TCER Working Paper Series

CAN SMALL CLASS POLICY CLOSE THE GAP? AN EMPIRICAL ANALYSIS OF CLASS SIZE EFFECTS IN JAPAN

Hideo Akabayashi
Ryosuke Nakamura

November 2012

Working Paper E-51
http://tcer.or.jp/wp/pdf/e51.pdf

TCER

TOKYO CENTER FOR ECONOMIC RESEARCH
1-7-10-703 Iidabashi, Chiyoda-ku, Tokyo 102-0072, Japan

Abstract

Can a smaller-class at school lead to a better educational outcome and more equality in achievement? We estimate the causal effects of class-size on achievement tests by using discontinuous changes in class-size under the Japanese public compulsory education system. We employ a value-added model using achievement test conducted at two different times during the same school year. Our results show a reduction in class-size has significant positive effects on Japanese language test scores in the sixth grade especially at schools in wealthy areas. That is, we find no evidence that a universal small-class policy closes the achievement gap among schools.

Hideo Akabayashi
TCER
and
Keio University
Department of Economics
Mita 2-15-45, Minato-ku, Tokyo 108-8345, Japan
hakab@econ.keio.ac.jp

Ryosuke Nakamura
Keio University
Department of Economics
Mita 2-15-45, Minato-ku, Tokyo 108-8345, Japan
ryonaka.email@gmail.com

Can Small Class Policy Close the Gap? An Empirical Analysis of Class Size Effects in Japan

By Hideo Akabayashi* and Ryosuke Nakamura

Keio University

First version: 2012/03/20

This version: 2012/09/06

Can a smaller-class at school lead to a better educational outcome and more equality in achievement? We estimate the causal effects of class-size on achievement tests by using discontinuous changes in class-size under the Japanese public compulsory education system. We employ a value-added model using achievement test conducted at two different times during the same school year. Our results show a reduction in class-size has significant positive effects on Japanese language test scores in the sixth grade especially at schools in wealthy areas. That is, we find no evidence that a universal small-class policy closes the achievement gap among schools.

JEL Classification Numbers: I21, I28.

---

* Corresponding author: Faculty of Economics, Keio University, 2-15-45, Mita, Minato-ku, Tokyo 108-8345, Japan
(e-mail. hakab@econ.keio.ac.jp; tel. +81-3-5427-1367; fax. +81-3-5427-1578)

1. Introduction

As of 2008, the average class size of Japanese public elementary schools was 28.0, while the average class size of OECD countries' was 21.6 (OECD 2010). It has been 30 years since the Japanese Government last reduced the upper class size limit from 45 to 40. In April 2011 a bill to reduce the class size upper limit for the first grade of elementary school from 40 to 35 was passed. The Ministry of Education, Culture, Sports, Science and Technology in Japan (hereafter MEXT) has predicted that one of the effects of reducing class size will be to improve students' academic abilities (MEXT 2009a).

Can a class size reduction actually have the effect of improving students' academic achievement? There are a great number of studies regarding the effects of class size on academic achievement around the world; however, results have not been consistent. For example, Hanushek (1996) finds that the effect of teacher-pupil ratio on students' performances is insignificant in almost all studies. Meanwhile, Angrist and Lavy (1999) and Krueger (1999) show that small class size has significant and positive effects on students' academic abilities. In Japan, Shimizu (2002), Miyake (2002), Shinozaki (2008), Hojo and Oshio (2010), and Hojo (2011) attempt to estimate the effect of class size but did not find significant effects.

The goals of this study are to identify the causal effects of a class size reduction on test scores by using discontinuous changes in class size and to investigate whether the effects on the scores are distributed equally among schools and areas. Until 2010, the class size upper limit in compulsory education was 40 students, as specified in the *Act on Standards for Class Formation and Fixed Number of School Personnel of Public Compulsory Education Schools(enactment:1958)* (hereafter ASCFF). In practice, this means that if a school has 40 students in a grade all 40 students will be in the same classroom, whereas if the same school has 41 students in a grade they will be divided between two classrooms. In short, ASCFF creates a situation where enrollment exceeding the upper limit threshold causes a sharp class size reduction. This is similar to the feature of class size in Israel analyzed by Angrist and Lavy (1999).

Our value-added model methodology departs from Angrist and Lavy (1999) as well as previous research using Japanese data. We use data from two separate standardized tests conducted each school year in Yokohama. The first test is the *National Assessment of Academic Ability* (hereafter NAAA), conducted at the beginning of the school year. The second test is the *Yokohama City Achievement Test* (hereafter YCAT), conducted at the end of the school year. We use school-grade level average test score data for the sixth grade in public elementary schools and the third grade in public junior high school because of the availability of the NAAA data.[1]

Our empirical results show that a class size reduction has significant positive effects on Japanese language test scores in the sixth grade. Our value-added model estimates using an instrumental variable method show that a 1 student reduction in class size increases standardized language test scores by 0.0112 standard deviations (Table 5; column 3 of Panel A). The instrumental variable we employ is the predicted class size derived from the class size upper limit that is developed by Angrist and Lavy (1999). In test score data from the third year of junior high school, however, we cannot find any statistically significant effects using a value-added model estimation. By dividing the sample by the initial test score and by the land prices, we find that the test score improvement due to small class size at elementary schools tends to appear only at schools within relatively wealthy areas. That is, we find no evidence that a universal small-class policy closes the achievement gap among schools in compulsory education.

The remainder of this paper is organized as follows. In Section II we review the literature regarding analysis of class size effects. In Section III we describe how we constructed our data set. In Section IV we explain our identification strategies and the Japanese class size limit in detail and present a graphical analysis. In Section V we

[1] Japanese compulsory education system consists of 6 years elementary school and 3 years junior high school, therefore, for simplicity we address the third grade in junior high school as the ninth grade.

report and discuss the results of our value-added model. In Section VI we summarize our results regarding the effects of class size on achievement tests.

2. Literature Review

Most empirical studies of class size effects in Japan are investigated from a pedagogical point of view and do not pay attention to causal relationship between class size and academic outcomes of students. Most Japanese research on class size has, until very recently, been based on subjective evaluations by teachers (Oshio and Seno 2003).

There are several economic studies of class size effects outside Japan that attempt to identify causality under plausible assumptions. Two types of identification strategies have been commonly employed.[2] First, there are studies that analyze the relationship between class size and students' outcome using discontinuous changes in class size caused by the exogenously determined administrative rules. Second, there are studies using randomized experiments to measure class size effects on outcomes.[3]

We are aware of five studies analyzing the effects of class size in Japan. Hojo and Oshio (2010) and Hojo (2011) analyze the causal effects of class size by using Angrist

---

[2] There are also papers that include class size or teacher-student ratio as one of school resources. Boozer and Rouse (2001), Lindahl (2005) and Bressoux, Kramarz and Prost (2009), etc., for example, tackle to estimate class size effect by unique identification strategies.

[3] Studies using randomized experiments such as the Project STAR (the Tennessee Student/Teacher Achievement Ratio experiment) include Krueger (1999), Finn and Achilles (1990), and Krueger and Whitmore (2001). These studies exploit different dependent variables but they find that small class has positive effects.

and Lavy's (1999) empirical strategy and TIMSS (Trends in International Mathematics and Science Study) data. In these cross-sectional studies they were not able to find significant effects of class size on test scores. The other studies analyze the class size effects using ANOVA or OLS. Shimizu (2002) compares the math test scores of fifth and seventh grade students between five categories of class size (under 20, 21-25, 26-30, 31-35, 36-40) using ANOVA, while Miyake (2002) analyzes science test scores using the same method. Shimizu and Miyake use the same dataset and do not find significant differences of test scores between the five categories at any grade level. Shinozaki (2008) estimates class size effects using OLS on Chiba prefecture's NAAA cross-sectional data. He finds positive significant effects on Japanese language and math test scores in elementary school but he notes that these results cannot be considered causation.

Angrist and Lavy (1999) estimate the effect of class size on academic ability in Israel by taking advantage of *Maimonides' rule*. Like the Japanese ASCFF, this rule sets the class size upper limit to 40. Angrist and Lavy (1999) find significant positive effects of reducing class size on reading and math scores in fifth grade and reading scores in fourth grade by using exogenous changes of class size caused by the rule. Expansions of Angrist and Lavy's (1999) study have been performed in a number of countries. In the

Netherlands, Dobbelsteen, Levin and Oosterbeek (2002) do not find the evidence that the reduction of class size improves academic ability. In Norway, Leuven, Oosterbeek and Rønning (2008) do not find significant effects with data from nationwide academic ability tests. In Denmark, Browning and Heinesen (2007) show that a class size reduction of 5% increases years of schooling by 0.02 and the probability of completion of secondary education by 0.4%.

Among the large literature that uses discontinuous changes in class size to identify the causal educational effect of a class size reduction, few studies have also incorporated a value-added approach. In the U.S., Hoxby (2000) implements a value-added model with adjacent cohorts that experienced a change in the number of classes at the same school and does not find that a class size reduction has significant positive effects on the score.[4] Bonesrønning (2003) regresses individual tenth grade test scores on class size and ninth grade class-level average scores and finds the coefficient of class size is significant and positive.

Previous studies tend to find that students of low socioeconomic status acquire more benefits from smaller classes. In a paper closely related to ours, Bressoux, Kramarz and Prost (2009) separate students into four groups by ascending order of initial test score

---

[4] Hoxby (2000) also uses *Maimonides'rule* and random variation of enrollments as instrumental variables but she concludes that the class size reduction does not improve students' academic achievements.

and find that the effects of class size on math scores are stronger for low achieving students but that on reading scores there is no difference between the groups. If the class size effect is smaller for more initially-disadvantaged schools, overall class size reductions would not necessarily close the gap between educational outcomes among schools.

In this paper we group the schools in our sample not only by the initial test score level but also by land prices where schools are located. By doing this we are able to examine whether the class size cap policy in Japan has any heterogeneous effects on students across schools with different initial average achievement and different average family wealth conditions. Examining the role of family wealth in the effects of education policy is important since families with greater wealth are more likely to have alternative investment opportunities for their children, such as outside tutoring, especially in the Japanese social context. Our contribution is that we examine heterogeneity of the class size effect on the students' achievement growth using multiple test scores within the same student cohort-year for the first time for Japan.

3. Data

Recently, Japan has begun to accumulate an increasing amount of education data at both the school and student levels. For example, MEXT has implemented the *National*

Assessment of Academic Ability (NAAA) for all students in the sixth and ninth grades since 2007. Many prefectures and large municipalities, such as Tokyo and Yokohama, have also implemented their own achievement tests for public school students. Although not all this data is publically available, opportunities for quantitatively evaluating education policies in Japan have expanded greatly.

Our data set is school-level data from the NAAA and the *Yokohama City Achievement Test* (YCAT), both of which included data from all public schools in Yokohama. Yokohama is the second largest city in Japan, which is populated by 3.6 million people based on 2010 Census, located about 60 km south of Tokyo.[5] The school system is managed by the city's Board of Education and included 345 public elementary schools and 146 public junior high schools in 2010.[6] The average taxable income per taxpayer in 2009 was 4.1 million yen ("Statistical Profile of Municipalities (*Toukei de Miru Shikuchouson no Sugata)* 2011"), higher than the national average of 3.3 million yen. The schools in Yokohama have many urban characteristics and thus cannot be understood to be representative of schools in Japan as a whole.

NAAA is a national standardized achievement test of Japanese language ability and

[5] Yokohama is one of the 20 "ordinance-designated cities" that have a greater authority—nearly equivalent to prefectures—than normal cities in Japan. .

[6] There were also 10 private elementary schools and 31 private junior high schools in the city, which are not included in our data set.

math that has been conducted annually at the beginning of school year in April since 2007. Participation in the NAAA is mandatory for all national and public elementary and junior high schools. It is optional for private schools. The test is given to students entering their sixth grade of elementary school or third year of junior high school (equivalent to ninth grade in the U.S.). We obtained school level average scores of language and math for the NAAA tests conducted in 2008 and 2009. Our second data set is the YCAT, which is independently administered by Yokohama City. Students in elementary schools take the YCAT in February, first and second year junior high school students take the test in March, and third year students take it in November. In other words, students take the YCAT at the end of their respective school years. Students in first and second grade take two subjects, language and math; students in third through sixth grade take four subjects, language, math, social studies, and science; and students in the seventh through ninth grades take five subjects, language, math, social studies, science, and English.

We obtained school level NAAA and YCAT test score data for 2008 and 2009 using a "*Act on Access to Information Held by Administrative Organs*", which is equivalent of Freedom of Information Act in U.S.[7] The following four characteristics of the two data

---

[7] At the time of our request to Yokohama school-level achievement data was not publicly available. The city administration of Yokohama has been quite proactive in its efforts to

sets are important to keep in mind: First, the raw data we were provided with included the grade-level average percentage of correct answers for each subject for each school. Second, we have converted the average percentage of correct answers into a standardized score with a mean of 50 and standard deviation of 10 within each year/subject. Third, the YCAT 2008 public data set is the school level average of test scores sampled at each school.[8] Finally, there are some schools that did not implement the tests due to special school events or because of temporary school closure due to influenza outbreaks.

We obtained official data on the number of classes and enrollment by school and grade level as of May 1st for the relevant year from the Yokohama City Board of Education's web site. The number of classes per school grade level and enrollment by sex is available only for the junior high schools. We calculated the grade level average class size by dividing school enrollment by the number of classes.

In order to control for the average socioeconomic status of students, we employ the school attendance zone-level average land prices per square meter from survey data

---

make government data available.

[8] In the YCAT2008, the calculation of the average test scores is based on randomly sampled students. The sampling rate differs by school size. For example, if the enrollment is 20-30, 20 students are extracted for the calculation. If the enrollment is 30-40, 40-50, 50-100 and 100-, then the number of students that is extracted for calculation is 30, 40, 50, 100, respectively.

obtained in compliance with the *Public Notice of Land Prices Act*. We calculated average land prices by matching data on land prices for residential lots in 2006 to school attendance zones. Some school attendance zones, however, cannot be perfectly matched to the land price measurement points. By including land prices in the analysis, we lose data for approximately 50 elementary schools and 4 junior high schools, representing about 14% and 2% of the original sample, respectively.

We include descriptive statistics of our two year school panel data in Table 1. [9] Average class size is 32.32 in the sixth grade and 35.71 in the ninth grade. Average female-ratio is just below 50 in the ninth grade and average land price is about 200,000 yen per square meter. As described below, 7% of schools introduced smaller classes in the sixth grade and 3% of schools introduced smaller classes in ninth grade.

4. Identification Strategies

We use discontinuous changes in class size caused by the class-size rule in order to identifying the causal effects of class size, as originally developed by Angrist and Lavy (1999). In Japan, the *Act on Standards for Class Formation and Fixed Number of School Personnel of Public Compulsory Education Schools* (ASCFF) sets the class size upper limit to 40. According to the act, the number of classes in a grade is one when

[9] The standardized test scores' mean is 50 but their standard deviation is 9.99 because of the calculation based on pooled two-year data.

enrollment is less than or equal to 40, two when enrollment is 41-80, and three when enrollment is 81-120. This relationship is expressed by int $[(n-1)/40]+1$, where int[•] returns the largest integer in the brackets and n is the total enrollment in a grade.

Although the national rule of the class size upper limit is set by the ASCFF, since 2001 prefectural Boards of Education have the ability to apply a different rule when they wish to lower the upper class size limit for some or all schools and grades. Kanagawa Prefecture, where Yokohama is located, has specified two special cases for reducing the upper limit to 35. One case applies to "experimental schools" where new pedagogy is developed and the other applies to "flexible class-formation schools," where a principal is allowed to request additional teachers in order to reduce class size.[10] As a result, there are 28 elementary schools (8%) with non-standard class-size rules in 2009 as shown in Table 1.

In Fig. 1, we show the predicted and observed class size distributions in Yokohama against grade enrollment for the sixth (Panel A) and ninth (Panel B) grades, separately. The solid lines plot the relationship between enrollment and class size as predicted by the ASCFF rule:

$$z_{ikt} = n_{ikt} / \{ \text{int}[(n_{ikt} - 1)/40] + 1 \}, \qquad (1)$$

[10] We, hereafter, address the school applied these cases the school with intended small classes.

In (1), the *i* indicates school, *k* indicates the neighboorhood, and *t* indicates time. The observed enrollment-class size relationships are shown with three different markers: Average Class Size (i) indicates the average class size in a standard school, Average Class Size (ii) indicates the average class size in an experimental school, and Average Class Size (iii) indicates the average class size in a flexible class-formation school. In Fig. 1-A for the sixth grade we find some schools with enrollment that is close to the upper limit set up small classes; in Fig. 1-B for the ninth grade we find almost all of the junior high schools follow the solid line.

In Fig. 2 and 3, we present a graphical analysis of the relationship between predicted class sizes and test scores in 2009.[11] In Fig. 2, we show the predicted class size and average test score "levels" measured in March 2010 by YCAT for sixth and ninth grades, while in Fig. 3 we show the predicted class size and the average test score "growth" measured by the difference between YCAT score in March 2010 and NAAA scores in April 2009. From Fig. 2, we find that test score levels tend to increase as enrollment increases, but we do not find evidence that the scores move up sharply at the thresholds in either grade. In Fig. 3, it appears that a sharp decline in the class size at the thresholds created by the ASCFF rule is associated with a higher growth in test score, especially in

[11] In this analysis, we categorize schools into groups based on every 10 enrollments and calculate average test scores and predicted class size within each group.

sixth grade, although the relationship is not always clear.

In this study, we estimate the causal effects of class size on test scores based on an education production function. A grade is our minimum observational unit and estimating equation is can be written as

$$Y_{ijkt} = \alpha + \boldsymbol{\beta X}_{ikt} + \delta C_{ikt} + \varepsilon_{ijkt} , \qquad (2)$$

In the above Y_{ijkt} indicates school i's standardized test score for subject j in neighborhood k at year t. Let C_{ikt} denote the grade-level average class size and $\boldsymbol{X}_{ikt}$ denote a vector consisting of individual school characteristics such as enrollment, female-ratio in the class, etc. ε_{ijkt} represents unobserved factors which potentially include school, neighborhood, year and neighborhood-year fixed effects.

Note that schools in Yokohama have the ability to reduce the average class size of a grade if they choose to do so. That means if a school chooses to reduce the class size for a high-ability cohort, the effects of reducing class size on test scores will be upward biased. Our simple solution to this potential bias is to introduce a dummy variable for schools that have chosen to make the class smaller than the rule and to drop these schools from our estimations. Since the number of such schools is so small, we believe these solutions give sufficient robustness checks for this issue.

We now examine one of the key identification assumptions that justify the use of predicted class size as defined in equation (1) as an instrumental variable for observed average class size. Table A1 in the Appendix shows the results of the first stage of the 2SLS estimations that correspond to the results in Tables 5, 7, 8, and 9, where the observed average class size is the dependent variable and predicted class size and other variables are controlled. Every column shows that predicted class size has significant positive effects. Angrist and Lavy (1999), moreover, point out the need to check the other important identification assumption that there is no manipulation of deterministic variable. We believe that the possibility of the manipulation of enrollment is extremely limited in Japan for two reasons. First, public schools in Japan have no ability to set student quotas.[12] Second, there is limited room for manipulation of enrollment by parents—it is very costly for families to move and even if they do so they are unable to control class size perfectly. [13]

5. Results and Discussions

In this section, first, we regress equation (2) using Ordinary Least Squares and typical instrumental variable (IV) regression methods to estimate the class size effects. Second,

---

[12] In Chilean case where private schools control their quotas, this manipulation caused bias on the class size effects (Urquiola and Verhoogen 2009).

[13] In the City of Yokohama, school administrators typically assign students to public schools on the basis of students' addresses. The class formation is determined using enrollment data only a few days before the school year begins, and therefore parents have virtually no ability to relocate by predicting class size before enrollment.

we introduce the concept of a value-added model and estimate this model using OLS and IV methods. We control for average class size and a constant in all estimations and additionally for female-ratio in junior high school cases. We include 3rd-order polynomials of enrollment, average land prices of school attendance zones, year dummies, neighborhood dummies, and year-neighborhood dummies as additional independent variables.[14] We use robust standard error clustered by school for statistical inferences.

A. Ordinary Least Squares and Typical Instrumental Variable Estimation

Tables 2 and 3 show the results estimated by OLS and IV for sixth grade and ninth grade, respectively. The dependent variables are Japanese language and math scores from YCAT. Each panel shows estimation results of OLS and IV, respectively. In these tables, columns (1) – (3) and (7) - (9) are estimations with all schools included and columns (4) – (6) and (10) – (12) are estimations without schools that intentionally introduced small classes.

The estimation results from these tables are summarized as below. First, in the sixth grade the class size effects for each panel in each subject are significantly positive without controlling for enrollment. When controlling for enrollment and its higher order

[14] The neighborhood dummies are defined by the 18 *ku* (wards) in Yokohama City.

polynomials, however, significant effects of class size disappear. Second, in the ninth grade the class size effects are not statistically significant except for in columns (1) and (8) of Panel A. Broadly speaking, when we control for enrollment the class size effects are smaller than that when we do not. Finally, comparison between OLS and IV estimators implies a positive correlation between class size and unobserved factors, which suggests that lower-achieving students are assigned to smaller classes.

B. Value-Added Model

The aim of using value-added model is to control for unobserved fixed effects. The idea is to add initial average test scores to the right hand side of equation (2). By controlling for initial ability we can control for unobserved family characteristics, the quality of teachers, and cohort effects and estimate the class size effects more precisely. Our data will minimize the effects of changing grade composition caused by student migration, because two tests are held in the same school year. In this section, we build two value-added models and estimate these models using OLS and IV. Furthermore, we examine these estimations with a restricted sample as a robustness check.

The first type of value-added model is based on estimating a single equation for each subject. In other words, the score of each subject in the YCAT is explained by the score of each subject in the NAAA. The estimating equation can be written as

$$Y_{ijkt} = \alpha + \boldsymbol{\beta}\boldsymbol{X}_{ikt} + \delta C_{ikt} + \gamma NAAA_{ijkt} + \varepsilon_{ijkt} \, , \qquad (3)$$

where *NAAA* indicates school *i*'s standardized test score of subject *j* in NAAA in neighborhood *k* at year *t*.

The second type of value-added model allows cross-subject effects pooling data from multiple subjects. Specifically, the scores of Japanese language and math on the YCAT are allowed to be affected by the scores of both language and math on the NAAA. The estimating equation for this model can be written as

$$Y_{ijkt} = \alpha + \boldsymbol{\beta}\boldsymbol{X}_{ikt} + \delta C_{ikt} + \gamma_1 NAAA_{i1kt} + \gamma_2 NAAA_{i2kt} + \varepsilon_{ijkt} \, , \qquad (4)$$

where $NAAA_{i1kt}$ and $NAAA_{i2kt}$ indicate school *i*'s standardized test score in language and math on NAAA in neighborhood *k* at year *t*, respectively. Y_{ijkt} indicates the language and math scores from the YCAT.

Tables 4 through 7 show the results of OLS and IV estimations for grades six and nine. In these tables, Panel A shows the results of estimations of language and Panel B shows that of math. In each panel, the first six columns use the full sample and the second six columns use a sample restricted to within ±5 around the discontinuity points in the class size rule (equation 1). The coefficients shown in the upper tables are class size and NAAA scores, while the other control variables appear in the lower tables. Female-ratio is controlled in ninth grade data only.

The results of the type 1 value-added model show that the effects of reducing class size on language score are significant and positive in the sixth grade. In Table 4 Panel A, which shows the results for language test score using OLS, the coefficients of class size are negative but not statistically significant. In Table 5 Panel A, which is the result estimated by IV, the class size effects on language score are negative and statistically significant at the 10% level in columns (2) – (5) and at the 5% level in column (6) where we control for the average land price of each school attendance zone as proxy variable for socioeconomic status. In columns (7) – (12) of the panel A, where we use the discontinuity sample, the class size effects become insignificant. In the case of math, shown in the panel B of each table, we cannot find any similar effects once we control for enrollment.

The results of our type 2 value-added model also imply that a class size reduction has positive effects on test scores in the sixth grade. The OLS estimation results of class size effects, shown in the panel A of Table 8, are negative but insignificant in both the full sample and the discontinuity sample cases. In the panel B of Table 8, the IV estimation results show marginally significant and negative class size effects in columns (1) – (5) and significant and negative effects at the 10% level in column (6). In the discontinuity sample cases, the effects of class size are negative and especially significant at the 10%

level in column (12).

In ninth grade, both types of value-added model do not suggest significant effects of class size. In tables 6 and 7, which show the results of our type 1 model, the class size effects are not significant in any columns and the sign of the class size coefficient is inconsistent. Table 9, which is the estimation result under our type 2 model, shows that the class size effects are not significant in both panels.

C. Can Small Classes Close the Education Gap between Schools?

Here we investigate whether the class size effect on educational outcomes differs among schools with different initial achievement and with different area characteristics. To see this, we separate the schools in the sample into two subgroups on the basis of the median of the initial test scores (the standardized NAAA scores) and the median of the local land prices.[15]

Table 10 shows our IV estimation results of class size effects on the four subsamples divided by the subjects and the initial test score using the type 1 value-added model. The table is arranged with Panel A for higher achieving schools and Panel B for lower achieving schools in the initial score. In columns (1)-(6) of Panels A and B of Table 10, the class size effects on Japanese language scores are significant and negative only for

[15] Though we also estimate the effects of class size for junior high school in subgroups, we cannot find significant effects of class size in almost all estimations.

the initially higher achieving schools. Meanwhile, the effects of reducing class size on math scores are insignificant in both the higher and lower achieving groups of Table 10. Contrary to Bressoux, Kramarz and Prost (2009), the higher ability group receives more benefit from small classes and these effects are significant at least at the 5% level.

Table 11 shows the IV estimation results of class size effects for the two subgroups divided by school attendance zone-level average land prices using the type 1 value-added model. Similarly to Table 10, Table 11 indicates that schools located in economically advantaged areas with higher land prices obtain more benefit from small classes in terms of improving both Japanese and math test scores.

Table 11 also shows that the initial test score gap in Japanese is 6.13 between economically advantaged and disadvantaged areas of the city. Column (6) of Table 11 indicates that a one-student decrease in class size would have a significant and positive effect on scores of 0.34 in wealthy areas without any statistically significant effect on scores in less wealthy areas. For this reason, a universal class size reduction from 40 to 30 students could be said to widen the school level achievement gap between the two groups by roughly 50%. Therefore, we conclude that an overall class size reduction does not necessarily close the achievement gap and may even enlarge the gaps that exist between areas for economic reasons.

D. Robustness Check Based on the Fixed Effects Model

Finally, we check the robustness of the coefficient estimates under the fixed effects estimation. Todd and Wolpin (2003) point out that a value-added model assumes strict conditions to obtain consistent estimators, and if a lagged achievement measure, or NAAA score in our paper, correlates with unobservables then the coefficients of both the lagged achievement measure and class size will be biased. Our estimators may be contaminated by this bias because the measure of lagged test score and the dependent variable, or YCAT score, are not designed to cover the exactly the same material.

Tackling this issue, we take account of Todd and Wolpin's (2003) point and estimate equation (3) using a fixed effects estimation. Here, school fixed effects μ_i are decomposed into two parts, μ_i^Y and μ_i^N. We redefine the education production function as below,

$$Y_{ijkt} = \alpha + \boldsymbol{\beta}\boldsymbol{X}_{ikt} + \delta C_{ikt} + \mu_i^Y + \theta t + \varepsilon_{ijkt}, \qquad (5)$$

$$NAAA_{ijkt} = \alpha^N + \boldsymbol{\beta}^{\boldsymbol{N}}\boldsymbol{X}_{ikt} + \delta^N C_{ikt} + \mu_i^N + \theta t + v_{ijkt}, \qquad (6)$$

where μ_i^Y and μ_i^N are the fixed effects of YCAT and NAAA, respectively. In this situation, even if we subtract equation (6) multiplied by γ from equation (5), school fixed effects, or $\mu_i = \mu_i^Y - \gamma\mu_i^N$, will not be zero, and NAAA will correlate μ_i and its coefficient may be upward bias. Table 11 shows the results of fixed effects estimations.

In Table 12 we find that the class size effects on language score remain negative and significant at the 10% level, but there are no class size effects on math scores. Incidentally, the coefficients of NAAA score are smaller than in Table 4, which implies that NAAA and μ_i^Y are positively correlated.[16]

E. Discussion

In this section, we interpret the estimated size of class size reduction effects on test scores in perspective. To give a sense of the size of our estimate, we can compare the estimated class size effects for elementary school with results from related literature. We calculate the effect size on Japanese language test scores in the sixth grade (column (3) of Table 5 Panel A), and obtain a coefficient of 0.0112 (0.1118 / 9.9927). This effect size is about one third as large as a comparable figure from Israeli (Angrist and Lavy's (1999) effect size on reading test score in the 5th grade is 0.035) and about half as large as one from the U.S. included in Krueger (1999).[17] Therefore, the class size effect that we find in this sample, if any, is much smaller than those found in the literature.

---

16 At the suggestion of a referee, we performed estimations excluding schools with extremely low Japanese language scores. When we estimated the Japanese language test score regressions for 6th grade excluding the two lowest-performing schools we saw a decrease in the significance level of all results. However, we did not observe any systematic influence on class size effects of including or excluding extremely low-performing schools.

17 We take into account Krueger's (1999) calculation of the effect size based on small classes (which are smaller than regular classes by 8 students) and we multiply our effect size by 8 to obtain 0.0896.

One puzzling result is that reducing class size has positive and significant effects on language test scores in the sixth grade but not on math scores. This is partly explained by the fact that recently an increasing number of schools tend to teach math under a different class arrangement than the standard class arrangement. MEXT (2009b) notes that 65% of schools teach math in groups smaller than the official class size in the sixth grade, while only 20% of schools use small groups in language instruction. This implies that the effective class size of math classes is smaller than the class size recorded in the data, and this source of measurement error may have caused a bias in the math score estimation.

Another puzzling result is that the class size effects are more difficult to identify in junior high schools. One potential reason is that the ninth grade students in the data have been exposed to the current class environment for a shorter period of time than the sixth grade students due to the different timing of the second test. Another potential reason is that ninth grade students in Japan, in preparation to take high school entrance exams, are more likely to attend private cram schools to improve their academic skills than sixth grade students. If attendance at cram schools explains the difference in the class size effects for the two grades, the effect of class size should be smaller for the schools where students are more likely to attend the crams schools. However, our regression results including the high land prices dummy interacted with class size show the opposite: The class size effects are larger at schools with a higher land prices. This is a puzzling and depressing result that can be interpreted in a number of ways.

Similarly to our results, Konstantopoulos (2008), after reanalyzing STAR data, finds that small class size has positive effects on test scores only in the first and second years of the study and that the effects are larger in high achievers. He postulates two mechanisms through which small class may benefit higher achievers more than lower achievers. One is that teachers are more likely to identify higher-achieving students in small classes and thus are more likely to provide effective teaching strategies. Second, it is possible that instructional practices in small classes benefit higher achievers more. He also suggests that cumulative effects of small classes diminish over time. Based on his hypotheses, class size effects for ninth grade may be much smaller than those for sixth grade and therefore more difficult to be distinguished in our small school level sample even when controlling for initial achievements.

Another possibility is suggested by the casual observation that at the secondary school level high achievers are those who are more likely to switch to attend private schools from public schools at the end of elementary school. If smaller class size has any reinforcing effect with the initial achievement, as Konstantopoulos (2008) and

Table 10 in our paper suggest, the class size reduction at public junior high schools, which most able students do not attend, may not have as strong an effect on achievement as it does at public elementary schools. Given the current data set that includes only public schools, it is difficult to find any further evidence to support this hypothesis. Ideally future research should use data sets that include both public and private schools.

6. Conclusion

In this paper, we identify the effects of reducing class size on test scores using discontinuous changes in class size induced by the Japanese class size rule, applying Angrist and Lavy's (1999) instrumental variable estimation. Additionally, to our knowledge this is the first paper that has identified the effects of class size with a value-added model taking into account class size decision rules at a municipality level.

Empirical results indicate that a class size reduction has positive effects on test scores, especially for Japanese language test scores in the sixth grade. In the ninth grade, however, we cannot find any significant class size effects. These results imply that the effects of reducing class size are more apparent in lower grade. By dividing the sample by the initial test score and by the land prices, we find that the test score improvement due to small class size in elementary schools tends to appear only at schools within relatively wealthy areas. That is, we find no evidence that a universal small class policy closes the achievement gap among schools. Therefore, for the purpose of reducing the achievement gap in compulsory education it may be necessary to use policy tools that target relatively disadvantaged children and schools when the small class policy is expanded.

A number of issues remain to be investigated in future research. In particular, recent reductions in the class size limit for the first grade of elementary school have raised a question of the priority of the class size reduction while offering the opportunity to investigate this issue for younger children. A study of the cost-effectiveness of policies that reduce class size is also warranted. Moreover since we use only school-level average data, our results do not have any direct implications to whether a universal small class policy closes the achievement gap *among students*. Finally, with additional administrative data it would be possible to control for the nonrandom assignment of teachers between and within schools.

Acknowledgements

We would like to thank the city of Yokohama for generously providing us with test score data. Mr. Hiroaki Kitajima gave us detailed information on the local class size rules in Yokohama City. We are also grateful to a referee, Souich Ohta, and those who participated in seminars at Keio, the 2011 JEA, 2011 Labor Conference (Awajishima), Hitotsubashi University, and MEXT. This research was financially supported by Keio University, the Grant-in-Aid for Scientific Research (A 20243020: Akabayashi) , and the Grant-in-Aid for JSPS Fellows (24-4938: Nakamura). The views expressed in this paper are those of the authors and do not necessarily reflect the positions or policies of the organizations mentioned above. All the remaining errors are ours.

References

[1] Angrist, J. D. and V. Lavy (1999) "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", *Quarterly Journal of Economics*, Vol. 114, No. 2, pp. 533-575.

[2] Bonesrønning, H. (2003) "Class Size Effects on Student Achievement in Norway: Patterns and Explanations", *Southern Economic Journal*, Vol. 69, No. 4, pp. 952-965.

[3] Boozer, M. and C. Rouse (2001) "Intraschool Variation in Class Size: Patterns and Implications", *Journal of Urban Economics*, Vol. 50, No. 1, pp. 163-189.

[4] Browning, M. and E. Heinesen (2007) "Class Size, Teacher Hours and Educational Attainment", *Scandinavian Journal of Economics*, Vol. 109, No. 2, pp. 415-438.

[5] Bressoux, P., F. Kramarz and C. Prost (2009) "Teachers' Training, Class Size and Students' Outcomes: Learning from Administrative Forecasting Mistakes", *Economic Journal*, Vol. 119, No. 536, pp. 540-561.

[6] Dobbelsteen, S., J. Levin and H. Oosterbeek (2002) "The Causal Effect of Class Size on Scholastic Achievement: Distinguishing the Pure Class Size Effect from the Effect of Changes in Class Composition", *Oxford Bulletin of Economics and Statistic*s, Vol. 64, No. 1, pp. 17-38.

[7] Finn, J. D. and C. M. Achilles (1990) "Answers and Questions about Class Size: a Statewide Experiment", *American Educational Research Journal*, Vol. 27, No. 3, pp. 557-577.

[8] Hanushek, E. A. (1996) "Measuring Investment in Education", *Journal of Economic Perspectives*, Vol. 10, No. 4, pp. 9-30.

[9] Hojo, M. (2011) "Education Production Function and Class-Size Effects in Japanese Public Schools", *Global COE Hi-Stat Discussion Paper Series*, No. 194.

[10] Hojo, M and T. Oshio (2010) "What factors determine student performance in East Asia? New evidence from TIMSS 2007", *PIE/CIS Discussion Paper*, No. 494.

[11] Hoxby, C. M. (2000) "The Effects of Class Size on Student Achievement: New Evidence from Population Variation", *Quarterly Journal of Economics*, Vo. 115, No. 4, pp.1239-1285.

[12] Konstantopoulos, S. (2008) "Do Small Classes Reduce the Achievement Gap between Low and High Achievers? Evidence from Project STAR", *Elementary School Journal*, Vol. 108, No. 4, pp. 275-291.

[13] Krueger, A. B. (1999) "Experimental Estimates of Education Production Functions", *Quarterly Journal of Economics*, Vol. 114, No. 2, pp. 497-532.

[14] Krueger, A. B. and D. M. Whitmore (2001) "The Effect of Attending a Small Class

in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR", *Economic Journal*, Vol. 111, No. 468, pp. 1-28.

[15] Leuven, E., H. Oosterbeek and M. Rønning (2008) "Quasi-experimental Estimates of the Effect of Class Size on Achievement in Norway", *Scandinavian Journal of Economics*, Vol. 110, No.4, pp. 663-693.

[16] Lindahl, M. (2005) "Home versus School Learning: A New Approach to Estimating the Effect of Class Size on Achievement", *Scandinavian Journal of Economics*, Vol. 107, No. 2, pp. 375-394.

[17] Ministry of Education,Culture,Sports,Science & Technology in Japan (2009a) *Heisei 21 nendo Monbu Kagaku Hakusyo*, in Japanese (*White Paper on Education, Culture, Sports, Science and Technology*), Oita: Saeki Printing.

[18] Ministry of Education,Culture,Sports,Science & Technology in Japan (2009b) *Heisei 19･20 nendo Zenkoku Gakuryoku･Gakusyu Jyokyo Chosa Tuika Bunseki Houkokusyo*, in Japanese (*Additional Research of National Assessment of Academic Ability in 2007 & 2008*), Tokyo: MEXT.

[19] Miyake, M. (2002) "Rika no Gakuryoku Chosa Kekka", in Japanese ("The Results of Science Achievement Test"), *NIER Research Bulletin*, No. 131, pp. 71-77.

[20] OECD (2010) *Education at a Glance 2009: OECD Indicators*, Paris: OECD

Publishing.

[21] Oshio, T. and W. Seno (2003) "Nihon no Kyoiku Keizaigaku: Jissyo Bunseki no Tenbou to Kadai", in Japanese ("The Economics of Education in Japan"), *ESRI Discussion Paper Series*, No. 69.

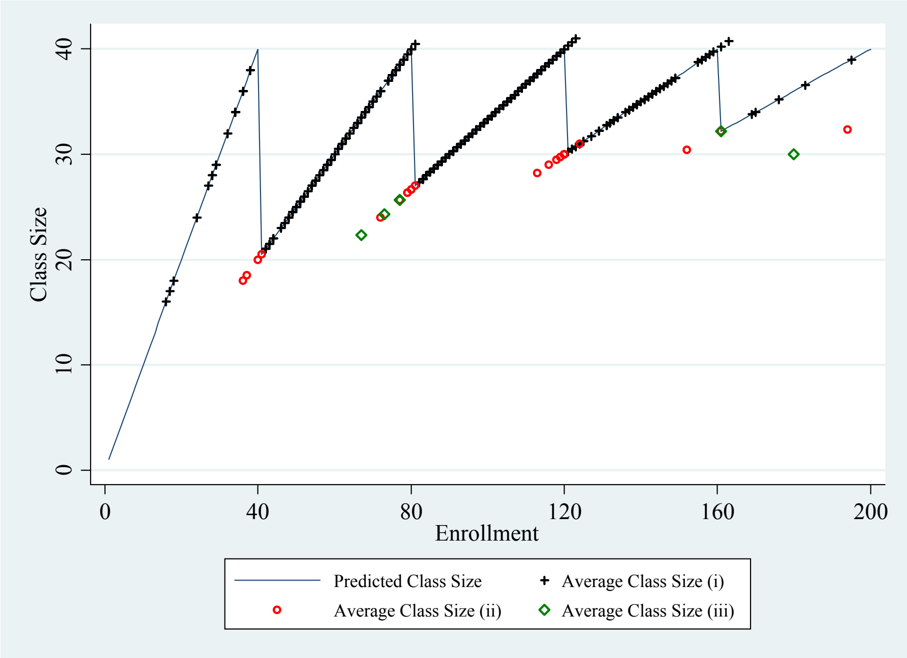
[22] Shimizu, K. (2002) "Sansu· Sugaku no Gakuryoku Chosa Kekka", in Japanese ("The Results of Math Achievement Test"), *NIER Research Bulletin*, No. 131, pp. 62-70.

[23] Shinozaki, T. (2008) "Kyoiku Shigen to Gakuryoku no Kankei", in Japanese ("The Relationship between Educational Resources and Academic Ability"), in Chiba-Ken Kensyo Kaizen Iinkai, ed., *Heisei 19 nendo "Zenkoku Gakuryoku･ Gakusyu Jyokyo Chosa" Bunseki Houkoku* (*Report on National Assessment of Academic Ability in 2007*), Tokyo: TamaType, pp. 73-97.

[24] Todd, P. E. and K. I. Wolpin (2003) "On the Specification and Estimation of the Production Function for Cognitive Achievement", *Economic Journal*, Vol. 113, No. 485, pp.F3-33.

[25] Urquiola, M. and E. Verhoogen (2009) "Class-Size Caps, Sorting, and the Regression-Discontinuity Design", *American Economic Review*, Vol. 99, No. 1, pp. 179-215.

A. Sixth grade in elementary school



B. Third grade in junior high school (ninth grade)

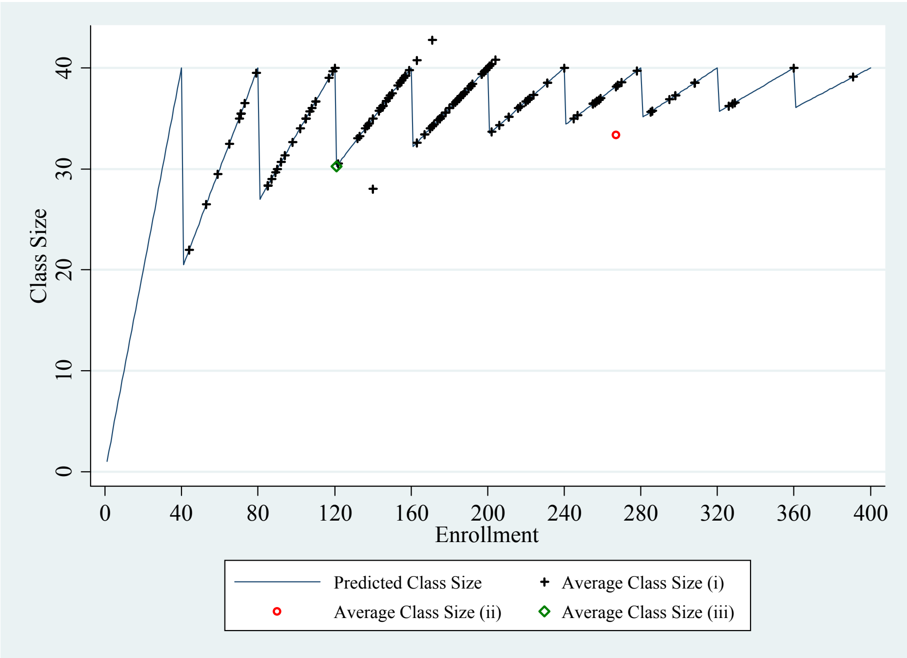
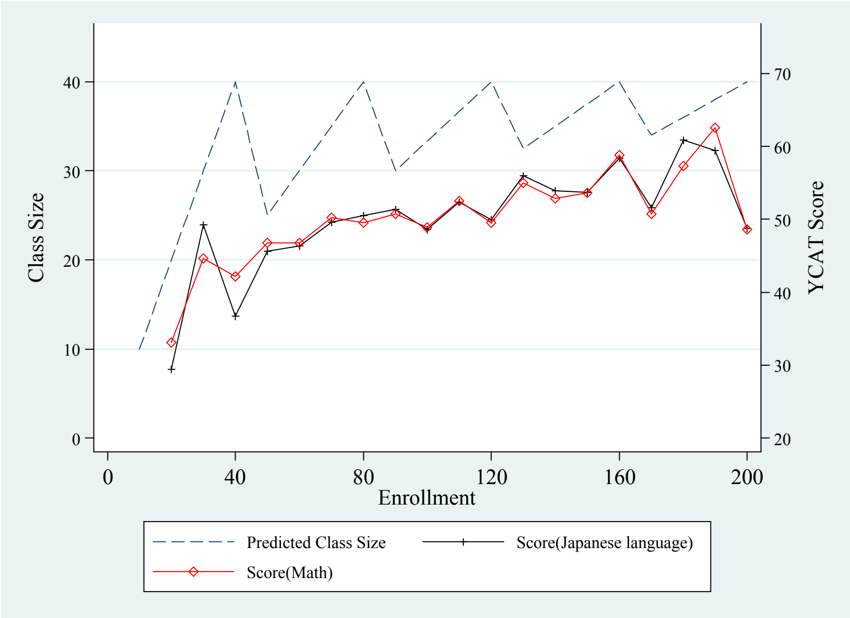


FIGURE 1.−Comparison between predicted and observed class sizes in 2009.

Note: Solid lines indicate predicted class size calculated by equation (1). The plus, circle, and diamond markers indicate the average class size of schools complying with ASCFF, experimental schools, and flexible class-formation schools, respectively.

A. Sixth grade in elementary school



B. Third grade in junior high school (ninth grade)

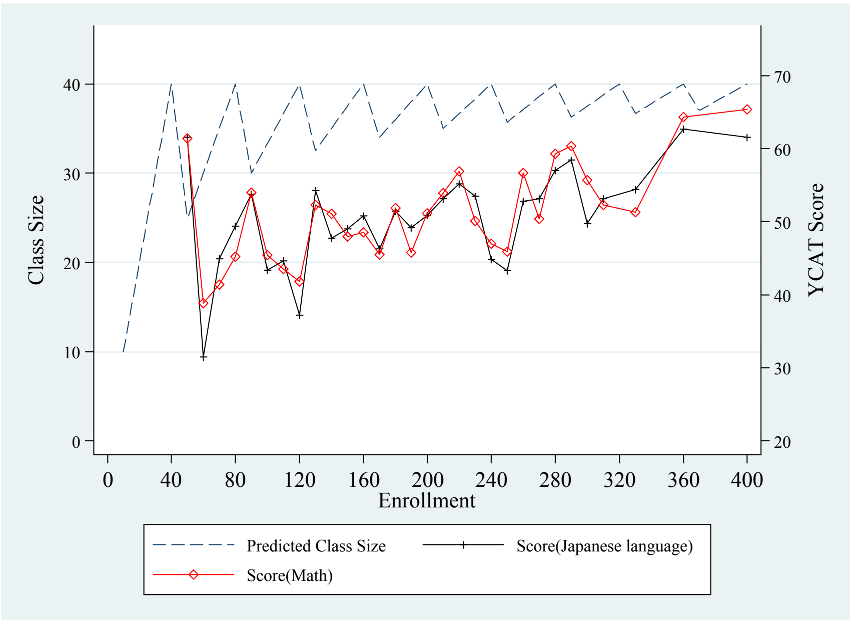
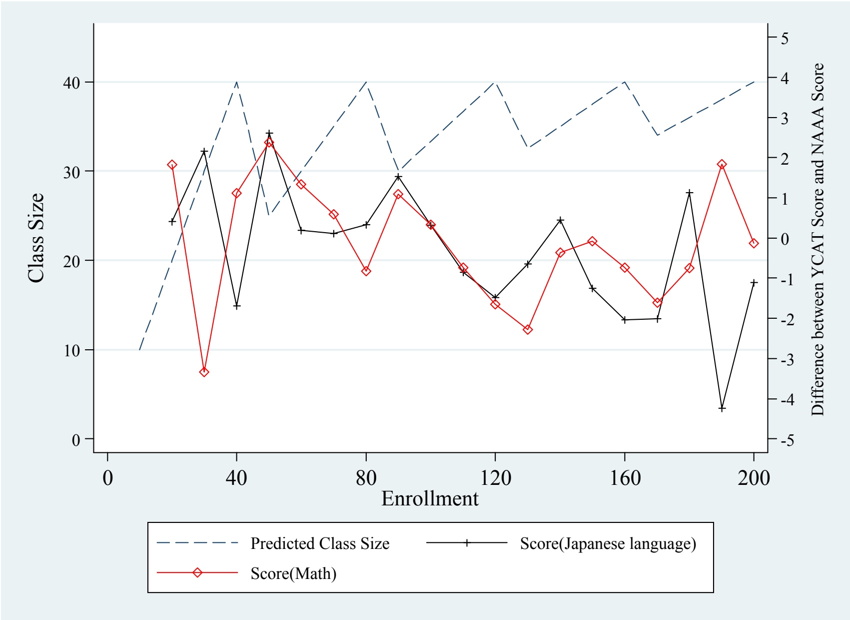


FIGURE 2.−Relationship between predicted class size and average test scores in 2009.

Note: YCAT is the Yokohama City Achievement Test. Dashed lines indicate predicted class size as calculated by equation (1). Plus and circle markers indicate average test scores over the range of 10 students in enrollment of Japanese language and math, respectively.

A. Sixth grade in elementary school



B. Third grade in junior high school (ninth grade)

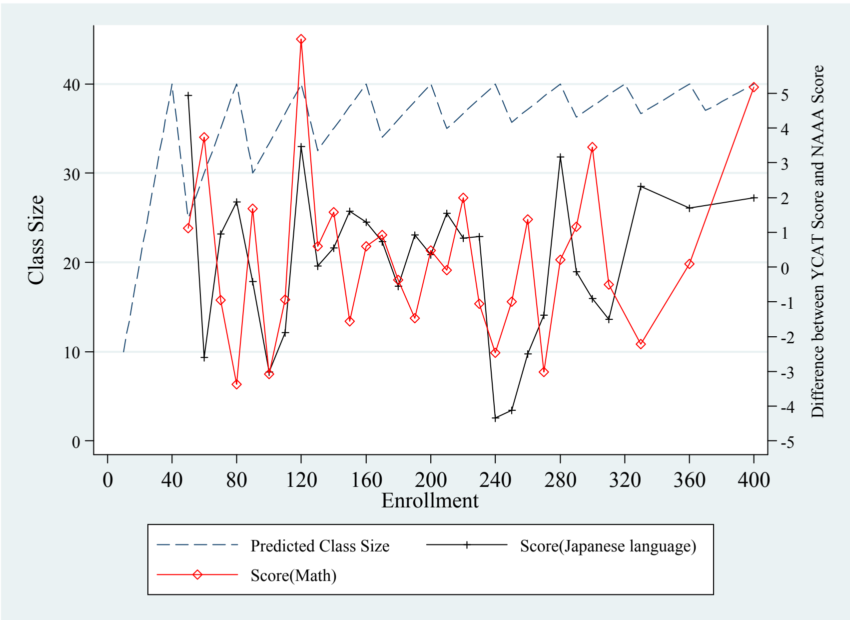


FIGURE 3−Relationship between predicted class size and difference YCAT average score and NAAA average score in 2009.

Note: YCAT is the Yokohama City Achievement Test and NAAA is the National Assessment of Academic Ability. Plus and circle markers indicate the difference between the average YCAT scores and average NAAA scores over the range of 10 students in enrollment of Japanese language and math, respectively. This difference denotes the change of score from the beginning to the end of the school year.

Table 1. Descriptive statistics

	Mean	S.D.	Min	Max	N
Sixth grade in elementary school					
Dependent variables					
YCAT Score (Japanese language)	50.00	9.99	5.56	75.81	692
YCAT Score (Math)	50.00	9.99	17.79	80.14	689
YCAT Score (Language and Math)	50.00	9.99	5.56	80.14	1381
Independent variables					
Average class size	32.32	4.72	16.00	41.00	692
NAAA Score (Japanese language)	50.00	9.99	12.70	74.24	692
NAAA Score (Math)	50.00	9.99	15.66	72.40	692
Enrollment	92.13	34.21	16.00	195.00	692
Land prices (unit: ¥1000/㎡)	205.49	36.51	131.33	311.00	594
Intended small class dummy	0.07	0.25	0.00	1.00	692
Third grade in junior high school (ninth grade)					
Dependent variables					
YCAT Score (Japanese language)	50.00	9.98	2.49	71.13	288
YCAT Score (Math)	50.00	9.98	15.63	74.38	288
YCAT Score (Language and Math)	50.00	9.97	2.49	74.38	576
Independent variables					
Average class size	35.71	3.57	13.33	42.75	290
NAAA Score (Japanese language)	50.00	9.98	14.18	68.92	290
NAAA Score (Math)	50.00	9.98	18.50	72.93	290
Enrollment	173.91	69.42	38.00	391.00	290
Land prices (unit: ¥1000/㎡)	204.04	36.48	131.33	311.00	282
Female Ratio (%)	47.82	3.96	31.25	59.09	290
Intended small class dummy	0.03	0.16	0.00	1.00	290

Note: YCAT is the Yokohama City Achievement Test and NAAA is the National Assessment of Academic Ability. Average class size, enrollment and female ratio come from administrative data published in Report on the State of Schools in Yokohama. "Land prices" is the school attendance zone-level average land prices in 2006. "Intended small class dummy" indicates whether a school is an experimental school or a flexible class-formation school.

Table 2. Estimations of the class size effects on test scores by OLS and IV (sixth grade in elementary school)

Panel A: Japanese language	OLS						IV					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	0.3359***	-0.1002	-0.1018	-0.1049	-0.1037	-0.0887	0.2994***	-0.1580	-0.1645	-0.1353	-0.1347	-0.1333
	(0.0886)	(0.1015)	(0.1052)	(0.1014)	(0.1055)	(0.1097)	(0.0955)	(0.1096)	(0.1125)	(0.1038)	(0.1065)	(0.1091)
Adjusted R^2	0.2493	0.3125	0.2976	0.3182	0.3029	0.3956	0.2490	0.3120	0.2971	0.3181	0.3028	0.3954
N	692	692	692	647	647	557	692	692	692	647	647	557

Panel B: Math	OLS						IV					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	0.2763***	-0.0268	-0.0195	-0.0331	-0.0231	0.0084	0.2418***	-0.0792	-0.0723	-0.0629	-0.0553	-0.0197
	(0.0787)	(0.1065)	(0.1094)	(0.1072)	(0.1104)	(0.1118)	(0.0846)	(0.1157)	(0.1168)	(0.1108)	(0.1121)	(0.1124)
Adjusted R^2	0.2590	0.2868	0.2740	0.2879	0.2755	0.3966	0.2587	0.2864	0.2736	0.2878	0.2753	0.3965
N	689	689	689	644	644	555	689	689	689	644	644	555
3rd-Order polynomials of enrollment		yes	yes	yes	yes	yes		yes	yes	yes	yes	yes
Intended small class excluded				yes	yes	yes				yes	yes	yes
Year dummies			yes		yes	yes			yes		yes	yes
Neighborhood dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year-Neighborhood dummies			yes		yes	yes			yes		yes	yes
Land prices						yes						yes

Note: The dependent variable is the Yokohama City Achievement Test. All estimations include a constant. We control enrollment, $enrollment^2/100$, and $enrollment^3/10000$ if the "third-order of polynomials of enrollment" is "yes". We exclude schools which implement intended small classes from the estimations if the "intended small class excluded" is "yes". The instrument variable for average class size is predicted class size as calculated by equation (1). Standard errors in parentheses are clustered by school. Significance level: *** 1%, ** 5%, * 10%.

Table 3. Estimations of the class size effects on test scores by OLS and IV (third grade in junior high school: ninth grade)

Panel A: Japanese language	OLS						IV					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	0.4566*	0.1733	0.2131	0.1869	0.2142	0.3375	0.5198*	0.1373	0.1710	0.0587	0.0591	0.2294
	(0.2443)	(0.2380)	(0.2492)	(0.2467)	(0.2616)	(0.2582)	(0.3139)	(0.3770)	(0.3774)	(0.3265)	(0.3360)	(0.3308)
Adjusted R^2	0.2694	0.2940	0.2623	0.3169	0.2850	0.3055	0.2690	0.2939	0.2622	0.3158	0.2834	0.3046
N	288	288	288	280	280	272	288	288	288	280	280	272

Panel B: Math	OLS						IV					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	0.2701	-0.0255	-0.0374	-0.0265	-0.0455	0.1175	0.2866	-0.1657	-0.1504	-0.1794	-0.2129	0.0230
	(0.2083)	(0.2027)	(0.2230)	(0.2151)	(0.2371)	(0.2207)	(0.2660)	(0.3278)	(0.3364)	(0.2806)	(0.2988)	(0.2788)
Adjusted R^2	0.2749	0.3130	0.2769	0.3326	0.2984	0.4014	0.2749	0.3116	0.2760	0.3310	0.2964	0.4007
N	288	288	288	280	280	272	288	288	288	280	280	272
3rd-Order polynomials of enrollment		yes	yes	yes	yes	yes		yes	yes	yes	yes	yes
Intended small class excluded				yes	yes	yes				yes	yes	yes
Year dummies			yes		yes	yes			yes		yes	yes
Neighborhood dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year-Neighborhood dummies			yes		yes	yes			yes		yes	yes
Land prices						yes						yes

Note: The dependent variable is the Yokohama City Achievement Test. All estimations include a constant and female-ratio. We control enrollment, enrollment²/100, and enrollment³/10000 if the "third-order of polynomials of enrollment" is "yes". We exclude schools which implement intended small classes from the estimations if the "intended small class excluded" is "yes". The instrument variable for average class size is predicted class size as calculated by equation (1). Standard errors in parentheses are clustered by school. Significance level: *** 1%, ** 5%, * 10%.

Table 4. Estimations of the class size effects on test scores with value-added model type 1 by OLS (sixth grade in elementary school)

Panel A: Japanese language	Full sample						Discontinuity Sample (±5)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	-0.0289	-0.0858	-0.0885	-0.0898	-0.0972	-0.1166*	-0.0366	-0.0452	-0.0003	-0.0437	-0.0038	-0.0756
	(0.0493)	(0.0603)	(0.0624)	(0.0598)	(0.0620)	(0.0682)	(0.0723)	(0.0794)	(0.0801)	(0.0828)	(0.0838)	(0.1009)
NAAA Score (Japanese language)	0.8788***	0.8658***	0.8712***	0.8687***	0.8750***	0.8550***	0.8590***	0.8464***	0.8297***	0.8543***	0.8240***	0.7924***
	(0.0227)	(0.0238)	(0.0241)	(0.0243)	(0.0248)	(0.0319)	(0.0423)	(0.0429)	(0.0469)	(0.0525)	(0.0572)	(0.0721)
Adjusted R^2	0.7648	0.7683	0.7694	0.7688	0.7694	0.7575	0.8341	0.8373	0.8409	0.8451	0.8488	0.8524
N	692	692	692	647	647	557	159	159	159	126	126	110

Panel B: Math	Full sample						Discontinuity Sample (±5)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	-0.0578	-0.0280	-0.0226	-0.0328	-0.0273	-0.0329	-0.1065	-0.0637	-0.0654	-0.0585	-0.0595	-0.0804
	(0.0421)	(0.0579)	(0.0579)	(0.0583)	(0.0582)	(0.0666)	(0.0818)	(0.0930)	(0.0993)	(0.0960)	(0.0958)	(0.0982)
NAAA Score (Math)	0.8860***	0.8935***	0.8990***	0.8845***	0.8899***	0.8615***	0.9029***	0.9275***	0.9266***	0.8985***	0.8536***	0.8350***
	(0.0214)	(0.0229)	(0.0228)	(0.0236)	(0.0235)	(0.0307)	(0.0357)	(0.0417)	(0.0487)	(0.0484)	(0.0513)	(0.0634)
Adjusted R^2	0.7990	0.7998	0.8038	0.7966	0.8010	0.7895	0.8381	0.8418	0.8362	0.8394	0.8526	0.8688
N	689	689	689	644	644	555	157	157	157	124	124	109
3rd-Order polynomials of enrollment		yes	yes	yes	yes	yes		yes	yes	yes	yes	yes
Intended small class excluded				yes	yes	yes				yes	yes	yes
Year dummies			yes		yes	yes			yes		yes	yes
Neighborhood dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year-Neighborhood dummies			yes		yes	yes			yes		yes	yes
Land prices						yes						yes

Note: The dependent variable is the Yokohama City Achievement Test. All estimations include a constant. We control enrollment, $\text{enrollment}^2/100$, and $\text{enrollment}^3/10000$ if the "third-order of polynomials of enrollment" is "yes". We exclude schools which implement intended small classes from the estimations if the "intended small class excluded" is "yes". Standard errors in parentheses are clustered by school. Significance level: *** 1%, ** 5%, * 10%.

Table 5. Estimations of the class size effects on test scores with value-added model type 1 by IV (sixth grade in elementary school)

Panel A: Japanese language	Full sample						Discontinuity Sample (±5)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	-0.0396	-0.1075*	-0.1118*	-0.1075*	-0.1144*	-0.1468**	-0.0630	-0.0829	-0.0439	-0.0723	-0.0289	-0.1307
	(0.0521)	(0.0624)	(0.0637)	(0.0613)	(0.0627)	(0.0685)	(0.0743)	(0.0760)	(0.0751)	(0.0787)	(0.0748)	(0.0867)
NAAA Score (Japanese language)	0.8801***	0.8657***	0.8712***	0.8687***	0.8750***	0.8553***	0.8633***	0.8490***	0.8336***	0.8573***	0.8273***	0.7971***
	(0.0223)	(0.0234)	(0.0235)	(0.0239)	(0.0240)	(0.0308)	(0.0398)	(0.0397)	(0.0407)	(0.0478)	(0.0486)	(0.0598)
Adjusted R^2	0.7648	0.7682	0.7693	0.7688	0.7693	0.7574	0.8339	0.8369	0.8403	0.8448	0.8486	0.8514
N	692	692	692	647	647	557	159	159	159	126	126	110

Panel B: Math	Full sample						Discontinuity Sample (±5)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	-0.0756*	-0.0575	-0.0482	-0.0461	-0.0422	-0.0426	-0.1612*	-0.1356	-0.1329	-0.0913	-0.1170	-0.1267
	(0.0451)	(0.0624)	(0.0619)	(0.0601)	(0.0590)	(0.0669)	(0.0928)	(0.0970)	(0.0940)	(0.0948)	(0.0868)	(0.0847)
NAAA Score (Math)	0.8878***	0.8935***	0.8991***	0.8845***	0.8899***	0.8616***	0.9114***	0.9322***	0.9337***	0.9015***	0.8613***	0.8398***
	(0.0212)	(0.0225)	(0.0221)	(0.0231)	(0.0228)	(0.0295)	(0.0338)	(0.0386)	(0.0418)	(0.0435)	(0.0433)	(0.0524)
Adjusted R^2	0.7989	0.7997	0.8037	0.7966	0.8010	0.7895	0.8372	0.8404	0.8349	0.8390	0.8515	0.8681
N	689	689	689	644	644	555	157	157	157	124	124	109
3rd-Order polynomials of enrollment		yes	yes	yes	yes	yes		yes	yes	yes	yes	yes
Intended small class excluded				yes	yes	yes				yes	yes	yes
Year dummies			yes		yes	yes			yes		yes	yes
Neighborhood dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year-Neighborhood dummies			yes		yes	yes			yes		yes	yes
Land prices						yes						yes

Note: The dependent variable is the Yokohama City Achievement Test. All estimations include a constant. We control enrollment, enrollment²/100, and enrollment³/10000 if the "third-order of polynomials of enrollment" is "yes". We exclude schools which implement intended small classes from the estimations if the "intended small class excluded" is "yes". The instrument variable for average class size is predicted class size as calculated by equation (1). Standard errors in parentheses are clustered by school. Significance level: *** 1%, ** 5%, * 10%.

Table 6. Estimations of the class size effects on test scores with value-added model type 1 by OLS (third grade in junior high school: ninth grade)

Panel A: Japanese language	Full sample						Discontinuity Sample (±5)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	0.1033	0.0803	0.0834	0.0966	0.1091	0.0345	0.0501	0.1813	0.2778	0.1691	0.3145	0.3128
	(0.1052)	(0.1080)	(0.1129)	(0.1132)	(0.1195)	(0.1128)	(0.1367)	(0.1699)	(0.2172)	(0.1685)	(0.2374)	(0.2458)
NAAA Score (Japanese language)	0.8853***	0.8822***	0.8802***	0.8728***	0.8703***	0.9169***	0.9303***	0.9027***	0.9731***	0.8893***	0.9742***	0.9701***
	(0.0401)	(0.0417)	(0.0438)	(0.0427)	(0.0448)	(0.0527)	(0.0763)	(0.0775)	(0.0879)	(0.0770)	(0.0932)	(0.0982)
Adjusted R^2	0.8201	0.8187	0.8126	0.8175	0.8117	0.8064	0.8509	0.8527	0.8392	0.8564	0.8437	0.8384
N	288	288	288	280	280	272	67	67	67	62	62	62

Panel B: Math	Full sample						Discontinuity Sample (±5)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	-0.0189	-0.0213	-0.0456	-0.0412	-0.0579	-0.0449	0.0140	0.1677	0.1175	0.1327	0.0692	0.0791
	(0.0878)	(0.0969)	(0.1034)	(0.0999)	(0.1089)	(0.1066)	(0.1677)	(0.2300)	(0.2425)	(0.2481)	(0.2739)	(0.2855)
NAAA Score (Math)	0.8708***	0.8612***	0.8663***	0.8500***	0.8549***	0.8369***	0.8737***	0.8248***	0.9013***	0.8151***	0.8819***	0.9134***
	(0.0351)	(0.0330)	(0.0351)	(0.0331)	(0.0355)	(0.0415)	(0.1022)	(0.1155)	(0.1369)	(0.1163)	(0.1397)	(0.1713)
Adjusted R^2	0.8217	0.8219	0.8203	0.8209	0.8191	0.8178	0.7747	0.7786	0.7680	0.7812	0.7720	0.7670
N	288	288	288	280	280	272	67	67	67	62	62	62
3rd-Order polynomials of enrollment		yes	yes	yes	yes	yes		yes	yes	yes	yes	yes
Intended small class excluded				yes	yes	yes				yes	yes	yes
Year dummies			yes		yes	yes			yes		yes	yes
Neighborhood dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year-Neighborhood dummies			yes		yes	yes			yes		yes	yes
Land prices						yes						yes

Note: The dependent variable is the Yokohama City Achievement Test. All estimations include a constant and female-ratio. We control enrollment, enrollment²/100, and enrollment³/10000 if the “third-order of polynomials of enrollment” is “yes”. We exclude schools which implement intended small classes from the estimations if the “intended small class excluded” is “yes”. Standard errors in parentheses are clustered by school. Significance level: *** 1%, ** 5%, * 10%.

Table 7. Estimations of the class size effects on test scores with value-added model type 1 by IV (third grade in junior high school: ninth grade)

Panel A: Japanese language	Full sample						Discontinuity Sample (±5)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	0.1330	0.1197	0.1384	0.0458	0.0492	-0.0554	0.0799	0.1752	0.3492	0.0569	0.2029	0.1938
	(0.1375)	(0.1645)	(0.1656)	(0.1474)	(0.1503)	(0.1417)	(0.1595)	(0.2222)	(0.2628)	(0.1706)	(0.1904)	(0.1929)
NAAA Score (Japanese language)	0.8837***	0.8818***	0.8794***	0.8733***	0.8710***	0.9204***	0.9299***	0.9028***	0.9722***	0.8881***	0.9708***	0.9661***
	(0.0386)	(0.0396)	(0.0400)	(0.0411)	(0.0415)	(0.0486)	(0.0623)	(0.0614)	(0.0597)	(0.0625)	(0.0638)	(0.0664)
Adjusted R^2	0.8200	0.8186	0.8124	0.8173	0.8114	0.8058	0.8507	0.8527	0.8384	0.8549	0.8419	0.8363
N	288	288	288	280	280	272	67	67	67	62	62	62

Panel B: Math	Full sample						Discontinuity Sample (±5)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	0.0042	-0.0151	-0.0049	-0.0057	-0.0209	-0.0119	0.1009	0.2366	0.2240	0.1766	0.0959	0.1303
	(0.1199)	(0.1517)	(0.1624)	(0.1286)	(0.1381)	(0.1374)	(0.1958)	(0.2809)	(0.2523)	(0.2621)	(0.2306)	(0.2314)
NAAA Score (Math)	0.8697***	0.8612***	0.8663***	0.8500***	0.8549***	0.8362***	0.8762***	0.8256***	0.9030***	0.8162***	0.8830***	0.9163***
	(0.0338)	(0.0314)	(0.0323)	(0.0316)	(0.0327)	(0.0380)	(0.0847)	(0.0914)	(0.0944)	(0.0934)	(0.0954)	(0.1152)
Adjusted R^2	0.8216	0.8219	0.8201	0.8208	0.8190	0.8178	0.7732	0.7781	0.7664	0.7810	0.7719	0.7667
N	288	288	288	280	280	272	67	67	67	62	62	62
3rd-Order polynomials of enrollment		yes	yes	yes	yes	yes		yes	yes	yes	yes	yes
Intended small class excluded				yes	yes	yes				yes	yes	yes
Year dummies			yes		yes	yes			yes		yes	yes
Neighborhood dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year-Neighborhood dummies			yes		yes	yes			yes		yes	yes
Land prices						yes						yes

Note: The dependent variable is the Yokohama City Achievement Test. All estimations include a constant and female-ratio. We control enrollment, $enrollment^2/100$, and $enrollment^3/10000$ if the "third-order of polynomials of enrollment" is "yes". We exclude schools which implement intended small classes from the estimations if the "intended small class excluded" is "yes". The instrument variable for average class size is predicted class size as calculated by equation (1). Standard errors in parentheses are clustered by school. Significance level: *** 1%, ** 5%, * 10%.

Table 8. Estimations of the class size effects on test scores with value-added model type 2 (sixth grade in elementary school)

Panel A: Japanese language and Math	OLS / Full sample						OLS / Discontinuity Sample (±5)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	-0.0533	-0.0608	-0.0595	-0.0651	-0.0658	-0.0779	-0.0854	-0.0627	-0.0432	-0.0594	-0.0407	-0.0797
	(0.0347)	(0.0500)	(0.0511)	(0.0499)	(0.0510)	(0.0572)	(0.0609)	(0.0677)	(0.0683)	(0.0691)	(0.0682)	(0.0770)
NAAA Score (Japanese language)	0.3711***	0.3646***	0.3631***	0.3672***	0.3638***	0.3602***	0.4020***	0.4060***	0.4089***	0.4130***	0.3776***	0.5056***
	(0.0447)	(0.0458)	(0.0449)	(0.0487)	(0.0480)	(0.0563)	(0.0778)	(0.0805)	(0.0808)	(0.1156)	(0.1035)	(0.1257)
NAAA Score (Math)	0.5365***	0.5424***	0.5481***	0.5353***	0.5426***	0.5259***	0.5069***	0.5118***	0.4956***	0.4898***	0.4827***	0.3204**
	(0.0445)	(0.0448)	(0.0438)	(0.0472)	(0.0462)	(0.0534)	(0.0759)	(0.0784)	(0.0806)	(0.1099)	(0.1054)	(0.1288)
Adjusted R^2	0.7910	0.7928	0.7957	0.7917	0.7945	0.7795	0.8447	0.8473	0.8470	0.8519	0.8589	0.8618
N	1381	1381	1381	1291	1291	1112	316	316	316	250	250	219

Panel B: Japanese language and Math	IV / Full sample						IV / Discontinuity Sample (±5)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	-0.0680*	-0.0871	-0.0847	-0.0800	-0.0813	-0.0970*	-0.1254*	-0.1173*	-0.0998	-0.0897	-0.0825	-0.1272*
	(0.0374)	(0.0533)	(0.0543)	(0.0517)	(0.0525)	(0.0584)	(0.0692)	(0.0708)	(0.0718)	(0.0712)	(0.0688)	(0.0749)
NAAA Score (Japanese language)	0.3726***	0.3638***	0.3624***	0.3667***	0.3635***	0.3601***	0.4037***	0.4061***	0.4037***	0.4140***	0.3735***	0.5056***
	(0.0442)	(0.0454)	(0.0441)	(0.0482)	(0.0471)	(0.0552)	(0.0752)	(0.0777)	(0.0759)	(0.1103)	(0.0960)	(0.1154)
NAAA Score (Math)	0.5367***	0.5431***	0.5488***	0.5357***	0.5430***	0.5262***	0.5118***	0.5158***	0.5071***	0.4920***	0.4929***	0.3251***
	(0.0441)	(0.0443)	(0.0431)	(0.0467)	(0.0454)	(0.0524)	(0.0740)	(0.0759)	(0.0768)	(0.1051)	(0.0993)	(0.1188)
Adjusted R^2	0.7909	0.7928	0.7957	0.7917	0.7945	0.7794	0.8443	0.8466	0.8462	0.8517	0.8584	0.8613
N	1381	1381	1381	1291	1291	1112	316	316	316	250	250	219
3rd-Order polynomials of enrollment		yes	yes	yes	yes	yes		yes	yes	yes	yes	yes
Intended small class excluded				yes	yes	yes				yes	yes	yes
Year dummies			yes		yes	yes			yes		yes	yes
Neighborhood dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year-Neighborhood dummies			yes		yes	yes			yes		yes	yes
Land prices						yes						yes

Note: The dependent variable is the Yokohama City Achievement Test. All estimations include a constant. We control enrollment, enrollment²/100, and enrollment³/10000 if the "third-order of polynomials of enrollment" is "yes". We exclude schools which implement intended small classes from the estimations if the "intended small class excluded" is "yes". The instrument variable for average class size is predicted class size as calculated by equation (1). Standard errors in parentheses are clustered by school. Significance level: *** 1%, ** 5%, * 10%.

Table 9. Estimations of the class size effects on test scores with value-added model type 2 (third grade in junior high school: ninth grade)

Panel A: Japanese language and Math	OLS / Full sample						OLS / Discontinuity Sample (±5)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	0.0385	0.0316	0.0189	0.0294	0.0258	-0.0026	0.0103	0.1493	0.1809	0.1411	0.1936	0.1962
	(0.0798)	(0.0758)	(0.0797)	(0.0803)	(0.0849)	(0.0806)	(0.0955)	(0.1121)	(0.1201)	(0.1171)	(0.1347)	(0.1366)
NAAA Score (Japanese language)	0.4246***	0.4229***	0.4387***	0.4135***	0.4310***	0.4125***	0.6317***	0.6492***	0.6420***	0.6299***	0.6238***	0.6106***
	(0.0562)	(0.0570)	(0.0588)	(0.0594)	(0.0615)	(0.0665)	(0.1438)	(0.1350)	(0.1525)	(0.1324)	(0.1522)	(0.1788)
NAAA Score (Math)	0.4682***	0.4642***	0.4499***	0.4631***	0.4466***	0.4831***	0.3034**	0.2471*	0.3326*	0.2589*	0.3459**	0.3651
	(0.0553)	(0.0556)	(0.0573)	(0.0571)	(0.0594)	(0.0685)	(0.1482)	(0.1404)	(0.1680)	(0.1367)	(0.1632)	(0.2231)
Adjusted R^2	0.8109	0.8107	0.8095	0.8087	0.8076	0.8011	0.8419	0.8496	0.8465	0.8483	0.8457	0.8441
N	576	576	576	560	560	544	134	134	134	124	124	124

Panel B: Japanese language and Math	IV / Full sample						IV / Discontinuity Sample (±5)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Average class size	0.0661	0.0589	0.0704	0.0281	0.0184	-0.0222	0.0556	0.1534	0.2429	0.0841	0.1236	0.1320
	(0.1063)	(0.1240)	(0.1287)	(0.1055)	(0.1099)	(0.1055)	(0.1360)	(0.1780)	(0.1948)	(0.1440)	(0.1431)	(0.1458)
NAAA Score (Japanese language)	0.4222***	0.4212***	0.4346***	0.4136***	0.4314***	0.4145***	0.6197***	0.6486***	0.6343***	0.6341***	0.6258***	0.6146***
	(0.0546)	(0.0553)	(0.0558)	(0.0580)	(0.0587)	(0.0633)	(0.1246)	(0.1164)	(0.1282)	(0.1196)	(0.1317)	(0.1539)
NAAA Score (Math)	0.4691***	0.4658***	0.4535***	0.4630***	0.4462***	0.4818***	0.3157**	0.2477**	0.3409**	0.2536**	0.3412**	0.3576*
	(0.0539)	(0.0543)	(0.0548)	(0.0558)	(0.0568)	(0.0653)	(0.1264)	(0.1199)	(0.1406)	(0.1237)	(0.1406)	(0.1919)
Adjusted R^2	0.8109	0.8106	0.8093	0.8087	0.8076	0.8010	0.8415	0.8496	0.8462	0.8480	0.8453	0.8437
N	576	576	576	560	560	544	134	134	134	124	124	124
3rd-Order polynomials of enrollment		yes	yes	yes	yes	yes		yes	yes	yes	yes	yes
Intended small class excluded				yes	yes	yes				yes	yes	yes
Year dummies			yes		yes	yes			yes		yes	yes
Neighborhood dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year-Neighborhood dummies			yes		yes	yes			yes		yes	yes
Land prices						yes						yes

Note: The dependent variable is the Yokohama City Achievement Test. All estimations include a constant and female-ratio. We control enrollment, enrollment²/100, and enrollment³/10000 if the "third-order of polynomials of enrollment" is "yes". We exclude schools which implement intended small classes from the estimations if the "intended small class excluded" is "yes". The instrument variable for average class size is predicted class size as calculated by equation (1). Standard errors in parentheses are clustered by school. Significance level: *** 1%, ** 5%, * 10%.

Table 10. Estimations of the class size effects on test scores in subgroup divided by initial test score with value-added model type 1 by IV (sixth grade in elementary school)

Panel A (Sample: More than median of NAAA)

	Japanese language						Math					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Coefficient of average class size	-0.1626***	-0.2220***	-0.2180***	-0.1993***	-0.2007***	-0.2260***	-0.1031	-0.1312	-0.1396	-0.1239	-0.1460	-0.1489
(standard error)	(0.0587)	(0.0737)	(0.0755)	(0.0721)	(0.0738)	(0.0844)	(0.0761)	(0.1004)	(0.0965)	(0.0950)	(0.0898)	(0.1040)
Initial NAAA mean score (full sample)	58.03						58.03					
(standard deviation)	(5.40)						(5.27)					
Adjusted R^2	0.6141	0.6148	0.6113	0.6262	0.6252	0.5975	0.6540	0.6625	0.6805	0.6636	0.6846	0.6682
N	346	346	346	324	324	288	345	345	345	322	322	283

Panel B (Sample: Less than median of NAAA)

	Japanese language						Math					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Coefficient of average class size	0.0074	-0.0659	-0.0758	-0.0690	-0.0755	-0.1331	-0.0304	-0.0091	0.0192	-0.0012	0.0299	0.0047
(standard error)	(0.0708)	(0.0944)	(0.0946)	(0.0918)	(0.0914)	(0.0996)	(0.0551)	(0.0775)	(0.0805)	(0.0764)	(0.0778)	(0.0874)
Initial NAAA mean score (full sample)	41.97						41.97					
(standard deviation)	(6.45)						(6.54)					
Adjusted R^2	0.5099	0.5206	0.5366	0.5183	0.5304	0.5131	0.5204	0.5211	0.5204	0.5155	0.5118	0.4700
N	346	346	346	323	323	269	344	344	344	322	322	272
3rd-Order polynomials of enrollment		yes	yes	yes	yes	yes		yes	yes	yes	yes	yes
Intended small class excluded				yes	yes	yes				yes	yes	yes
Year dummies			yes		yes	yes			yes		yes	yes
Neighborhood dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year-Neighborhood dummies			yes		yes	yes			yes		yes	yes
Land prices						yes						yes

Note: The dependent variable is the Yokohama City Achievement Test. All estimations include a constant. We control enrollment, enrollment^2/100, and enrollment^3/10000 if the "third-order of polynomials of enrollment" is "yes". We exclude schools which implement intended small classes from the estimations if the "intended small class excluded" is "yes". The instrument variable for average class size is predicted class size as calculated by equation (1). Standard errors in parentheses are clustered by school. Significance level: *** 1%, ** 5%, * 10%.

Table 11. Estimations of the class size effects on test scores in subgroup divided by land prices with value-added model type 1 by IV (sixth grade in elementary school)

Panel A (Sample: More than median of land prices)	Japanese language						Math					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Coefficient of average class size	-0.1302	-0.3132***	-0.3117***	-0.3281***	-0.3380***	-0.3378***	-0.2226***	-0.2303***	-0.2213***	-0.2365***	-0.2437***	-0.2511***
(standard error)	(0.1085)	(0.0911)	(0.0881)	(0.0866)	(0.0838)	(0.0835)	(0.0568)	(0.0814)	(0.0812)	(0.0783)	(0.0780)	(0.0778)
Initial NAAA mean score (full sample)	53.51						53.3					
(standard deviation)	(9.86)						(9.83)					
Adjusted R^2	0.7614	0.7734	0.7678	0.7687	0.7629	0.7619	0.7993	0.7995	0.8041	0.7956	0.8000	0.8023
N	298	298	298	280	280	280	298	298	298	280	280	280

Panel B (Sample: Less than median of land prices)	Japanese language						Math					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Coefficient of average class size	0.0696	0.0061	-0.0101	0.0107	-0.0095	-0.0075	0.0808	0.1320	0.1383	0.1470	0.1466	0.1421
(standard error)	(0.0597)	(0.0852)	(0.0873)	(0.0844)	(0.0869)	(0.0852)	(0.0678)	(0.0967)	(0.0962)	(0.0939)	(0.0929)	(0.0926)
Initial NAAA mean score (full sample)	47.38						47.44					
(standard deviation)	(8.43)						(8.56)					
Adjusted R^2	0.6862	0.6923	0.6926	0.6995	0.7015	0.7006	0.7015	0.7068	0.7098	0.7109	0.7155	0.7162
N	296	296	296	277	277	277	294	294	294	275	275	275
3rd-Order polynomials of enrollment		yes	yes	yes	yes	yes		yes	yes	yes	yes	yes
Intended small class excluded				yes	yes	yes				yes	yes	yes
Year dummies			yes		yes	yes			yes		yes	yes
Neighborhood dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year-Neighborhood dummies			yes		yes	yes			yes		yes	yes
Land prices						yes						yes

Note: The dependent variable is the Yokohama City Achievement Test. All estimations include a constant. We control enrollment, enrollment²/100, and enrollment³/10000 if the "third-order of polynomials of enrollment" is "yes". We exclude schools which implement intended small classes from the estimations if the "intended small class excluded" is "yes". The instrument variable for average class size is predicted class size as calculated by equation (1). Standard errors in parentheses are clustered by school. Significance level: *** 1%, ** 5%, * 10%.

Table 12. Estimations of the class size effects on test scores with fixed effects model (sixth grade in elementary school)

	Japanese language					Math				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Average class size	-0.1212	-0.1394*	-0.1445*	-0.1524*	-0.1492*	0.0111	0.0139	0.0319	-0.0015	0.0308
	(0.0844)	(0.0755)	(0.0802)	(0.0792)	(0.0840)	(0.0834)	(0.0876)	(0.0829)	(0.0923)	(0.0854)
NAAA Score	0.6069***	0.6055***	0.6266***	0.5701***	0.5952***	0.6793***	0.6784***	0.7071***	0.6488***	0.6800***
(Corresponding dependent variable)	(0.0484)	(0.0475)	(0.0470)	(0.0478)	(0.0452)	(0.0492)	(0.0493)	(0.0468)	(0.0552)	(0.0528)
Adjusted R^2	0.3082	0.3151	0.3359	0.2863	0.3156	0.3629	0.3633	0.4087	0.3256	0.3804
N	692	692	692	647	647	689	689	689	644	644
3rd-Order polynomials of enrollment		yes	yes	yes	yes		yes	yes	yes	yes
Intended small class excluded				yes	yes				yes	yes
Year dummies			yes		yes			yes		yes
Year-Neighborhood dummies			yes		yes			yes		yes

Note: The dependent variable is the Yokohama City Achievement Test. All estimations include a constant. We control enrollment, enrollment^2/100, and enrollment^3/10000 if the "third-order of polynomials of enrollment" is "yes". We exclude schools which implement intended small classes from the estimations if the "intended small class excluded" is "yes". Standard errors in parentheses are clustered by school. Significance level: *** 1%, ** 5%, * 10%.

Appendix Table A1. The first-stage results of instrumental variable estimations by OLS

	Sixth grade in elementary school						Third grade in junior high school (ninth grade)					
Panel A (First stage of value-added model type 1 for Japanese language)												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Predicted class size	0.9287***	0.8678***	0.8676***	0.9659***	0.9659***	0.9626***	0.7887***	0.6487***	0.6614***	0.8363***	0.8444***	0.8328***
	(0.0188)	(0.0277)	(0.0277)	(0.0133)	(0.0131)	(0.0154)	(0.0600)	(0.0814)	(0.0785)	(0.0560)	(0.0533)	(0.0555)
Adjusted R^2	0.8836	0.8918	0.8928	0.9470	0.9471	0.9427	0.6236	0.6744	0.6726	0.7075	0.7043	0.6935
N	692	692	692	647	647	557	288	288	288	280	280	272
Panel B (First stage of value-added model type 1 for math)												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Predicted class size	0.9297***	0.8664***	0.8664***	0.9655***	0.9657***	0.9624***	0.7913***	0.6495***	0.6625***	0.8388***	0.8472***	0.8367***
	(0.0189)	(0.0281)	(0.0280)	(0.0134)	(0.0132)	(0.0155)	(0.0599)	(0.0818)	(0.0791)	(0.0562)	(0.0535)	(0.0560)
Adjusted R^2	0.8817	0.8905	0.8915	0.9462	0.9463	0.9422	0.6233	0.6741	0.6720	0.7082	0.7049	0.6941
N	689	689	689	644	644	555	288	288	288	280	280	272
Panel C (First stage of value-added model type 2 for Japanese language and math)												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Predicted class size	0.9366***	0.8675***	0.8674***	0.9656***	0.9657***	0.9625***	0.7889***	0.6468***	0.6579***	0.8417***	0.8492***	0.8375***
	(0.0176)	(0.0275)	(0.0273)	(0.0133)	(0.0129)	(0.0152)	(0.0593)	(0.0804)	(0.0767)	(0.0525)	(0.0496)	(0.0517)
Adjusted R^2	0.8831	0.8929	0.8954	0.9475	0.9484	0.9446	0.6373	0.6883	0.6985	0.7206	0.7280	0.7193
N	1381	1381	1381	1291	1291	1112	576	576	576	560	560	544
3rd-Order polynomials of enrollment		yes	yes	yes	yes	yes		yes	yes	yes	yes	yes
Intended small class excluded				yes	yes	yes				yes	yes	yes
Year dummies			yes		yes	yes			yes		yes	yes
Neighborhood dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year-Neighborhood dummies			yes		yes	yes			yes		yes	yes
Land prices						yes						yes

Note: The dependent variable is actual average class size. Each panel includes NAAA score of corresponding subjects for the dependent variable in the second stage. Panels A and B correspond to the results in Tables 5 and 7, and Panel C corresponds to Tables 8 and 9. All estimations include a constant and the estimations of junior high school also include female-ratio. We control enrollment, $\text{enrollment}^2/100$, and $\text{enrollment}^3/10000$ if the “third-order of polynomials of enrollment” is “yes”. We exclude schools which implement intended small classes from the estimations if the “intended small class excluded” is “yes”. Standard errors in parentheses are clustered by school. Significance level: *** 1%, ** 5%, * 10%.