The system serves diverse needs, including veteran teachers seeking flexible, lower-burden positions and early-career teachers aiming to re-enter the profession or to build experience. At the same time, challenges are becoming evident. In a survey targeting principals, 64% reported that the supply teacher system has an adverse impact on student behavior and 61% reported negative impacts on learning outcomes. Consequently, whether in Chiba or nationwide, any consideration of a mechanism similar to the supply teacher system should be accompanied by rigorous quantitative evaluation to determine its effects on student outcomes and on teachers'working conditions.

3.2 Research Design

In selecting schools for the RCT, we leveraged data from the 2022 National Assessment of Academic Ability to identify public elementary schools in each of the five administrative districts in Chiba Prefecture with relatively low academic performance and that had no more than four classes of either third and fourth grade students in the 2023 academic year. From among these schools, we randomly selected, within each district, four schools to receive a mathematics specialist teacher (mathematics schools), four schools to receive a science specialist teacher (science schools), and four schools to serve as a control group without specialist teacher deployment (control schools).

This resulted in a total of 12 schools per district. Ultimately, 20 mathematics schools, 20 science schools, and 20 control schools were selected. The deployment of mathematics and science specialists occurred in separate sets of schools, and no school received specialists for both subjects simultaneously.

Deployment of part-time teachers began in April, 2023, however, only 55.0% of schools had their assigned specialist in place at the beginning of the academic year, which starts in April in Japan. The remaining schools received their assigned teachers after April in which part-time subject-specialists were successfully deployed in 13 mathematics schools (65.0%) and 15 science schools (75.0%) by the baseline testing. Eventually, part-time subject-specialists were assigned to 92.5% of mathematics schools and 97.5% of science schools by the end of the academic year.

4 Data

4.1 Standardized Achievement Tests and Surveys

This study draws on data from standardized achievement tests and surveys administered to fourth grade students in the 2023 academic year, as well as questionnaire responses from their homeroom teachers.[3] Standardized assessments in the Japanese language, mathematics, and science, along with student surveys were conducted during regular class hours. The student survey was the "Saitama Prefecture Academic Achievement and Learning Conditions Survey," which has been implemented in Saitama Prefecture. A distinctive feature of this survey is the inclusion of psychological scales, such as those measuring self-control, that allow for the assessment of students' non-cognitive skills. The teacher questionnaire was developed and administered by the research team. Surveys were distributed to homeroom teachers of the 111 classes included in the impact evaluation, with responses received from 91 teachers (response rate: 81.9%).

[3]Although part-time subject-specialists were also deployed in third grade classes, no appropriate standardized test was available for that grade level. Given the constraints, the impact evaluation focused solely on fourth graders. In addition, standardized achievement tests were administered to fifth graders for the purposes of conducting placebo tests and comparing the intervention's effects with other education policies, such as small-size classrooms.

The baseline survey was conducted in May 2023, just before the deployment of part-time subject-specialists, and the endline survey was administered one year later, in May 2024. It is important to note that data from some of the schools included in the evaluation could not be used. Some schools did not participate in the standardized tests due to internal circumstances, and some others mistakenly assigned different identifiers to students for the baseline and endline tests. As a result, the response rate for the baseline survey was 98.5%, and the response rate for the endline survey was 95.2%.[4] Table1 summarizes the number of schools for which data were available and the actual implementation status of part-time subject-specialists. Appendix Table A1 compares the average achievement scores of one mathematics school and one science school that dropped out after the baseline with those of the remaining schools. The results indicate that attrition had little to no impact on the analysis results.

Table 1: Status of Program Implementation

	Mathematics Schools（T1）	Science Schools（T2）	Control（C）
Number of Assigned Schools	20	20	20
Number of Classrooms （= Number of Homeroom Teachers）	40	36	35
Number of Schools Where Assignment Took Place by Baseline Testing	13	15	20
Number of Schools Where Assignment Was Completed by April	8	14	-
Number of Schools Responding to Baseline Survey	19	20	20
Number of Schools Responding to Endline Survey	18	19	20
Number of Students Enrolled at Time of Baseline Survey	1,021	852	913
Number of Students with Available Data at Baseline Survey	989	852	913
Number of Students with Available Data at Endline Survey	977	764	913

[4]Response rates were calculated by dividing the number of students who participated in the achievement tests by the total number of enrolled students as of April 2, 2023.

4.2 Instruments

The primary outcome variable in this study is students' academic outcomes measured through the standardized achievement test. The test scores used in mathematics, Japanese language, and science were used, which were normalized to have a mean of 0 and a variance of 1. Although subject-specialist teachers were deployed only in mathematics and science, Japanese language assessments also were conducted to test for possible spillover effects. We tested whether reductions in homeroom teachers' workloads due to an assigned specialist may have allowed more preparation time for other subjects and thereby improved student performance beyond mathematics and science.

The student survey includes measurements on learning strategies and non-cognitive skills. Learning strategies refer to deliberate actions taken by students to enhance the effectiveness of their learning. The survey measured five types of learning strategies: (1) flexible (adapting study methods to one's circumstances); (2) planning (systematic approaches to study); (3) task-oriented (learning through concrete activities such as note-taking); (4) cognitive (learning attitude aiming to deepen understanding); and (5) effort regulation (managing emotions, such as frustration to maintain motivation). These scales were developed by Sato and Arai (1998) and were computed based on students' responses to the survey. Higher scores indicate maturity with each learning strategy.

Non-cognitive skills were measured using a self-control scale and a self-efficacy scale, both measured through eight items. The self-control scale is a Japanese translation of Tsukayama et al. (2013), while the self-efficacy scale is a translation of the "Self-Efficacy for Learning and Performance" subscale from Pintrich et al. (1991). Higher scores indicate lower self-control and self-efficacy.

A separate teacher survey was administered through the Chiba Prefectural Board of Education in February 2024. This survey collected information on teachers' gender, age, educational background, years of teaching experience, working hours, lesson preparation time, perceived workload, instructional practices, mental health, and growth mindset. Variables, such as working hours, are impacted by the intervention, since the survey was conducted after the deployment of specialist teachers.

Perceived workload was measured using eight items (e.g., “The class size is too large”), with responses collected on a four-point Likert scale ranging from “strongly agree”to “strongly disagree.” These items match to those used in the teacher survey for the 2019 Trends in International Mathematics and Science Study (TIMSS). Instructional practices were assessed across three domains: (1) Lecture-based instruction – listening to teacher explanations of lesson content or problem-solving methods. (2) Pattern practice instruction – memorizing formulas or practicing problem-solving independently. (3) Active and dialogue-based instruction – engaging in class-wide activities under teacher guidance, presenting one's own ideas, applying previously learned concepts to new problems, and participating in peer discussions. Responses were collected on a four-point scale ranging from “always or almost always” to “never.” These items were also drawn from the 2019 TIMSS teacher questionnaire. Mental health was measured using the WHO-5 Well-Being Index. Respondents rated five items (e.g., “I felt cheerful and in good spirits”) on a six-point scale ranging from 5 (“at all times”) to 0 (“not at all”), with higher scores indicating better mental well-being. Finally, teachers'growth mindset, the belief that intellectual ability can be improved through effort and effective learning strategies, was measured using two statements adapted from Yeager et al. (2022): “To become a top-performing student in science or mathematics, special innate talent is required, and it is not something that can be developed through teaching.” and “People possess a certain fixed level of intelligence, and it cannot be significantly changed.” Responses were collected on a six-point scale from 6 (“strongly agree”) to 1 (“strongly disagree”). Previous research shows that students taught by teachers with a strong growth mindset demonstrate significantly greater improvements in mathematics achievement following educational interventions (Yeager et al., 2022).

4.3 Descriptive Statistics

Table 2 presents the baseline descriptive statistics for student-level variables. Column (1) reports the mean values for mathematics treatment schools, Column (2) for the science treatment schools, and Column (3) for the control group schools. Column (4) shows the mean differences between the mathematics treatment schools and the control schools, while Column (5) presents the differences between the science treatment schools and the control schools.

No statistically significant differences are observed in baseline test scores for science, mathematics, or Japanese language among the three groups. Similarly, there are no significant differences in student characteristics, learning strategies, or non-cognitive skills, indicating that the treatment and control groups are balanced prior to the deployment of specialist teachers.

Table 2: Student-level Descriptive Statistics and Baseline Balance

	(1) Mathematics Schools N		(2) Science Schools N		(3) Control Schools N		(1)-(3)	(2)-(3)
Science (standardized score)	983	0.138	854	0.239	910	0.154	-0.016	0.085
Mathematics (standardized score)	988	0.002	861	0.063	914	-0.013	0.015	0.076
Japanese (standardized score)	986	-0.037	861	0.077	913	0.002	-0.040	0.074
Learning Strategies: Flexible	975	0.033	854	0.031	897	0.038	-0.005	-0.007
Learning Strategies: Planning	974	0.029	857	0.007	896	0.022	0.007	-0.014
Learning Strategies: Task-Oriented	973	0.099	855	0.016	897	0.027	0.072	-0.011
Learning Strategies: Cognitive	975	-0.000	859	0.000	902	0.014	-0.015	-0.014
Learning Strategies: Effort Regulation	975	-0.006	858	0.023	897	0.029	-0.036	-0.006
Self-Efficacy	979	0.048	859	0.094	905	0.103	-0.055	-0.009
Self-Control	986	-0.032	859	0.040	901	-0.003	-0.028	0.043
Gender (Share Male)	997	0.488	869	0.503	920	0.467	0.021	0.035
Number of Students	20	51.050	20	42.600	20	45.650	5.400	-3.050
Number of 3rd Grade Classes	20	1.900	20	1.850	20	1.800	0.100	0.050
Number of 4th Grade Classes	20	1.950	20	1.650	20	1.650	0.300	0.000

Note: Columns (1)–(3) report the mean values of each variable for the mathematics treatment schools, science treatment schools, and control schools, respectively. Columns (1−3) and (2−3) show the mean differences between the mathematics treatment and control schools, and between the science treatment and control schools, respectively. ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively. In the tests of mean differences, we control for district fixed effects and the number of classes in each school, and use standard errors clustered at the school level.

5 Empirical Strategy

To assess the effect of deploying subject-specialist teachers on students' academic performance, we estimated a value-added (VA) model using students'test scores as the dependent variable. Our first step was to estimate the Intention-to-Treat (ITT) effect based on the following specification 1.

$$Y_{ist} = \alpha + \beta\, Y_{ist-1} + \tau_1\, D^{T1}_{is} + \tau_2\, D^{T2}_{is} + \boldsymbol{\gamma}\mathbf{X}_{is} + \varepsilon_{ist} \tag{1}$$

In specification 1, Y_{ist} denotes the academic achievement of student i in school s at time t (endline), while Y_{ist-1} represents the baseline achievement at time $t-1$. The variable D^{T1}_{is} is a dummy that equals 1 if school s is assigned a mathematics specialist teacher, and D^{T2}_{is} is a dummy that equals 1 if school s is assigned a science specialist teacher. The vector $\mathbf{X}_{is}$ includes a set of control variables such as student gender, class size, and total enrollment.

As shown in Table 1, specialist teachers were deployed in only 13 mathematics schools and 15 science schools by the baseline testing. Due to this partial compliance, we further estimated the Local Average Treatment Effect (LATE) by employing random assignment as an instrumental variable (IV). The estimation followed a standard two-stage least squares (2SLS) approach, with the first-stage equations specified as:

$$D^{T1}_{is} = \pi_1\, Z^{T1}_{is} + \pi_2\, Z^{T2}_{is} + \boldsymbol{\pi}_3\mathbf{X}_{is} + u_{is} \tag{2}$$

$$D^{T2}_{is} = \rho_1\, Z^{T1}_{is} + \rho_2\, Z^{T2}_{is} + \boldsymbol{\rho}_3\mathbf{X}_{is} + v_{is} \tag{3}$$

The second-stage equation is then given by:

$$Y_{ist} = \alpha + \beta\, Y_{ist-1} + \gamma_1\, \widehat{D}^{T1}_{is} + \gamma_2\, \widehat{D}^{T2}_{is} + \boldsymbol{\gamma}\mathbf{X}_{is} + \varepsilon_{ist}. \tag{4}$$

Here, Z_{is}^{T1}、 Z_{is}^{T2} are dummy variables indicating whether school s was randomly assigned to the mathematics or science treatment group, respectively. The terms D_i^{T1}, D_i^{T2} indicate whether a mathematics or science specialist teacher was actually deployed. By using random assignment as an instrument for actual deployment, the 2SLS estimation isolates the causal impact of the specialist teacher intervention under conditions of imperfect treatment compliance.

6 Results

6.1 ITT

Table 3 reports the results of the ITT estimation based on the VA model. All specifications include fixed effects for educational districts (Strata FE). Standard errors are clustered at the school level. The coefficient on D_{is}^{T2} is positive and statistically significant for both science and mathematics. This indicates that the deployment of science specialist teachers led to an improvement in students' academic performance in schools assigned to the science school treatment. Specifically, students in these schools experienced an increase of 0.166 SD in science test scores and 0.119 SD in mathematics test scores. The latter result suggests the presence of a spillover effect: improvements in science performance can enhance mathematics outcomes, a pattern documented in prior studies (e.g., Judson and Sawada, 2000). This implies that the assignment of science specialists not only boosted achievement in science but also indirectly contributed to gains in mathematics.

In addition, among the various learning strategies measured, *flexible strategies*, defined as students' ability to adapt their learning approaches to suit their circumstances, showed a statistically significant positive effect. No statistically significant changes were observed for non-cognitive skills. By contrast, in schools assigned a mathematics specialist (D_{is}^{T1}), there were no statistically significant effects on student achievement in mathematics, science, Japanese language, or on learning strategies or non-cognitive outcomes. The absence of a clear effect on mathematics achievement suggests that no spillover benefits were observed for science. Furthermore, in both science and mathematics treatment schools, there is no evidence of spillover effects on Japanese language achievement.

Table 3: ITT

	Academic Achievement			Learning Strategies					Non-cognitive Skills	
	Science	Math	Japanese Language	Flexible	Planning	Task-oriented	Cognitive	Effort Regulation	Self-Efficacy	Self-Control
D^{T1} (Math Treatment school)	-0.038	0.015	0.022	0.070	0.046	0.057	0.027	-0.063	-0.019	0.014
	(0.071)	(0.064)	(0.046)	(0.046)	(0.065)	(0.058)	(0.071)	(0.078)	(0.058)	(0.072)
D^{T2} (Science Treatment school)	0.166**	0.119*	0.052	0.103*	0.069	0.110	0.126	-0.043	0.032	-0.132
	(0.060)	(0.053)	(0.047)	(0.051)	(0.066)	(0.056)	(0.064)	(0.066)	(0.061)	(0.067)
Y_{it-1}	0.014***	0.722***	0.721***	0.421***	0.421***	0.428***	0.407***	0.449***	0.481***	0.241***
	(0.003)	(0.027)	(0.034)	(0.028)	(0.024)	(0.024)	(0.026)	(0.023)	(0.024)	(0.029)
Strata FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Covariates	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
N	2,395	2,394	2,402	2,334	2,334	2,333	2,203	2,335	2,350	2,353
R^2	0.096	0.513	0.531	0.183	0.184	0.200	0.167	0.203	0.238	0.098

Note: ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively. Standard errors are clustered at the school level. In addition to strata fixed effects, the regressions control for student gender, number of classes, and total student enrollment as covariates.

6.2 LATE

Table 4 presents the LATE estimates, which closely mirror the ITT results reported in Table 3. In schools assigned to the science school treatment, the deployment of specialist teachers led to an improvement of 0.178 SD in science test scores and an improvement of 0.127 SD in mathematics test scores. Moreover, students in the science school treatment exhibited a 0.110 SD increase in the use of *flexible learning strategies* and a 0.142 SD improvement in *self-control*. None of these effects were observed in schools assigned to the mathematics school treatment. In both treatment groups, there was no evidence of significant effects on Japanese language achievement. Overall, the LATE estimates are highly consistent with the ITT results in terms of both statistical significance and effect size. Moving forward, we focus primarily on the LATE results.

Table 4: LATE

	Academic Achievement			Learning Strategies					Non-cognitive Skills	
	Science	Math	Japanese Language	Flexible	Planning	Task-oriented	Cognitive	Effort Regulation	Self-Efficacy	Self-Control
D^{T1} (Math Treatment school)	-0.047	0.018	0.023	0.080	0.053	0.066	0.030	-0.074	-0.022	0.018
	(0.082)	(0.074)	(0.054)	(0.052)	(0.075)	(0.068)	(0.081)	(0.088)	(0.067)	(0.084)
D^{T2} (Science Treatment school)	0.178**	0.127*	0.051	0.110*	0.073	0.117	0.134	-0.045	0.034	-0.142*
	(0.059)	(0.053)	(0.050)	(0.055)	(0.071)	(0.062)	(0.071)	(0.069)	(0.064)	(0.070)
Y_{it-1}	0.274***	0.722***	0.726***	0.421***	0.421***	0.428***	0.407***	0.448***	0.481***	0.241***
	(-0.054)	(0.026)	(0.033)	(0.028)	(0.023)	(0.023)	(0.026)	(0.023)	(0.023)	(0.028)
Strata FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Covariates	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
F-statistic	35.37	39.48	37.98	44.71	44.83	44.34	44.68	43.73	43.94	43.70
P-value	0.646	0.642	0.668	0.620	0.530	0.412	0.457	0.340	0.225	0.979
N	2,395	2,394	2,399	2,334	2,334	2,333	2,203	2,335	2,350	2,353
R^2	0.098	0.514	0.532	0.183	0.183	0.199	0.166	0.204	0.238	0.099

Note: ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively. Standard errors are clustered at the school level. In addition to strata fixed effects, the regressions control for student gender, number of classes, and total student enrollment as covariates.

6.3 Effects on Science Achievement by Subfield

Table 5 presents the estimated effects of the intervention on science achievement across different subfields. The assessment distinguishes five domains, *basic*, *applied*, *knowledge/skills*, *inquiry/reasoning/expression*, and *proactive attitude*, as defined by the organization responsible for administering the standardized test. The results show that the intervention in science treatment schools had positive effects on all domains except *proactive attitude*, with particularly pronounced improvements observed in the *knowledge and skills* domain (0.201 SD).

Table 5: Effects on Science Achievement by Subfield (LATE)

	Basic	Applied	Knowledge/Skills	Inquiry/Reasoning/Expression	Proactive Attitude
D^{T1} (Math Treatment school)	-0.035	-0.071	-0.039	-0.045	-0.010
	(0.084)	(0.072)	(0.085)	(0.074)	(0.072)
D^{T2} (Science Treatment school)	0.172**	0.157*	0.201***	0.131*	0.096
	(0.059)	(0.064)	(0.058)	(0.065)	(0.057)
Y_{it-1}	0.260***	0.170***	0.256***	0.214***	0.150***
	(0.050)	(0.039)	(0.048)	(0.047)	(0.036)
Strata FE	YES	YES	YES	YES	YES
Covariates	YES	YES	YES	YES	YES
F-statistic	35.35	35.47	35.27	35.47	35.23
P-value	0.805	0.837	0.462	0.710	0.578
N	2,395	2,395	2,395	2,395	2,395
R^2	0.093	0.046	0.093	0.060	0.027

Note: ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively. Standard errors are clustered at the school level. In addition to strata fixed effects, the regressions control for student gender, number of classes, and total student enrollment as covariates.

6.4 Attrition

Appendix Table A2 examines the potential effects of attrition by estimating a model in which the dependent variable is a dummy variable equal to 1 if a student dropped out of the sample at either the baseline or endline, and 0 otherwise. While students who dropped out were more likely to have lower academic performance in $t-1$, there is no statistically significant correlation between attrition and the assignment of subject-specialist teachers.

Appendix Table A3 further explores this issue by estimating the LATE using an instrumental variable equal to 1 if the specialist teacher was assigned by April and 0 if the assignment occurred later. The results show that the coefficient for science treatment schools is larger than in Table 3, and this difference is statistically significant. This suggests that schools where specialist teachers were deployed earlier experienced stronger effects, indicating that the duration of specialist teacher assignment may play a critical role in determining the magnitude of the intervention's effect.

6.5 Placebo Test

Although the intervention targeted fourth grade students in the 2023 academic year, standardized achievement tests were also administered to fifth grade students who were not part of the intervention. Using this data, we conducted a placebo test by estimating the coefficient on a dummy variable equal to 1 if a fifth grade student attended a school where the fourth grade cohort had been assigned to either the mathematics or science treatment group, and 0 otherwise. The results, reported in Table 6, show no significant effects on academic achievement for fifth grade students in either treatments. This finding supports the robustness of our main results by confirming that the observed effects are not driven by unobserved school level factors unrelated to the intervention.

Table 6: Placebo Test（LATE)

	Academic Achievement			Learning Strategies					Non-cognitive Skills	
	Science	Math	Japanese Language	Flexible	Planning	Task-oriented	Cognitive	Effort Regulation	Self-Efficacy	Self-Control
D^{T1} (Math-focused school)	-0.019	-0.117	-0.041	-0.072	-0.029	-0.066	-0.057	-0.122*	-0.089	-0.151**
	(0.071)	(0.094)	(0.059)	(0.062)	(0.062)	(0.061)	(0.058)	(0.057)	(0.054)	(0.055)
D^{T2} (Science-focused school)	-0.009	-0.040	-0.032	-0.011	0.041	-0.095	-0.002	-0.091	0.032	-0.072
	(0.051)	(0.055)	(0.043)	(0.064)	(0.066)	(0.062)	(0.059)	(0.052)	(0.060)	(0.056)
Y_{it-1}	0.362***	0.771***	0.773***	0.511***	0.480***	0.505***	0.488***	0.474***	0.611***	0.517***
	(0.061)	(0.022)	(0.021)	(0.018)	(0.023)	(0.020)	(0.022)	(0.025)	(0.016)	(0.023)
Strata FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Covariates	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
F-statistic	26.97	34.08	32.94	35.87	36.23	36.25	35.20	36.03	35.82	35.98
P-value	0.937	0.621	0.538	0.422	0.147	0.324	0.096	0.097	0.087	0.098
N	2,360	2,375	2,389	2,348	2,343	2,344	2,228	2,348	2,350	2,349
R^2	0.153	0.573	0.597	0.261	0.229	0.271	0.244	0.228	0.379	0.266

Note: ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively. Standard errors are clustered at the school level. In addition to strata fixed effects, the regressions control for student gender, number of classes, and total student enrollment as covariates.

6.6 Cost-effectiveness Analysis: Comparison with Class Size Reduction

We next compare the effects of the intervention with those of other educational policies, using data from fifth grade students in 2023 who were not part of the intervention. As noted earlier, under national education policy, subject-specialist teaching had already begun to be introduced for fifth and sixth grade classes. Among the fifth-grade cohort examined here, approximately 55.2% of schools had implemented subject-specialist teaching in science, and 10.3% in math. It is important to note, however, that this type of subject-specialist teaching differs from the intervention analyzed in this study. In these schools, instead of maintaining a homeroom-based system and assigning part-time subject-specialists to specific subjects, the conventional model involves full departmentalization without additional part-time staffing.

Table 7 presents the results of OLS estimations in which a dummy variable, taking the value of 1 if departmentalization is implemented and 0 otherwise, is regressed on various academic outcomes. The results indicate that departmentalization exerts little to no effect on overall academic achievement. However, schools that introduced departmentalization teaching in science show a statistically significant negative impact on students' mathematics performance (−0.135 SD). This analysis revealed that implementing subject specialization in the core curriculum is associated with a negative effect on mathematics achievement. This finding is consistent with prior international evidence (e.g., Fryer Jr., 2018), suggesting that a simple implementation of subject specialization does not necessarily yield positive outcomes.

Table 7: Impact of Departmentalization on Cognitive Outcomes

	Science						Math	Japanese Language
		Basic	Applied	Knowledge/Skills	Inquiry/Reasoning/Expression	Proactive Attitude		
Math Departmentalization	0.021	0.051	-0.037	0.056	0.004	0.071	-0.019	0.036
	(0.084)	(0.080)	(0.090)	(0.054)	(0.117)	(0.051)	(0.066)	(0.054)
Science Departmentalization	0.065	0.075	0.038	0.096	0.043	0.034	-0.135*	-0.006
	(0.046)	(0.046)	(0.046)	(0.051)	(0.047)	(0.051)	(0.055)	(0.037)
Class Size	-0.009*	-0.009*	-0.006	-0.008	-0.010*	-0.006	-0.006	-0.003
	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)	(0.003)
Y_{it-1}	0.318***	0.297***	0.252***	0.271***	0.274***	0.147***	0.777***	0.780***
	(0.048)	(0.045)	(0.036)	(0.042)	(0.042)	(0.029)	(0.017)	(0.016)
Covariates	YES	YES	YES	YES	YES	YES	YES	YES
N	3,451	3,451	3,451	3,451	3,451	3,451	3,591	3,601
R^2	0.130	0.121	0.080	0.105	0.100	0.045	0.606	0.621

Table 8 examines the effects of departmentalization on students' learning strategies and non-cognitive skills. The results show that departmentalization in mathematics enhances students' use of planning strategies (approaches that emphasize systematic and organized learning). However, although not statistically significant, departmentalization is associated with a decline in strategies in other categories, and no consistent pattern emerges across outcomes.

Table 8: Impact of Departmentalization on Non-cognitive Outcomes

	Flexible	Planning	Task-oriented	Cognitive	Effort Regulation	Self-Efficacy	Self-Control
Math Departmentalization	0.014	0.156*	-0.045	0.006	-0.032	0.018	-0.058
	(0.061)	(0.064)	(0.050)	(0.074)	(0.050)	(0.064)	(0.058)
Science Departmentalization	0.064	0.041	0.056	0.024	-0.036	-0.011	-0.065
	(0.048)	(0.043)	(0.049)	(0.043)	(0.036)	(0.044)	(0.045)
Class Size	-0.003	-0.001	0.000	0.001	0.000	-0.002	-0.001
	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
Y_{it-1}	0.494***	0.467***	0.499***	0.471***	0.471***	0.594***	0.502***
	(0.015)	(0.018)	(0.015)	(0.017)	(0.020)	(0.014)	(0.018)
Covariates	YES	YES	YES	YES	YES	YES	YES
N	3,555	3,552	3,551	3,376	3,560	3,560	3,559
R^2	0.261	0.232	0.280	0.246	0.235	0.376	0.265

Table 7 illustrates the effects of class size reduction as well. In the case of science, reducing the number of students per class by ten leads to an increase in achievement of 0.09 SD. According to Ito et al. (2020), an analysis of class size reduction in one Japanese municipality, covering students from fourth grade in elementary school through ninth grade in junior high school, estimated that a reduction of ten students per class yields at most a 0.03 SD improvement in academic achievement. The key issue here lies in the policy's cost-effectiveness.

As shown in Appendix B.2, the cost per student of achieving a 1 SD improvement in science achievement is approximately 79,413 JPY (539 USD) under the current intervention with part-time subject-specialists, compared with 966,446 JPY (6,555 USD) under a small class size policy. These numbers demonstrate that implementing part-time specialist teachers is substantially more cost-effective.

6.7 Heterogeneity

We next examine heterogeneity by students' gender, baseline achievement, and socioeconomic status (SES). At the baseline, there is no statistically significant gender gap in average achievement in mathematics, whereas girls outperform boys in science. Given such heterogeneity, Appendix Table A5 tests whether the deployment of part-time subject-specialists has differential effects for boys and girls. The LATE estimates, including the interaction between the intervention indicator and gender, show no systematic gender-specific patterns.

Appendix Table A6 investigates whether treatment effects vary by students' baseline achievement. As it is also clear from Table 4, the coefficient on baseline achievement, Y_{ist}, is smaller in science than in mathematics or Japanese. This suggests that science achievement is less about accumulation and changes in rank are more likely to occur. Classifying students into top, middle, and bottom terciles of baseline achievement, we find that in science schools the scores of students who were in the top tercile at baseline decline by -0.316 SD, consistent with returns to the mean. By contrast, the interaction between the science-school treatment indicator and a dummy for top tercile baseline achievement is 0.565 $s.d$ and statistically positive, offsetting the decline among initially high-achieving students. This pattern indicates that for students who were already strong in science, the intervention was particularly effective in maintaining or improving science achievement.

Appendix Tables A7 and A8 examine whether the intervention effects vary according to SES and the classroom teacher's years of experience. The results show no significant heterogeneity by SES in science schools. In mathematics schools, however, students from lower-SES backgrounds exhibit declines in flexible learning strategies, planning strategies, and effort regulation strategies.

In contrast, self-control improves among students from lower-SES households. Regarding teacher experience, the results indicate that longer teaching experience is associated with larger gains in science achievement, although the coefficient is not large in magnitude. Overall, these analyses suggest that, particularly in science schools where the intervention produced notable improvements in cognitive outcomes, there is little evidence of heterogeneity that raises concern when considering the potential scale-up of the program.

6.8 Impact on Homeroom Teachers

Next, we estimate the impact of the intervention on homeroom teachers using data from the teacher survey conducted in February 2024. Table 9 reports the results of balance tests based on teacher characteristics determined prior to the start of the intervention. Given the relatively small sample size, homeroom teachers in mathematics and science schools tend to have longer teaching experience compared to those in control schools. In addition, science schools have a higher proportion of teachers with a university degree or higher. These variables are controlled for as covariates in the following analyses.

Table 9: Descriptive Statistics (Teacher Characteristics Before Intervention)

	(1) Math Treatment		(2) Science Treatment		(3) Control		(1)-(3)	(2)-(3)
	N		N		N			
Gender (Female = 1)	34	0.559	22	0.409	35	0.457	0.102	-0.048
Age	34	37.588	22	38.909	35	34.686	2.903	4.223
Years of teaching experience	34	12.647	22	14.364	35	8.743	3.904*	5.621**
Experience outside teaching (Yes = 1)	34	0.147	22	0.136	35	0.229	-0.082	-0.092
Educational attainment (Bachelor's or above = 1)	34	0.941	22	1.000	35	0.914	0.027	0.086*
Possesses secondary school teaching license (Yes = 1)	34	0.353	22	0.545	35	0.514	-0.161	0.031

Note: ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively.

Table 10 shows the estimated effects of assigning specialist teachers on homeroom teachers. First, and most importantly, the deployment of specialist teachers did not significantly affect homeroom teachers' time spent at school, time for lesson preparation, or perceived workload. For instructional practices, repetitive instruction declined in mathematics schools, while lecture-based instruction decreased in science schools. However, there is no evidence of a corresponding increase in student-centered instructional approaches, leaving the precise nature of the changes in teaching practices somewhat unclear. Moreover, the absence of clear spillover effects on Japanese language instruction suggests that the presence of specialist teachers did not substantially influence homeroom teachers' instructional methods or overall teaching quality. On the other hand, in both mathematics and science schools, teachers' growth mindsets showed notable improvement.

Table 10: Impact on Homeroom Teachers (LATE)

	Working Hours	Lesson Preparation Time	Perceived Workload	Lecture-Based Instruction	Repetitive Instruction	Student-Centered Instruction	Mental Health (WHO-5)	Growth Mindset
Math Schools	0.082	-0.232	-0.047	-0.272	-0.412*	-0.066	0.995	0.836**
	(0.273)	(0.423)	(0.090)	(0.148)	(0.161)	(0.135)	(1.663)	(0.290)
Science Schools	0.602	0.189	0.065	-0.370*	-0.159	0.041	-0.434	0.747**
	(0.353)	(0.389)	(0.072)	(0.160)	(0.154)	(0.113)	(1.629)	(0.271)
F-statistic	43.59	43.50	40.67	92.70	43.70	43.59	42.78	42.78
P-value	0.168	0.152	0.993	0.403	0.109	0.0902	0.970	0.722
N	91	90	88	90	89	91	90	90
R^2	0.137	0.103	0.181	0.084	0.147	0.108	0.113	0.317

Note: ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively. Standard errors are clustered at the school level. Covariates include teacher gender, age, educational attainment, years of teaching experience, possession of a secondary school level teaching license, experience outside the teaching profession, number of classes, and total student enrollment.

6.9 Implication from Results

In this section, we interpret the results. A key question is why the deployment of specialist teachers significantly improved student achievement only in science schools. Science is a subject with a high level of specialized knowledge, and elementary school teachers in Japan are often reported to experience considerable anxiety about teaching science due to insufficient subject-specific knowledge. According to the 2008 Survey on the Status of Science Education in Elementary Schools conducted by the Japan Science and Technology Agency, approximately 50% of homeroom teachers sampled from public elementary schools nationwide reported a lack of confidence in teaching science overall. Many teachers felt particularly unprepared in the areas of physics and earth sciences, and this lack of confidence was especially pronounced among teachers with fewer than ten years of teaching experience (Hanagami et al., 2009). Moreover, about 70% of respondents rated their knowledge and skills related to science experiments as low, and a large proportion reported a lack of confidence in guiding students' independent research.

More recent municipal-level surveys have found similar trends, such that elementary school teachers continue to feel anxious about teaching science (Namikoshi, 2025). An analysis of a survey of 488 science teachers by Nakajima and Kusaka (2020) shows that anxiety about teaching science tends to decrease with teaching experience, as teacher self-efficacy increases. A study surveying both homeroom and specialist science teachers (Hayashi and Misaki, 2017) shows that homeroom teachers perceive themselves as less specialized and report insufficient time for lesson preparation compared to specialist teachers. In addition, homeroom teachers value the subject expertise and thorough preparation that subject-specialist teachers bring to science instruction.

Studies from overseas point to similar challenges of teaching science in elementary schools. The 2018 National Survey of Science and Mathematics Education in the United States (Banilower et al., 2018), which surveyed approximately 7,000 teachers, found that only 31% of teachers reported feeling "very well prepared"to teach science, compared with 77% for language arts, 73% for mathematics, and 42% for social studies. A study by Cantrell et al. (2003), which tracked changes in science teaching self-efficacy among 268 pre-service teachers throughout their teacher training programs, found that practical teaching experience, such as teaching practices, significantly enhanced self-efficacy. Similarly, Menon and Sadler (2016) examined 62 pre-service teachers in a physics course and found significant improvements in science teaching self-efficacy following the course, suggesting that deepening subject knowledge plays a key role in building teacher confidence. Taken together with studies showing that teacher self-efficacy has a positive impact on student academic performance (Caprara et al., 2006; Mojavezi and Tamiz, 2012), in the case of science, teachers' subject-specific knowledge and experience has a particularly strong influence on students' learning outcomes.

The specialist teachers deployed in this intervention were, on average, older, as many were rehired after reaching their retirement age. It is therefore possible that their extensive teaching experience contributed to the delivery of high-quality instruction and the observed learning gains. Through group interviews with specialist teachers in elementary schools, Okada (2024) shows that teachers with substantial prior homeroom teacher experience tend to incorporate classroom goals and overall classroom management policies into their instructional practices after transitioning to a specialist role. Since all of the specialist teachers assigned in this intervention had prior experience as homeroom teachers, it is plausible that they engaged in similarly comprehensive forms of instruction, as suggested by Okada (2024). Notably, no comparable evidence has been identified in the context of mathematics.

Meanwhile, both mathematics and science schools exhibited increases in teachers'growth mindsets, suggesting that the deployment of specialist teachers may have had a positive influence on homeroom teachers' beliefs about their capacity to develop professionally. Although the precise mechanism remains unclear, it is possible that communication and collaboration with more experienced specialist teachers facilitated this shift in mindset. Prior research conducted in other countries has shown that improvements in teachers' growth mindsets have a positive impact on student achievement (Yeager et al., 2022).

With the cooperation of the Chiba Prefectural Board of Education, the authors conducted three rounds of interviews with both homeroom and specialist teachers. Although purely anecdotal, several science specialist teachers, more so than those teaching mathematics, reported that they were able to devote more time to lesson planning and assessment than they could while in full-time employment. In detailed discussions, they explained that, for example, in the unit on “Seasons and Living Things,” they incorporated outdoor activities, and in the unit on “Objects’ Temperature and Volume,” they designed experiments to enhance students' engagement. They described how they were able to devote more time to fostering students' thinking and observational skills by integrating experiments, audiovisual materials, and hands-on model construction. These accounts suggest a more active shift toward inquiry-based learning. Science, in particular, is a subject that requires extensive preparation time, and teachers noted that it had been difficult to devote sufficient time to such exploratory activities while working full-time.

However, there were also reports that specialist teachers tended to extend class time, sometimes starting lessons earlier or shortening break times, which created additional coordination burdens for homeroom teachers. Some homeroom teachers mentioned that when a positive working relationship was established, they learned a great deal from the specialist teachers' instructional methods and approaches to interacting with students. At the same time, they noted that assessments needed to be carefully aligned between homeroom and specialist teachers, and that this coordination effort meant the deployment of specialist teachers did not immediately translate into a reduction in workload. These field observations broadly align with the empirical findings that the presence of specialist teachers did not significantly affect homeroom teachers'time at school, lesson preparation time, or perceived workload.

7 Conclusion

This study has investigated the causal effects of deploying part-time specialist teachers in upper-elementary grades on students'academic achievement and learning attitudes, based on evidence from a cluster-randomized controlled trial conducted in Chiba Prefecture, Japan. The analysis revealed that in schools where specialist teachers were assigned to science classes, students' science achievement improved significantly by approximately 0.18 SD, and positive spillover effects also were observed in mathematics (about 0.13 SD In contrast, the deployment of specialist teachers in mathematics had no statistically significant effects on achievement in either mathematics or science. No spillover effects were observed in Japanese language achievement for either treatment arms.

Furthermore, students in science schools exhibited improvements in flexible learning strategies (ability to adapt learning approaches to different contexts) and in self-control. This suggests that high-quality, subject-specific instruction delivered by specialist teachers may have encouraged more autonomous learning behaviors in students. Although there were no significant changes in homeroom teachers' lesson preparation time or allocation of instructional time across subjects, their growth mindset improved significantly, indicating that collaboration with specialist teachers may have contributed to teachers' own learning and professional development.

Placebo tests confirmed that no similar effects were observed among fifth graders, who were not part of the intervention, supporting the robustness of the findings. In addition, schools that received specialist teachers earlier tended to show larger treatment effects, suggesting that the duration of specialist deployment plays an important role in improving learning outcomes. A comparison with other educational interventions using the same dataset further demonstrated that the specialist teacher program was highly cost-effective: The cost per student for a 1 standard deviation increase in achievement was 79,413 JPY (539 USD), approximately one-twelfth the cost of class size reduction policies.

These findings suggest that the effectiveness of departmentalization is not uniform across subjects and underscore the importance of designing systems that account for subject characteristics and teacher expertise. In particular, deploying experienced retired teachers as part-time specialists in highly specialized subjects such as science has the potential to enhance both student achievement and the professional practice of teachers. Future work should explore sustainable and effective models of departmentalization, including the optimal forms of collaboration between specialist and homeroom teachers and the integration of cross-curricular learning approaches.

References

Anderson, R. C. (1962). The case for teacher specialization in the elementary school. *The Elementary School Journal*, *62*(5), 253–260.

Asada, E., & Nakanishi, K. (2018). *Chuuichi gyappu no kaishou ni muketa shougakkou ni okeru kyoukatanninsei no kouka no kentou - gakkyuu tanninsei de manabu jidou tono ishikichousa no hikaku wo tooshite. A Study on the Effects of Departmentalization in Elementary Schools to Resolve the "Grade 7 Gap": A Comparison by Attitude Survey with Pupils Taught under the Homeroom-Teacher System* [Doctoral dissertation, International Pacific University] [In Japanese. Available at https://cir.nii.ac.jp/crid/1390853649310018816].

Backes, B., Cowan, J., & Goldhaber, D. (2024). Departmentalized instruction and elementary school effectiveness. working paper no. 298-0424. *National Center for Analysis of Longitudinal Data in Education Research (CALDER)*.

Backes, B., Cowan, J., & Goldhaber, D. (2025). Building bridges to middle school? elementary school departmentalization and academic achievement in the upper grades. working paper no. 315-0325. *National Center for Analysis of Longitudinal Data in Education Research (CALDER)*.

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). Report of the 2018 nssme+. *Horizon Research, Inc.*

Bastian, K. C., & Fortner, C. K. (2020). Is less more? subject-area specialization and outcomes in elementary schools. *Education Finance and Policy*, *15*(2), 357–382.

Bastian, K. C., Fortner, C. K., & Caton, K. (2023). Subject-area specialization and teacher retention: An elementary school story. *The Elementary School Journal*, *124*(2), 343–366.

Cantrell, P., Young, S., & Moore, A. (2003). Factors affecting science teaching efficacy of preservice elementary teachers. *Journal of Science Teacher Education*, *14*(3), 177–192.

Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of school psychology*, *44*(6), 473–490.

Central Council for Education, MEXT. (2024, August). *"reiwa no nihon-gata gakkō kyōiku" o ninau shitsu no takai kyōshi no kakuho no tame no kankyō seibi ni kansuru sōgō-teki na hōsaku ni tsuite (tōshin): Comprehensive policies for improving the environment to secure high-quality teachers responsible for the "reiwa-style japanese school education" (report)* (Chukyoshin No. 251) (Adopted at the 139th General Meeting, August 27, 2024). Ministry of Education, Culture, Sports, Science and Technology (MEXT). Tokyo, Japan. https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/079/sonota/1412985_00006.htm

Chan, T. C., & Jarman, D. (2004). Departmentalize elementary schools. *Principal*, *84*(1), 70–72.

Chiba Prefecture. (2022, December). *Chiba kenmin dayori reiwa 4-nen 12 gatsu-gō. chiba prefectural newsletter december 2022* (tech. rep.) (https://www.pref.chiba.lg.jp/kouhou/kenmindayori/r4/r4-12/index.html). Chiba Prefecture. Chiba, Japan.

Chiba Prefecture Board of Education. (2025, October). *Kouritsu gakkou rinjiteki nin'yō kyōshokuin / kaikei-nen'yō nin'yō shokuin / ninki-tsuki kyōshokuin no tōroku ni tsuite —taigū ni tsuite. conditions of employment for temporary and fixed-term teaching staff at public schools in chiba prefecture* (tech. rep.) (https://www.pref.chiba.lg.jp/kyouiku/syokuin/saiyou/koushi/taiguu.html). Chiba Prefecture Board of Education. Chiba, Japan.

Department for Education. (2024). *Use of supply teachers in schools: Research report* (tech. rep. No. RR1449) (https://www.gov.uk/government/publications/use-of-supply-teachers-in-schools). United Kingdom Government Department for Education.

Fryer Jr., R. G. (2018). The "pupil" factory: Specialization and the production of human capital in schools. *American Economic Review*, *108*(3), 616–656.

Hanagami, K., Matsumoto, M., Shimojo, T., Tanaka, H., Kiba, Y., Shindou, A., Hasegawa, H., Sato, A., Watanabe, R., & Ogura, Y. (2009). 2g3-e3 shou・chugakkou rikakyouiku jittaichousa houkoku i: Rika ni taisuru kyouin no ishikitou ni tsuite (kagakukyouiku jinzaiyousei (3), ippannhappyou, jisedai no kagakuryoku wo sodateru). 2g3-e3 survey report on the status of science education in elementary and junior high schools i: Teachers' attitudes toward science (science education human resource development (3), general research presentations, fostering the scientific capabilities of the next generation) [In Japanese. Available at https://doi.org/10.14935/jssep.33.0_443]. *Proceedings of the Annual Conference of the Japan Society for Science Education 33*, 443–444.

Hayashi, K., & Misaki, T. (2017). *Gakkyutannin to rikasenka no rikajyugyou ni taisuru ishiki. perceptions of science lessons by homeroom teachers and science specialists* [Doctoral dissertation, Shinshu University Library] [In Japanese. Available at https://soar-ir.repo.nii.ac.jp/records/18759].

Hwang, N., & Kisida, B. (2022). Spread too thin: The effect of specialization on teaching effectiveness. *Educational Evaluation and Policy Analysis*, *44*(4), 593–607.

Ibaraki Prefecture Board of Education. (2023, February). *Shougakkou kou-gakunen kara no kyouka tannin-sei no dounyuu ni tsuite. introduction of subject teacher system from upper grades of elementary school* (tech. rep. No. 210222-2) (https://kyoiku.pref.ibaraki.jp/wp-content/uploads/2023/02/210222-2.pdf). Ibaraki Prefecture Board of Education. Ibaraki, Japan.

Ito, H., Nakamuro, M., & Yamaguchi, S. (2020). Effects of class-size reduction on cognitive and non-cognitive skills. *Japan and the World Economy*, *53*, 100977.

Judson, E., & Sawada, D. (2000). Examining the effects of a reformed junior high school science class on students' math achievement. *School Science and Mathematics*, *100*(8), 419–425.

Kajita, H., & Doho, A. (2022). Shogakkou ni okeru kyoukatanninsei ni muketeno kongo - jissenkou ni okeru ishikichousa kara. toward the introduction of the departmentalization in elementary schools: From a survey of mindsets in practicing schools [In Japanese. Available at https://hijiyama-u.repo.nii.ac.jp/records/371]. *Review of Teacher Education Curriculum, Hijiyama University / Hijiyama University Junior College*, *8*, 24–33.

McGrath, C. J., & Rust, J. O. (2002). Academic achievement and between-class transition time for self-contained and departmental upper-elementary classes. *Journal of Instructional Psychology*, *29*(1).

Menon, D., & Sadler, T. D. (2016). Preservice elementary teachers' science self-efficacy beliefs and science content knowledge. *Journal of Science Teacher Education*, *27*(6), 649–673.

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2021, June). *Hyōjun jugyō jisū ni tsuite. overview of standard class hours* (tech. rep.) (https://www.mext.go.jp/content/20210629-mxt_kyoiku01-000016453_4.pdf). Ministry of Education, Culture, Sports, Science and Technology (MEXT). Tokyo, Japan.

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2023, March). *Shougakkou kou-gakunen ni okeru kyouka tannin-sei ni kansuru jirei-shuu. casebook on subject-teacher system in upper grades of elementary school* (tech. rep.) (https://www.mext.go.jp/content/20230310-

mext_zaimu-000027939_1.pdf). Ministry of Education, Culture, Sports, Science and Technology (MEXT). Tokyo, Japan.

Miyazaki Prefecture Board of Education. (2023, February). *Shougakkou kou-gakunen ichibu kyouka tannin-sei [tebiki]. guide to partial subject-teacher system in upper grades of elementary school* (tech. rep.) (https://sites.google.com/g.miyazaki-c.ed.jp/gimu/小学校教科担任制). Miyazaki Prefecture Board of Education. Miyazaki, Japan.

Mojavezi, A., & Tamiz, M. P. (2012). The impact of teacher self-efficacy on the students'motivation and achievement. *Theory and Practice in Language Studies*, *2*(3), 483–491.

Nakajima, A., & Kusaka, T. (2020). Shogakkou kyouin no rika shidou ni taisuru fuan, kyoushi kouryokukan, gakushudouki no kentou. a study of elementary school teachers'anxiety about science instruction, teacher efficacy, and learning motivation [In Japanese. Available at: https://doi.org/10.15077/jjet.43044]. *Japan Journal of Educational Technology*, *44*(1), 1–9.

Namikoshi, K. (2025). Shogakkou rikakyouiku jyujitsu no tameno jizokukanou na taisei no kouchiku: Rika jyugyou dukuri sapo-tosaito no sakusei to katsuyou ni mukete. building a sustainable system for enhancing elementary school science education: Development and utilization of a science lesson support website [In Japanese. Available at https://cir.nii.ac.jp/crid/1520023637927134592]. *Bulletin of the Nara Prefectural Institute of Educational Research*, (7), 18–40.

Oita Prefecture Board of Education. (2025, May). *Shougakkou kyouka tannin-sei ni kakaru tebiki. guide for introduction of subject-teacher system in elementary school* (tech. rep.) (https://www.pref.oita.jp/s Oita Prefecture Board of Education. Oita, Japan.

Okada, K. (2024). Shogakkou senkakyoushi no gakkyuunei ni taisuru ishiki to sono jissai - shougakkou senkakyoushi no guru-pu・intabyu- kara. specialist elementary school teachers'perceptions and practices of classroom management: Insights from group interviews [In Japanese. Available at https://yamato-u.repo.nii.ac.jp/records/2000076]. *Bulletin of the Faculty of Sociology, Yamato University*, *3*, 23–34.

Pintrich, P., Smith, D., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for learning questionnaire (mslq)* (Technical Report No. NCRIPTAL-91-B-004) (ERIC Document Reproduction Service No. ED338122 / Contract OERI-86-0010). National Center for Research to Improve Postsecondary Teaching and Learning, University of Michigan. Ann Arbor, MI. https://files.eric.ed.gov/fulltext/ED338122.pdf

Saitama City. (2025, May). (reiwa 7-nen 5-gatsu 26-nichi happyou) reiwa 9-nendo kara shougakkou kyouka tanninsei "saitama moderu" wo chuugakunen demo zenkou jisshi shimasu. announcement on may 26, 2025: From fy2027, the 'saitama model' subject-specialist teacher system will be implemented in all elementary schools for middle grades as well [https://www.city.saitama.lg.jp/006/

Sato, J., & Arai, H. (1998). Gakushu houryaku no shiyou to tassei mokuhyou oyobi genin kizoku tono kankei. the relation between the use of learning strategies learning goals and causal attributions [In Japanese]. *Tsukuba Psychological Research*, *20*, 115–124. https://cir.nii.ac.jp/crid/1050282677523203456

Tokoro, S., Suzuki, K., Tamura, T., & Kato, M. (2022). *Shogakkou kyoukatanninsei no unyou ni kansuru ichiteiann: Iwateken ni okeru kyoukatanninsei no genjyouhaaku to kadaikaiketsu no tedaten no tansaku kara. A Proposal for the Operation of the Departmentalization in Elementary Schools: From Surveying the Current State and Exploring Solutions in Iwate*

Prefecture [Doctoral dissertation, Iwate University] [In Japanese. Available at https://iwate-u.repo.nii.ac.jp/records/15669].

Tottori Prefecture Board of Education. (2025, March). *Shougakkou kou-gakunen ni okeru kyouka tannin-sei ni tsuite. on the subject-teacher system in upper-grade elementary school* (tech. rep.) (https://www.pref.tottori.lg.jp/273629.htm). Tottori Prefecture Board of Education. Tottori, Japan.

Tsukayama, E., Duckworth, A. L., & Kim, B. (2013). Domain-specific impulsivity in school-age children. *Developmental Science*, *16*(6), 879–893.

Yeager, D. S., Carroll, J. M., Buontempo, J., Cimpian, A., Woody, S., Crosnoe, R., Muller, C., Murray, J., Mhatre, P., Kersting, N., et al. (2022). Teacher mindsets help explain where a growth-mindset intervention does and doesn't work. *Psychological science*, *33*(1), 18–32.

Yomiuri Shimbun. (2025). Byoki ya kazoku no kanbyou nado de yasumu kyouin wo kaba-, kyouino-bra wo tanki haken suru shikumi dukuri he... eikoku wo sankou ni. creating a system to hire retired teachers short-term to cover teachers who are absent due to illness or caring for family members —inspired by the system in the uk [https://www.yomiuri.co.jp/kyoiku/kyoiku/news/20250813-OYT1T50004/]. *The Yomiuri Shinbun*.

A Tables

Table A1: Impact of Attrition

		Non-dropped(a)	Dropped from Sample(b)	(a)-(b)
Math Schools（T1）	Math	-0.022	0.140	-0.162
	N	977	11	
	Science	-0.045	0.340	-0.385
	N	970	11	
	Japanese Language	-0.049	-0.071	0.022
	N	975	11	
Science Schools（T2）	Math	0.064	-0.149	0.214
	N	777	84	
	Science			
	N			
	Japanese Language	0.088	-0.149	0.238*
	N	777	84	

Table A2: Relationship of Attrition and Subject-specialist Deployment

	Science	Math	Japanese Language
D^{T1} （Math School）	-0.107	-0.125	-0.124
	(0.080)	(0.082)	(0.082)
D^{T2} （Science School）	-0.137	-0.034	-0.034
	(0.085)	(0.114)	(0.114)
Y_{it-1}	-0.010	-0.026***	-0.020**
	(0.005)	(0.007)	(0.007)
Strata FE	YES	YES	YES
Covariates	YES	YES	YES
N	2,631	2,732	2,732
R^2	0.109	0.126	0.123

Note: ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively. Standard errors are clustered at the school level. In addition to strata fixed effects, the regressions control for number of classes and total student enrollment as covariates.

Table A3: Effect of Subject-specialist Teacher Deployment (IV: whether deployment happened from the beginning of the new academic year or not)

	Science	Math	Japanese Language
D^{T1} (Math School)	-0.083	0.028	0.041
	(0.149)	(0.130)	(0.095)
D^{T2} (Science School)	0.247*	0.177*	0.071
	(0.098)	(0.077)	(0.068)
Y_{it-1}	0.274***	0.724***	0.727***
	(0.053)	(0.026)	(0.033)
F-statistic	8.911	8.617	8.435
P-value	0.221	0.766	0.917
Strata FE	YES	YES	YES
Covariates	YES	YES	YES
N	2,395	2,394	2,399
R^2	0.093	0.514	0.532

Note: ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively. Standard errors are clustered at the school level. In addition to strata fixed effects, the regressions control for number of classes and total student enrollment as covariates.

Table A4: Effect of Subject-specialist Teacher Deployment (Using IRT Theta Score)

	Math	Japanese Language
D^{T1} （Math School）	0.002	0.021
	(0.068)	(0.049)
D^{T2} （Science School）	0.111*	0.050
	(0.049)	(0.046)
Y_{it-1}	0.717***	0.713***
	(0.028)	(0.034)
Strata FE	YES	YES
Covariates	YES	YES
F-statistic (Weak instrument)	35.96	36.03
P-value (Exogeneity)	0.578	0.560
N	2,445	2,418
R^2	0.513	0.524

Note: ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively. Standard errors are clustered at the school level. In addition to strata fixed effects, the regressions control for number of classes and total student enrollment as covariates.

Table A5: Effect of Subject-specialist Teacher Deployment (Heterogeneity: Gender)

	Academic Achievement			Learning Strategies					Non-cognitive Skills	
	Science	Math	Japanese Language	Flexible	Planning	Task-oriented	Cognitive	Effort Regulation	Self-Efficacy	Self-Control
D^{T1}（Math School）	-0.066	0.020	0.078	0.048	-0.008	0.064	0.005	-0.036	-0.064	0.011
	(0.091)	(0.067)	(0.075)	(0.068)	(0.098)	(0.078)	(0.106)	(0.104)	(0.080)	(0.097)
D^{T2}（Science School）	0.180**	0.117*	0.070	0.074	0.037	0.148*	0.135	-0.080	-0.046	-0.176
	(0.064)	(0.058)	(0.067)	(0.066)	(0.095)	(0.072)	(0.096)	(0.087)	(0.072)	(0.098)
Female	-0.026	-0.020	0.048	-0.159**	-0.161**	0.108	-0.136	0.022	-0.190***	0.242***
	(0.045)	(0.046)	(0.044)	(0.052)	(0.062)	(0.061)	(0.072)	(0.057)	(0.051)	(0.065)
$D^{T1}\times$ Female	0.040	-0.005	-0.112	0.067	0.126	0.002	0.051	-0.078	0.087	0.013
	(0.132)	(0.085)	(0.085)	(0.094)	(0.087)	(0.092)	(0.113)	(0.103)	(0.082)	(0.109)
$D^{T2}\times$ Female	-0.003	0.018	-0.038	0.071	0.071	-0.060	-0.000	0.066	0.157*	0.066
	(0.084)	(0.059)	(0.064)	(0.077)	(0.094)	(0.101)	(0.102)	(0.101)	(0.080)	(0.117)
F-statistic	15.01	17.74	16.41	21.12	21.45	21.16	23.53	20.29	20.33	19.78
P-value	0.645	0.637	0.649	0.628	0.534	0.415	0.459	0.335	0.229	0.983
N	2,395	2,394	2,399	2,334	2,334	2,333	2,203	2,335	2,350	2,353
R^2	0.098	0.514	0.532	0.183	0.185	0.199	0.166	0.204	0.240	0.099

Note: ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively. Standard errors are clustered at the school level. In addition to strata fixed effects, the regressions control for the baseline test score, number of classes, and total student enrollment as covariates.

Table A6: Effect of Subject-specialist Teacher Deployment (Heterogeneity: Baseline Test Score)

	Academic Achievement			Learning Strategies					Non-cognitive Skills	
	Science	Math	Japanese Language	Flexible	Planning	Task-oriented	Cognitive	Effort Regulation	Self-Efficacy	Self-Control
D^{T1}（Math School）	-0.034	0.109	0.155	0.190	0.114	0.178	0.054	-0.063	-0.007	0.112
	(0.177)	(0.127)	(0.134)	(0.116)	(0.128)	(0.109)	(0.130)	(0.150)	(0.109)	(0.130)
D^{T2}（Science School）	-0.036	0.056	0.069	0.159	0.071	0.110	0.083	-0.043	-0.015	-0.194
	(0.139)	(0.061)	(0.061)	(0.092)	(0.110)	(0.087)	(0.111)	(0.123)	(0.093)	(0.129)
Bottom tercile	-0.125	-0.177*	-0.202**	0.011	0.005	-0.173	0.059	-0.010	-0.055	-0.468***
	(0.141)	(0.074)	(0.068)	(0.121)	(0.093)	(0.096)	(0.099)	(0.098)	(0.075)	(0.088)
Middle tercile	-0.209	-0.207*	-0.059	0.153	-0.022	-0.198*	-0.002	-0.021	-0.104	-0.331***
	(0.147)	(0.088)	(0.073)	(0.100)	(0.073)	(0.101)	(0.109)	(0.085)	(0.078)	(0.082)
Top tercile	-0.316*	0.168	0.062	0.052	-0.105	-0.163	0.207	0.080	-0.021	0.163
	(0.154)	(0.110)	(0.098)	(0.109)	(0.086)	(0.116)	(0.109)	(0.083)	(0.098)	(0.096)
$D^{T1}\times$ middle tercile	-0.021	-0.097	-0.187	-0.224	-0.085	-0.176	0.090	0.017	-0.004	-0.152
	(0.198)	(0.129)	(0.143)	(0.161)	(0.120)	(0.117)	(0.127)	(0.143)	(0.094)	(0.127)
$D^{T1}\times$ top tercile	-0.016	-0.229	-0.251	-0.123	-0.109	-0.205	-0.190	-0.064	-0.052	-0.177
	(0.289)	(0.192)	(0.247)	(0.210)	(0.163)	(0.161)	(0.172)	(0.153)	(0.153)	(0.156)
$D^{T2}\times$ middle tercile	0.157	0.127	-0.067	-0.134	-0.030	-0.038	0.180	-0.037	0.072	0.058
	(0.200)	(0.082)	(0.080)	(0.110)	(0.098)	(0.115)	(0.126)	(0.119)	(0.082)	(0.152)
$D^{T2}\times$ top tercile	0.565*	0.122	0.051	-0.007	0.053	0.085	-0.032	0.035	0.083	0.153
	(0.270)	(0.080)	(0.077)	(0.138)	(0.106)	(0.112)	(0.139)	(0.146)	(0.080)	(0.157)
F-statistic	10.46	11.85	9.602	11.85	14.06	16.14	16.02	14.24	14.91	15.56
P-value	0.639	0.659	0.734	0.608	0.545	0.411	0.466	0.329	0.220	0.996
N	2,395	2,394	2,399	2,334	2,334	2,333	2,203	2,335	2,350	2,353
R^2	0.111	0.527	0.532	0.183	0.184	0.202	0.165	0.205	0.239	0.134

Note: ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively. Standard errors are clustered at the school level. In addition to strata fixed effects, the regressions control for baseline test score, the number of classes, and total student enrollment as covariates.

Table A7: Effect of Subject-specialist Teacher Deployment (Heterogeneity: Household SES)

	Academic Achievement			Learning Strategies					Non-cognitive Skills	
	Science	Math	Japanese Language	Flexible	Planning	Task-oriented	Cognitive	Effort Regulation	Self-Efficacy	Self-Control
D^{T1} (Math School)	-0.044	-0.090	0.123	0.399*	0.466*	0.329	0.221	0.255	0.069	0.442**
	(0.224)	(0.197)	(0.141)	(0.185)	(0.196)	(0.177)	(0.173)	(0.158)	(0.171)	(0.153)
D^{T2} (Science School)	0.083	-0.100	-0.046	0.181	0.239	0.179	0.188	-0.092	-0.072	0.002
	(0.207)	(0.138)	(0.123)	(0.193)	(0.211)	(0.203)	(0.191)	(0.186)	(0.158)	(0.181)
Low SES	0.000	0.001	0.036	0.011	0.051	0.011	0.031	0.052	0.007	0.064***
	(0.029)	(0.029)	(0.042)	(0.022)	(0.027)	(0.024)	(0.021)	(0.029)	(0.021)	(0.017)
$D^{T1}\times$ low SES	0.002	0.036	-0.030	-0.105	-0.126*	-0.086	-0.057	-0.092	-0.027	-0.125**
	(0.068)	(0.059)	(0.042)	(0.054)	(0.056)	(0.054)	(0.051)	(0.047)	(0.049)	(0.044)
$D^{T2}\times$ low SES	0.030	0.074	0.012	-0.024	-0.048	-0.021	-0.016	0.026	0.036	-0.034
	(0.061)	(0.042)	(0.017)	(0.054)	(0.063)	(0.058)	(0.054)	(0.061)	(0.050)	(0.054)
F-statistic	16.49	20.30	17.55	30.73	31.46	30.51	35.92	30.24	30.22	27.67
P-value	0.392	0.344	0.494	0.357	0.237	0.228	0.208	0.280	0.280	0.374
N	2,395	2,394	2,402	2,334	2,334	2,333	2,203	2,335	2,350	2,353
R^2	0.097	0.515	0.532	0.185	0.187	0.202	0.168	0.207	0.239	0.103

Note: ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively. Standard errors are clustered at the school level. In addition to strata fixed effects, the regressions control for number of classes and total student enrollment as covariates. Low SES is based on the ratio of students receiving school attendance support.

Table A8: Effect of Subject-specialist Teacher Deployment (Heterogeneity: Homeroom Teacher's Year of Experience)

	Academic Achievement			Learning Strategies					Non-cognitive Skills	
	Science	Math	Japanese Language	Flexible	Planning	Task-oriented	Cognitive	Effort Regulation	Self-Efficacy	Self-Control
D^{T1}（Math School）	-0.005	-0.007	-0.002	0.146	0.087	0.194	0.089	-0.067	-0.013	0.099
	(0.155)	(0.107)	(0.097)	(0.123)	(0.163)	(0.146)	(0.159)	(0.156)	(0.141)	(0.147)
D^{T2}（Science School）	-0.023	0.172*	0.119	0.250*	0.188	0.168	0.314*	-0.036	-0.067	-0.255
	(0.102)	(0.085)	(0.086)	(0.107)	(0.151)	(0.125)	(0.133)	(0.144)	(0.141)	(0.180)
Year	-0.009	-0.006	-0.008	-0.004	-0.003	0.003	0.012	-0.006	-0.002	-0.010
	(0.007)	(0.009)	(0.007)	(0.007)	(0.013)	(0.010)	(0.011)	(0.012)	(0.009)	(0.011)
$D^{T1} \times year$	-0.000	0.004	0.004	-0.006	-0.004	-0.015	-0.011	-0.001	-0.002	-0.006
	(0.014)	(0.011)	(0.010)	(0.012)	(0.014)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)
$D^{T2} \times year$	0.025**	0.002	0.001	-0.007	-0.008	-0.007	-0.021	0.002	0.008	0.016
	(0.009)	(0.010)	(0.009)	(0.010)	(0.013)	(0.011)	(0.013)	(0.013)	(0.012)	(0.014)
F-statistic	11.46	13.10	12.06	16.80	16.93	16.59	17.75	16.60	16.49	15.96
P-value	0.607	0.992	0.451	0.558	0.493	0.614	0.472	0.358	0.0752	0.658
N	2,180	2,168	2,173	2,114	2,113	2,112	1,994	2,115	2,129	2,133
R^2	0.104	0.504	0.527	0.179	0.176	0.205	0.160	0.202	0.240	0.107

Note: ***, **, and * indicate statistical significance at the 0.1%, 1%, and 5% levels, respectively. Standard errors are clustered at the school level. In addition to strata fixed effects, the regressions control for student gender, number of classes, and total student enrollment as covariates.The year variable indicates the homeroom teacher's years of experience.

B Appendix

B.1 Examples of Departmentalization in Japan

Ibaraki Prefecture piloted departmentalization in a limited number of public elementary schools in 2020 and implemented it across all 469 public elementary schools in the prefecture (including public integrated elementary-middle schools) in 2021. The departmentalization occurred primarily in mathematics, science, and English for upper-grade students (Ibaraki Prefectural Board of Education, 2023). In this system, teachers holding middle school teaching licenses are assigned to elementary schools and assume responsibility for three to five lessons per class per week in these subjects. This allows, for instance, teachers with a middle school mathematics license to teach mathematics at the elementary school level, thereby strengthening subject-specialization instruction.

In Saitama City, where integrated elementary and middle school education is promoted, departmentalization was first introduced in 10 pilot schools in fiscal year 2021 and expanded to all 104 public elementary schools in 2023. The city also announced plans to extend the program to middle schools starting in fiscal year 2027 (Saitama City, 2025). There are also cases where middle school instructors teach classes in the upper grades of elementary school, thereby leveraging the benefits of integrated education. Oita Prefecture implemented departmentalization in 27 elementary schools since the fiscal year 2020. Initially, specialist instruction was provided in science and foreign languages; from 2022 onward, homeroom teachers began exchanging responsibility among each other for the five core subjects (Japanese, social studies, mathematics, science, and foreign languages) (Oita Prefectural Board of Education, 2025). In other words, teachers exchange responsibility for certain subjects to make better use of their expertise. Tottori Prefecture began piloting departmentalization in 2016 and has since developed implementation models suited to different school sizes (Tottori Prefecture Board of Education, 2025). For example, in schools with only one class per grade, fifth and sixth grade teachers exchange subjects with each other. In schools with two classes per grade, subject exchanges occur within the grade. In schools with three or more classes per grade, part-time subject-specialists or non-homeroom teachers are added to the system. Miyazaki Prefecture introduced departmentalization in 19 pilot schools in 2020 and expanded it to 43 schools in 2022 (Miyazaki Prefecture Board of Education, 2023). The basic approach is the exchange of subjects among homeroom teachers within a school, but some schools combine this with an "add-on" model, in which additional specialist teachers are assigned to certain subjects or grades.

In summary, the methods of deploying subject-specialist teachers and the subjects targeted for departmentalization vary widely across municipalities. A casebook, "Casebook on Departmentalization in Upper Grades of Elementary Schools" (MEXT, 2023), compiled by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) introduces several approaches for departmentalization. This includes hiring new specialist teachers, subject exchanges among homeroom teachers, and utilizing middle school teachers so that different teachers are responsible for each of the five core subjects (Japanese, social studies, mathematics, science, and English). During the early stages of implementation, English, science, mathematics, and physical education were often prioritized as "focus subjects" for departmentalization. More recently, however, some municipalities have expanded the practice to cover all five core subjects, including Japanese and social studies. In integrated elementary-middle schools, there are cases where middle school subject-specialists also teach classes at the elementary level. This approach not only provides specialized instruction from the early stages of schooling, it also helps ensure a smoother transition to middle school, mitigating what is often referred to as the "grade seven gap," which students experience when enrolling in middle school.

B.2 Cost–Effectiveness Calculation

This section presents the basis for calculating the cost-effectiveness of deploying specialist science teachers and reducing class size.

Cost-Effectiveness of Deploying Specialist Science Teachers The total annual cost of deploying specialist science teachers was calculated as: 「Total Cost = (Number of Schools with Science Specialists) × (Number of Classes per School) × (Standard Instruction Hours) × (Hourly Wage Rate)」. In this calculation, the number of schools refers to the actual number of schools where specialist science teachers were assigned (15 schools), and the number of classes refers to the average number of fourth grade classes in those schools (1.85 classes). The standard number of instruction hours (105 hours per year) and the hourly wage rate for teachers (3,100 JPY per hour) were taken from guidelines issued by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) (2021) and the Chiba Prefecture Board of Education (2025), respectively. Based on these figures, the total cost of deploying specialist science teachers across the 15 schools was calculated to be 9,032,625 JPY.

Given that the average number of fourth grade students in these schools was 42.6, the total number of students who received instruction from specialist science teachers was 639. Accordingly, the cost per student was calculated to be 14,135.56 JPY. Since the deployment of specialist science teachers increased science test scores by 0.178 SD (see Table 4, LATE estimate), the cost of achieving a one standard deviation increase in test scores is estimated to be 79,413.28 JPY.

Cost-Effectiveness of Class Size Reduction The cost of class size reduction is estimated based on the expenditure required to reduce the number of students per teacher by one. Since the marginal cost of reducing class size increases as class sizes become smaller, this approach yields a conservative estimate which is likely to be lower than the actual cost of implementing a small class size policy.

Given that the average number of fifth grade classes per school is 1.65 and the average number of students per class is 26.6, reducing the class size by one student would require the equivalent of hiring an additional 0.064 teachers per school. With the average teacher salary in Chiba Prefecture being 5,923,000 JPY per year (Chiba Prefecture, 2022), the cost of reducing class size by one student is estimated to be 381,755.86 JPY per school. As the average number of fifth grade students per school is 43.9, the per-student cost of implementation is calculated to be 8,698.01 JPY. Since reducing class size by one student increases test scores by 0.009 SD (see Table 7), the cost of achieving a one standard deviation increase in test scores is estimated to be 966,446.06 JPY.

Taken together, these calculations indicate that the deployment of specialist science teachers is substantially more cost-effective than reducing class size. Using the estimates derived from the Chiba Prefecture data, the specialist teacher intervention is approximately 12.2 times more cost-effective. Considering that the cost estimate for class size reduction is conservative, the true difference in cost-effectiveness is likely to be even larger.