Japan & The World Economy 53 (2020) 100977

Contents lists available at ScienceDirect

Japan & The World Economy

journal homepage: www.elsevier.com/locate/jwe

Effects of class-size reduction on cognitive and non-cognitive skills☆

Hirotake Ito[a], Makiko Nakamuro[a], Shintaro Yamaguchi[b]

[a] Graduate School of Media and Governance, Keio University, 5322 Endo,Fujisawa-shi, Kanagawa 252-0882, Japan
[b] Faculty of Economics, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

ARTICLE INFO

Keywords:
Regression discontinuity design
Longitudinal data
Maimonides' rule

ABSTRACT

We estimate the effects of class-size reduction by exploiting exogenous variation caused by Maimonides' rule, which requires that the maximum class size is 40 students and that classes be split into two when 41 students are enrolled. Our data cover all fourth to ninth graders in 1064 public schools in an anonymous prefecture of Japan for three years. We find that the effects of class-size reduction on academic test scores are statistically and/or economically insignificant when school fixed effects are controlled. We find no evidence that small class size improves non-cognitive skills.

1. Introduction

The general public and education administrators appear to believe that smaller class sizes contributes to greater learning and better experiences for students. However, this is far from obvious. Indeed, the literature on the economics of education has long sought evidence for the effectiveness of class-size reductions. The analysis of data from a randomized controlled trial (RCT) has established positive effects of smaller class size in the context of famous Project STAR (e.g. Krueger, 1999), which was implemented in the state of Tennessee in the United States in the 1980s. Nevertheless, its external validity must be verified in each country's context.

Because conducting RCTs can be expensive and politically controversial, many researchers have attempted to estimate the effects of class-size reduction using a natural experiment. Angrist and Lavy (1999) was the first study to exploit the exogenous variation of class size generated by Maimonides' rule, using Israeli data. That is, regulations require that the maximum class size is 40 students and that classes be split into two when 41 or more students are enrolled in a given grade. Because Maimonides' rule is applied in many other countries outside of Israel, including Japan, many researchers have estimated the effects of class-size reduction following the approach developed by Angrist and Lavy (1999).

There are several studies on class size that use Japanese data (for example, see (Niki, 2013; Hojo, 2013; Senoh et al., 2014; Akabayashi and Nakamura, 2014; Senoh and Hojo, 2016; Ito et al., 2017)). So far, the evidence from Japan is mixed, presumably owing to differences in the statistical methods and samples chosen. Some studies report insignificant results, whereas others report positive significant effects of class-size reduction, although the effect size is typically small. From our reading of the literature, there does not appear to be any consensus about the effects of class-size reduction in Japan.

We contribute to this literature by providing an additional piece of evidence based on large scale data from an anonymous prefecture[1] in Japan. Our research has the following three key features. First, our data set is large and covers all students in grades four to nine in 1064 public schools in an anonymous prefecture for three years. Although our data set is smaller than the one used by Senoh et al. (2014), which covers all sixth and ninth graders in Japanese public schools, it is much larger than the data set used in studies by, e.g., Akabayashi and Nakamura (2014), Ito et al. (2017), both of which examined data from municipality, which is a subdivision of a prefecture. Second, our data set includes non-cognitive skill measures, including conscientiousness, self-control, and self-efficacy. Evidence for the effects on non-cognitive skills is relatively scarce. Third, we control for school fixed effects to address possible omitted variable biases. Many previous studies use data from a single year, which prevents them from controlling for school fixed effects. Our large data set enables us to overcome this

☆ We thank an anonymous prefecture for providing data and research fund. This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (A), 2016-2021 (16H02022, Makiko Nakamuro) and conducted as a part of the Project Research on the Improvement in Resource Allocation and Productivity among the Healthcare and Education Service Industries undertaken at Research Institute of Economy,Trade and Industry (RIETI).

E-mail addresses: itouhrtk@keio.jp (H. Ito), makikon@sfc.keio.ac.jp (M. Nakamuro), syamaguchi@e.u-tokyo.ac.jp (S. Yamaguchi).

[1] Prefectures are the first level of jurisdiction and administrative division. There are 47 prefectures in Japan and they are somewhat comparable to provinces and states in other countries.

https://doi.org/10.1016/j.japwor.2019.100977
Received 31 January 2019; Received in revised form 22 August 2019; Accepted 26 August 2019
Available online 16 October 2019
0922-1425/

limitation.

Our estimates indicate that the effects of class-size reduction on Japanese and math test scores are statistically and/or economically insignificant for students in grades four to eight. In our preferred specification with school fixed effects, the only statistically significant estimate is for math test scores in grade six. This estimate implies that a 10-student reduction improves the test score by 0.03 standard deviations, which we consider very minimal. Further, we find no evidence that class-size reduction improves students' non-cognitive skills.

We also examine the heterogeneity of class-size effects by students' socio economic status (SES). Our estimates suggest that the effects of class-size reduction may be slightly stronger for students who do not attend a private tutoring school. However, it should be noted that our measures of SES may be endogenous, and hence, our results are suggestive at best.

The rest of the paper is structured as follows. We review the literature briefly in Section 2 and describe the data and present descriptive statistics in Section 3. We outline our identification strategy in Section 4. The empirical results are presented in Section 5. We conclude in Section 6.

2. Literature review

A potential endogeneity bias arises from the correlation between class size and unobserved school characteristics. Most of the prior research addressed this problem by taking one of the following approaches: (i) randomized experiments, such as Project STAR (e.g. Krueger, 1999); (ii) natural experiments, using the situation where schools have a single class per grade and a monopoly in their area of influence (e.g. Urquiola, 2006) or using variations in enrollment driven by cohort sizes across different years (e.g. Hoxby, 2000); and (iii) the regression discontinuity design exploiting exogenous and discontinuous variations in class size around the cutoffs (e.g. Angrist and Lavy, 1999).

Earlier studies on class-size effects have reported large effects on students' academic performance. The STAR experiment in the US yielded an effect size of 0.13–0.27 standard deviations for an eight-student reduction and an effect size of 0.16–0.33 standard deviations for a 10-student reduction (see Finn and Achilles [1999, Table 5]). In Israel, a 10-student reduction in class size increased the standardized test scores by 0.13–0.25 standard deviations for students in grades four and five (Angrist and Lavy, 1999).[2]

Angrist and Lavy (1999) used a large Israeli sample from 2002 to 2011 and found that the class-size effect was nearly zero, with small standard errors. Interestingly, Angrist and Lavy (1999) stated that the large significant estimates for class-size effects previously reported by Angrist and Lavy (1999) may have been “a chance finding.” They argued that “it seems fair to say that the 1991 results are unusual in showing strong class-size effects,” essentially dismissing the conclusion that smaller class size improves students' performance in Israel.

There are only a few studies that examine the effects on outcomes beyond academic test scores. Dee and West (2011) found that smaller class size was associated with improvements in school engagement. Fredriksson et al. (2013) evaluated the longer-run effects of class-size reduction on cognitive and non-cognitive outcomes. They reported that small classes for students aged 10–13 years improved several outcomes, including effort motivation, aspirations, self-confidence, sociability, absenteeism, and for ages 13–16 years, improved anxiety. Chetty et al. (2011) linked the experimental data from the STAR project to administrative records and found that students in smaller classes were significantly more likely to attend college. Furthermore, they exhibited statistically significant improvements in outcomes concerning home ownership, savings, mobility rates, college graduate, and marital status.

Most existing estimates from Japanese data indicate that the effects of class-size reduction are statistically insignificant or small. Using a nationally representative sample of students from the 2003 Trend in International Mathematics and Science Study (TIMSS), Niki (2013) found insignificant effects of class size on mathematics and science test scores as well as on non-cognitive skill measures for eighth graders. Hojo (2013) drew a sample of fourth graders from the 2003 TIMSS and found marginally positive significant effects of small class size.

Senoh et al. (2014) analyzed the data from the 2009 National Assessment of Academic Ability (NAAA) study that covered Japanese and math test scores for virtually all students in grade six and nine in Japanese public schools. They found the effects of class-size reduction were insignificant across subjects and grades except for Japanese test scores for sixth graders.

Senoh and Hojo (2016) updated Senoh et al. (2014) using family background information added to the 2009 NAAA via a follow-up survey. They found that the effects of a small class were positive and significant for ninth graders once students' SES was controlled for. Although their estimates are greater than the previous estimates from Japanese data, the effect size was still less than half of the estimates from the Project STAR. Importantly, they found that the effects of class size were stronger in schools attended by many students from low SES families.

Combining data from the 2007 NAAA and Yokohama City Achievement Test, Akabayashi and Nakamura (2014) estimated the effects of class size on sixth and ninth graders in the city of Yokohama. In addition to exploiting the exogenous variation arising from Maimonides' rule, they controlled for unobserved heterogeneity using a value-added model and controlling for school fixed effects. They estimated class-size effects on Japanese and math tests for sixth and ninth graders and found insignificant effects across subjects and grades except for Japanese test scores for sixth graders.

Ito et al. (2017) used data from an anonymous city for nine years and estimated class-size effects on cognitive and non-cognitive outcomes. By demeaning variables at the school level, they exploited within-school variation as well as the exogenous variation arising from Maimonides' rule. Note that their demeaning is essentially equivalent to controlling for school fixed effects. By pooling observations from grades four to nine, they estimated a random effect model and found positive effects of class size on a wide array of cognitive and non-cognitive outcomes. Nevertheless, the effect size was modest and comparable to the results of Senoh and Hojo (2016).

3. Data

3.1. Overview

Our data are drawn from the standardized achievement tests on Japanese and math from an anonymous prefecture in 2016–2018. Students who took the exams also completed a series of questionnaires to measure some non-cognitive skills and were asked about their life.

The data cover all public school students in grades four to nine in the prefecture. Although there are some private and national schools, the majority of schools in the prefecture are public. According to official statistics, 99.3% of elementary schools and 93.0% of junior high schools in this prefecture are public. Our data cover approximately 300,000 students in 1064 public schools (708 elementary schools and 356 junior high schools) in 62 municipalities.[3]

[2] Note that the reported regression coefficients in Angrist and Lavy (1999) are normalized using the standard deviation of class-average scores, instead of individual student's scores. We discuss their estimates adjusted to the standard deviation at the individual level.

[3] Unfortunately, we do not have access to information for students who were absent on the test day. Around 2–3% of students were absent in the survey period. This is similar number to the fraction of absent students in NAAA, which was administered nationwide by the Ministry of Education, Culture, Sports,

Table 1
Descriptive statistics by grade.

	Variable	Year	Grade 4	5	6	7	8
School	Grade enrollment	2016	88.195 (47775, 36.158)	90.157 (49118, 36.690)	91.726 (47454, 37.117)	166.338 (47807, 60.826)	165.155 (48048, 59.978)
		2017	91.661 (49308, 38.592)	88.119 (48106, 35.967)	89.521 (46377, 36.709)	164.487 (47441, 60.964)	163.596 (47306, 59.631)
		2018	90.970 (49095, 36.986)	91.775 (49465, 38.754)	88.061 (45489, 36.169)	161.144 (46198, 59.594)	161.844 (46813, 59.831)
	Class size	2016	31.646 (47775, 5.245)	31.854 (49118, 5.074)	32.115 (47454, 5.030)	33.935 (47807, 3.474)	34.247 (48048, 3.273)
		2017	32.041 (49308, 5.113)	31.625 (48106, 5.109)	31.797 (46377, 4.998)	33.887 (47441, 3.600)	34.115 (47306, 3.534)
		2018	32.070 (49095, 5.276)	32.215 (49465, 5.097)	31.527 (45489, 5.142)	33.767 (46198, 3.535)	34.261 (46813, 3.566)
SES	No Books	2016	0.119 (49441, 0.324)	0.097 (48471, 0.295)	0.084 (49748, 0.277)	0.122 (48428, 0.328)	0.135 (48118, 0.342)
		2017	0.122 (48966, 0.327)	0.093 (49844, 0.291)	0.081 (48908, 0.274)	0.113 (47996, 0.317)	0.120 (48191, 0.325)
		2018	0.127 (46793, 0.333)	0.094 (49601, 0.292)	0.082 (50095, 0.274)	0.112 (46391, 0.315)	0.118 (46285, 0.322)
	Private Tutoring	2016	0.519 (48337, 0.500)	0.503 (47703, 0.500)	0.508 (49519, 0.500)	0.508 (48363, 0.500)	0.588 (48006, 0.492)
		2017	0.603 (47066, 0.489)	0.564 (48458, 0.496)	0.563 (47999, 0.496)	0.565 (47117, 0.496)	0.618 (47784, 0.486)
		2018	0.617 (46749, 0.486)	0.569 (48529, 0.495)	0.567 (49294, 0.495)	0.566 (46078, 0.496)	0.593 (46249, 0.491)

Note: The unit of observation is students. Means are reported in each cell along with the number of observations and standard deviations in parentheses (in this order). The variable "No Books" is a dummy variable that takes one if a student has no book at home and takes zero otherwise. The variable "Private Tutoring" is a dummy variable that takes one if a student goes to a private tutoring school and takes zero otherwise.

Students took the achievement tests and responded to the survey in the second week of April in each of the study years because April is the beginning of the academic year in Japan. Because it may take several months for class size to affect students' outcomes, we examine the relationship between the class size in the current year and outcomes at the beginning of the following year. For example, to estimate the effects of class size in grade g in year t, we use test scores in grade $g + 1$ in year $t + 1$, which means that we can estimate the class-size effects for the fourth to eighth graders, but not for the ninth graders.

Table 1 presents the descriptive statistics of selected variables. The average grade size is about 90 students for elementary school for grades four to six. For junior high school, average grade size is about 170 students, exceeding that of elementary schools. The average class size is 32 students for elementary school, whereas it is 34 for junior high school.

3.2. Cognitive and non-cognitive skills

Our measures of cognitive skills are test scores for Japanese and math. They are available for virtually all students from grade four to grade nine in all survey years. To facilitate interpretation, we normalize the test scores so that the mean is zero and the standard deviation is one for each grade.

We measure the following non-cognitive skills by a 40-min questionnaire: (i) self-control (Duckworth et al., 2013), (ii) self-efficacy (Pintrich et al., 1991), and (iii) conscientiousness (Barbaranelli et al., 2003). A large literature suggests that non-cognitive skills are an important determinant of subsequent outcomes in adulthood as well as academic performance (Borghans et al., 2008; Heckman et al., 2006; Heineck and Anger, 2010; Carneiro et al., 2007; Mueller and Plug, 2006).

Only one of the three non-cognitive skills is measured for each cohort to reduce the burden on students in answering questions. For example, the cohort that was in grade four in 2015 answered questions about self-control in surveys during 2015–2018. Hence, we track the self-control skills of for this cohort, but we do not track their other non-cognitive skills. Similarly, the cohort that was in grade five in 2015 answered questions about self-efficacy in 2015–2018. This survey structure prevents us from comparing different cohorts in terms of non-cognitive skills.

(*footnote continued*)
Science and Technology. We consider that the absent students may have transferred to a school outside of this prefecture or may have been sick.

Self-control is a psychological state defined as "the tendency to regulate impulses and resist immediately rewarding temptations in the service of long-term goals." (Duckworth et al., 2013) Our data employ an eight-item questionnaire, originally developed by Tsukayama et al. (2013). A self-rated scale on a five-point frequency scale, ranging from 5 (= very frequently) to 1 (= almost never), is averaged such that a higher score indicates higher self-control (see the appendix of Ito et al. (2019) for details). In computing the scale, several items, for example, intentionally negatively worded items required that their rating be reversed. If an item has to be reversed, a student who chose 1 received a score of 5.

To avoid controversy regarding the stability of measures across different situations (sometimes referred to as the "person-situation" debate), Tsukayama et al. (2013) developed a self-control scale specific to school-aged children in academic and school contexts and demonstrated that this scale successfully predicted their academic performance, hours of studying, and hours of watching television.

The second measure of non-cognitive skills is self-efficacy, defined as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances" (Bandura, 1986, p. 391). Individuals with low self-efficacy consider themselves incapable of accomplishing tasks and try to avoid them, whereas those with high self-efficacy would attempt to accomplish tasks and remain engaged in them over the long run, even if the tasks are challenging.

Our measure of self-efficacy was originally developed by Pintrich et al. (1991). An eight-item questionnaire is based on a self-rated scale on a five-point Likert scale from 5 (= very true of me) and 1 (= not at all true of me). The scale is constructed by taking the means of all items, with reverse-coded items properly reflected (see the appendix of Ito et al. (2019) for details).

This is also the scale developed specifically for learning and task performance in academic settings. Pintrich and De Groot (1990) showed that the scale is strongly correlated with the final grade over the school year. Other studies that employed this type of scale demonstrated that self-efficacy particularly predicted math performance (Pajares and Miller, 1994; Pajares and Kranzler, 1995; Pajares and Graham, 1999).

Conscientiousness is well known as one of the big five personality traits and the most robust predictor of academic achievement among them. Poropat's (2009) meta-analysis on the relationship between the big five personality traits revealed that both the raw and the partial correlations between conscientiousness and grade point average were almost as large as those between IQ and grade point average.

Our measure of conscientiousness is based on the method developed

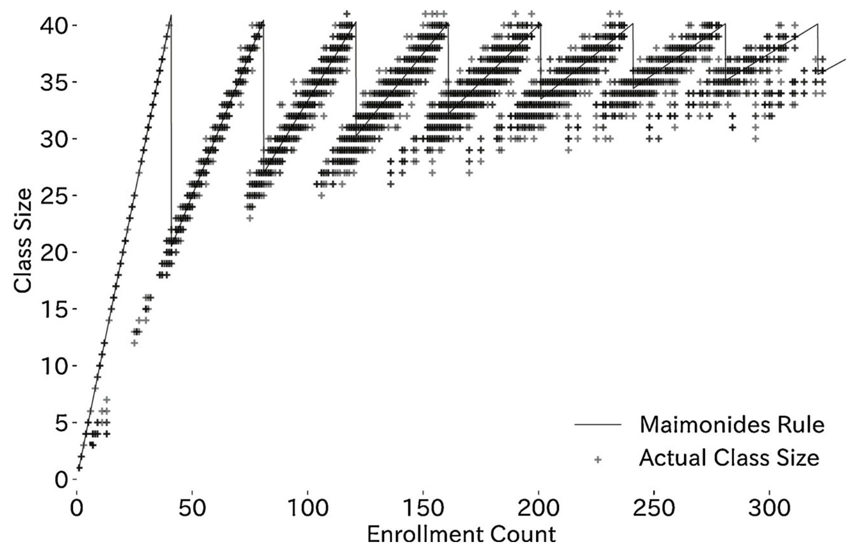

Fig. 1. Predicted vs actual class size. Note: The plot shows the relationship between enrollment count and class size (upper panel: elementary school, lower panel: junior high school). Markers represent the actual class size and solid lines indicate predicted class size calculated by Maimonides rule.

by Barbaranelli et al. (2003). It assesses students' dependability, orderliness, precision, and fulfilling of commitments in school environments. By taking an average of 13 items from the self-rated scale on a five-point Likert scale, ranging from 5 (= very true of me) to 1 (= not at all true of me), we constructed a single measure of conscientiousness (See the appendix of Ito et al. (2019) for details).

Similarly to the cognitive skill measures, we normalize the non-cognitive skill measures so that they have zero mean and unit standard deviations to facilitate interpretation.

3.3. Socio-economic status

We use two proxies for students' SES. One is an indicator for having no books at home, which is a proxy often used in education research when parents' incomes or occupations are unknown. Using internationally comparable standardized test scores, such as TIMSS and the Programme for International Student Assessment (PISA), prior studies showed a strong correlation between the number of books at home and parental SES. In addition, Kawaguchi (2016,2017) showed that the number of books at home is a good predictor of parental income and educational backgrounds in Japan.

Another proxy is the indicator of attending a private tutoring school. Because the average monthly fee to attend a private tutoring school is about 40,000 Japanese yen (equivalent to 400 US dollars), students from wealthier families are more likely to attend to improve their test scores than are students from low income families.

Table 1 indicates that 8–15% of students who have no books at home. Around 50–60% in grade four attend private tutoring school and the proportion increases with grade, reaching about 70% in grade nine (not in the table).

4. Identification strategy

We exploit exogenous and discontinuous changes in class size to estimate the causal effects on cognitive and non-cognitive skills. The maximum class size is capped at 40 students (or 35 students for those in grade one since 2011), which is legally determined by the Act on Standards for Class Formation and Fixed Number of School Personnel of Public Compulsory Education Schools. Therefore, students in a grade with up to 40 students are assigned to a single class, but a grade with 41 students is divided into two classes.

In practice, many schools in our data form classes that are smaller than the legally mandated class size. Hence, we cannot simply compare schools with grade enrolments near the cutoff. Our approach is to exploit the discontinuity of the predicted class size, rather than the actual class size.

The predicted class size in a given grade is formulated as follows:

$$z_{\mathrm{st}} = \frac{n_{\mathrm{st}}}{[(n_{\mathrm{st}} - 1)/40] + 1}, \tag{1}$$

where z_{st} is the predicted class size in school s at time t, n_{st} is grade enrolment, and the square bracket is an operator that takes the integer part of the division. This rule is known as Maimonides' rule in the literature.

We take the predicted class size as an instrument for observed class size in the following equation:

$$y_{\mathrm{ist}} = c_{\mathrm{ist}}\beta + x'_{\mathrm{ist}}\gamma + \epsilon_{\mathrm{ist}}, \tag{2}$$

where y_{ist} is student i's outcome in school s at time t, c_{ist} is the class size, and x_{ist} is a set of control variables, including grade enrollment and school fixed effects. The error term is ϵ_{ist} and it is uncorrelated with any of the observed variables. We allow for correlation of the error term within the school and hence, standard errors are clustered at the school level.

We find controlling for enrollment (= number of students in a given grade) is necessary to avoid an endogeneity bias. Note that the predicted class size z_{st} tends to increase in enrollment n_{st}, as illustrated in Fig. 1, although it is not monotonic. In fact, the actual class size and enrollment are positively correlated in our sample.

Figs. 2 and 3 provide evidence that enrollment and the SES of students' families are positively correlated. In Fig. 2, we plot enrollment and the fraction of students with no books at home. Recalling that this is a proxy for low SES, the negative regression slope in Fig. 2 implies a positive correlation between enrollment and students' SES. In Fig. 3, we plot enrollment and the fraction of students going to a private tutoring school, a proxy for higher SES. The evidence suggests that if enrollment is not controlled for in the regressions, the estimated coefficients for class size are likely to be upward biased.

In addition to enrollment, we control for school fixed effects to address a possible correlation between the predicted class size and unobserved school characteristics. Note that the identifying variation for a model with school fixed effects is the variation of the predicted class size across different cohorts (or years, equivalently) within school and grade.

However, a few concerns remain. If teachers are systematically assigned according to class size, teacher characteristics are likely to be

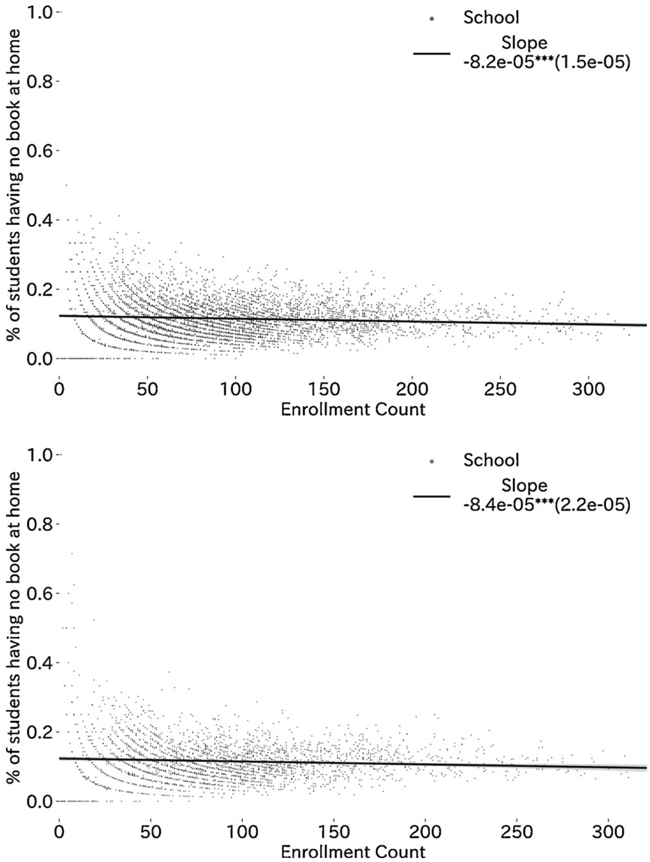

Fig. 2. Relationship between enrollment and low SES indicator.

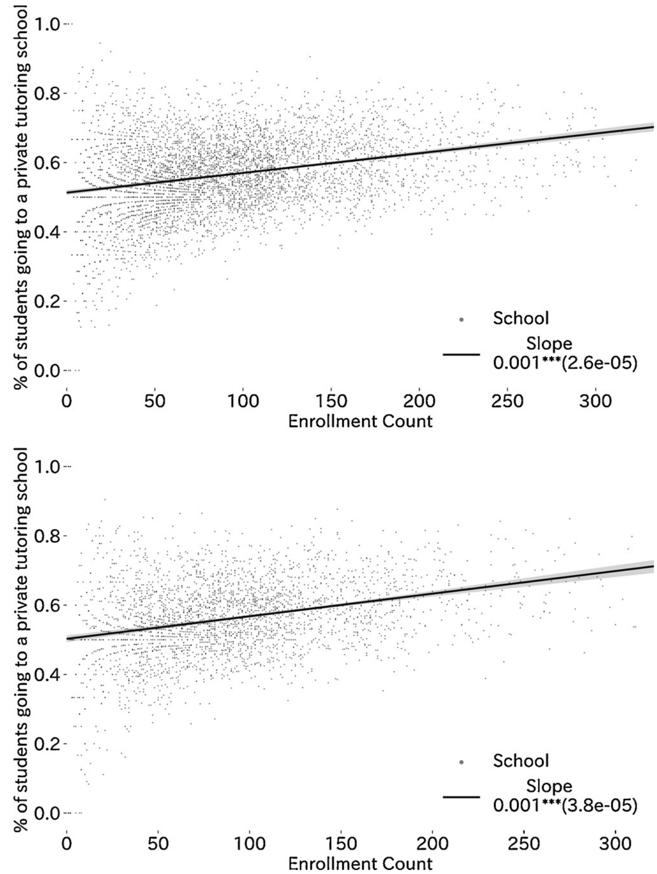

Fig. 3. relationship between enrollment and high SES indicator.

correlated with the predicted class size. Unfortunately, we are unable to control for teacher fixed effects because teacher ID is not in the data. We note that, as far as we know, no previous studies control for teacher fixed effects.

The extent of the bias may be different between elementary and junior high schools. In elementary schools, homeroom teachers teach nearly all subjects, whereas, in junior high school, subject teachers teach each subject. Because of this teaching practice, we expect that the biases arising from endogenous teacher assignment are smaller for junior high schools than for elementary schools.

Predicting the direction of the bias from omitting teacher fixed effects is difficult. On the one hand, highly skilled teachers may be assigned to a larger class because they are able to teach more students effectively. On the other hand, highly skilled teachers may be assigned to a smaller class as a non-monetary reward because teaching a small class may require less effort.

Another issue is that some junior high schools adopt ability tracking for mathematics classes. Unfortunately, we do not know whether the size of mathematics classes is different from that of other subjects. While we are unable to predict how this issue may bias our estimates, we note that the estimation results for mathematics in junior high schools may not be reliable.

Yet another issue is the teachers' effort. If teachers assigned to a large class exert more effort to improve students' learning experiences, the class-size effects are downward biased. Unfortunately, we do not observe the teachers' efforts.

5. Results

5.1. First-stage regression

We first show graphical evidence that Maimonides' rule has strong predictive power for the actual class size. Fig. 1 plots the actual class size and the predicted class size following Maimonides' rule. The top panel is for elementary school (grade six), whereas the bottom panel is for junior high school (grades from seven to nine). Maimonides' rule implies that class size sharply drops at the cutoffs that are multiples of 40. The observed class size closely follows this discontinuous change, although some deviate from the predictions.

Next, we show more formal evidence that the predicted class size using Maimonides' rule is strongly correlated with the observed class size. In Table 2, we present the estimates from the first-stage regression. The first set of results (labeled Model 3) is from the first-stage regression in which sex, year, and enrollment are controlled for. The coefficients for the predicted class size across different grades are between 0.418 and 0.823 and are statistically significant. The corresponding F-values are high at or above 114. We soundly reject the hypothesis that the instrument is weak for this specification on the ground that they are much higher than the threshold value of 10 (see (Stock et al., 2002)).

The second set of results (labeled Model 4) is from the first-stage regression in which school fixed effects are controlled for in addition to the basic controls in Model 3. The coefficients and the F-values are smaller than those of the model without school fixed effects because some between-school variations are absorbed by the school fixed effects. Nevertheless, the instruments are strong enough for all grades.

5.2. Effects on cognitive and non-cognitive skills

Table 3 reports estimates of Eq. (2) across different specifications for Japanese and math test scores for each grades from four to eight. The columns labeled Model 1 provide the ordinary least squares (OLS) estimates controlled for sex and year effects. The coefficients for observed class size vary across subjects and grades. Many coefficients are positive, and they are large and statistically significant for junior high

Table 2
First stage regression.

		Grade4	Grade5	Grade6	Grade7	Grade8
Model-3 (IV)	Predicted class size	0.823*** (0.021)	0.815*** (0.022)	0.798*** (0.021)	0.418*** (0.039)	0.513*** (0.037)
	R^2	0.775	0.764	0.748	0.393	0.441
	F-value	1527.697	1389.871	1392.296	114.022	187.737
	N	146004	146594	139219	141303	142028
Model-4 (IV + School Fixed Effect)	Predicted class size	0.800*** (0.025)	0.776*** (0.027)	0.756*** (0.027)	0.292*** (0.043)	0.385*** (0.045)
	R^2	0.862	0.858	0.840	0.680	0.655
	F-value	1017.473	831.049	760.295	46.412	74.247
	N	146004	146594	139219	141303	142028
Control	School enrollment	✓	✓	✓	✓	✓
	Female Ratio	✓	✓	✓	✓	✓
	Year	✓	✓	✓	✓	✓

Source: Authors' calculations
Note: The table shows first stage regression results for two-stage least squares (2SLS). The unit of observations is student. Model 3 refers to the IV regression model without school fixed effects and Model 4 refers to the model with school fixed effects. Standard errors are in parentheses and clustered by the school. "***", "**", and "*" represent 1 percent, 5 percent, and 10 percent significance level, respectively.

school students (grades seven and eight).

Columns labeled Model 2 show instrumental variables (IV) estimates in which the observed class size is instrumented by the predicted class size. The control variables are the same as Model 1. One might expect that the observed class size is positively associated with unobserved school resources because schools in urban areas are larger and their students are from wealthier families. This implies that the OLS estimates are upward biased. However, the OLS and IV estimates are very similar in our sample.

In columns labeled Model 3, we also control for enrollment. Otherwise, the specification remains the same as in Model 2. Because enrollment is positively correlated with the SES of students as well as observed class size, the estimates from Model 3 are much smaller than those from Model 2. This implies that controlling for school enrollment is essential for avoiding omitted variable bias. The estimated coefficients for observed class size for Model 3 are small and negative for both Japanese and math for elementary school students (grades four to six) and they are precisely estimated. The estimated effects of class size for junior high school students are close to zero and statistically insignificant. It is worth noting that additionally controlling for the quadratic and cubic terms of enrollment does not essentially change the results (see appendix of Ito et al. (2019)).

To address the possible correlation between the predicted class size and unobserved school characteristics, we additionally control for school fixed effects. With school fixed effects, our identifying variation is only from within the school. The estimates are presented in the columns labeled Model 4. Although the estimates do not differ greatly from those for Model 3, the estimated effects of class size are insignificant, except for math test score in grade six. Note that standard errors are also similar between Models 3 and 4.

We do not find evidence that class-size reduction significantly improves students' academic performance for junior high school students (grades seven and eight). For elementary school students, class-size reduction may improve test scores only in grade six, but the estimated effects are very small.

Table 4 reports estimates across different specifications for the effects of class-size reduction on non-cognitive skills for each grade. Models 1 and 2 involve OLS and IV regressions, respectively, without any controls for grade enrollment. Because controlling for grade enrollment is important for endogeneity bias, we focus on the results from the IV regression that controls for grade enrollment labeled Model 3. Note that we cannot control for school fixed effects when we estimate class-size effects on non-cognitive skills owing to the survey structure (see Section 3.2 for details).

We find almost no effects of class-size reduction on self-efficacy and conscientiousness across grades. In fact, the effects of class size on self-control are positive and statistically significant for grades six to eight, which implies that class-size reduction worsens students' self-control skills.

5.3. Robustness

We address two issues regarding the robustness of the main results. Although we control for possible confounders by including enrollment and school fixed effects in the regressions, observations with different distances from the cutoffs for Maimonides' rule may not be comparable. We address this issue by taking a subsample near the cutoffs. A drawback of this approach is a loss of sample size, and hence, standard errors are typically greater than those in the main results.

Another issue is the use of school-level data. Some of the previous studies including Akabayashi and Nakamura (2014) use variables at the school level instead of the student level. This approach eliminates variations within the school–grade–year combinations. Hence, if there is a correlation between unobserved classroom and/or teacher characteristics and the instrument, this approach avoids omitted variable bias. In practice, however, they are uncorrelated because the predicted class size is the same within the school-grade-year combinations. Although we do not see any clear advantage or disadvantage in using school-level variables in terms of bias, the standard errors may be larger because there is less variation in the school-level data.

We extensively discuss the estimation results for these two issues in the appendix of Ito et al. (2019). The estimation results indicate that our results are robust to these two issues.

5.4. Heterogeneity by socio economic status

Some previous studies, such as Krueger and Whitmore (2001), found that the effect sizes are substantially larger for black students and students eligible for the free lunch program, an indicator of low SES. A possible explanation for the heterogeneous class-size effect is that disadvantaged students are less likely to receive educational investment from their families, and hence, they are more susceptible to their school environment.

To examine how effects of class size vary by individual students' SES, we split the sample by whether the students have books at home or not, which is a proxy for their SES. In an alternative specification, we split the sample by whether a student attends a private tutoring school, which is also a proxy for SES. Note that our exercises provide what may be considered suggestive evidence because the number of books at home and attendance at a private tutoring may not be exogenous.

Tables 5 and 6 report estimates for cognitive and non-cognitive skills when SES is measured by the number of books at home. We do not

Table 3
Effects of class size on cognitive skills.

Model type		Japanese Model-1	Japanese Model-2	Japanese Model-3	Japanese Model-4	Math Model-1	Math Model-2	Math Model-3	Math Model-4
Grade	grade4	0.004*** (0.001)	0.005*** (0.001)	−0.001 (0.001)	0.001 (0.001)	0.002 (0.001)	0.004*** (0.001)	−0.001 (0.001)	0.002 (0.002)
	grade5	0.002** (0.001)	0.003** (0.001)	−0.003** (0.002)	0.000 (0.001)	0.000 (0.001)	0.001 (0.001)	−0.004*** (0.001)	−0.001 (0.001)
	grade6	0.001 (0.001)	0.001 (0.001)	−0.005*** (0.001)	−0.002 (0.001)	−0.002* (0.001)	−0.002 (0.001)	−0.006*** (0.001)	−0.003** (0.001)
	grade7	0.009*** (0.002)	0.015*** (0.004)	0.002 (0.006)	−0.005 (0.006)	0.009*** (0.002)	0.012*** (0.004)	0.002 (0.007)	−0.010 (0.006)
	grade8	0.008*** (0.002)	0.010*** (0.003)	−0.001 (0.005)	−0.005 (0.004)	0.008*** (0.002)	0.010*** (0.003)	0.003 (0.006)	0.000 (0.005)
Fixed effect	School				✓				✓
Control variable	School Enrollment			✓	✓			✓	✓
	Sex	✓	✓	✓	✓	✓	✓	✓	✓
	Year	✓	✓	✓	✓	✓	✓	✓	✓

Note: The coefficients for class size are reported. The unit of observations is student. Columns labeled as Model 1 show OLS estimates. Columns labeled as Model 2–4 show IV estimates. All dependent variables are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors are in parentheses and clustered by the school. “***”, “**”, and “*” represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

Table 4
Effects of class size on non-cognitive skills.

Model type		Self-control Model-1	Self-control Model-2	Self-control Model-3	Self-efficacy Model-1	Self-efficacy Model-2	Self-efficacy Model-3	Conscientiousness Model-1	Conscientiousness Model-2	Conscientiousness Model-3
Grade	grade4	0.000 (0.001)	0.000 (0.002)	−0.001 (0.002)	0.001 (0.001)	0.000 (0.002)	0.000 (0.002)			
	grade5	0.000 (0.001)	−0.001 (0.002)	−0.001 (0.002)	0.000 (0.001)	0.000 (0.002)	−0.001 (0.002)	−0.003** (0.001)	−0.003* (0.002)	−0.002 (0.002)
	grade6	0.001 (0.002)	0.001 (0.002)	0.004* (0.002)	−0.002 (0.002)	−0.002 (0.002)	0.000 (0.002)	−0.006*** (0.002)	−0.007*** (0.002)	−0.004* (0.002)
	grade7	−0.001 (0.003)	0.009** (0.005)	0.014* (0.008)	−0.004* (0.002)	−0.003 (0.004)	−0.004 (0.006)	−0.010*** (0.003)	−0.005 (0.005)	0.002 (0.008)
	grade8	0.002 (0.003)	0.008* (0.004)	0.012** (0.006)	0.001 (0.002)	−0.002 (0.003)	−0.009 (0.005)	−0.005** (0.002)	−0.005 (0.004)	0.000 (0.005)
Fixed Effect	School									
Control Variable	School Enrollment			✓			✓			✓
	Sex	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Year	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: The coefficients for class size are reported. The unit of observations is student. Columns labeled as Model 1 show OLS estimates. Columns labeled as Model 2–3 show IV estimates. All dependent variables are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors are in parentheses and clustered by the school. “***”, “**”, and “*” represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

Table 5
Heterogenous effects on cognitive skills across SES groups (1).

Model type		Japanese Books	Japanese No books	Math Books	Math No books
Grade	grade4	0.000 (0.002)	−0.002 (0.002)	−0.001 (0.002)	−0.001 (0.003)
	grade5	−0.003* (0.002)	−0.003 (0.003)	−0.003 (0.002)	−0.005* (0.003)
	grade6	−0.005*** (0.002)	−0.002 (0.003)	−0.006*** (0.002)	−0.004 (0.003)
	grade7	0.000 (0.008)	0.001 (0.008)	0.001 (0.008)	0.000 (0.009)
	grade8	0.002 (0.006)	−0.008 (0.007)	0.002 (0.006)	−0.009 (0.007)
Fixed effect	School				
Control	School Enrollment	✓	✓	✓	✓
	Sex	✓	✓	✓	✓
	Year	✓	✓	✓	✓

Note: The unit of observations is student. The coefficients for class size are reported. Columns labeled as Model 1 show OLS estimates. Columns labeled as Model 2–3 show IV estimates. All dependent variables are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors are in parentheses and clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

Table 6
Heterogenous effects on non-cognitive skills across SES groups (1)

Model Type		Self-control Books	Self-control No books	Self-efficacy Books	Self-efficacy No books	Conscientiousness Books	Conscientiousness No books
Grade	grade4	−0.002 (0.002)	0.001 (0.004)				
	grade5	−0.001 (0.002)	−0.002 (0.004)	−0.001 (0.002)	−0.002 (0.005)		
	grade6			0.000 (0.002)	−0.005 (0.005)	−0.005** (0.002)	0.002 (0.005)
	grade7	0.012 (0.009)	0.015 (0.012)			0.002 (0.009)	−0.003 (0.010)
	grade8	0.014** (0.006)	−0.002 (0.009)	−0.011* (0.006)	−0.002 (0.009)		
Fixed Effect	School						
Control	School Enrollment	✓	✓	✓	✓	✓	✓
	Sex	✓	✓	✓	✓	✓	✓
	Year	✓	✓	✓	✓	✓	✓

Note: The unit of observations is student. The coefficients for class size are reported. Columns labeled as Model 1 show OLS estimates. Columns labeled as Model 2–3 show IV estimates. All dependent variables are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors are in parentheses and clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

Table 7
Heterogenous effect on cognitive skills across SES groups (2).

Model type		Japanese No Private tutoring	Japanese Private tutoring	Math No private tutoring	Math Private tutoring
Grade	grade4	−0.002 (0.002)	0.000 (0.002)	−0.001 (0.002)	0.000 (0.002)
	grade5	−0.005*** (0.002)	−0.002 (0.002)	−0.004** (0.002)	−0.001 (0.002)
	grade6	−0.005*** (0.002)	−0.004** (0.002)	−0.008*** (0.002)	−0.004** (0.002)
	grade7	0.002 (0.007)	−0.002 (0.009)	0.000 (0.008)	−0.002 (0.009)
	grade8	−0.006 (0.007)	0.004 (0.006)	−0.007 (0.007)	0.004 (0.006)
Fixed Effect	School				
Control	School Enrollment	✓	✓	✓	✓
	Sex	✓	✓	✓	✓
	Year	✓	✓	✓	✓

Note: The unit of observations is student. The coefficients for class size are reported. Columns labeled as Model 1 show OLS estimates. Columns labeled as Model 2–3 show IV estimates. All dependent variables are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors are in parentheses and clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

find clear evidence that the effects of class-size reduction vary by individual SES, as measured by the number of books, because estimated effects are stronger for low SES students in some grades, but they are weaker in other grades.

Table 7 shows estimates for cognitive skills when SES is measured by attendance at a private tutoring school. The estimate tend to be smaller for students who do not attend a private tutoring school, implying that the effects of class-size reduction tend to be stronger for these students. Table 8 reports the estimates for non-cognitive skills. They indicate that there is no class-size effects across subgroups.

5.5. Comparison with previous estimates in Japan

Our preferred estimates based on the IV regressions with school fixed effects for Japanese and math test scores are insignificant except for math in grade six. We add a note of caution the readers that even this statistically significant result might be a false positive given that we examine 10 academic outcomes in total (two subjects for five grades). Even if this estimate is a true positive, it is very small and implies that a 10-student reduction improves math test scores by only 0.03 standard deviations.

Our results are consistent with many previous studies that use Japanese data to estimate class-size effects on many academic outcomes (multiple subjects for multiple grades). Consistent with our study, these studies found statistically insignificant estimates, except for one or two outcomes and the effect size of the statistically significant estimates is typically small.

An exception is Ito et al. (2017), who reported statistically significant estimates for a wide array of cognitive and non-cognitive outcomes, which implies that their estimates are unlikely to be false

Table 8
Heterogenous effects on non-cognitive skills across SES groups (2).

Model type		Self-control No Private tutoring	Self-control Private tutoring	Self-efficacy No private tutoring	Self-efficacy Private tutoring	Conscientiousness No private tutoring	Conscientiousness Private tutoring
Grade	grade4	−0.001 (0.002)	−0.001 (0.002)				
	grade5	−0.002 (0.002)	0.000 (0.002)	−0.001 (0.003)	−0.002 (0.003)		
	grade6			−0.001 (0.003)	−0.001 (0.003)	−0.003 (0.003)	−0.006** (0.003)
	grade7	0.009 (0.008)	0.016 (0.011)			0.004 (0.008)	−0.001 (0.010)
	grade8	0.006 (0.006)	0.015** (0.007)	−0.014** (0.007)	−0.008 (0.006)		
Fixed Effect	School						
Control	School Enrollment	✓	✓	✓	✓	✓	✓
	Sex	✓	✓	✓	✓	✓	✓
	Year	✓	✓	✓	✓	✓	✓

Note: The unit of observations is student. The coefficients for class size are reported. Columns labeled as Model 1 show OLS estimates. Columns labeled as Model 2–3 show IV estimates. All dependent variables are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors are in parentheses and clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

positives. Their estimated effect sizes are smaller than those from the STAR project, but still sizable.

Ito et al. (2017) argued that previous papers reported small or insignificant effects of class size because they relied on between-school comparisons, and hence, failed to control for unobserved school characteristics. Controlling for school fixed effects is important in their study, the data for which comes from an anonymous city in the Chubu region. However, controlling for school fixed effects does not seem to bias estimates substantially in our data from an anonymous prefecture. Akabayashi and Nakamura (2014) also found that their preferred estimates without school fixed effects are robust to the inclusion of school fixed effects, using data from the city of Yokohama. It is possible that the source of the data influences the importance of controlling for school fixed effect, as we discuss below.

Another issue raised by Ito et al. (2017) is the choice of functional form. They criticized previous studies for including only a linear term of grade enrollment and failing to account for a possible nonlinear relationship between grade enrollment and unobserved factors, which results in biased estimates. We agree that controlling grade enrollment is necessary to avoid endogeneity bias, but this functional form issue appears to be irrelevant for our data because including the third-order polynomial of grade enrollment as a regressor does not affect the estimates.

Ito et al. (2017) also claimed that the inclusion of schools that do not strictly follow Maimonides' rule is inappropriate and would make estimates "instable", although they did not clarify what they meant by this term. However, our instruments are strongly correlated with the endogenous variable (i.e., class size) and hence, the estimates are unlikely to suffer biases from weak instruments. In addition, the standard errors for the estimates are small. We argue that including schools that do not strictly follow Maimonides' rule does not lead to a biased or imprecise estimate when class size is instrumented by the predicted class size using Maimonides' rule.

We have just recently learned that Hojo and Senoh (2019) found that class-size reduction had a larger effect on cognitive ability of economically disadvantaged students, although the overall effect of class size on cognitive ability of ninth-grade students was small. Given that the anonymous prefecture that we study is larger and wealthier than the average, our economically insignificant effects of class-size reduction is consistent with the finding by Hojo and Senoh (2019) who used NAAA that cover the whole country.

In conclusion, we argue that the differences in the estimates are likely to reflect the fact that samples are taken from different locations, rather than differences in the statistical methods.

6. Conclusion

We estimated the effects of class size on students' academic test scores and non cognitive skills, using data that cover all public school students in grades four to nine in an anonymous prefecture. Our identification approach is based on Maimonides' rule for class size.

We find that the effects of a class-size reduction on Japanese and math test scores are largely insignificant. The only statistically significant estimate is for math in grade six, but the effect size remains very small. The estimate implies that a 10-student reduction improves math test scores in grade six by 0.03 standard deviations. In addition, we also find no evidence that class-size reduction improves non cognitive skills.

In terms of model specifications, including grade enrollment is necessary to avoid endogeneity bias. However, controlling for school fixed effects or accounting for possible nonlinearity in grade enrollment seems to have little influence on our estimates. Hence, we argue that the different estimates found in the existing studies on Japan arise from the fact that they are based on different population groups, rather than different statistical methods. Comprehensive research on identifying subpopulation groups for which class-size reduction is effective is an interesting future research topic.

An important limitation of our paper as well as of other existing studies on Japan is a lack of teacher information. If skilled teachers are assigned to smaller classes as a non pecuniary reward or if hiring more teachers lowers the average teacher quality, teacher quality is correlated with class size. Another interesting question is whether class-size reduction is effective under particular circumstances or not. For example, teachers could take advantage of smaller class size if they are allowed to choose what to teach depending on students' skills. To address these issues in future studies, data on teacher characteristics and their teaching environment are necessary.

References

Akabayashi, H., Nakamura, R., 2014. Can Small class policy close the gap? An empirical analysis of class size effects in Japan. Japanese Economic Review 65 (3), 253–281.

Angrist, J.D., Lavy, V., 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. The Quarterly Journal of Economics 114 (2), 533–575.

Bandura, A., 1986. Social Foundations of Thought and Action. Englewood Cliffs, NJ, pp. 1986.

Barbaranelli, C., Caprara, G.V., Rabasca, A., Pastorelli, C., 2003. A questionnaire for measuring the Big Five in late childhood. Personality and Individual Differences 34 (4), 645–664.

Borghans, L., Duckworth, A.L., Heckman, J.J., Weel, B.T., 2008. The economics and psychology of personality traits. Journal of Human Resources 43 (4), 972–1059.

Carneiro, P., Crawford, C., Goodman, A., 2007. The Impact of Early Cognitive and Non-cognitive Skills on Later Outcomes. London School of Economics, pp. 92.

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W., Yagan, D., 2011. How does your kindergarten classroom affect your earnings? Evidence from project star. The Quarterly Journal of Economics 126 (4), 1593–1660.

Dee, T.S., West, M.R., 2011. The non-cognitive returns to class size. Educational Evaluation and Policy Analysis 33 (1), 23–46.

Duckworth, A.L., Tsukayama, E., Kirby, T.A., 2013. Is it really self-control? Examining the predictive power of the delay of gratification task. Personality and Social Psychology Bulletin 39 (7), 843–855.

Finn, J., Achilles, C., 1999. Tennessee's Class Size Study: Findings, Implications, Misconceptions. Educational Evaluation and Policy Analysis 21 (2), 97–109.

Fredriksson, P., Öckert, B., Oosterbeek, H., 2013. Long-term effects of class size. The Quarterly Journal of Economics 128 (1), 249–285.
Heckman, J.J., Stixrud, J., Urzua, S., 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. Journal of Labor Economics 24 (3), 411–482.
Heineck, G., Anger, S., 2010. The returns to cognitive abilities and personality traits in Germany. Labour Economics 17 (3), 535–546.
Hojo, M., 2013. Class-size effects in Japanese schools: A spline regression approach. Economics Letters 120 (3), 583–587.
Hojo, M., Senoh, W., 2019. Do the disadvantaged benefit more from small classes? Evidence from a large-scale survey in Japan. Japan and the World Economy In press.
Hoxby, C.M., 2000. The effects of class size on student achievement: New evidence from population variation. The Quarterly Journal of Economics 115 (4), 1239–1285.
Ito, H., Nakamuro, M., Yamaguchi, S., 2019. Effects of Class-Size Reduction on Cognitive and Non-Cognitive Skills. https://doi.org/10.2139/ssrn.3326288.
Ito, H., Hamada, M., Murayama, Y., Nobuya, T., Nomura, K., Myogan, M., Tsuji, M., 2017. Effects of class size on academic achievement and emotional and behavioral problems. The Japanese Journal of Educational Psychology 65 (4), 451–465.
Krueger, A.B., 1999. Experimental estimates of education production functions. Quarterly Journal of Economics 114 (2), 497–532.
Krueger, A., Whitmore, D., 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. The Economic Journal 111 (468), 1–28.
Mueller, G., Plug, E., 2006. Estimating the effect of personality on male and female earnings. ILR Review 60 (1), 3–22.
Niki, M., 2013. The effects of class size on cognitive and non-cognitive abilities. Economic Analysis 186, 30–49.
Pajares, F., Kranzler, J., 1995. Self-efficacy beliefs and general mental ability in mathematical problem-solving. Contemporary Educational Psychology 20 (4), 426–443.
Pajares, F., Graham, L., 1999. Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. Contemporary Educational Psychology 24 (2), 124–139.
Pajares, F.M., Miller, D., 1994. Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. Journal of Educational Psychology 86 (2), 193.
Pintrich, P.R., De Groot, E.V., 1990. Motivational and self-regulated learning components of classroom academic performance. Journal of Educational Psychology 82 (1), 33.
Pintrich, P.R., et al., 1991. A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).
Poropat, A., 2009. A meta-analysis of the five-factor model of personality and academic performance. Psychological Bulletin 135 (2), 322–338.
Senoh, W., Hojo, M., 2016. Does the class size affect the academic achievement of junior high school students? An empirical study with National Assessment of Academic Ability of Japan. NIER Research Bulletin 145, 119–128.
Senoh, W., Hojo, M., Shinozaki, T., Sano, S., 2014. Estimation of Class-size Effects with a Regression Discontinuity Design: An Analysis Focusing on Public Elementary and Lower Secondary Schools in Japan. NIER Research Bulletin 143, 89–101.
Stock, J.H., Wright, J.H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business & Economic Statistics 20 (4), 518–529.
Tsukayama, E., Duckworth, A.L., Kim, B., 2013. Domain-specific impulsivity in school-age children. Developmental Science 16 (6), 879–893.
Urquiola, M., 2006. Identifying class size effects in developing countries: Evidence from rural bolivia. Review of Economics and Statistics 88 (1), 171–177.