RIETI Discussion Paper Series 24-E-004

How do Classmates Matter for the Class-size Effects?

TANAKA, Ryuichi
RIETI

WANG, Tong
Ritsumeikan University

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 24-E-004

January 2024

How do Classmates Matter for the Class-size Effects?[1]

Ryuichi Tanaka
University of Tokyo / RIETI
Tong Wang
Ritsumeikan University

Abstract

This paper studies the effect of class-size reduction on students' academic outcomes with a special focus on its heterogeneity based on classmates' characteristics. We estimate the causal effects of class-size reduction on students' mathematics and language test scores, controlling student-teacher fixed effects and applying the predicted class size with a cap as an instrument for the actual class size. Using rich panel data on Japanese primary school students, we find that the average effect of class size reduction is positive and robust for math test scores and that classes with high-ability classmates benefit even more from class size reduction. We find that the effect of class size reduction depends positively on the ability of the student with the lowest rank in a class. In addition, we find that classes with a high share of female students benefit more from class size reduction. Our findings provide strong support for the theoretical framework of Lazear (2001).

Keywords: Education, test scores, class-size reduction, ability, heterogeneity.
JEL classification: J13, J18, N35

[1]This study is conducted as a part of the project "Microeconometric Analysis of Education Policy" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). The draft of this paper was presented at a DP seminar of the Research Institute of Economy, Trade and Industry (RIETI). The authors are grateful for valuable comments from professors Daiji Kawaguchi, Masayuki Morikawa, and Shujiro Urata at the seminar. The authors acknowledge financial support for this project from the Japan Society for the Promotion of Science (Grant Numbers 20H05629 for Ryuichi Tanaka).

1 Introduction

Class-size reduction is one of the typical policy measures to improve the quality of educational environments. It is believed that small classes are equipped with richer educational resources per student and thus produce better educational outcomes. However, the effectiveness of class-size reduction is difficult to be observed empirically to a large extent (e.g., Hanushek (1986), Hanushek (2003), Hanushek (2006)). Although many papers found that small class produces better academic outcomes (e.g., Angrist and Lavy (1999); Krueger (1999), Krueger (2003)), the magnitudes of the *average* effects are limited and sometimes insignificant (e.g., Angrist et al. (2019)).

One potential reason for the small or insignificant effects of class size reduction is the heterogeneity of the class compositions: since class size reduction may be better for classes of some types but worse off for others, the average effect could be small or insignificant. Similar arguments apply to heterogeneity within a class: class size reduction may benefit some students meanwhile hurt other students. Therefore, it is important to understand how the class-size reduction effects depend on the composition of students in the class.

The component of classmates matters for the effectiveness of class-size reduction for at least two reasons. First, even without interaction among classmates, if the class-size effect is heterogeneous by students' characteristics such as baseline ability and gender, the average effect varies by the composition of students. More importantly, in the context of schooling where students interact with each other, the distribution of students' characteristics (for example, gender and ability of classmates) matters for the class-size reduction effects through various channels, such as peer effects. In the model of Lazear (2001) where schooling is modeled as a joint production by students, one student's failure to behave well leads to the disruption of the entire education process in the class, deteriorating the educational outcomes of all classmates. If schooling is a process with such strong complementarity among classmates, the component of classmates generates heterogeneity of class size effects.

We study the heterogeneous effect of class-size reduction on students' academic outcomes. As the source of heterogeneity of the classes, we focus on the distribution of the baseline academic outcomes measured by test scores in the previous grade (hereafter, call it ability) and the gender of students. In the spirit of Lazear's model, the individual probability of students disturbing the education process is one of the fundamental determinants of educational outcomes. We focus on the ability and gender of students as proxies for this probability. Generally, students' ability and misbehavior in classrooms are negatively correlated (e.g., Myers et al. (1987)). On the gender of students, boys may have a higher probability of disturbing classes than girls because, for example, boys are more likely to bully and cyberbully than girls (Li (2006)), and because boys are more likely to be referred, diagnosed, and treated for Attention-Deficit Hyperactivity Disorder (ADHD) symptoms than girls (e.g., Gaub and Carlson (1997); Gershon and Gershon (2002)). We estimate the heterogeneous effect of class-size reeducation by (1) the average, the maximum, and the minimum ability of classmates, and (2) the share of female students in a class.

Our source of the data is administrative data of primary school students in a large municipality in the Tokyo metropolis of Japan collected from the year 2010 to 2016. The data contain the panel information of academic performance measured by a standardized test, linked with the information of teachers and socioeconomic background such as eligibility for school financial assistance (similar to free lunch program) of students in the second to sixth grades.

To identify the causal effects of class-size reduction on students' outcomes, it is necessary to use either random variations of class size within a school (e.g., Krueger (1999)) or quasi-random variations created by an institutional setting (e.g., Angrist and Lavy (1999)). In the absence of randomized controlled formation of class size in our data, we follow the latter approach with quasi-experimental variations of class size. Japanese primary schools are subject to the rule on the class size set by the Ministry of Education, the central authority of education policy in Japan, that the class size is capped at 35 for the second graders and 40 for third to sixth graders. Similar to the identification strategy of Angrist and Lavy (1999), we use the predicted class size from the grade size (so-called Maimonides' rule) as an instrument for the actual class size. In addition, we control for the teacher-student fixed effects to address the issue of the potential threat to the identification of the class-size reduction effect due to endogenous matching between students and teachers. In Japanese primary schools, it is typical that class composition, as well as classroom teachers, are shuffled and reassigned when students progress to the next grade, generating variations of class size even when a student is taught by the same teachers in consecutive grades. Controlling the teacher-student fixed effects, we identify the causal effect of class-size reduction from the change of class size for the students who are taught by the same teachers. This strategy, combined with the quasi-random variation of class size generated by the class size cap, makes our identification strategy solid.

As a result of our estimation analyses, we confirmed that smaller classes are better for students' academic performance on average, and it is stronger for mathematics than Japanese subjects. A marginal reduction in class size can lead to an increase in 0.00938 standard deviations in mathematics scores and 0.0029 standard deviations in Japanese scores. We also find that students in classes with higher average baseline academic performance benefit more from class size reduction. There is an unambiguously significant heterogeneity effect of class-size reduction on the classes with different mean baseline test scores - classes with higher baseline tests will benefit more from class size reduction for both mathematics and Japanese. In addition, our finding shows that classes composed of a higher percentage of female students benefit more from class size reduction. Moreover, the effect of class size reduction is more sensitive to the ability of the bottom student than the ability of the top student. Classes whose bottom student has a better baseline score will benefit significantly more from class size reduction than the classes whose bottom students have a lower baseline score.

The remainder of the paper is as follows. In Section 2, we review the literature and state our contributions. Section 3 explains the institutional background and our empirical strategy. Section 4 explains the data. Section 5 reports our findings. Section 6 discusses the interpretation of our findings. Section 7 concludes.

2 Literature Review and Our Contribution

There is a strand of literature on the relationship between class size and student performances using observational data. With the experimental data from Project STAR in Tennessee of the U.S., Krueger (1999) finds a positive effect of small class size. Although it is unclear how percentile scores test scores map into tangible outcomes, relative to the standard deviation of the average percentile score, the effect sizes are .20 in kindergarten, .28 in first grade, .22 in second grade, and .19 in third grade. In addition, a series of meta-analyses by Hanushek (1986), Hanushek (2003), Hanushek (2006) find no consistent relationship between class size and student outcomes without a quasi-experimental setting.

As the first paper estimating the class size effect with observational data in a quasi-experimental setting, Angrist and Lavy (1999) find a negative relationship between class size and student academic performances measured by test scores in Israeli public primary schools. Israeli public primary schools are subject to the regulation on the maximum class size ('Maimonides' Rule), generating a discontinuous change of class size around the multiples of the class size cap. They show that a regression discontinuity design is an effective identification strategy for the causal effect of class size. Following that, there are many papers applying a regression discontinuity design to observational data and finding a positive effect of class size reduction (e.g., Urquiola (2006), Browning and Heinesen (2007); Akabayashi and Nakamura (2014); Hojo and Senoh (2019); and Gilraine (2020) among many).[1] Hoxby (2000)

[1] A recent study by Angrist et al. (2019) show that the effect was insignificant, which cast doubt on the effectiveness of class size reduction. Similarly, Ito et al. (2020) find no effect of class size reduction in Japanese compulsory schools.

used a long panel data set in which the class size change is driven by idiosyncratic variation, and also applied the 'Maimonides' Rule, and found that class size has an insignificant effect on students' academic performance. Whilst these papers contribute to the discussion on the effect of class size reduction, none has analyzed how the effect varies for different types of classes with different student compositions. Applying the same identification strategy to observational data, our paper contributes to the literature by showing how the class size effect varies by the distribution of student ability.

Our paper is closely related to the papers on the heterogeneous effects of class size reduction with different identification strategies. Using a rich data set at a university in the UK, Bandiera et al. (2010) showed a heterogeneous marginal effect of class size reduction when the composition of students' ability in the class varies. Their identification strategy relies on the control of student and teacher fixed effects, respectively. Controlling student and teacher fixed effects, however, may suffer from a potential endogeneity of matching between students and teachers. In addition, there is no effective class size cap in the university, and thus the application of a regression discontinuity design is infeasible. In our paper, we conquer the identification issue by applying a regression discontinuity design and controlling for the fixed effects for the matched pair of student and teacher. Kedagni et al. (2021) applied structural estimation methodologies to Greek administrative data and found that the effect of class size on academic achievement is hump-shaped. They quantified the cost and benefit of hiring and firing a teacher; there is no analysis of the heterogeneity in the effect of class size reduction by students' characteristics. With experimental data, Ding and Ding and Lehrer (2011) showed that there is a heterogeneous effect of class size reduction by teachers' characteristics. While they used experimental data, they simply grouped the classes into big classes and small classes, which made it difficult to analyze the marginal change in the class size. Applying a solid identification strategy to observational data, our paper contributes to the literature by showing the heterogenous effects of class size reduction on students' academic performances.

We also provide discussions on the interpretation of our findings based on a theoretical model built on Lazear (2001). The differences in the results of the class size effects can be explained by the heterogeneity of class composition and the heterogeneity of the teachers. To explain the mechanism of how class size affects students' academic performance, Lazear (2001) interprets the education production process as a combat against students' disturbance behavior. The larger the class, the higher possibility that the education process is disturbed by a student. To our knowledge, our findings provide the first supportive evidence for the theoretical framework of Lazear (2001).

3 Institutional Background

Japanese compulsory education is based on the Constitution, specifically, the Fundamental Law of Education, promulgated in 1947. Compulsory education consists of six years of primary education in elementary school and three years of lower secondary education in middle school. Children who are six years old start first grade in elementary school on April 1 and receive schooling for nine years in total compulsory. The majority of elementary and middle schools are publicly financed and run by the education board of the local municipality. Each municipality establishes school districts and assigns students to designated public schools based on their residential addresses.[2]

Public elementary (grades one to six) schools in Japan have an upper limit of class size set by the Act on Standards for Class Formation and Fixed Number of School Personnel of Public Compulsory Education Schools. The law allows education boards of governments to set an original upper limit to class size as long as the limit is below the national standard. Primary schools have an upper limit of thirty-five students for the first and second grades and forty students for the other grades (third, fourth, fifth, and sixth grades) before 2021. Students experience changes in the upper limit of class sizes and class reshuffle simultaneously. Students who were previously in the same class for some years get reshuffled, which commonly happens when students go from second to third grade and when students get promoted from elementary to middle school.

All teachers in Japanese elementary schools are required to obtain a license as a teacher regardless of whether the school is national, public, or private. Teachers are assigned to schools by the education board of the prefectural government with authority over teachers' personnel issues. Since teachers with a range of three to seven years of teaching in a school are transferred to another school, public school teachers cannot self-select into schools for a long duration. The assignment of teachers to classes within schools is at the discretion of the school principal, and thus the matching between teachers and students/classes can be endogenous. To account for the potential endogeneity of matching between students and teachers, we control teacher-student fixed effects in the empirical analysis. The assigned teachers teach all subjects to students in the class.

[2]Students are also allowed to attend private or public schools run by the national government. In the case of school attendance at private and national schools, students need to take entrance examinations.

4 Empirical Strategy

We estimate the effect of class size reduction using the following regression model:

$$Y_{ijcgst} = \beta_0 + \beta_1 C_{jcgst} + \gamma X_{ijcgst} + f(E_{gst}) + d_{ij} + d_g + d_s + d_t + \epsilon_{ijcgst} \quad (1)$$

where Y_{ijcgst} is the outcome variable (i.e. student academic performance) for student i, in class c taught by teacher j, grade g, school s, and year t. C_{jcgst} is the size of class c taught by teacher j, in grade g at school s in year t. X_{ijcgst} is the vector of observable characteristics of students (e.g. gender and socioeconomic status of their household), class characteristics (e.g. average academic performance of peers), teacher characteristics (e.g. teaching experience). $f(E_{gst})$ is polynomial of enrollment in grade g at school s in year t, E_{gst}, and we include upto the third polynomial of grade size. d_g, d_s, d_t are the grade, school and year fixed effects, respectively. d_{ij} represents the fixed effects for student-teacher pairs. ϵ_{ijcgst} is the idiosyncratic error term.

Inclusion of the student-teacher fixed effects is crucial for the identification of the class size effects, because it rules out the effect of sorting between teachers and students based on unobservable factors. If we control student fixed effects and teacher fixed effects separately, we control for the unobserved characteristics of the students and teachers, separately. However, since we have discussed in the background, teachers and students are reshuffled every year and how teachers should be matched to classes is determined by the manager of the school (school principal). Moreover, the rule of teacher assignment to classes is highly school specific, leading to endogenous matching based on unobservables of teacher and student characteristics. We identify the effect of class size employing the variation of class size within a pair of teacher and student: when we observe a student taught by the same teacher, but with different class size across grades, the variation in the academic outcome is caused by the variation of class size. In the next section, we will show that there are enough variations in the class size in order to estimate the effect of class size reduction.

Another threat against the identification of class size effect, i.e. the coefficient β_1, is the potential endogeneity of class size. For example, the districts with higher average student academic performance may be endowed with rich educational environment out of schools such as cram schools and supply of private tutoring, which attracts more families to locate there, and therefore, more students will enroll into schools, resulting in larger class size. To handle this potential endogeneity problem, we follow Angrist and Lavy (1999), using the cap of class size (so called Maimonides' rule) to calculate a predicted class size as an instrument for the (possibly endogeneous) actual class size variable. Given the number of students enrolled in grade g at school s in year t, assuming that classes are divided almost equally, we have

$$\hat{C}_{jcgst} = \frac{E_{gst}}{int[\frac{E_{gst}-1}{\bar{C_{gt}}}]+1} \tag{2}$$

where $\bar{C_{gt}}$ is the maximum possible number of students of a class. In our data, $\bar{C_{gt}} = 35$ for first and second graders since year 2012, and $\bar{C_{gt}} = 40$ for the rest grade and before 2012.

To study the heterogeneity of class size effect by the baseline ability of classmates, we include interaction terms between class size and the average baseline score of the peers in a class ($\bar{Y}_{-i,jcgs,t-1}$), i.e. the average score of all students except student i herself, in class c taught by teacher j in grade g at school s in time period $t-1$:

$$Y_{ijcgst} = \beta_0+\beta_1 C_{jcgst}+\beta_2 C_{jcgst}\bar{Y}_{-i,jcgs,t-1}+X_{ijcgst}\gamma_1+\bar{Y}_{-i,jcgs,t-1}\gamma_2+f(E_{gst})+d_{ij}+d_g+d_s+d_t+\epsilon_{ijcgst}. \tag{3}$$

In addition, to explore further how the distribution of classmates' ability affects the class size effects, we estimate the model with the interaction term between cladss size and the maximum and minimum baseline score of a class, $Y^{max}_{jcgs,t-1}$ and $Y^{min}_{jcgs,t-1}$, respectively.

$$Y_{icgst} = \beta_0+\beta_1 C_{jcgst}+\beta_2 Y^{max}_{jcgs,t-1}+\beta_3 Y^{max}_{jcgs,t-1}C_{jcgst}+\beta_4 Y^{min}_{jcgs,t-1}+\beta_5 Y^{min}_{jcgs,t-1}C_{jcgst}+f(E_{gst})+d_{ij}+d_g+d_s+d_t+\epsilon_{ic} \tag{4}$$

Furthermore, we investigate the heterogeneity of class size effect by the percentage of female students in class:

$$Y_{ijcgst} = \beta_0+\beta_1 C_{jcgst}+\beta_2 C_{jcgst}P_{jcgst}+X_{ijcgst}\gamma_1+P_{jcgst}\gamma_2+f(E_{gst})+d_{ij}+d_g+d_s+d_t+\epsilon_{ijcgst} \tag{5}$$

where P_{jcgst} is the percentage of female students in a class. We estimate the models with interaction terms, using the predicted class size interacted with these variables as instruments for these interaction terms.

5 Data

This paper utilizes the administrative data collected by the Education Board of a city from year 2010 to 2016.[3] The city is a large municipality in the Tokyo Metropolis with more than 300,000 households and a population of more than 600,000 people, as of 2015. We collected data from all public primary schools (74 schools) in total from the city. Our data set includes students from second grade to sixth grade.

Our data comprises of three parts: student academic test data, student socioeconomic status data and teacher survey data. Student academic test data includes standard tests

[3]Due to the confidentiality agreement, we cannot identify the city by name in our paper.

that evaluates students' academic achievement. For students from second to sixth grade, we have their Japanese and mathematics scores. Students' socioeconomic status data displays students' conditions of receiving financial aids from the government. Students with lower income and/or family tragedies (e.g. parents' divorces) will receive government aids. In the data, a student's socioeconomic status is indicated by the dummy variable of whether she receive any type of aid.

Table 1 presents the summary statistics of students' test scores, class size, and grade size. Students' mathematics scores and Japanese scores are standardized with mean 0 and standard deviation 1 within grade and year. The average grade size (total student of a grade at a school) is around 79 students, and the average class size is about 31 students. The standard deviation of grade size is about 28 students, which shows that schools vary quite a lot in the grade size.

Table 1: *Summary Statistics*

VARIABLES	mean	std deviation	min	max
Math score	0.009	0.987	-5.941	1.782
Japanese score	0.0096	0.987	-5.787	2.051
Class size	31.482	4.364	10	41
Grade size	78.776	28.238	10	217
Female share	0.493	0.4999	0	1
Share of school financial assistance receivers	0.349	0.477	0	1
Total teaching experience	11.776	10.398	1	44
Tenure at the current school	3.458307	2.052	0	17
Teacher age	37.268	10.248	22	65
Baseline ability of language	-0.002	0.266	-1.511	0.877
Baseline ability of math	-0.001	0.274	-1.237	0.861
Observations	149,727			

Next, we will then show the summary statistics of the class size variation of the same student taught by the same teacher. This is important because our identification of the effect of class size reduction can only be achieved if the students taught by the same teacher across different years experience variation in the class size. In Table 2, in each row, we look at the difference of class size between the current year and the previous year, the current year and two and three years ago, respectively. The first row shows that the class size change in the consecutive two grades is 0.635 on average conditional on that students are taught by the same teacher in these grades. The average change of class size is small because many experiences no change in the class size for the consecutive two grades. However, the class size cap generates a big change of class size: some students experienced reduction of class size by 13 students, and others experienced increase of class size by 21 students. Increasing the consecutive years from two to a larger numbers, we can see the mean of the class size variation is increasing with the number of years back. The standard deviation and the minimum and maximum number of class size changes within a pair of student and teacher are source of identification in our setting.

Table 2: *Summary statistics of class size of students taught by the same teacher*

Changes of class size	obs.	mean	std. dev.	min	max
1-year difference	22,265	0.635	2.271	-13	21
2-year difference	2,382	1.228	3.234	-4	20
3-year difference	670	1.978	3.610	-2	20
4-year difference	169	2.349	3.047	-2	8

Table 3: *Summary statistics of total number of students taught by the same teacher*

Observed year	Number of Students	Percent
1 year	104,413	69.74%
2 years	40,440	27.01%
3 years	4,110	2.74%
4 years	744	0.50%
5 years	20	0.01%

Although the class compositions are typically reshuffled when students move from one grade to another, the chance of being taught by the same teacher in the consecutive two grade is far above zero. Table 3 shows the number of students taught by the same teacher in multiple years (not necessarily consecutive). As is shown, the vast majority of students in the sample changed teachers every year, which comprised of 69.74% of all students in the sample. This implies that our identification of the class size reduction effect is achieved from the rest 31.68% of students. Among the students that experienced the same teacher for multiple years, the majority of them experienced the same teacher for two years, which comprised of 27.01% of the total students.

As we have mentioned, the classes in primary schools are shifted each year, so the students in each class may change from year to year. Since there are no explicit rules to assign to students to classes, we emphasize here that the differences in the average baseline scores are insignificant across classes within a grade of a school. To show this point, we conducted a pair-wise t-test, i.e. for each year in each grade of a school, we test whether the mean baseline score of each class is equal. In the pairwise t-tests, we use the Welch method with a Bonferroni correction to the significance level. Our main focus is the t-statistics.

We test the balance in the assignment of classes in the year-school-grade level. In any year-school-grade, if there is only one class, we don't need to test. So we focus on those with more than one class, and the maximum number of classes a year-school-grade had was 8. Row 1 to 7 of Column 1 of Table 4 show the number of year-school-grades with 2 to 8 classes, respectively. Column 2 of Table 4 shows the number of class pairs that have the absolute value of t-statistics greater than 2. We can see that very few percent of the pairs of classes are assigned with unevenly distributed baseline test scores. Therefore, we can assume that when we include the mean of the peers' (excluding the student herself) baseline test scores in our regressions, we can largely avoid the endogeneity in the class assignment. Due to the small fraction of the number of peers with t-statistics greater than 2, we can assume the unbalanced pairs were generated randomly.

Table 4: *Pairwise t-test for balancing of the baseline class test scores*

	Number	t-stat > 2
2 classes	1308	1%
3 classes	752	3%
4 classes	87	4%
5 classes	11	5%
6 classes	5	2%
7 classes	2	10%
8 classes	1	0

6 Results

We apply our benchmark model (1) to evaluate the effect of class size reduction, and we apply model (3), (4) and (5) to evaluate different types of heterogeneity in class size reduction. This section displays the results of analysis.

Table 5 shows the result of benchmark model. It shows that in general, class size has a negative and significant impact on students' academic achievement, i.e. class size reduction is good for students' academic achievement, for both math and Japanese subject. Column 1 and 2 show the results of the model that includes student fixed effect and teacher fixed effect, rather than including the student-teacher fixed effect. That is, in this model, we don't control the unobserved heterogeneity of student-teacher matching rules. Column 1 use Japanese scores, as the dependent variable, and Column 2 use math scores as the dependent variable. We can see that when class size is increased (decreased) by 1 student, the students' average Japanese and math score will decrease (increase) by 0.00451 and 0.00705 standard deviation, respectively. Column 3 and 4 show the results of the model that includes student-teacher fixed effect, with the dependent variables of Japanese score and math score, respectively. It is shown that when class size is increased (decreased) by 1 student, the students' average Japanese and math score will decrease (increase) by 0.0029 and 0.00938 standard deviation, respectively. The results are robust for different specifications of fixed effects, i.e. whether controlling for the unobserved teacher-student matching rule. Although class size reduction has a positive effect on both Japanese and mathematics scores, interestingly, the effect is significant for mathematics but not for Japanese when we include student-teacher fixed effect. Moreover, the size of the class size reduction effect is larger for mathematics than for Japanese. The reason might be that math is a scientific subject and its instruction requires more techniques, so more concentrated instructions can lead to better performance. However, Japanese subject does not require too much technique, and students can even benefit from communicating with each other. Moreover, language is more of an art so it is not necessarily to have right or wrong answers. As a result, more concentrated teaching is less important for Japanese than for math.

Table 6 shows the effect of class size reduction on classes with different performance of classmates. Column (1) and (2) show the effect of class size reduction on students' Japanese scores and math scores, respectively, in classes with different average peer math scores. Column (3) and (4) show the effect of class size reduction on students' Japanese scores and math scores, respectively, in classes with different average peer Japanese scores. Column (5) and (6) show the effect of class size reduction on students' Japanese scores and math scores, respectively, in classes with different baseline scores of the top student and bottom student. All the columns show that the main effect of class size reduction, i.e. the coefficient associated with class size, is negative. The interpretation is, conditioning on the mean math score, mean Japanese score, maximum/minimum Japanese/math scores, etc. are zero, the impact on one's test score if the class size is increased by 1. A negative coefficient means a positive impact of class size reduction on test scores. Column (1) and (2) show that when the mean mathematics test score of the class peers is zero standard deviation from the mean, increasing the class size by one student will lead to a decrease in 0.00255 standard deviation in Japanese score and 0.00908 standard deviation in mathematics score. To measure the heterogeneity effect, we include the interaction term of class size and mean mathematics performance, the coefficient associated to which is negative and significant. That is, classes with higher average mathematics scores will benefit more from class size reduction. Note that when we calculate the average performance, we excluded the student's own score in order to avoid endogeneity, so to be more precise, the average performance is average peer performance. The results show that the effect of increasing the class size by one student will be 0.0107 standard deviation more negative for one's Japanese score in a class with one percentage point standard deviation higher in the mean mathematics score. Moreover, the effect of increasing the class size by one student will be 0.00948 standard deviation more negative for one's mathematics score in a class with one percentage point standard deviation higher in the mean mathematics score. For both mathematics and Japanese test score, the heterogeneity in the effect of class size reduction is significant for the classes with different mean math scores. Column (3) and (4) shows the heterogeneity in the effect of class size reduction for classes with different mean Japanese scores, and we can see the coefficients associated with the class size and mean Japanese score are also negative, with -0.011 for Japanese score and -0.0135 for mathematics score, respectively, both of which are significant. As we can see, class size reduction is more effective in enhancing students' test scores for the classes with higher ability students. Therefore, a more effective allocation of students should be assigning higher academic performance students together in small classes.

Column (5) and (6) of Table 6 shows another interesting result on heterogeneity in class size reduction. Here we interact the maximum and the minimum baseline score of a class with the class size. When we include these interactions, firstly, the sign and significance of the main effect of class size reduction does not change. For both Japanese score and mathematics score, the main effects are significant and negative, -0.0188 standard deviation for Japanese score and -0.0161 standard deviation for mathematics score. Moreover, the highest baseline score of a class does not influence the heterogeneity of the effect of class size reduction as significantly as the lowest baseline score does. This is true for both mathematics and Japanese score. However, the interaction term between minimum score of the class and the class size is negative and significant, which means that the heterogeneity of class size reduction is more sensitive to the performance of the student with the lowest baseline score in the class. Especially, increasing the class size by one student has significantly more negative effect for students in a class with higher bottom-student's baseline ability in both mathematics and Japanese, for mathematics scores, -0.00341 standard deviation when lowest baseline mathematics score is increased by one percentage point standard deviation, and -0.0024 when lowest baseline Japanese score is increased by one percentage point standard deviation. The same heterogeneity in the class size reduction effects on Japanese scores follow a similar pattern, but only the coefficient associated with the lowest baseline score of

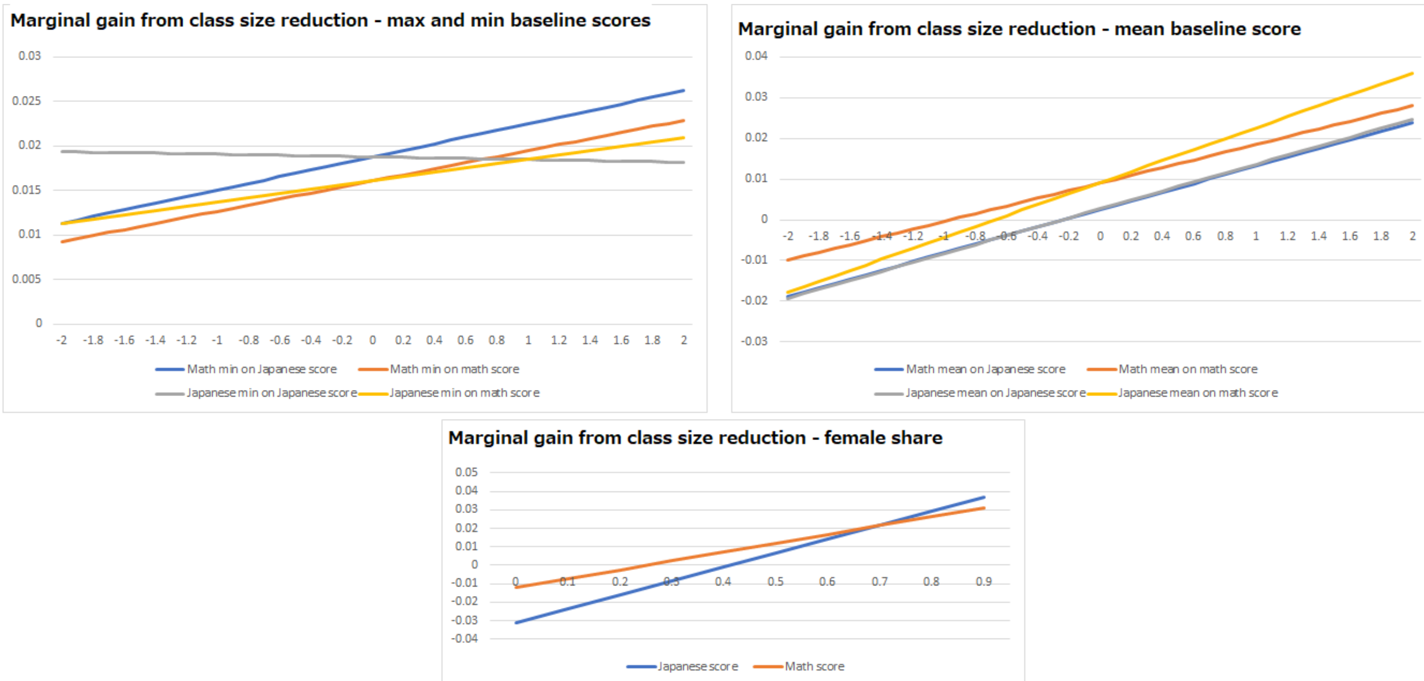

Figure 1: *Graph heterogeneity*

mathematics is significant. The reason why class size reduction has a significantly heterogeneous effect when the bottom-student's baseline score changes particularly interest and can be explained by Lazear's (2001) model. We will discuss it in the next section.

Table 7 shows the heterogeneity in the class size reduction effect about gender. Column (1) and (2) show the heterogeneous impact of class size reduction on students with different genders, on Japanese scores and mathematics scores, respectively. The coefficient associated with class size times female dummy is insignificant, which indicates that class size reduction has similar effects on male and female students. Column (3) and (4) show the effect of class size reduction on classes with different percent of female students (excluding the student herself) on Japanese score and mathematics score, respectively. It shows that the coefficient associated with the interaction term of class size and percent of female student in the class is negative and significant for both Japanese and mathematics, although the coefficient associated with class size is positive. As we can see, for the class with zero percent of female student, increasing the class size by one will lead to an increase in the Japanese score by 0.0309 standard deviation, and mathematics score by 0.0121 standard deviation. But one percentage point's increase in the percent of female students in the class will lead to the increase of one student in the class becoming negative, and this number is 0.0755 standard deviation for Japanese score and 0.048 standard deviation for mathematics score. The total impact of the impact of class size is $0.0309 - 0.0755 * femalepercent$ on Japanese score, and $0.0121 - 0.048 * femalepercent$ on mathematics score. We can see that class size reduction has a negative total effect on Japanese scores for classes with female students lower than 41 percent, and will have a positive effect on Japanese score when the percent of female students is higher than 41 percent. Similarly, class size reduction will have a positive impact on one's mathematics score when female is higher than 25 percent.

Figure 1 summarizes our results in heterogeneity in class size reduction in a graphical manner. When we interact class size with other variables, the effect of class size reduction takes a linear form. Therefore, we graph these linear relationships. We look at the heterogeneity in the maximum and minimum baseline scores of a class, the mean baseline score of a class and the female share of a class.

7 Discussion

To explain how class size impacts education activities, Lazear (2001) provides a theoretical framework with great explanatory power, which is also highly applicable. He assigns each student with a possibility of being well-behaved, not disturbing other students, p. Without being disturbed, the total value of education production in a class is denoted by V. In a classroom, as long as a student is not well-behaved, the teacher need to handle this misbehave, and henceforth, the education production process will be temporarily stopped. That is, each student's disturbing behavior has an negative externality that impacts the education production of the whole classroom.

Lazear's model can be extended in various dimensions, and our results provide strong support to Lazear's model in an extended dimension. The heterogeneity in students' gender or socioeconomic status are factors that can determine their value within a unit of time and/or the possibility of being well-behaved. If we kick out a student with a higher possibility of disturbing other students, the possibility of disturbing others will be reduced by much, which can even be viewed as a benefit for the education activity. However, if we kick out a student with a lower possibility of disturbing others, that will be a lost for the education because such a student could contribute more to class education by helping others or asking thought-provoking questions.

In our result, when the average baseline academic performance is higher, class size reduction is more effective. Our explanation is that when the students have higher academic performance, they are less likely to disturb the class, and also their education production within a time unit is larger. But when the class is disturbed, the loss of education production is higher.

Our finding on the heterogeneity in female percentile also fits into Lazear's model. Classes with higher percent of female students will benefit more from class size reduction for both Japanese and mathematics subject, because classes with higher percent of female students are less likely to be disturbed. According to the investigation report by the Japanese Ministry of Education, in 2021, among all recognized primary school bully cases, 246211 are carried out by male, and 174686 by female, which shows the males are more likely to carry out problematic behaviors that can disturb the education process.

Another interesting finding that can be explained by Lazear's model, and also can be viewed as evidence to support Lazear's model is the heterogeneous effect of class size reduction in the classes with different baseline scores of the bottom student. We view baseline score of a student as her ability, which may reflect her level of intelligence, cognitive skills and non-cognitive behaviors. It is natural to assume that the bottom student of a class is more likely to disturb the class education than other students, and the higher a student's baseline score is, the less likely she is going to disturb the education process, e.g. she is less likely to ask a questions which everyone knows the answer. Therefore, the lower baseline score the bottom student of a class has, when reducing this type of student, the possibility that the education process in that class would be disturbed is reduced more largely than under the situation when the bottom student's baseline score is higher. The detailed discussion with formal mathematical model is put into the Appendix.

Our results of the average effect of class size reduction are comparable to other literature on class size reduction such as Bandiera et al. (2010) and Urquiola (2006) in terms of the magnitude of the coefficient, although smaller in terms of size. Bandiera et al. (2010) found that one standard deviation reduction in class size will increase students' test scores by 0.074 standard deviation, while Urquiola (2006) found this increase was up to 0.3 standard deviation. Angrist and Lavy (1999) found that the size was 0.13 to 0.27 standard deviation for pupils. We measure the class size by the number of students in the class, rather than the standard deviation in class size as the papers above did, but we know that the standard deviation of class size in our data set is 4.364 students, so the effect of one standard deviation decrease in class size can be calculated as approximately 0.04 standard deviation in mathematics scores and about 0.02 standard deviation in Japanese scores.[4] Our estimates are slightly smaller than the estimates in the literature partly because of the inclusion of teacher-student fixed effects.[5]

[4]Hojo and Senoh (2019) report that the class size reduction by one student results in the improvement of math score by 0.018 and Japanese score by 0.014 in the Japanese context.

[5]One may be interested in the cost-benefit analysis based on our estimates. However, for the full calculation of the cost-benefit ratio, it is required to obtain the estimates of the long-run effects on not only

Another important point that make our paper unique to others is the methodology. Different from other projects on class size reduction, we not only apply the Maimonides' rule, but also control for student-teacher fixed effect, rather than controlling student fixed effect and teacher fixed effect separately. If we control student fixed effect and teacher fixed effect separately, we control for the students' and teachers' time-invariant unobservable characteristics. By controlling student-teacher fixed effect, besides controlling students' and teachers' unobservable characteristics, we also control for the unobservable sorting rule between teachers and students. As we have explained in previous sections, in Japanese primary schools, the school principal has the discretion of assigning teachers to classes, and usually these assignment rules are not clearly documented, and henceforth not observable and would also be difficult to model. It is also difficult to assume that the assignment of teachers to classes is random, so it is important to control for the effect of the sorting rule in order to achieve unbiased estimates.

Other papers also made significant efforts in getting the unbiased estimate of the impact of class size reduction, but still not as sufficient as we did. The data set used by Angrist and Lavy (1999) was class-level data, which made it impossible to add student fixed effect, and they didn't get teacher data either. To make the methodology consistent with Angrist and Lavy (1999), their new project (Angrist et al. (2019)), even using individual level data, did not add student fixed effect. Bandiera et al. (2010), although added student fixed effect, would not be able to apply the Maimonides' rule because Maimonides' rule does not apply in college education. Although they controlled for teacher and student observable characteristics, and student fixed effect and teacher fixed effect, yet the endogeneity of class size and the sorting of teacher and classes may still exist. The reason that they could not include student-teacher fixed effect as we did is that in the college, every course was taught by the same teacher across years, so within a teacher, there is no variation in the class size across years. In our paper, however, by adding student-teacher fixed effect, and applying the Maimonides' rule, we better solve the endogeneity issue in class formation. We avoid the following case: when a grade has 41 students, and has to be divided into two classes with class size of 20 and 21, schools tend to assign teachers with certain unobserved traits to larger classes.

Interestingly, Kedagni et al. (2021), applying both reduced-form and structural analysis, also found a non-linear shape of the effect of class size on students' academic achievement. Our results also suggest such a non-linear shape due to the heterogeneity of the classes.

academic outcomes but also other skills such as non-cognitive skills in addition to a bunch of assumptions in the specific context of education, which is beyond our scope for the current paper. We would keep it as future research.

8 Conclusion

In this paper, we mainly estimate heterogeneous effects of class size reduction on academic outcomes by class average baseline test scores, percent of female students in class, and the baseline score of the top student and bottom student in a class. We found that the main effect of class size reduction has a positive effect on students' academic performance. We also found that smaller classes are better for students in classes with higher peer baseline scores, and it is stronger for mathematics than for Japanese subject, which has an important policy implication of class assignment that it would be effective to assign students with higher ability to smaller classes. Another interesting finding is that a class whose bottom student has a higher baseline score will benefit more from class size reduction; meanwhile, the baseline score of the top student of a class does not generate significant heterogeneity in class size reduction. Therefore, teachers are advised to pay more attention to the bottom students of the classes. All of our findings can be interpreted in the framework of Lazear (2001).

References

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

Angrist, J., V. Lavy, J. Leder-Luis, and A. Shany (2019): "Maimonides' Rule Redux," *American Economic Review: Insights*, 1, 309–24.

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

Bandiera, O., V. Larcinese, and I. Rasul (2010): "Heterogeneous Class Size Effects: New Evidence from a Panel of University Students," *Economic Journal*, 120, 1365–1398.

Browning, M. and E. Heinesen (2007): "Class Size, Teacher Hours and Educational Attainment*," *Scandinavian Journal of Economics*, 109, 415–438.

Ding, W. and S. Lehrer (2011): "Experimental estimates of the impacts of class size on test scores: robustness and heterogeneity," *Education Economics*, 19, 229–252.

Gaub, M. and C. L. Carlson (1997): "Gender Differences in ADHD: A Meta-Analysis and Critical Review," *Journal of the American Academy of Child Adolescent Psychiatry*, 36, 1136–1139.

Gershon, J. and J. Gershon (2002): "A Meta-Analytic Review of Gender Differences in ADHD," *Journal of Attention Disorders*, 5, 143–154, pMID: 11911007.

Gilraine, M. (2020): "A Method for Disentangling Multiple Treatments from a Regression Discontinuity Design," *Journal of Labor Economics*, 38, 1267–1311.

Hanushek, E. A. (1986): "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature*, 24, 1141–1177.

——— (2003): "The Failure of Input-Based Schooling Policies," *Economic Journal*, 113, 64–98.

——— (2006): "School Resources," in *Handbook of the Economics of Education*, ed. by E. Hanushek and F. Welch, Elsevier, vol. 2 of *Handbook of the Economics of Education*, chap. 14, 865–908.

Hojo, M. and W. Senoh (2019): "Do the disadvantaged benefit more from small classes? Evidence from a large-scale survey in Japan," *Japan and the World Economy*, 52.

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

Kedagni, D., K. Krishna, R. Megalokonomou, and Y. Zhao (2021): "Does class size matter? How, and at what cost?" *European Economic Review*, 133, 103664.

Krueger, A. (1999): "Experimental Estimates of Education Production Functions," *The Quarterly Journal of Economics*, 114, 497–532.

Krueger, A. B. (2003): "Economic Considerations and Class Size," *Economic Journal*, 113, 34–63.

Lazear, E. P. (2001): "Educational Production," *The Quarterly Journal of Economics*, 116, 777–803.

Li, Q. (2006): "Cyberbullying in Schools: A Research of Gender Differences," *School Psychology International*, 27, 157–170.

Myers, D. E., A. M. Milne, K. Baker, and A. Ginsburg (1987): "Student Discipline and High School Performance," *Sociology of Education*, 60, 18–33.

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

Table 5: *Class size effect - benchmark model*

VARIABLES	(1) z_jpn	(2) z_math	(3) z_jpn	(4) z_math
Class size	-0.00451***	-0.00705***	-0.00290	-0.00938***
	(0.00140)	(0.00131)	(0.00336)	(0.00313)
School financial assistance	-0.0144	-0.00609	0.00254	-0.000376
	(0.00884)	(0.00847)	(0.0185)	(0.0166)
Total teaching experience	-0.00426	-0.00126	0.00976**	0.00353
	(0.00264)	(0.00218)	(0.00466)	(0.00367)
Tenure at the current school	-0.000474	0.00197	0.00253	0.00594
	(0.00184)	(0.00170)	(0.00570)	(0.00536)
Teacher age	0.0373**	-0.0340**	0.0628***	-0.0110
	(0.0147)	(0.0138)	(0.0238)	(0.0209)
Baseline ability of language	-0.0203	0.0404***	-0.187***	0.0299
	(0.0164)	(0.0153)	(0.0251)	(0.0227)
Baseline ability of math	-0.0128	-0.0500***	-0.0105	-0.165***
	(0.0157)	(0.0148)	(0.0240)	(0.0220)
Percentage of female students	0.262***	0.0300	-0.0340	-0.0490
	(0.0943)	(0.0891)	(0.184)	(0.169)
Share of financial assistance receivers	-0.00873	-0.0656*	-0.0180	-0.0999
	(0.0355)	(0.0335)	(0.0691)	(0.0634)
maximum Japanese score	-0.0443***	-0.00199	-0.0426*	0.0149
	(0.0143)	(0.0133)	(0.0220)	(0.0200)
maximum math score	-0.0134	0.000441	0.0316	0.0743**
	(0.0220)	(0.0202)	(0.0345)	(0.0304)
minimum Japanese score	0.000143	0.0113***	-0.000466	0.0127**
	(0.00358)	(0.00332)	(0.00592)	(0.00515)
minimum math score	0.00256	-0.00188	-0.00333	-0.00603
	(0.00335)	(0.00313)	(0.00599)	(0.00556)
Constant	-0.885	2.212***	-2.433***	0.594
	(0.749)	(0.686)	(0.943)	(0.832)
Observations	149,727	149,727	149,727	149,727
	126,193	126,193	126,193	126,193
	YES	YES	YES	YES
year FE	YES	YES	YES	YES
grade FE	YES	YES	YES	YES
student FE	YES	YES	NO	NO
teacher FE	YES	YES	NO	NO
student-teacher FE	NO	NO	YES	YES

All specifications include school fixed effects, year fixed effects, grade fixed effects, and student fixed effects and teacher fixed effects separately for the columns (1) and (2), and student-teacher pair fixed effects for the columns (3) and (4) in addition to the variables listed in the table. Student's own status is excluded from the calculation of the class means. Standard errors for the columns (1) and (2) arerobust to heteroskedasticity, and those for the columns (3) and (4) are clustered at the student-teacher pair level.
*Significant at 10%; **Significant at 5%; ***Significant at 1%

Table 6: *Heterogeneous class size effect - peer performance*

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	z_jpn	z_math	z_jpn	z_math	z_jpn	z_math
Class size	-0.00255	-0.00908***	-0.00270	-0.00915***	-0.0188*	-0.0161
	(0.00337)	(0.00313)	(0.00336)	(0.00313)	(0.0112)	(0.0103)
x mean math score	-0.0107***	-0.00948***				
	(0.00380)	(0.00338)				
x mean Japanese score			-0.0110***	-0.0135***		
			(0.00423)	(0.00372)		
x maximum math score					-0.00255	-0.0110*
					(0.00679)	(0.00606)
x maximum Japanese score					0.00814**	0.00401
					(0.00404)	(0.00353)
x minimum math score					-0.00372***	-0.00341***
					(0.00129)	(0.00119)
x minimum Japanese score					0.000294	-0.00240**
					(0.00131)	(0.00115)
Constant	-2.520***	0.516	-2.494***	0.518	-2.279**	0.530
	(0.941)	(0.832)	(0.942)	(0.831)	(0.988)	(0.872)
Observations	149,727	149,727	149,727	149,727	149,727	149,727
Number of student-teacher pairs	126,193	126,193	126,193	126,193	126,193	126,193

All specifications include the class mean, maximum, and minimum baseline math score, the class mean, maximum, and minimum baseline Japanese score, the class share of female students, the class share of school financial assistance receivers, teacher's total teaching experience, teacher's tenure at the current school, teacher's age, student's own status of school financial assistance, school fixed effects, year fixed effects, grade fixed effects, and student-teacher fixed effects in addition to the variables listed in the table. Student's own status is excluded from the calculation of the class means. Class size and its interaction terms are instrumented by the predicted class size and its interactions. Standard errors are clustered at the student-teacher pair level. *Significant at 10%; **Significant at 5%; ***Significant at 1%

Table 7: *Heterogeneous class size effect - female student*

VARIABLES	(1) z_jpn	(2) z_math	(3) z_jpn	(4) z_math
Class size	-0.00320	-0.00884**	0.0309**	0.0121
	(0.00434)	(0.00368)	(0.0144)	(0.0126)
x share of female students			-0.0755**	-0.0480*
			(0.0305)	(0.0269)
x female dummy	0.000672	-0.00123		
	(0.00537)	(0.00510)		
Constant	-2.434***	0.596	-3.322***	0.0279
	(0.942)	(0.832)	(1.002)	(0.893)
Observations	149,727	149,727	149,727	149,727
Number of student-teacher pairs	126,193	126,193	126,193	126,193

All specifications include the class mean of baseline math score, the baseline Japanese score, female students, school financial assistance receivers, teacher's total teaching experience, teacher's tenure at the current school, teacher's age, own status of school financial assistance, school fixed effects, year fixed effects, grade fixed effects, and student-teacher fixed effects in addition to the variables listed in the table. Standard errors are clustered at the school level. Their status is excluded from the calculation of the class means. Class size and its interaction terms are instrumented by the predicted class size and its interactions. Standard errors are clustered at school level. *Significant at 10%; **Significant at 5%; ***Significant at 1%

Appendix

Illustration of Education Process

We illustrate our results in the heterogeneity effects of class size reduction by an extended version of the model by Lazear (2001). In our formal model, in a class, there are n students, each of which has a type $t \in \{1, 2, ..., T\}$. Each type t has n_t students, and $\sum_{t=1}^{t=T} n_t = n$, and we suppose $n_t > 0$ for any t. The possibility that each student with type t will well-behave in class is p_t, and w.l.o.g. we assume $1 > p_1 > p_2 .... > p_T > 0$. Like Lazear (2001) model, let the value of a unit of learning be V, which is determined by the market value of human capital and the likelihood that a student is focusing on learning during the given instant. We assume that $V > 0$, i.e. human capital has a strictly positive value. Therefore, in a time unit, the expected value of education is specified as follow:

$$\pi = V \times \Pi_{j=1}^{j=T} p_j^{n_j}$$

In our model, π can be viewed as the students' academic performance, measured by test scores. The possibility of each type being well-behaved, p_j, is determined by its baseline academic performance, i.e. students with higher baseline scores have higher possibility of being well-behaved. It is relatively intuitive that students with higher baseline scores are less likely to disturb other students, for example, less likely to ask questions which all other students know the answer.

We first want to show that reducing the class size will have a positive effect on the education production, regardless of the type of students whose number is reduced. We take the derivative of π with respect to n_j for any given j, we get

$$\frac{\partial \pi}{\partial n_j} = V ln(p_j) \Pi_{j=1}^{j=T} p_j^{n_j}$$

Because $V > 0$, $\Pi_{j=1}^{j=T} p_j^{n_j} > 0$ and $ln(p_j) < 0$, $\frac{\partial \pi}{\partial n_j} < 0$, for any type j. This can explain our result that the main effect of class size reduction is positive since the first derivative of education production with respect to class size is always negative.

Take the derivative of π with respect to p_k, we have

$$\frac{\partial \pi}{\partial p_k} = (V \Pi_{j \neq k} p_j) \times \frac{\partial p_k^{n_k}}{\partial p_k} = (V \Pi_{j \neq k} p_j) n_k p_k^{n_k - 1}$$

Since all elements in the formula, V, p_j and n_j are strictly positive, we have $\frac{\partial \pi}{\partial p_k} > 0$. Therefore, the education production is a strictly increasing function of the possibility of any type of students being well-behaved, which is also quite intuitive.

We are now going to provide some possible theoretical explanation to the heterogeneous effect of class size reduction. In particular, we explain why class size reduction has a significantly positive effect for classes whose bottom student is better. To do this, we take the derivative to $\frac{\partial \pi}{\partial n_k}$ with respect to p_k. That is, we look at the effect of class size reduction by reducing the number of a type of students with the possibility of well-behave of p_k.

$$\frac{\partial^2 \pi}{\partial n_k \partial p_k} = V\frac{1}{p_k}\Pi_{j=1}^{j=T} p_j^{n_j} + \frac{1}{p_k} V ln(p_k) n_k \Pi_{j=1}^{j=T} p_j^{n_j} = V\Pi_{j=1}^{j=T} p_j^{n_j}[1 + n_k ln(p_k)] \times \frac{1}{p_k}$$

We can see that the sign of $\frac{\partial^2 \pi}{\partial n_k \partial p_k}$ depends on the sign of the term $[1 + n_k ln(p_k)]$. If p_k is large and n_k is small, it would be more likely that the term is positive, which means the effect of class size reduction is less positive only if we reduce the number of the type of students whose possibility of well-behave is high enough, and meanwhile, there are enough number of this type of students.

If we put some specific numbers into the term $[1 + n_k ln(p_k)]$, we can get a sense of how large p_k and how small n_k should be, in order to make it positive:

If $n_k = 10$ and $p_k = 0.9$, we have $ln(0.9)$=-0.105 and $n_k ln(p_k)$=-1.05, so $[1 + n_k ln(p_k)]$=-0.05, still slightly smaller than zero. And if we reduce n_k by 1, i.e. $n_k = 9$, we get $[1 + n_k ln(p_k)]$=0.05, which is slightly greater than zero. When p_k=0.8, if $n_k = 5$, $[1 + n_k ln(p_k)] = -0.1 < 0$, and if $n_k = 4$, $[1 + n_k ln(p_k)] = 0.12 > 0$. We can see that when p_k is larger, for the term to be positive, the number of students in this type k can be allowed to be larger than when p_k is smaller.

This theoretical result can explain our result that the heterogeneity in the effect of class size reduction is not sensitive to the baseline score of the top student in the class, but is sensitive to the baseline score of the bottom student of the class. Class size reduction has a significantly higher effect when the class' bottom student has a better baseline score, but insignificant when the top student's bottom score varies.

For the student with the worst baseline score in a class, p_k is relatively small, so $\frac{\partial^2 \pi}{\partial n_k \partial p_k} < 0$ is more likely to hold. Therefore, when reducing the number of this type of students n_k, and increase p_k, we can see that the class size reduction effect on education production is even more positive. However, for the top student, p_k is high, and when the number of this type is smaller enough, $\frac{\partial^2 \pi}{\partial n_k \partial p_k} > 0$ is more likely to hold. When we reduce this type of students, the effect of class size reduction is less obvious, and that's why the coefficient associated with the interaction term of class size and the score of bottom student in the class is negative and significant, but the coefficient associated with the class size and the score of the top student is not significant.