Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

The Effect of Class-size Reduction on Students' Well-Being in School

杉田 壮一朗

2025 年 6 月 1 日
DP2025-012
https://ies.keio.ac.jp/publications/26202/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
1 June, 2025

The Effect of Class-size Reduction on Students' Well-Being in School

杉田壮一朗

IES Keio DP2025-012
2025年6月1日
JEL Classification: I21, I28, H52
キーワード: ウェルビーイング、少人数学級、操作変数法

【要旨】

The empirical literature on the causal effects of class-size reduction on academic outcomes is crowded and yields mixed results, with particularly limited and inconsistent findings in the Japanese context. This study examines whether smaller class sizes enhance classroom climate and student-teacher relationships, using large-scale panel data from a student achievement survey conducted in a Japanese prefecture. Employing an instrumental variable approach based on the Maimonides rule, I find that a reduction of 10 students per class yields modest improvements—up to 0.07 standard deviations—in measures of teacher-student relationships. The analysis does not provide sufficient evidence to conclude that class-size reduction is effective in improving students' well-being.

杉田壮一朗
慶應義塾大学大学院政策・メディア研究科後期博士課程
sugita3520@keio.jp

謝辞：本稿は指導教授の中室牧子教授の承認を得て投稿している。また、中室教授の他、現サイバーエージェントの伊藤寛武氏には本稿の執筆の上で多くのアドバイスをいただいた。記して感謝したい。また、本稿にありうべき主張、誤りの一切の責任は筆者個人に帰する。

The Effect of Class-size Reduction on Students' Well-Being in School

Soichiro Sugita *

June 1, 2025

Abstract

The empirical literature on the causal effects of class-size reduction on academic outcomes is crowded and yields mixed results, with particularly limited and inconsistent findings in the Japanese context. This study examines whether smaller class sizes enhance classroom climate and student-teacher relationships, using large-scale panel data from a student achievement survey conducted in a Japanese prefecture. Employing an instrumental variable approach based on the Maimonides rule, I find that a reduction of 10 students per class yields modest improvements—up to 0.07 standard deviations—in measures of teacher-student relationships. The analysis does not provide sufficient evidence to conclude that class-size reduction is effective in improving students' well-being.

JEL Classification: I21, I28, H52

Keywords: Classroom climate, Class size reduction

1 Introduction

Class-size reduction has long been a topic of interest in the economics of education, primarily regarding its impact on academic achievement. However, empirical findings have been mixed,

*Graduate School of Media and Governance, Keio University. Email: sugita3520@keio.jp

and evidence from Japan remains particularly limited. Moreover, a growing body of research has begun to explore its effects on broader educational outcomes, such as student well-being, mental health, and classroom environment.

Within the traditional framework of the education production function[1],class size has been considered a key input, as reducing it typically requires the additional hiring of teachers, thereby increasing instructional resources (Hanushek, 2003). This policy implication is grounded in the assumption that additional teachers can deliver higher-quality instruction and individualized support. However, many empirical studies report that smaller classes have limited or modest effects on academic performance. This has led some researchers to question the effectiveness of input-based policies like class-size caps (Hanushek and Woessmann, 2017).

In response, recent studies have shifted their focus toward non-academic outcomes, including socio-emotional development, bullying, and absenteeism (Niki 2013; Hojo 2021; Oikawa et al. 2022; Nakamuro 2017). These developments align with international policy perspectives, notably those advocated by the OECD, which frames student well-being —including perceptions of classroom safety, social support, and teacher relationships— as core outcomes of effective schooling (OECD, 2017, 2021).

This study builds on that shift by asking whether smaller class sizes improve students' subjective evaluations of their school experience—such as classroom calmness, teacher attentiveness, and emotional support. The objective is to explore whether class-size reduction holds policy value beyond traditional academic metrics.

This study contributes in two key ways. First, it draws on individual-level panel data from a prefectural achievement survey conducted between 2016 and 2019. Unlike previous studies that rely on cross-sectional data (e.g., Sugita, 2025), this dataset allows for the estimation of student and school fixed effects, thereby enabling us to control for time-invariant unobserved heterogeneity and to exploit within-student variation in class size exposure. This micro-level panel structure enhances the credibility of causal inference.

[1]The education production function applies the concept of a production function from economics to education, modeling the relationship between inputs such as school and family resources and outputs such as students' academic achievement. Since the publication of the Coleman Report (Coleman, 1968), there has been an ongoing debate regarding the effects of school resource inputs.

Second, this study provides new empirical evidence on the effect of class-size reduction on subjective measures of well-being. These include classroom climate and the quality of student–teacher interactions. I use an instrumental variable approach based on the Maimonides rule, which sets a legal cap on class size in Japan.

However, estimating the effect of class size is complicated by endogeneity concerns. For instance, problematic students may be more likely to be assigned to smaller classes, or more experienced teachers may be deliberately assigned to classes with struggling students. When such endogenous selection occurs, simple regression analyses that treat class size as exogenous may yield biased estimates.

To identify the causal effect of class size reduction, this study follows the approach of (Angrist and Lavy, 1999), applying a two-stage least squares (2SLS) method that uses predicted class size based on grade-level enrollment as an instrumental variable. In addition, I exploit the panel nature of the data to control for both school and individual fixed effects(FE), thereby accounting for unobserved confounders.

The 2SLS estimates indicate that class size reduction has a statistically significant positive effect on student well-being, but the magnitude is modest—up to approximately 0.07 standard deviations. These findings are consistent with effect sizes reported in previous studies, including those using similar data and specifications. Moreover, the results are robust across various model specifications, including those with conservative estimates. The largest effect is observed for the outcome measuring a calm learning environment, suggesting that smaller class sizes may be particularly beneficial in fostering a more focused and orderly classroom atmosphere. Although I also explored potential heterogeneity in the treatment effect, no clear patterns emerged. Based on the results, I find no strong evidence that class size reduction is an effective policy for improving students' well-being in school life.

The remainder of this paper is organized as follows. Section 2 reviews the prior literature on class-size effects. Section 3 describes the data and presents summary statistics. Section 4 outlines the empirical strategy. Section 5 reports the estimation results and presents additional analyses, including robustness checks and subgroup comparisons. Section 6 concludes with a summary of the findings, a discussion of the study's limitations, and directions for future research.

2 Literature Review

This paper contributes to the empirical literature on class-size effects. From a broader perspective, the literature on class size belongs to the wider context of research on the impact of school resources. Since the publication of the Coleman Report (Coleman, 1968), there has been ongoing debate over whether and how schools affect student outcomes. However, the empirical findings remain inconclusive, and the accumulation of evidence in Japan is particularly limited. For example, Hanushek and Woessmann (2017) report that input measures such as school spending and class size—commonly used in education production functions—have little to no effect on academic achievement.

A large number of empirical studies have investigated the effects of class size reduction, especially in countries outside Japan where the evidence base is much more extensive. Although relatively limited in number, there are also studies from Japan that address this topic. This chapter reviews the literature on class size reduction in sequence, beginning with international studies and followed by those conducted in Japan, and then discusses the novelty and contribution of the present study

2.1 International Literature

The international literature has employed a variety of methodological approaches to address the endogeneity of class size in estimating its causal effect on cognitive and non-cognitive outcomes. Broadly speaking, these approaches can be categorized into three types: randomized controlled trials (RCTs) such as the STAR experiment, instrumental variable approaches based on discontinuities like that of Angrist and Lavy (1999), and natural experiments leveraging exogenous variation such as single-grade classrooms or fluctuations in cohort size. However, the overall literature remains inconclusive as to whether class size reduction contributes meaningfully to improved student achievement.

2.1.1 Experimental Studies

A well-known example of an RCT is the STAR project, a large-scale social experiment conducted in Tennessee in the 1980s to evaluate the effect of class size reduction. Krueger

(1999) showed that class size reduction had a positive impact on academic performance in the first year of the experiment. Chetty et al. (2011) further analyzed the STAR data and reported statistically significant effects on non-cognitive outcomes such as earnings, college attendance, and homeownership at age 27.

However, the findings from the STAR project have also been subject to critical review. Hanushek (1999) questioned the external validity of the study and pointed out that the effects were only evident in kindergarten and first grade, and that the heterogeneity in effects for disadvantaged students raises concerns about the equity of such policies.

Beyond the STAR project, Bettinger et al. (2017) conducted an RCT in the context of online education and found that increasing class size did not significantly affect outcomes such as course completion rates.

2.1.2 Quasi-Experimental Approaches Using the Maimonides Rule

Angrist and Lavy (1999), in contrast, proposed a quasi-experimental approach that addresses the endogeneity of class size. They exploited an institutional rule in Israel requiring that when the number of students in a grade exceeds 40, the grade must be split into two classes[2]. Using this discontinuity, they employed an instrumental variable strategy based on a regression discontinuity design and found a statistically significant negative relationship between class size and students' math scores in fifth grade. Subsequent studies have applied similar identification strategies. For instance, Fredriksson et al. (2013) estimated the effect of class size on both cognitive skills (math and language test scores) and non-cognitive traits such as effort, motivation, aspirations, and self-confidence for students in grades 4 to 6, finding statistically significant results.

However, Angrist et al. (2017), expressed a more skeptical view, questioning whether the statistically significant relationship between class size and academic achievement still holds today.

[2]This institutional rule is often referred to as the Maimonides Rule. For example, when there are 41 students in a grade, they are split into two classes of 20 and 21 students, respectively.

2.1.3 Other Approaches

In addition to the quasi-experimental designs above, several studies have exploited naturally occurring variation in class size to address endogeneity concerns. Hoxby (2000), for example, used changes in class size induced by cohort population fluctuations and found no effect of smaller classes on academic achievement. She also suggested that earlier positive findings based on the STAR experiment (e.g., Krueger, 1999) may reflect the influence of the Hawthorne effect[3]. Another approach involves using single-grade classrooms, where class size is fixed and not subject to endogenous sorting. Urquiola (2006) analyzed single-class schools in rural Bolivia and reported a statistically significant positive impact of smaller class sizes on academic performance.

2.2 Literature in Japan

In Japan as well, a number of empirical studies have examined the effects of class size reduction, but their findings remain inconsistent, and academic debate continues.

Regarding academic achievement, the estimated effects of class size reduction vary across studies. Akabayashi and Nakamura (2014), using data from the Yokohama City Student Survey and the National Assessment of Academic Ability (NAAA), applied an estimation strategy following (Angrist and Lavy, 1999). Their results show no statistically significant effect of smaller classes on academic performance, except for Japanese language scores among sixth graders. In contrast, Senoh and Hojo (2016), using NAAA data, found that class size reduction improves test scores, especially in schools attended by students from low socioeconomic status (SES) backgrounds.

Several Japanese studies have also examined the effects of class size on non-cognitive outcomes, though these results are similarly mixed. Ito et al. (2017), using panel data from grades 4 through 9 in a city in central Japan, estimated a model accounting for classroom and school-level hierarchies. Their results indicated significant effects not only on academic scores in Japanese and mathematics, but also on some non-cognitive outcomes. On the other hand, Ito et al. (2020) used individual-level panel data from a prefectural achievement

[3]The Hawthorne effect refers to changes in participants' behavior simply because they are being observed.

survey and found no statistically significant effects of class size reduction on either academic or non-cognitive outcomes. Importantly, these null results persisted even after controlling for school fixed effects, which absorb unobserved, school-specific factors.

Other studies have used international data such as TIMSS to examine the effects of class size reduction. Niki (2013), analyzing the 2003 wave of TIMSS, applied a methodology similar to (Angrist and Lavy, 1999) to estimate the effects of class size on both academic outcomes (math and science test scores) and non-cognitive skills. The analysis found no significant effects on academic performance and only limited evidence for effects on non-cognitive outcomes. Hojo (2013), using the same TIMSS dataset, reported statistically significant effects on both math and science scores, though the estimated magnitudes were small.

Among the limited literature examining outcomes beyond cognitive and non-cognitive skills, Nakamuro (2017) stands out. Using data from Japanese elementary schools, she reported that class size reduction helped reduce school absenteeism and that the presence of additional teachers was effective in this regard.

2.3 Recent Trends in Research

Since the 2010s, a growing body of literature has investigated the effects of class size reduction on outcomes beyond academic achievement. These studies include analyses of class size effects on non-cognitive skills—viewed as a form of human capital that cannot be fully captured by test scores(Chetty et al., 2011; Fredriksson et al., 2013; Niki, 2013; Ito et al., 2017, 2020). Other studies have examined its impact on teacher mental health (Hojo, 2021), the role of smaller classes in preventing the spread of infectious diseases (Oikawa et al., 2022), and effects on school life, including bullying and student absenteeism (Nakamuro, 2017; Hojo, 2023; Sugita, 2025).

Whereas earlier studies prior to the 2000s focused primarily on the effects of school inputs such as class size on test scores—and often found limited or no robust effects—recent research has sought to identify alternative channels through which class size reduction may matter. In particular, studies such as Nakamuro (2017), Hojo (2023), and Sugita (2025), which focus on students' broader school experiences, are closely aligned with the motivation of the present study. Notably, Sugita (2025), using cross-sectional data from TIMSS 2011, finds positive effects of smaller classes on students' satisfaction with school life. The present study contributes to this line of research by examining how class size reduction affects students' subjective evaluations of their school experiences, providing new evidence on the potential well-being effects of class size policies.

3 Data

3.1 Data Description

In this study, I use individual-level panel data from the Saitama Prefectural Achievement Survey (hereafter, the Saitama Survey), conducted between 2016 and 2019. Some outcome variables were introduced in 2017; for those variables, I use data from the three years between 2017 and 2019. The specific variables introduced in 2017 are described later.

A notable feature of the Saitama Survey is that it is a census-based assessment covering all public elementary and junior high school students from Grade 4 through Grade 9 in every municipality of the prefecture, excluding the prefectural capital. Each cohort consists of approximately 300,000 students, ensuring a large sample size. Because the dataset includes student identifiers, individual students can be tracked over time during the survey period.

The Saitama Survey adopts item response theory (IRT) to generate test scores, allowing for meaningful comparisons of academic performance across different survey years. In addition to standardized test scores, the dataset includes student questionnaire responses on socioeconomic background (e.g., participation in private tutoring, number of books at home) as well as perceptions of classroom climate and student-teacher relationships.

3.1.1 Notes on Data Processing

Since the Saitama Survey is administered at the beginning of each school year (in April), I use the class size from year $t-1$ to estimate its effect on outcomes in year t. This is based on the assumption that the class size a student experienced in the previous school year affects the outcomes measured at the beginning of the current school year. Accordingly, variables such as grade size, school size, and school ID (from the previous year) are also lagged by one year.

Due to this data structure, the analysis excludes Grade 4 and 5 students. This is because the analysis includes individual and school fixed effects, and thus restricts the sample to students who participated in the survey at least twice.

This sampling structure and processing procedure are consistent with that of Ito et al. (2020), who also use the same data set.

Survey items based on a 4-point Likert scale were reverse-coded so that 4 represents the most affirmative response and 1 the least. These items were then standardized to have a mean of 0 and standard deviation of 1 for intuitive interpretation. Regarding students' socioeconomic status, I define a dummy variable for not attending cram school (coded as 1 if the student does not attend) and a dummy for having no books at home (coded as 1 if the number of books is zero).

3.1.2 Outcome Variables

The outcome variables in this study are derived from the student questionnaire in the Saitama Survey, capturing classroom dynamics, student–teacher relationships, and interpersonal interactions. The following is the list of outcome variables used in the analysis. Each outcome is labeled with an abbreviated name in brackets, which is used throughout the paper.

1. Did you enjoy your life in the classroom? [**Fun Class Life**]
2. Was your classroom a calm and focused place for learning? [**Calm Learning Env**]
3. Do you think your class worked together as a team in various activities (including events like sports day and field trips)? [**Class Events**]
4. Did your teachers recognize your strengths and good qualities? [**Teacher Recog. (Good)**]
5. Did your teachers listen to your worries and concerns? [**Teacher Listened**]
6. Did your classmates recognize your strengths and good qualities? [**Student Recog. (Good)**]

7. Did your teacher explain things you didn't understand in class or on tests until you fully understood them? [**Teacher Explained**]

These are self-reported subjective measures provided by students on a 4-point Likert scale. Responses are reversed so that a score of 4 corresponds to the most affirmative answer and 1 to the least [4].

Among the outcome variables, items 3, 5, 6, and 7 were first measured in 2017. Therefore, these are analyzed using data from 2017 to 2019 (three time points).

3.2 Summary Statistics

Table 1 presents descriptive statistics for the main variables used in the analysis. These statistics are based on the pooled sample of students in Grades 6 through 9, which serves as the estimation sample in this study.

The average class size is 32.77, while the average predicted class size is 33.55. The maximum predicted class size is 40, consistent with the Maimonides Rule–based calculation.

Among student-level covariates, the no cram school dummy and no books at home dummy serve as proxies for socioeconomic status. Since the data do not contain direct information on parental income, occupation, or education level, these binary variables are used instead. Approximately 32% of students do not attend private tutoring, and around 11% have no books at home.

4 Empirical Strategy

This section describes the empirical framework used to estimate causal effects. (Angrist and Lavy, 1999) considered the endogeneity of class size—that is, the possibility that class size may be correlated with unobserved factors in the error term—and exploited the institutional rule that when the number of students in a grade exceeds 40, schools typically split the cohort into two classes of 20 and 21 students, respectively. This rule, commonly referred to

[4] Kaiser and Oswald (2022) justify the use of subjective ordinal measures in quantitative analysis. They show that such measures—used to assess feelings like happiness and life satisfaction—have greater predictive power for future behavior than objective socioeconomic variables.

Table 1: Summary Statistics

Variables	Observations	Units	Mean	Median	SD	Min	Max
Variable of Interest							
Class Size	942599	393601	32.84	34.00	4.60	1	42
Instrumental Variable							
Predicted Class Size (IV)	942599	393601	33.57	34.33	4.63	1	40
Covariates							
Grade Cohort Size	942599	393601	119.39	109.00	59.30	1	356
School Size	942599	393601	356.18	325.00	175.53	4	961
Female (Dummy)	1448907	506308	0.49	0.00	0.50	0	1
No Private Tutoring (Dummy)	1134242	447857	0.32	0.00	0.47	0	1
No Books at Home (Dummy)	1148082	449263	0.11	0.00	0.31	0	1
Outcomes							
Fun Class Life	1430552	504980	3.55	4.00	0.78	1	4
Calm Learning Env	1442988	505181	2.99	3.00	0.89	1	4
Class Events	856704	393309	3.54	4.00	0.70	1	4
Teacher Recog. (Good)	1440739	504751	3.40	4.00	0.74	1	4
Teacher Listened	1437581	504499	3.38	4.00	0.79	1	4
Student Recog. (Good)	855205	392841	3.46	4.00	0.70	1	4
Teacher Explained	855268	393128	3.42	4.00	0.76	1	4

as the "Maimonides Rule," implies that a discrete reduction in class size occurs when grade enrollment exceeds a multiple of 40. Since these discontinuous changes in class size are driven by unpredictable fluctuations in grade size, class size can be regarded as exogenously determined in the neighborhood of each multiple of 40. To utilize this source of variation, (Angrist and Lavy, 1999) constructed a predicted class size based on grade size using the following formula:

$$PS_{ts} = \frac{GS_{ts}}{\text{int}\left(\frac{GS_{ts}-1}{40}\right) + 1} \tag{1}$$

Here, the subscript t denotes the time (year), and s denotes the school. PS is the predicted class size, GS is the total grade size, and `int()` returns the greatest integer less than or equal to the given real number (i.e., the floor function).

Since the predicted class size PS is positively correlated with actual class size, (Angrist and Lavy, 1999) treated PS as an instrumental variable and conducted two-stage least squares (2SLS) estimation to identify the causal effect of class size reduction.

This study adopts a similar design. In Japan, the maximum class size is set at 40 students according to the "Act on Standards for Class Formation and Staffing Levels at Public Compulsory Education Schools"[5]. Therefore, a cohort of 40 students is typically assigned to a single class, while a cohort of 41 students is split into two classes. This study estimates the following system of equations :

$$CS_{tsg} = \theta_0 + \theta_1 PS_{tsg} + \theta_2 GS_{tsg} + \theta_3 GS_{tsg}^2 + \boldsymbol{\pi}^T \boldsymbol{X}_{itsg} + \varepsilon_{icsg} \tag{2}$$

$$Y_{itsg} = \delta_0 + \delta_1 \widehat{CS}_{tsg} + \delta_2 GS_{tsg} + \delta_3 GS_{tsg}^2 + \boldsymbol{\gamma}^T \boldsymbol{X}_{itsg} + \omega_{itsg} \tag{3}$$

I apply two-stage least squares (2SLS) with instrument PS_{tsg} under the standard IV conditions:

1. **Relevance**: $Cov(PS_{sgt},\ C_{isgt}) \neq 0$ (first-stage $F > 10$).
2. **Exogeneity**: $E[PS_{sgt} u_{isgt}] = 0$, ensured by exogenous enrollment fluctuations at thresholds.

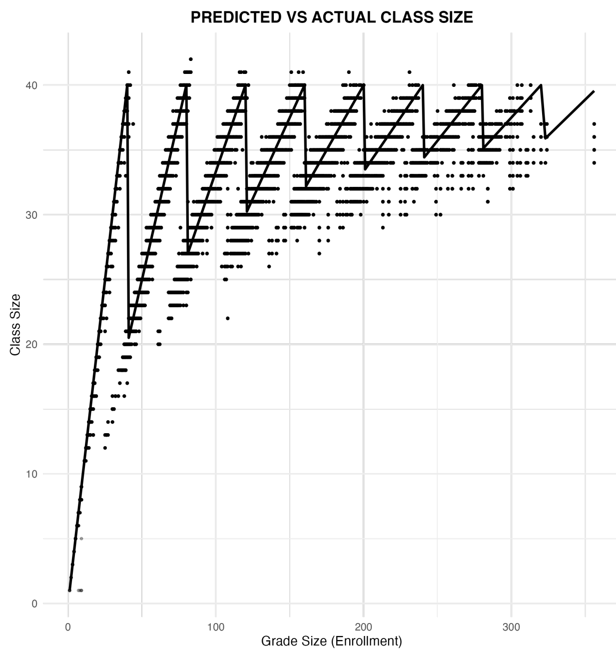
The subscripts i, t, s, and g indicate student, year, school, and grade, respectively. [6] Y represents outcome variables related to classroom climate and teacher–student relationships. $Enroll$ refers to school size, and $\boldsymbol{X}$ is a vector of student- and school-level control variables, including gender, no-books dummy, no-tutoring dummy, as well as grade and year fixed effects. Based on this empirical strategy, I conduct 2SLS estimation under three model specifications: (1) controlling only for student-level covariates, (2) adding school fixed effects, and (3) adding both school and student fixed effects. Assuming the validity of the IV strategy, the IV coefficient δ_1 identifies the causal effect of a one-student reduction in class size on standardized well-being outcomes. For a 10-student reduction, the magnitude is $10\delta_1$ in SD units.

Figure 1 illustrates the relationship between predicted class size (based on Equation 1), actual class size, and grade size. The vertical axis represents class size, while the horizontal axis represents grade size. As grade size increases, predicted class size also increases. The

[5] Since 2011, however, the class size cap for first-grade classes has been reduced to 35.

[6] Although the empirical analysis links class size from year $t-1$ to outcomes in year t, I use t as the time subscript throughout for notational clarity, to avoid confusion with lagged regressions. That is, the class size affecting outcomes in year t is denoted with subscript t.

Figure 1: Relationship between Grade Size and Class Size



Note:
The vertical axis shows the number of students, and the horizontal axis shows grade enrollment. The solid line represents the predicted class size based on Maimonides' rule. The plotted points indicate the actual observed class sizes.

graph suggests a positive correlation between class size and grade size, which is why the empirical models control for grade size when using 2SLS.

5 Results

5.1 2SLS: First Stage

Table 2: First Stage Estimates

	(1) IV	(2) IV (School FE)	(3) IV (School FE + Student FE)
Predict Size	0.823***	0.687***	0.451***
	(0.001)	(0.014)	(0.025)
$F(H_0 : \theta_1 = 0)$	1 965 743.0	2286.0	313.1
Num. Obs.	931 816	931 816	931 816
R2	0.640	0.714	0.891

Note: This table reports the first-stage estimates from two-stage least squares (2SLS) regressions. The dependent variable is actual class size, and the independent variable is the predicted class size calculated based on the Maimonides rule. Standard errors, clustered at the school level, are shown in parentheses. The F-statistic tests the null hypothesis that the coefficient on the instrumental variable equals zero ($H_0 : \theta_1 = 0$). Column (2) includes only school fixed effects, while Column (3) includes both school and student fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 5.1 presents the first-stage estimation results. I estimate three specifications: (1) including only student-level covariates, (2) adding school fixed effects, and (3) adding both school and student fixed effects. In all specifications, the predicted class size (used as the instrumental variable) is significantly correlated with the endogenous variable, actual class size. The results of the F-test for the null hypothesis that the coefficient on the instrument is zero indicate that the instrument satisfies the Stock and Yogo (2005) threshold for a strong instrument ($F > 10$)[7]. Based on these results, I conclude that the predicted class size is a strong instrument and proceed with 2SLS estimation.

Table 3: 2SLS Estimates

Variable	(1) Fun Class Life	(2) Calm Learning Env	(3) Class Events	(4) Teacher Recog. (Good)	(5) Teacher Listened	(6) Student Recog. (Good)	(7) Teacher Explained
Class Size	-0.003***	-0.006***	-0.003***	-0.004***	-0.005***	-0.004***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Grade Size	0.001***	0.000	-0.000	0.001***	0.000	0.000**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Grade Size^2	-0.000***	-0.000**	-0.000	-0.000***	-0.000*	-0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.066***	0.088***	0.083***	0.181***	0.145***	0.134***	0.056***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
No Tutor	-0.029***	-0.061***	-0.026***	-0.032***	-0.017***	-0.009***	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
No Books	-0.082***	-0.062***	-0.084***	-0.140***	-0.096***	-0.125***	-0.103***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	927670	927508	692931	926342	924879	691643	691863

Note:
Standard errors clustered at the school level are shown in parentheses. The table reports second-stage estimates from two-stage least squares (2SLS) regressions. The unit of observation is the student. The estimation sample pools data from 6th and 9th grade students. All regressions include baseline covariates but do not control for school or student fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

5.2 2SLS: Second Stage

Table 3 shows the 2SLS estimation results. Each column presents the estimated effect on one of the outcome variables. In this specification, I control for grade size, grade size squared, gender, cram school attendance, and number of books at home. School and student fixed effects are not included here; results including those effects are shown in Section 5.3.

The coefficients on class size in the first row are all statistically significant. The signs are negative across all outcomes, suggesting that smaller class sizes have positive effects on classroom climate and teacher-student relationships. The magnitude of the effects ranges from 0.02 to 0.06 standard deviations (SD) for a 10-student reduction in class size. The largest effects are found in Column (2): Calm Learning Environment and Column (7): Teacher Explained, both approximately 0.06 SD.

Ito et al. (2020), using the same dataset, report estimates ranging from 0.01 to 0.07 SD for academic outcomes. Senoh and Hojo (2016) report an estimate of about 0.18 SD for a 10-student class size change. The effects estimated in this study are smaller in comparison.

Overall, while class size reductions yield statistically significant improvements in student well-being measures, the magnitudes—less than 0.1 SD—suggest modest intervention effects.

[7] The $F > 10$ rule of thumb is derived under homoskedasticity (Stock and Yogo, 2005). Montiel Olea and Pflueger (2013) propose a more stringent criterion of $F > 23$ under heteroskedasticity when there is a single instrument. The first-stage estimates in this paper exceed both thresholds.

5.3 Estimates Including School and Student Fixed Effects

In this section, I test the robustness of the results from Section 5.2 by estimating alternative specifications: (1) controlling for school fixed effects, and (2) controlling for both school and student fixed effects.

Table 4: 2SLS Estimates (School FE)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variable	Fun Class Life	Calm Learning Env	Class Events	Teacher Recog. (Good)	Teacher Listened	Student Recog. (Good)	Teacher Explained
Class Size	-0.003***	-0.007***	-0.002**	-0.003***	-0.004***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Grade Size	0.000	-0.000	-0.001***	0.001*	0.000	0.000	-0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Grade Size^2	-0.000*	-0.000	0.000**	-0.000***	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.065***	0.087***	0.082***	0.181***	0.145***	0.133***	0.055***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
No Tutor	-0.028***	-0.059***	-0.026***	-0.030***	-0.016***	-0.008***	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
No Books	-0.081***	-0.059***	-0.084***	-0.138***	-0.095***	-0.123***	-0.103***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	927670	927508	692931	926342	924879	691643	691863

Note:
Standard errors clustered at the school level are shown in parentheses. The table reports second-stage estimates from two-stage least squares (2SLS) regressions. The unit of observation is the student. The estimation sample pools data from 6th and 9th grade students. All regressions control for school fixed effects.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

The estimates in Section 5.2 do not account for fixed effects. Here, I re-estimate the models controlling for school and student fixed effects to test whether the results remain stable.

Table 4 reports the results including school fixed effects. This specification uses variation across cohorts within the same school. All coefficients on class size remain statistically significant. The estimated effects range from 0.02 to 0.07 SD for a 10-student class size change. As before, the largest effect appears in Column (2): Calm Learning Environment.

Next, Table 5 reports the results including both school and student fixed effects. In this specification, variation comes from changes across grades for the same student in the same school. This controls for both school- and individual-level unobserved time-invariant factors, and thus provides a conservative benchmark.

While most coefficients remain negative, statistical significance disappears for Column (3): Class Events and Column (7): Teacher Explained. The effect size is largest for Column (2): Calm Learning Environment, approximately 0.07 SD. These results are similar in magnitude to the school fixed effects model, indicating that including student fixed effects does not substantially alter the estimated effects.

Table 5: 2SLS Estimates (School FE + Student FE)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variable	Fun Class Life	Calm Learning Env	Class Events	Teacher Recog. (Good)	Teacher Listened	Student Recog. (Good)	Teacher Explained
Class Size	-0.004**	-0.007**	-0.004	-0.004***	-0.005***	-0.005**	-0.004
	(0.002)	(0.003)	(0.003)	(0.001)	(0.002)	(0.002)	(0.003)
Grade Size	0.001	-0.002	-0.002	0.001	-0.000	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Grade Size^2	-0.000*	0.000	0.000	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.005	-0.015	-0.011	0.037*	0.028	0.021	-0.032
	(0.020)	(0.024)	(0.027)	(0.019)	(0.024)	(0.027)	(0.028)
No Tutor	-0.005**	-0.007**	-0.000	-0.003	-0.003	-0.003	-0.001
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
No Books	-0.026***	-0.029***	-0.024***	-0.033***	-0.033***	-0.027***	-0.033***
	(0.005)	(0.005)	(0.006)	(0.004)	(0.005)	(0.006)	(0.006)
Observations	927670	927508	692931	926342	924879	691643	691863

Note:
Standard errors clustered at the school level are shown in parentheses. The table reports second-stage estimates from two-stage least squares (2SLS) regressions. The unit of observation is the student. The estimation sample pools data from 6th and 9th grade students. All regressions control for school fixed effects and student fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

5.4 Estimates Using Aggregated School-Level Data

Table 7 presents 2SLS estimates using school-level aggregated data[8]. Table 6 shows the first-stage estimates for these models. All specifications satisfy the weak IV criterion of $F > 10$.

All coefficients are negative and significant at the 1% level. The estimated effects range from 0.04 to 0.11 SD for a 10-student reduction in class size. The overall pattern of results is consistent across specifications.

The largest effect (0.11 SD) is observed in Column (2): Calm Learning Environment, and is greater than the estimates based on individual-level data. Hanushek et al. (1996) caution that estimates using aggregated data can be upward biased compared to those using student-level data. Our findings are consistent with that concern.

5.5 Discussion

5.5.1 Heterogeneity

Table 8 reports the estimates of interaction terms between class size and student characteristics, including gender, absence of books at home, and absence of private tutoring. Overall, I find no strong evidence of heterogeneous treatment effects.

There are some suggestive patterns, such as a positive interaction between being female

[8] A similar robustness check was conducted in Ito et al. (2020).

Table 6: First Stage Estimates (School-Level Data)

	(1) IV	(2) IV (School FE)	(3) IV (School FE + Student FE)
Predicted Class Size	0.894***	0.738***	0.738***
	(0.008)	(0.016)	(0.016)
F($H_0 : \theta_1 = 0$)	13 463.6	2039.5	2039.7
Num. Obs.	11 714	11 714	11 714
R^2	0.831	0.878	0.878

Note: This table reports the first-stage estimates from two-stage least squares (2SLS) regressions using school-level data. The dependent variable is the average class size at the school level, and the independent variable is the predicted class size calculated based on the Maimonides rule. Standard errors, clustered at the school level, are shown in parentheses. The F-statistic tests the null hypothesis that the coefficient on the instrumental variable equals zero ($H_0 : \theta_1 = 0$). Column (2) includes school fixed effects, and Column (3) includes both school and student fixed effects. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: 2SLS Estimates of Class Size Effects (School-Level Data)

Outcome	(1) Fun Class Life	(2) Calm Learning Env	(3) Class Events	(4) Teacher Recog. (Good)	(5) Teacher Listened	(6) Student Recog. (Good)	(7) Teacher Explained
Class Size	-0.004***	-0.011***	-0.004***	-0.004***	-0.005***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	11709	11709	8898	11709	11711	8898	8895

Note: Standard errors clustered at the school level are shown in parentheses. The table reports second-stage estimates from two-stage least squares (2SLS) regressions. The unit of observation is the school. The estimation sample pools data from 6th and 9th grade students. All regressions control for school fixed effects.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

and teacher recognition, and between lacking books at home and enjoying class life. However, these findings are difficult to interpret.

Table 8: Heterogeneous Effects of Class Size by Student Covariates

	Fun Class Life	Calm Learning Env	Class Events	Teacher Recog. (Good)	Teacher Listened	Student Recog. (Good)	Teacher Explained
Female	-0.007*** (0.002)	-0.001 (0.001)	0.008*** (0.001)	-0.112*** (0.005)	0.001 (0.001)	-0.002*** (0.001)	-0.000 (0.000)
No Book	-0.045*** (0.002)	0.047*** (0.011)	0.002*** (0.001)	0.001 (0.001)	0.008*** (0.000)	0.001 (0.001)	0.003*** (0.000)
No Tutor	0.002*** (0.000)	0.002*** (0.000)	0.000 (0.000)	0.013*** (0.001)	0.001 (0.000)	0.002*** (0.000)	-0.000 (0.000)

Note:
Estimated coefficients for class size × covariate interactions from 2SLS regressions. Standard errors clustered at the school level in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

5.5.2 Distribution and Variation of Outcomes

This subsection examines the distribution and variability of the well-being outcomes related to students' school life used in this study. Figure 2 presents histograms showing the distribution of each outcome variable. The distributions are generally skewed toward higher values, suggesting that many students responded with either a 3 or 4 on the 4-point Likert scale.

Figure 2: Distribution of Outcomes

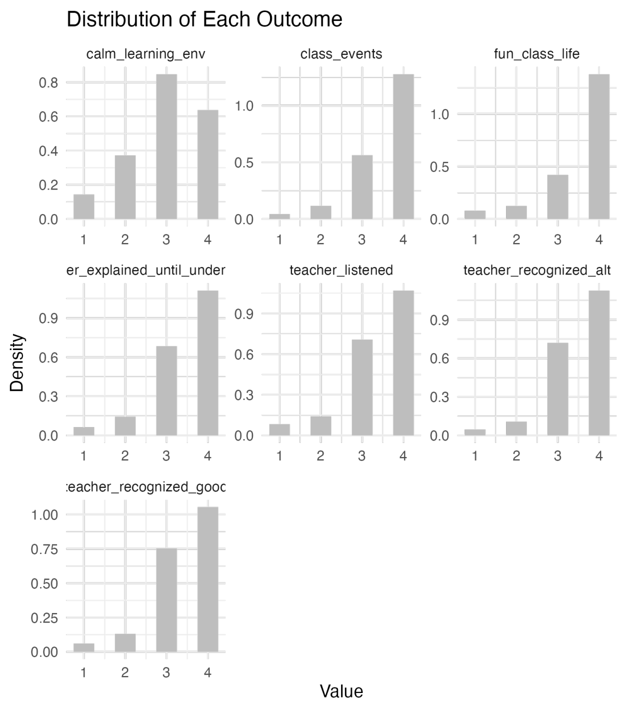
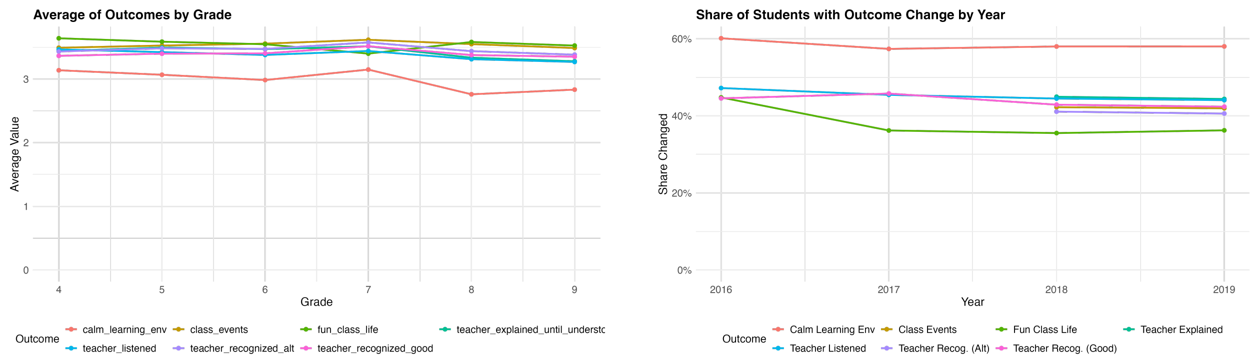


Figure 5 illustrates the trends of each outcome by school grade. The results suggest limited variation across grades for most outcomes. However, for *Calm Learning Env*, a noticeable decline is observed from Grade 8 onward, indicating greater variability compared to the other measures.

Figure 3 plots the yearly share of students whose responses changed from the previous year for each outcome variable. For six of the seven outcomes (excluding *Calm Learning Env*), approximately 40% of students changed their responses year over year, while the remaining 60% did not. In contrast, *Calm Learning Env* exhibited more variation, with more than 60% of students reporting a change in their response each year—suggesting that this outcome is relatively more sensitive to year-to-year fluctuations.

Taken together, the descriptive analyses indicate that the well-being variables used in this study exhibit limited dispersion and only modest temporal variation. These characteristics imply that detecting large treatment effects using these data may be inherently difficult.



(a) Variation by Grade

(b) Share of Students with Outcome Change by Year

Figure 3: Variation of Outcomes

6 Conclusion

This paper has examined the effect of class-size reduction on classroom climate and teacher–student relationships using data from Japan. Leveraging an instrumental variable approach based on the Maimonides rule—which stipulates a maximum class size of 40 students—I estimate that a reduction of 10 students per class leads to an improvement of up to 0.07 standard deviations in well-being outcomes related to the classroom environment and teacher relationships. Compared to prior studies in Japan that use the same dataset and estimation strategy, this effect size of 0.07 SD is relatively modest. Therefore, our findings do not provide strong evidence that class-size reduction substantially improves students' well-being.

This study contributes to the literature in two main ways. First, it adds new empirical evidence using Japanese data. Second, it examines the impact of class-size reduction on non-academic outcomes, using large-scale micro-level panel data at the student level. From a policy perspective, it is important to evaluate whether class-size reduction has significance beyond the traditional focus on academic achievement. Future policy discussions should adopt a broader perspective that considers diverse outcomes of class-size reduction, such as its effects on the school environment and student well-being.

In this regard, accumulating further evidence—including studies that assess the interaction between class-size reduction and other educational policies will be essential to fully understand the potential benefits and limitations of small-class policies.

References

Akabayashi, H. and R. Nakamura (2014) "Can Small Class Policy Close the Gap? An Empirical Analysis of Class Size Effects in Japan", *Japanese Economic Review*, **65** (3), pp. 253–281.

Angrist, J. D. and V. Lavy (1999) "Using Maimonides' Rule to Estimate the Effect of Class Size on Student Achievement", *The Quarterly Journal of Economics*, **114** (2), pp. 533–575.

——— , E. Battistin, and D. Vuri (2017) "In a Small Moment: Class Size and Moral Hazard in the Italian Mezzogiorno", *American Economic Journal: Applied Economics*, **9** (4), pp. 216–249.

Bettinger, E. P., L. Fox, S. Loeb, and E. S. Taylor (2017) "Virtual Classrooms: How Online College Courses Affect Student Success", *American Economic Review*, **107** (9), pp. 2855–2875.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011) "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR", *The Quarterly Journal of Economics*, **126** (4), pp. 1593–1660.

Coleman, J. S. (1968) "Equality of Educational Opportunity", *Integrated education*, **6** (5), pp. 19–28.

Fredriksson, P., B. Öckert, and H. Oosterbeek (2013) "Long-Term Effects of Class Size", *Journal of Human Resources*, **48** (1), pp. 1–29.

Hanushek, E. A. (1999) "Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects", *Educational Evaluation and Policy Analysis*, **21** (2), pp. 143–163.

——— (2003) "The failure of input-based schooling policies", *The economic journal*, **113** (485), pp. 64–98.

——— and L. Woessmann (2017) "Cognitive Abilities and Educational Outcomes", in *Cognitive Abilities and Educational Outcomes: A Festschrift in Honour of Jan-Eric Gustafsson*: Springer, pp. 149–171.

Hojo, M. (2013) "Class-size Effects in Japanese Schools: A Spline Regression Approach", *Economics Letters*, **120** (3), pp. 583–587.

——— (2021) "Association between student-teacher ratio and teachers' working hours and workload stress: evidence from a nationwide survey in Japan", *BMC public health*, **21**, pp. 1–8.

——— (2023) *The Economics of Small Class Sizes: A Vision for Evidence-Based Education Policy*, Tokyo, Japan: Keio University Press, in Japanese.

Hoxby, C. M. (2000) "The Effects of Class Size on Student Achievement: New Evidence from Population Variation", *The Quarterly Journal of Economics*, **115** (4), pp. 1239–1285.

Ito, H., M. Nakamuro, and S. Yamaguchi (2020) "Effects of Class-size Reduction on Cognitive and Non-cognitive Skills", *Japan and the World Economy*, **53**, p. 100977.

Ito, H., M. Hamada, Y. Murayama, N. Takayanagi, K. Nomura, M. Myogan, and M. Tsuji (2017) "Effects of Class Size on Academic Achievement and Emotional and Behavioral Problems: Causal Inference Using a Natural Experiment and Multilevel Modeling", *The Japanese Journal of Educational Psychology*, **65** (4), pp. 451–465, In Japanese.

Kaiser, C. and A. J. Oswald (2022) "The Scientific Value of Numerical Measures of Human Feelings", *Proceedings of the National Academy of Sciences*, **119** (42), p. e2210412119.

Krueger, A. B. (1999) "Experimental estimates of education production functions", *The quarterly journal of economics*, **114** (2), pp. 497–532.

Nakamuro, M. (2017) "Do Small Classes Reduce Bullying, Violence, and Truancy?", RIETI Discussion Paper Series 17-J-074, Research Institute of Economy, Trade and Industry (RIETI), pp. 1–37, in Japanese.

Niki, M. (2013) "The Effects of Class Size on Cognitive and Non-Cognitive Abilities", *Economic Analysis*, **186**, pp. 30–49, in Japanese.

OECD (2017) *PISA 2015 Results (Volume III): Students' Well-Being*, Paris: OECD Publishing, PISA.

——— (2021) *Beyond Academic Learning: First Results from the Survey of Social and Emotional Skills*, Paris: OECD Publishing.

Oikawa, M., R. Tanaka, S. ichiro Bessho, and H. Noguchi (2022) "Do Class Size Reductions Protect Students from Infectious Diseases? Lessons for COVID-19 Policy from a Flu Epidemic in the Tokyo Metropolitan Area", *American Journal of Health Economics*, **8** (4), pp. 449–476.

Senoh, W. and M. Hojo (2016) "Does the class size affect the academic achievement of junior high school students? : an empirical study with "National Assessment of Academic Ability" of Japan", *NIER Research Bulletin*, **145**, pp. 119–128.

Stock, J. H. and M. Yogo (2005) "Testing for Weak Instruments in Linear IV Regression", in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*: Cambridge University Press, pp. 80–108.

Sugita, S. (2025) "The Effect of Class-Size Reduction on Student Satisfaction with School Life", *KEIO SFC JOURNAL*, **24** (2), pp. 161–171, In Japanese.

Urquiola, M. (2006) "Identifying Class Size Effects in Developing Countries: Evidence from Rural Bolivia", *The Review of Economics and Statistics*, **88** (1), pp. 171–177.

A Appendix

A.1 Results for Binary Outcomes

Table 9 presents the 2SLS estimation results using binary outcome variables for student well-being. In this specification, each outcome is coded as 1 if the student responded with a 3 or 4, and 0 otherwise. Since we estimate a linear probability model using these binary variables, the coefficients represent percentage point changes in the probability of reporting a high level of well-being in response to a one-student reduction in class size.

As shown in Table 9, the pattern of coefficient signs and statistical significance remains largely unchanged. A class-size reduction of 10 students is associated with an increase of approximately 1 to 3 percentage points in the likelihood of students reporting high well-being. According to the results in Section 5.5.2, *Calm Learning Environment* is the outcome with the greatest year-to-year variation. This is consistent with the result in this section, where the largest treatment effect is also observed for *Calm Learning Environment*.

Table 9: 2SLS Estimates (School FE, Binary Outcomes)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variable	Fun Class Life	Calm Learning Env	Class Events	Teacher Recog. (Good)	Teacher Listened	Teacher Recog. (Alt)	Teacher Explained
Class Size	-0.001**	-0.003***	-0.000	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Grade Size	0.000	-0.000	-0.000***	0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Grade Size^2	-0.000	-0.000	0.000*	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.016***	0.037***	0.024***	0.053***	0.043***	0.042***	0.022***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
No Tutor	-0.008***	-0.024***	-0.004***	-0.006***	-0.002***	0.001	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
No Books	-0.024***	-0.022***	-0.022***	-0.042***	-0.032***	-0.033***	-0.034***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	927670	927508	692931	926342	924879	691643	691863

Note:
Standard errors clustered at the school level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

A.2 Local Linear Regression (Bandwidth ± 5)

This supplementary analysis reports the estimation results using a sample restricted to within five students of the cutoff. A similar robustness check using the same data is conducted in Ito et al. (2019).

Table 10 shows the results from a local linear regression using only the sample within five students of the cutoff. Because comparisons between observations far from the cutoff may not be valid, the estimation is conducted using samples as close to the cutoff as possible. Statistical significance disappears across all outcomes, and the signs of the estimated coefficients for (6) Teacher Recog. (Alt) and (7) Teacher Explained turn positive. As discussed later, it is possible that the outcome variables used in this paper exhibit limited variation. Therefore, restricting the sample to within ± 5 students of the cutoff substantially reduces the sample size, which likely lowers the precision of the estimates. In fact, in the results shown in Table 5, where school and student fixed effects are included, the estimate in column (7) also loses statistical significance.

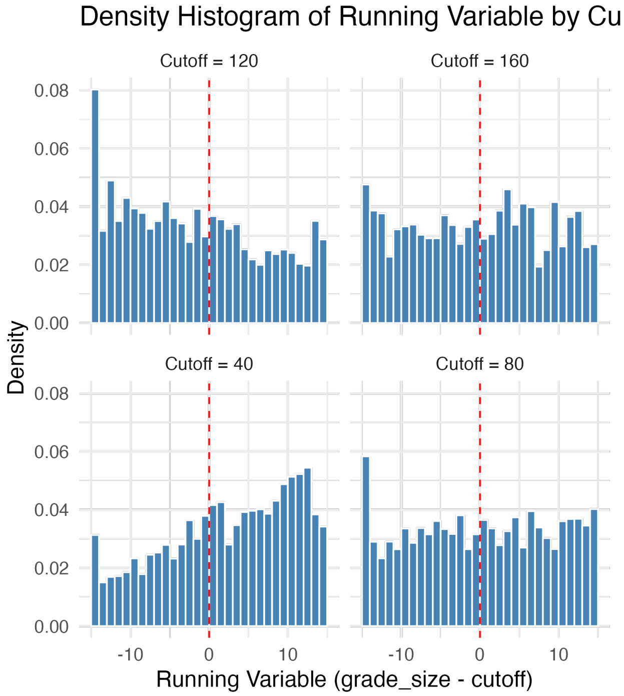
Table 10: Local Linear RD Estimates (± 5 Bandwidth)

	(1) Fun Class Life	(2) Calm Learning Env	(3) Class Events	(4) Teacher Recog. (Good)	(5) Teacher Listened	(6) Teacher Recog. (Alt)	(7) Teacher Explained
Treatment Effect	-0.004 (0.013)	-0.008 (0.028)	-0.021 (0.019)	-0.002 (0.013)	-0.008 (0.016)	0.008 (0.017)	0.022 (0.021)
Observations	140351	140346	105611	140135	139969	105412	105467

Notes: Standard errors clustered at the school level in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

A.3 Checking for Manipulation of the Running Variable

Figure 4: Relationship between Grade Size and Class Size



Note:
The dashed lines indicate cutoff points at multiples of 40. The histogram is based on samples with g
students of each cutoff.

A.4 Mechanism: Teacher and Students' Well-Being

Figure 5: Distribution of Teacher Outcomes by Class Size Group