Effects of Class-Size Reduction
On Cognitive and Non-cognitive Skills

Hirotake Ito
Keio University

Makiko Nakamuro
Keio University

Shintaro Yamaguchi
The University of Tokyo

January 2019

Effects of Class-Size Reduction
On Cognitive and Non-cognitive Skills[*]

Hirotake Ito[†]   Makiko Nakamuro[‡]   Shintaro Yamaguchi[§]

January 31, 2019

Abstract

We estimate the effects of class-size reduction by exploiting exogenous variation caused by Maimonides' rule that requires the maximum class size be 40 and class be split when 41 students are enrolled. Our data cover all fourth to ninth graders in 1,064 public schools in an anonymous prefecture for three years. We find that the effects of class-size reduction on academic test scores are small on average, but slightly stronger for students not going to a private tutoring school. We find no evidence that small class size improves non-cognitive skills. Our substantive conclusion does not change when school fixed effects are controlled.

[†]Graduate School of Media and Governance, Keio University, 5322 Endo, Fujisawa-shi, Kanagawa 252-0882 Japan, Email: itouhrtk@keio.jp.

[‡]Graduate School of Media and Governance, Keio University, 5322 Endo, Fujisawa-shi, Kanagawa 252-0882 Japan, Email: makikon@sfc.keio.ac.jp.

[§]Faculty of Economics, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan, Email: syamaguchi@e.u-tokyo.ac.jp

1

1 Introduction

The general public and education administrators seem to believe that smaller class size contributes to more learning and better experience for students, but this is far from obvious. Indeed, the literature on the economics of education has long sought the evidence for the effectiveness of class-size reduction. Although the analysis of data from a randomized controlled trial (RCT) in the famous Project STAR has established positive effects of smaller class size (e.g. Krueger, 1999), its external validity must be verified in each country's context.

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

There are several studies on class size using Japanese data. Examples include, but not limited to, Niki (2013), Hojo (2013), Senoh et al. (2014), Akabayashi and Nakamura (2014), Senoh and Hojo (2016), and Ito et al. (2017). Evidence from Japan so far is mixed presumably due to differences in their statistical methods and samples chosen. Some report insignificant results, while others report positive significant effects of class-size reduction, although the effect size is typically small. In our reading of the literature, there seems no consensus about the effects of class-size reduction in Japan.

We contribute to this literature by providing an extra piece of evidence using a large scale data from an anonymous prefecture[1] in Japan. Our re-

---

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

search has the following three key features. First, our data set is large and covers all students in grades four through nine in 1,064 public schools in an anonymous prefecture for three years. Although our data set is smaller than the one used by Senoh et al. (2014) that cover all students in Japanese public schools, it is much larger than data from a municipality (e.g. Akabayashi and Nakamura, 2014; Ito et al., 2017), which is a subdivision of a prefecture. Second, our data include 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 and student fixed effects to address possible omitted variable biases. Many previous studies use data from a single year, which prevents them from controlling for school and/or student fixed effects. Our large data set enables us to overcome this limitation.

Our estimates indicate that the effects of class-size reduction on Japanese and math test scores are small, if any. The estimated effects of a 10-student reduction vary by grades from 0.01 to 0.07 standard deviations in our preferred specification. When we control for school fixed effects, the estimates tend to be insignificant mostly due to larger standard errors and point estimates are in the same ballpark with or without school fixed effects.

By contrast to the estimates for test scores, we find no evidence that class-size reduction improves non-cognitive skills across grades and econometric specifications.

We also examine heterogeneity of the class-size effects by students' characteristics. Our estimates indicate that the effects of class-size reduction are slightly stronger for students not going to a private tutoring school.

The rest of the paper is structured as follows. We review the literature briefly in section 2. Then we describe 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.

---

states in other countries.

2   Literature review

Most prior research addressed a potential endogeneity bias arising from correlation between class size and unobserved school characteristics by taking one of the following approaches: (i) randomized experiments, such as Project STAR implemented in the state of Tennessee in 1980s (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 the effect size of 0.13-0.27 standard deviations for a 8-student reduction or 0.16-0.33 standard deviations for a 10-student reduction, equivalently (see Finn and Achilles [1990, Table 5]). In Israel, a 10-student reduction in class size increased the standardized test scores by 0.13-0.25 standard deviations in grades four and five (Angrist and Lavy, 1999).[2]

Angrist et al (2017) used large Israeli sample from 2002-2011 and found that the class-size effect was nearly zero with small standard errors. Interestingly, Angrist et al (2017) stated that the large significant estimates for class-size effects previously reported by Angrist and Lavy (1999) may have been "a chance finding" and "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 be-

---

[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.

yond academic test scores. Dee and West (2011) found that smaller class size was associated with improvement in school engagement. Fredriksson, Öckert and Oosterbeek (2013) evaluated the longer-run effects of class-size reduction on cognitive and non-cognitive outcomes. They reported that small class in ages 10-13 improved several outcomes including effort motivation, aspirations, self-confidence, sociability, absenteeism, and anxiety in ages 13-16. Chetty et al. (2011) linked the experimental data from the STAR project to administrative record and found that students in smaller classes were significantly more likely to attend college. Furthermore, they exhibited statistically significant improvement in 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, if any. Using a nationally representative sample of students from Trend in International Mathematics and Science Study (TIMMS) 2003, Niki (2013) found insignificant effects of class size on mathematics and science test scores as well as non-cognitive skill measures for eighth graders. Hojo (2013) drew a sample of fourth graders from TIMMS 2003 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) that covered Japanese and math test scores for virtually all students in grade six and nine in Japanese public schools. They found effects of class-size reduction are insignificant across subjects and grades except for Japanese test scores for sixth graders.

Senoh and Hojo (2016) updated Senoh et al. (2014) by using family background information added by a follow-up survey for the 2009 NAAA. They found that effects of small class were positive significant for ninth graders once students' SES was controlled. Although their estimates are greater than 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 in which many students from low SES families attended.

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 City of Yokohama. In addition to exploiting the exogenous variation by Maimonides' rule, they controlled for unobserved heterogeneity by using a value-added model and controlling for school fixed effects. The 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 de-meaning variables at the school level, they exploited within-school variation as well as the exogenous variation by Maimonides' rule. Note that their de-meaning 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 those 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 filled a series of questionnaires to measure some non-cognitive skills and were asked about students' life. The data also include students' IDs which enable us to keep track of the students throughout the survey period.

The data cover all public-school students in grades 4 through 9 in the

prefecture. While there are some private and national schools, most schools in the prefecture are public: according to official statistics, 99.3 percent of elementary schools and 93.0 percent of junior high schools are public in this prefecture. Our data include about 300,000 students in 1,064 public schools (708 elementary, 356 junior high) in 62 municipalities.[3] We restrict our sample include students who took the exam at least twice in order to control for individual fixed effects. As a result, we exclude fourth and fifth graders.

Notice that we examine the association between students' outcomes in the current year and school characteristics such as class size in the last year. This is because the outcomes are measured in beginning of school year (April) and they are likely to be affected by class size and other school environment in the previous year, if any.

Table 1 presents descriptive statistics of selected variables. The average grade size is about 90 students for elementary school for grades four through six. In the table, grade enrollment is about 90 for grades from five to seven for the reason we mentioned above. The grade enrollment of junior high school is greater and 160-165 students. The average class size is 31-32 for elementary school. It is greater for junior high school and about 34.

3.2   Cognitive and non-cognitive skills

Our measures of cognitive skills are Japanese and math test scores. To facilitate interpretation, we normalize 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-minutes questionnaire: (i) self-control (Duckworth, Tsukayama and Kirby, 2013), (ii) self-

---

[3]Unfortunately, we do not have an access to information for students who are absent on the test day. About 2-3% of students were absent in the survey period. This number is about the same as the fraction of absent students in NAAA, which has been administered nationwide by Ministry of Education, Culture, Sports, Science and Technology. We consider that the absent students may have transferred to a school outside of this prefecture or may have been sick.

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, Stixrud and Urzua, 2006; Heineck and Anger, 2010; Carneiro, Crawford and Goodman, 2007; Mueller and Plug, 2006).

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, Tsukayama and Kirby, 2013).

Our data employ an 8-item questionnaire, originally developed by Tsukayama, Duckworth and Kim (2013): a self-rated scale on a 5-point frequency scale where 5 (=very frequently) and 1 (=almost never) is averaged such that a higher score indicates higher self-control (See 14 for details). In computing the scale, several items, for example, intentionally negative worded items, are rated reversely. If an item has to be reversed, a student who choose 1 receives a score of 5.

To avoid controversy of the stability of measures across different situations (sometimes, referred to as "person-situation" debate), Tsukayama, Duckworth and Kim (2013) developed the 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 would believe that they are not capable of accomplishing tasks and then avoid doing them, while ones with high self-efficacy would try to accomplish tasks and be engaged in them for the long-run, even if they are very challenging.

Our measure of self-efficacy is originally developed by Pintrich et al.

(1991). An 8-item questionnaire is based on a self-rated scale on a 5-point Likert scale from 5 (=very true of me) and 1 (=not at all true of me). The scale is constructed by taking the mean of all items and reverse-coded items are properly reflected (See 14 for details).

This is also the scale developed specifically to learning 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 and other studies that employed this type of scale also 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. For example, Poropat (2009)'s meta-analysis on the relationship between the Big Five personality traits revealed that the both raw and partial correlations between conscientiousness and GPA were almost as large as those between GPA and IQ.

Our measure of conscientiousness is based on the method developed by Barbaranelli et al. (2003) and hence, assesses students' dependability, orderliness, precision, and the fulfilling of commitments in school environments. By taking an average of 13 items that are based on the self-rated scale on a 5-point Likert scale from 5 (=very true of me) and 1 (=not at all true of me), we construct a single measure of conscientiousness (See 14 for details).

Like 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 book at home, which is often used in education research when parents' income or occupations is unknown. By using internationally comparable standard-

ized test scores, such as TIMSS and Programme for International Student Assessment, prior studies show a strong correlation between the number of books at home and parental SES. Kawaguchi (2016) and Kawaguchi (2017) also show that the number of books at home is a good predictor of parental income and educational backgrounds in Japan.

Another proxy is an indicator for going to a private tutoring school. Because the average monthly fee is about 40,000 Japanese yen (or 400 US dollars, equivalently), students from wealthier families are more likely to go to a private tutoring school to improve their test scores.

Table 1 indicates that the average fraction of students receiving subsidies is 11-14% and it generally increases with grade. The average fraction of students who have no books at home is about 10-15%. The fraction of students going to a private tutoring school is 50-60% for grade four. It increases with grade and reaches about 70% in grade nine.

Table 1: DESCRIPTIVE STATISTICS BY GRADE

	Variable	year	Grade 4	5	6	7	8	9
School	Grade Enrollment Last Year	2016		88.195	90.157	91.726	166.338	165.155
				(47775, 36.158)	(49118, 36.690)	(47454, 37.117)	(47807, 60.826)	(48048, 59.978)
		2017		91.661	88.119	89.521	164.487	163.596
				(49308, 38.592)	(48106, 35.967)	(46377, 36.709)	(47441, 60.964)	(47306, 59.631)
		2018		90.970	91.775	88.061	161.144	161.844
				(49095, 36.986)	(49465, 38.754)	(45489, 36.169)	(46198, 59.594)	(46813, 59.831)
	Class Size Last Year	2016		31.646	31.854	32.115	33.935	34.247
				(47775, 5.245)	(49118, 5.074)	(47454, 5.030)	(47807, 3.474)	(48048, 3.273)
		2017		32.041	31.625	31.797	33.887	34.115
				(49308, 5.113)	(48106, 5.109)	(46377, 4.998)	(47441, 3.600)	(47306, 3.534)
		2018		32.070	32.215	31.527	33.767	34.261
				(49095, 5.276)	(49465, 5.097)	(45489, 5.142)	(46198, 3.535)	(46813, 3.566)
SES	No Books	2016	0.119	0.097	0.084	0.122	0.135	0.148
			(49441, 0.324)	(48471, 0.295)	(49748, 0.277)	(48428, 0.328)	(48118, 0.342)	(48764, 0.355)
		2017	0.122	0.093	0.081	0.113	0.120	0.132
			(48966, 0.327)	(49844, 0.291)	(48908, 0.274)	(47996, 0.317)	(48191, 0.325)	(48544, 0.339)
		2018	0.127	0.094	0.082	0.112	0.118	0.132
			(46793, 0.333)	(49601, 0.292)	(50095, 0.274)	(46391, 0.315)	(46285, 0.322)	(47425, 0.339)
	Private Tutoring	2016	0.519	0.503	0.508	0.508	0.588	0.687
			(48337, 0.500)	(47703, 0.500)	(49519, 0.500)	(48363, 0.500)	(48006, 0.492)	(48559, 0.464)
		2017	0.603	0.564	0.563	0.565	0.618	0.706
			(47066, 0.489)	(48458, 0.496)	(47999, 0.496)	(47117, 0.496)	(47784, 0.486)	(48202, 0.455)
		2018	0.617	0.569	0.567	0.566	0.593	0.702
			(46749, 0.486)	(48529, 0.495)	(49294, 0.495)	(46078, 0.496)	(46249, 0.491)	(47196, 0.457)

Note: The unit of observation is students. The average scores are reported in each cell, and the number of observation and standard deviations are reported in parentheses. As school resource, we calculate the number of students in a class("Class Size Last Year") and grade enrollment("Grade Enrollment Last Year"). These scores reported are values of the school each student belonged to last year. As social economic status(SES), we measure the numbers of books at home ("No Books", dummy variable about whether to have books at home) and whether to go private tutoring school or not ("Private Tutoring", dummy variable). These scores are measured in April every year.

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 (35 only for students in grade one since 2011) in Japan, which is legally determined by 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 into a single class, but a grade with 41 students is divided into two classes.

In practice, many schools in our data form classes smaller than the legally mandated class size. Hence, we cannot simply compare schools with grade enrollment 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.

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

where z_{st} is the predicted class size in school s at time t, n_{st} is grade enrollment, and the square bracket is an operator that takes an integer part of 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,

$$y_{ist} = c_{ist-1}\beta + x'_{ist}\gamma + \varepsilon_{ist},\tag{2}$$

where y_{ist} is student i's outcome in school s at time t, c_{ist-1} is the class size for student i in the previous academic year, and x_{ist} is a set of control variables including enrollment in a given grade, school fixed effects, and student fixed effects. Note that school characteristics including school fixed effects are those of the schools in which students attended in the previous year. This is because the outcomes observed in the beginning of the year are likely to be

affected by school characteristics in the previous year. The error term is ε_{ist} and uncorrelated with any of the observed variables. We allow for correlation of the error term within 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 Figure 1, although it is not monotonic. In fact, the actual class size and enrollment are positively correlated in our sample.

Figures 2 and 3 provide evidence that enrollment and the SES of students' families are positively correlated. In Figure 2, we plot enrollment and the fraction of students having no books at home. Remember that having no books at home is a proxy for low SES, and hence, the negative regression slope in Figure 2 implies a positive correlation between enrollment and students' SES. In Figure 3, we plot enrollment and the fraction of students going to a private tutoring school. Remember that going to a private tutoring school can be interpreted as a proxy for higher SES. The evidence suggests that, if enrollment is not controlled in regressions, the estimated coefficient for class size are likely to be upward biased.

In addition to enrollment, we also control for school fixed effects to address a possible correlation between the predicted class size and unobserved school characteristics. Note that 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.

For added safety, we additionally control for student fixed effects, which has not been controlled by previous studies as far as we are aware of. If students self-select into a smaller class through parents' school choice, unobserved students' and their family characteristics are likely to be correlated with the predicted school size. This is possible if parents are aware of Maimonides' rule and know the expected enrollment. Controlling for stu-

Figure 1: PREDICTED VS ACTUAL CLASS SIZE
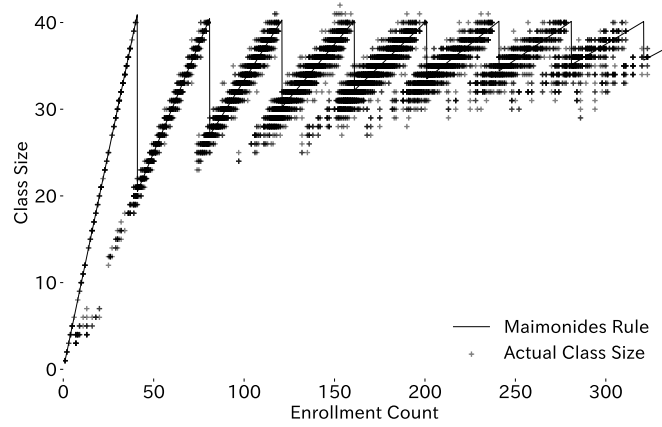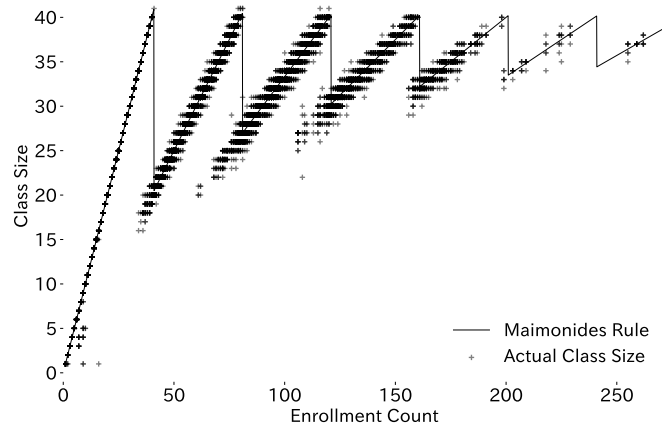
Note: The plot shows the relationship between enrollment count and class size(upper figure: elementary school, lower figure: junior high school). Each marker represents the actual class size of schools. Solid lines indicate predicted class size calculated by Maimonides rule.

Figure 2: CORRELATION BETWEEN ENROLLMENT AND SES (1)
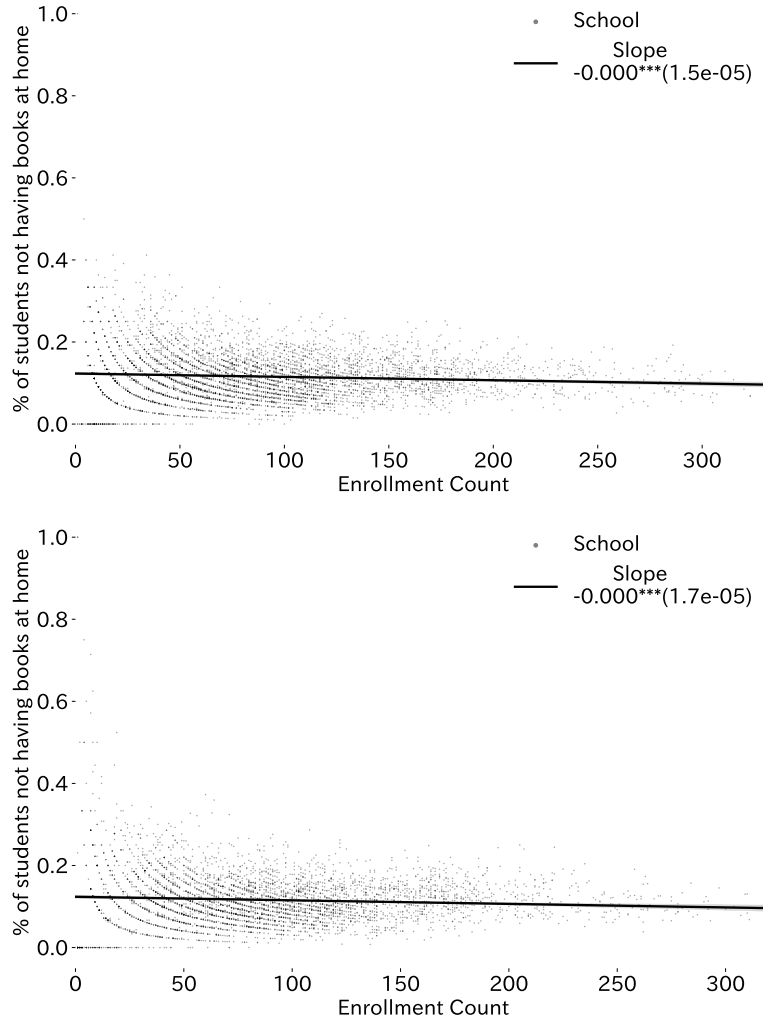
Note: The socio-economic status at a given school district is defined as the numbers of books at home.

Figure 3: CORRELATION BETWEEN ENROLLMENT AND SES (2)
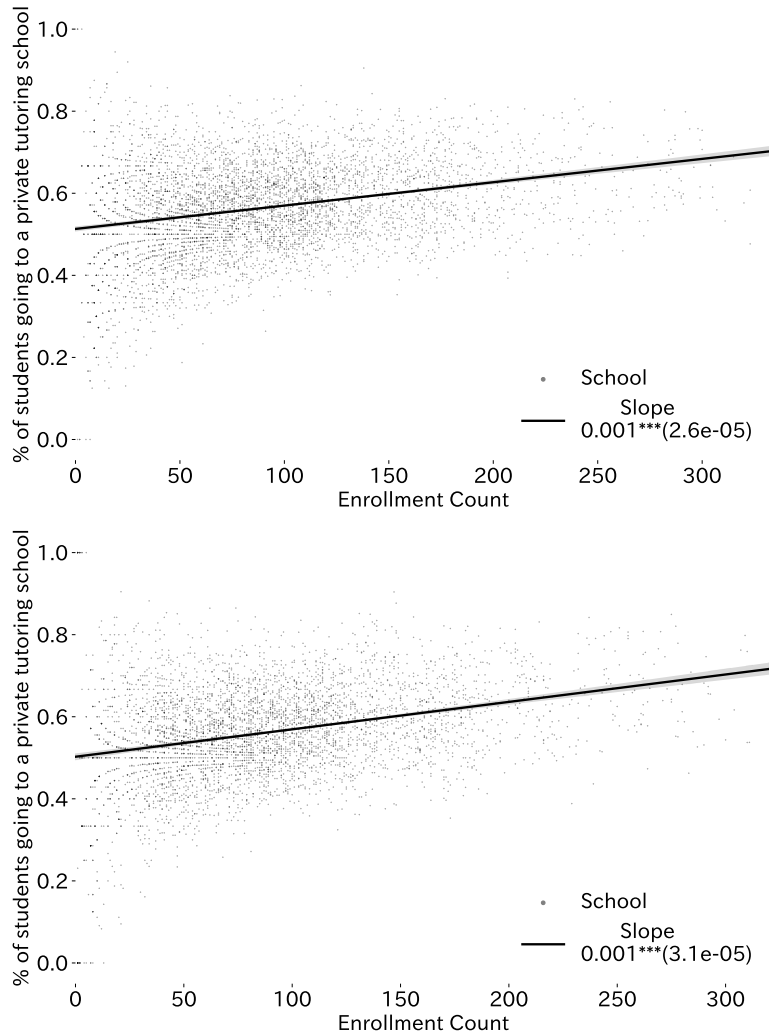
Note: The socio-economic status at a given school district is defined as the numbers of hours spent at private cramming schools outside of formal schooling.

dent fixed effects enables us to avoid this endogeneity bias arising from self-selection.

Although we control for more unobserved characteristics than previous studies, there is a remaining threat for identification, which is teacher characteristics. If teachers are systematically assigned according to class size, teacher characteristics are likely to be 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.

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 need less effort.

5   Results

5.1   First-stage regression

We first show graphical evidence that Maimonides' rule has a strong predictive power for the actual class size. Figure 1 plots the actual class size and the predicted class size followed by Maimonides' rule. The top panel is for elementary school (grade six), while the bottom panel is for middle 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.

We next show more formal evidence that the predicted class size by Maimonides' rule is strongly correlated with observed class size. In Table 2, we present the estimates from the first-stage regression. The first set of results (labeled as Model 3) is from the first-stage regression in which sex, year, and

17

enrollment are controlled. The coefficients for the predicted class size across different grades are high between 0.66 and 0.88 and statistically significant. The corresponding F-values are also high, ranging from 508 to 2534. We soundly reject the hypothesis that the instrument is weak for this specification because they are much higher than the threshold value of 10 (see (Stock, Wright and Yogo, 2002)).

The second set of results (labeled as Model 4) is from the first-stage regression in which school fixed effects are controlled in addition to the basic controls in Model 3. The coefficients substantially smaller than those of the model without school fixed effects. Consequently, the F-values are also small. With school fixed effects, instruments are strong enough for grades seven and eight and the pooled sample.

The third set of results (labeled as Model 5) is from the first-stage regression in which student fixed effects are controlled in addition to school fixed effects and the basic controls. Controlling for student fixed effects further reduces the predictive power of Maimonides' rule. In this specification, the instrument is strong only for grade seven and the pooled sample.

In the following analysis, we focus on specifications and samples in which the instrument is strong to avoid biases arising from a weak instrument.

Table 2: FIRST STAGE RESULTS

Wave(grade in 2018)		Grade6	Grade7	Grade8	Grade9	ALL
Model3 (IV)	Predicted Class Size	0.875***	0.768***	0.745***	0.655***	0.767***
		(0.023)	(0.027)	(0.025)	(0.029)	(0.015)
	R2	0.825	0.718	0.674	0.584	0.702
	Fvalue	1421.322	795.514	882.162	507.896	2534.427
	N	98677	141233	141581	141562	523053
Model4 (IV + School Fixed Effect)	Predicted Class Size	0.209**	0.266***	0.334***	-0.075*	0.682***
		(0.101)	(0.057)	(0.089)	(0.045)	(0.019)
	R2	0.959	0.918	0.954	0.901	0.799
	Fvalue	4.298	21.904	14.049	2.791	1257.768
	N	98677	141233	141581	141562	523053
Model5 (IV + School Fixed Effect + Student Fixed Effect)	Predicted Class Size	0.209	0.264*** (0.07)	0.332***	-0.075	0.372***
		(0.143)		(0.111)	(0.055)	(0.034)
	R2	0.967	0.926	0.962	0.913	0.912
	Fvalue	2.119	14.153	8.924	1.863	121.424
	N	98677	141233	141581	141562	523053
Control	School Enrollment			✓	✓	✓
	Sex			✓	✓	✓
	Year			✓	✓	✓
	Grade					✓

Source: Authors' calculations

Note: The table shows first stage results of two-stage least squares(2SLS) estimates. The unit of observations is student. The estimation is made for each wave("Grade6"-"Grade9") and pooled sample("ALL"). The label of the wave is named by grade in 2018, so "Grade6" wave represents the sample of six grade students in 2018. "Model3", "Model4", and "Model5" represents estimated specification("Model1" and "Model2" are omitted. "Model1" is not two-stage least squares estimates and "Model2" is very similar specification to "Model3"). The dependent variables of 2SLS are math scores. Standard errors in parentheses are clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

5.2 Effects on cognitive and non-cognitive skills

Table 3 reports estimates of equation (2) across different specifications for Japanese and math scores of students in grades six through nine separately.

Columns labeled as Model 1 show OLS estimates with controlling for sex and year effects. The coefficients for observed class size vary across subjects and grades. Some are positive significant, some are negative significant, and others are insignificant. We do not find a clear pattern here.

Columns labeled as Model 2 show IV estimates in which the observed class size is instrumented by the predicted class size. Control variables are same as Model 1. One might expect that 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 as Model 3, we additionally control for enrollment. Otherwise, the specification remains the same as in Model 2. Because enrollment is positively correlated with socio-economic status 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 negative for both Japanese and math for all grades and have small standard errors. It is worth noting that additionally controlling for the quadratic and cubic terms of enrollment does not essentially change the results (results available on request).

To address 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 from within school, instead of between schools.

The estimates are presented in columns labeled as Model 4. We focus on grades seven and eight and the pooled sample because the instrument is weak

for other subsamples. The coefficients are insignificant for both Japanese and math for all grades and the pooled sample, except for math in grade eight.

We further control for student fixed effects. We focus on grade seven and the pooled sample, because the instrument is weak for other subsamples. Columns labeled as Model 5 indicate that the coefficients for Japanese are close to zero and insignificant. The coefficients for math are small negative, but statistically significant only for the pooled sample.

As estimates vary by specifications and samples, we do not find strong evidence that class-size reduction substantially improves students' academic performance. While estimates are statistically significant in some specifications and samples, the point estimates are typically small.

Table 3: THE EFFECT OF CLASS SIZE ON COGNITIVE SKILLS

Model Type		Japanese Model 1	Japanese Model 2	Japanese Model 3	Japanese Model 4	Japanese Model 5	Math Model 1	Math Model 2	Math Model 3	Math Model 4	Math Model 5
Wave (grade in 2018)	grade6	0.003*	0.005**	-0.001			0.001	0.003	-0.001		
		(0.002)	(0.002)	(0.002)			(0.002)	(0.002)	(0.002)		
	grade7	0.001	0.001	-0.003*	0.001	0.002	-0.001	-0.001	-0.004**	-0.005	-0.004
		(0.001)	(0.002)	(0.002)	(0.004)	(0.005)	(0.001)	(0.001)	(0.002)	(0.005)	(0.006)
	grade8	-0.001	0.0005	-0.005**	-0.008		-0.003**	-0.003*	-0.007***	-0.012*	
		(0.001)	(0.002)	(0.002)	(0.005)		(0.001)	(0.001)	(0.002)	(0.007)	
	grade9	0.009***	0.010**	-0.006			0.010***	0.011**	-0.006		
		(0.002)	(0.004)	(0.006)			(0.003)	(0.004)	(0.007)		
	ALL	0.003***	0.004***	-0.002**	0.000	-0.001	0.001	0.002	-0.004***	-0.001	-0.005**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Fixed Effect	School				✓	✓				✓	✓
	Student					✓					✓
Control Variable	School Enrollment		✓	✓	✓	✓		✓	✓	✓	✓
	Sex	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Year	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Grade(Wave = ALL)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: The table shows the results of the estimation of the effect of class size on cognitive skills. The unit of observations is student. The coefficient of class size is reported in each cell in the rows of "Wave". The estimation is made for each wave("grade6"-"grade9") and pooled sample("ALL"). The label of the wave is named by grade in 2018, so "grade6" wave represents the sample of six grade students in 2018. "Model1"-"Model5" represents estimated specification. Columns of "Model1" show OLS estimates. Columns of "Model2"-"Model5" show two-stage least square estimates. The difference between "Model2"-"Model5" are control factor. What factor each estimates control are shown in the rows of "Fixed Effect" and "Control Variable". All dependent variable measures are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors in parentheses are clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

22

Table 4 reports estimates across different specifications for the effects of class-size reduction on non-cognitive skills for each grade. We do not report estimates for models with a weak instrument.

Columns labeled as Model 1 and Model 2 show that OLS estimates are negative, but they are statistically significant only for conscientiousness. The column labeled as Model 3 presents IV estimates. The coefficient for self-control in grade nine is positive significant, while others are all insignificant.

We control for school fixed effects for conscientiousness and self-efficacy. Once school fixed effects are controlled, the coefficients are insignificant in the column labeled as Model 4. Additionally controlling for student fixed effect does not change the results.

Overall, we do not find evidence that class-size reduction improves students' non-cognitive skills.

Table 4: THE EFFECT OF CLASS SIZE ON NON-COGNITIVE SKILLS

Model Type	Conscientiousness Model 1	Conscientiousness Model 2	Conscientiousness Model 3	Conscientiousness Model 4	Conscientiousness Model 5
Wave (grade in 2018)					
grade6					
grade7	-0.005*** (0.001)	-0.005*** (0.002)	-0.003 (0.002)		
grade8				-0.006 (0.008)	
grade9					
ALL					
Fixed Effect					
School					
Student				✓	✓
Control Variable					
School Enrollment				✓	✓
Sex	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓
Grade(Wave = ALL)	✓	✓	✓	✓	✓

Model Type	Selfcontrol Model 1	Selfcontrol Model 2	Selfcontrol Model 3	Selfcontrol Model 4	Selfcontrol Model 5
Wave (grade in 2018)					
grade6	-0.0001 (0.001)	-0.001 (0.001)	-0.001 (0.002)		
grade7					
grade8					
grade9	-0.0002 (0.002)	0.008** (0.004)	0.014** (0.006)		
ALL					
Fixed Effect					
School					
Student				✓	✓
Control Variable					
School Enrollment				✓	✓
Sex	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓
Grade(Wave = ALL)	✓	✓	✓	✓	✓

Model Type	Self-efficacy Model 1	Self-efficacy Model 2	Self-efficacy Model 3	Self-efficacy Model 4	Self-efficacy Model 5
Wave (grade in 2018)					
grade6	-0.001 (0.001)	-0.001 (0.002)	-0.0002 (0.002)	0.006 (0.006)	0.007 (0.007)
grade7					
grade8					
grade9					
ALL					
Fixed Effect					
School					
Student				✓	✓
Control Variable					
School Enrollment				✓	✓
Sex	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓
Grade(Wave = ALL)	✓	✓	✓	✓	✓

Note: The table shows the results of the estimation of the effect of class size on non-cognitive skills. The unit of observations is student. The coefficient of class size is reported in each cell in the rows of "Wave". The estimation is made for each wave("grade6"-"grade9") and pooled sample("ALL"). The label of the wave is named by grade in 2018, so "grade6" wave represents the sample of six grade students in 2018. "Model1"-"Model5" represents estimated specification. Columns of "Model1" show OLS estimates. Columns of "Model2"-"Model5" show two-stage least square estimates. The difference between "Model2"-"Model5" are control factor. What factor each estimates control are shown in the rows of "Fixed Effect" and "Control Variable". All dependent variable measures are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors in parentheses are clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

5.3 Alternative specifications

5.3.1 Schools near the cutoff

Although we control for possible confounders by including enrollment and school and student 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. We do not estimate models with a weak instrument.

Table 5 presents the estimates for the subsamples around the cutoffs. The coefficients are small and statistically insignificant in all outcomes except for the one for Japanese for grade seven. We find that the results from the subsamples near the cutoff are similar to those from our main results.

Table 5: DISCONTINUITY SAMPLES ±5

		Dependent Variable				
		Japanese	Math	Selfcontrol	Self-efficacy	Conscientiousness
Model Type		limited sample(+-5)	limited sample(+-5)	limited sample(+-5)	limited sample(+-5)	limited sample(+-5)
Wave(grade in 2018)	grade6	0.014 (0.011)	-0.002 (0.009)			
	grade7	-0.025** (0.011)			0.003 (0.010)	
	grade8		-0.016 (0.011)			0.002 (0.011)
	grade9					
Fixed Effect	School	✓	✓	✓	✓	✓
	Student					
Control	School Enrollment	✓	✓	✓	✓	✓
	Sex	✓	✓	✓	✓	✓
	Year	✓	✓	✓	✓	✓

Source: Authors' calculations

Note: The table shows the results of the estimation of the effect of class size on cognitive skills and non-cognitive skills in the discontinuity samples. The unit of observations is student. The coefficient of class size is reported in each cell in the rows of "Wave". The estimation is made for each wave("grade6"-"grade9") and pooled sample("ALL"). The label of the wave is named by grade in 2018, so "grade6" wave represents the sample of six grade students in 2018. All estimated specification are two-stage least square estimates which control school fixed effect. All dependent variable measures are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors in parentheses are clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

5.3.2 Aggregating to the school level

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 school-grade-year combination. Hence, if there is correlation between unobserved class-room 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 same within school-grade-year combination.

While we do not see any clear advantage or disadvantage of the use of school-level variables in terms of bias, the standard errors may be larger because there is less variation in the school-level data.

We examine how estimates change if we aggregate the student-level variables to the school-level variables. Table 6 shows the first stage regression estimates. The F-values indicate that the instrument is strong for all grades when school fixed effects are not controlled, but it is marginally strong for grade seven when school fixed effects are controlled.

Table 7 reports estimates for test scores when the aggregated variables are used. Columns labeled as Model 1 and Model 2 show OLS and IV estimates respectively, with control variables including sex ratio, year, enrollment, and grade dummies (only for the pooled sample). Coefficients are small negative except for those for grade nine.

Columns labeled as Model 3 present IV estimates with school fixed effects to avoid possible endogeneity bias arising from unobserved school characteristics. The identifying variation for this specification is variation within school across different years. We show results for grade seven and for the pooled sample because the instrument is strong for these samples only. The point estimates are not very different from those from the model without school fixed effects, although standard errors are larger because of omitting between-school variations.

Table 8 presents the estimates for non-cognitive skills. The results are

very similar to those from the student-level data. Namely, we do not find evidence that class-size reduction improves students' non-cognitive skills.

Table 6: FIRST STAGE RESULTS(AGGREGATE)

Wave(grade in 2018)		Grade6	Grade7	Grade8	Grade9	ALL
Model3 (IV)	Predicted Class Size	0.92***	0.837***	0.847***	0.787***	0.854***
		(0.019)	(0.025)	(0.021)	(0.026)	(0.013)
	R2	0.92	0.854	0.851	0.789	0.858
	Fvalue	2258.024	1127.914	1591.231	885.236	4080.363
	N	1420	2131	1780	1425	6756
Model4 (IV + School Fixed Effect)	Predicted Class Size	0.262	0.292***	0.34**	-0.061	0.727***
		(0.189)	(0.084)	(0.163)	(0.077)	(0.023)
	R2	0.987	0.964	0.986	0.976	0.908
	Fvalue	1.937	12.002	4.376	0.623	1004.366
	N	1420	2131	1780	1425	6756
Control	School Enrollment		✓	✓	✓	✓
	Female Ratio		✓	✓	✓	✓
	Year		✓	✓	✓	✓
	Grade					✓

Source: Authors' calculations

Note: The table shows first stage results of two-stage least squares(2SLS) estimates. The unit of observations is school. The estimation is made for each wave("Grade6"-"Grade9") and pooled sample("ALL"). The label of the wave is named by grade in 2018, so "Grade6" wave represents the sample of six grade students in 2018. "Model3", "Model4", represents estimated specification("Model1" and "Model2" are omitted. "Model1" is not two-stage least squares estimates and "Model2" is very similar specification to "Model3"). The dependent variables of 2SLS are math scores. Standard errors in parentheses are clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

Table 7: THE EFFECT OF CLASS SIZE ON COGNITIVE SKILLS (AGGREGATE)

Model Type		Japanese Model1	Japanese Model2	Japanese Model3	Math Model1	Math Model2	Math Model3
Wave (grade in 2018)	grade6	-0.004	-0.002		-0.004	-0.003	
		(0.003)	(0.003)		(0.002)	(0.002)	
	grade7	-0.003	-0.003	0.001	-0.002	-0.002	-0.002
		(0.002)	(0.003)	(0.005)	(0.002)	(0.002)	(0.006)
	grade8	-0.004*	-0.005**		-0.006***	-0.006***	
		(0.002)	(0.002)		(0.002)	(0.002)	
	grade9	0.007	0.007		0.008	0.005	
		(0.005)	(0.007)		(0.005)	(0.008)	
Fixed Effect	School			✓			✓
	Student						
Control Variable	School Enrollment	✓	✓	✓	✓	✓	✓
	Famale Ratio	✓	✓	✓	✓	✓	✓
	Year	✓	✓	✓	✓	✓	✓

Source: Authors' calculations

Note: The table shows the results of the estimation of the effect of class size on cognitive skills. The unit of observations is school. The coefficient of class size is reported in each cell in the rows of "Wave". The estimation is made for each wave("grade6"-"grade9") and pooled sample("ALL"). The label of the wave is named by grade in 2018, so "grade6" wave represents the sample of six grade students in 2018. "Model1"-"Model5" represents estimated specification. Columns of "Model1" show OLS estimates. Columns of "Model2"-"Model5" show two-stage least square estimates. The difference between "Model2"-"Model5" are control factor. What factor each estimates control are shown in the rows of "Fixed Effect" and "Control Variable". All dependent variable measures are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors in parentheses are clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

Table 8: THE EFFECT OF CLASS SIZE ON NON-COGNITIVE SKILLS (AGGREGATE)

Model Type		Conscientiousness Model1	Conscientiousness Model2	Conscientiousness Model3	Selfcontrol Model1	Selfcontrol Model2	Selfcontrol Model3	Self-efficacy Model1	Self-efficacy Model2	Self-efficacy Model3
Wave (grade in 2018)	grade6				0.001 (0.002)	0.001 (0.002)				0.005 (0.007)
	grade7	-0.002 (0.002)						-0.001 (0.002)	-0.002 (0.002)	
	grade8		-0.002 (0.002)							
	grade9				0.004 (0.004)	0.011** (0.005)				
Fixed Effect	School			✓			✓			✓
	Student									
Control Variable	School Enrollment	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Famale Ratio	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Year	✓	✓	✓	✓	✓	✓	✓	✓	✓

Source: Authors' calculations

Note: The table shows the results of the estimation of the effect of class size on non-cognitive skills. The unit of observations is student. The coefficient of class size is reported in each cell in the rows of "Wave". The estimation is made for each wave("grade6"-"grade9") and pooled sample("ALL"). The label of the wave is named by grade in 2018, so "grade6" wave represents the sample of six grade students in 2018. "Model1"-"Model5" represents estimated specification. Columns of "Model1" show OLS estimates. Columns of "Model2"-"Model5" show two-stage least square estimates. The difference between "Model2"-"Model5" are control factor. What factor each estimates control are shown in the rows of "Fixed Effect" and "Control Variable". All dependent variable measures are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors in parentheses are clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

31

5.3.3 Schools strictly following Maimonides' rule

We take a subsample of schools that follow Maimonides' rule strictly and estimate the effects of class size. Because we use the predicted class size by Maimonides' rule as an instrumental variable, whether observed class size in a given school follows Maimonides' rule strictly or not does not matter to identification. The identifying variation for an IV regression is precisely the predicted class size by Maimonides' rule.

Nevertheless, we report our estimates based on a subsample of schools that follow Maimonides' rule strictly for skeptic readers. Due to the sample size being so small, we are unable to control for school fixed effects. Our estimates in Table 9 indicate that the effects of class size are insignificant. They differ from the estimates from our main sample perhaps by chance due to large standard errors, perhaps because the schools that follow Mainonides' rule strictly are not representative, or both.

Table 9: CLASS-SIZE EFFECT ON COGNITIVE SKILLS AND NONCOGNITIVE SKILLS UNDER MAIMONIDES RULE COMPLETELY

		Dependent Variable		
		Japanese	Math	Self-efficacy
	grade7	-0.001	0.004	-0.001
		(0.009)	(0.004)	(0.005)
Fixed Effect	School			
	Student			
Control	School Enrollment	✓	✓	✓
	Sex	✓	✓	✓
	Year	✓	✓	✓
	Observations	758	758	752

Source: Authors' calculations

Note: The table shows the results of the estimation of the effect of class size on cognitive skills and non-cognitive skills in the samples under Maimonides rule completely. The unit of observations is a student. The coefficient of class size is reported in each cell in the rows of "grade7". "grade7" represents the sample of six grade students in 2018. All estimated specification are two-stage least square estimates. All dependent variable measures are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors in parentheses are clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

33

5.4 Heterogeneity by socio-economic status

We examine how effects of class size vary by individual students' SES, rather than by schools. We split the sample by whether they have no books at home or not, which is a proxy for low SES. Tables 10 and 11 report estimates for cognitive and non-cognitive skills. We do not find clear evidence that the effects of class-size reduction varies by individual SES measured by the number of books.

Finally, we estimate heterogeneity by whether a student goes to a private tutoring school, which is a proxy for higher SES because students from wealthier families are more likely to go to an expensive private tutoring school. Table 12 shows estimates for cognitive skills. They are smaller for students not going to a private tutoring school, implying that effects of class-size reduction tend to be stronger for them. Table 13 report estimates for non-cognitive skills. They indicate that there is no class-size effects across subgroups.

Table 10: HETEROGENOUS EFFECT OF SES ON COGNITIVE SKILLS (1)

		Japanese		Math	
		Books	No Books	Books	No Books
Wave(grade in 2018)	grade6	-0.001	-0.004	0.000	-0.006
		(0.002)	(0.003)	(0.002)	(0.004)
	grade7	-0.003	-0.001	-0.004**	-0.002
		(0.002)	(0.003)	(0.002)	(0.003)
	grade8	-0.004*	-0.003	-0.005***	-0.008***
		(0.002)	(0.003)	(0.002)	(0.003)
	grade9	-0.003	-0.003	-0.004	-0.003
		(0.003)	(0.004)	(0.003)	(0.004)
Fixed Effect	School				
	Student				
Control	School Enrollment	✓	✓	✓	✓
	Sex	✓	✓	✓	✓
	Year	✓	✓	✓	✓
ALL	Model3 (IV)	-0.002**	-0.003*	-0.004***	-0.005***
		(0.001)	(0.002)	(0.001)	(0.002)
	Model4 (IV + School Fixed Effect)	0.000	-0.001	-0.002	-0.001
		(0.001)	(0.002)	(0.001)	(0.002)
Control	School Enrollment	✓	✓	✓	✓
	Sex	✓	✓	✓	✓
	Year	✓	✓	✓	✓
	Grade	✓	✓	✓	✓

Source: Authors' calculations

Note: The tables show the results of the estimation of the effect of class size on cognitive skills. The unit of observations is a student. The sample is divided into either low SES student or not by the level of books at home. If a student doesn't have books at home, the student is classified into low SES students, otherwise. The estimation is made for each wave("grade6"-"grade9") and pooled sample("ALL"). In upper tables, The estimated coefficient of class size for each wave is reported. All estimated specification are two-stage least square estimates. In lower tables, The estimated coefficient of class size for the pooled sample are reported. All estimated specification are two-stage least square estimates. "Model3" and "Model4" represents estimated specification. "Model3" specification is two-stage least squares(2SLS), and "Model4" specification is also 2SLS which control school fixed effect. All dependent variable measures are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors in parentheses are clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

Table 11: HETEROGENOUS EFFECT OF SES ON NON-COGNITIVE SKILLS (1)

		Dependent Variable					
		Selfcontrol		Self-efficacy		Conscientiousness	
		Books	No Books	Books	No Books	Books	No Books
Wave(grade in 2018)	grade6	-0.001 (0.002)	0.000 (0.003)				
	grade7			0.000 (0.002)	-0.003 (0.003)		
	grade8	0.006** (0.002)				-0.003* (0.002)	0.000 (0.003)
	grade9		0.001 (0.004)				
Fixed Effect	School						
	Student						
Control	School Enrollment	✓	✓	✓	✓	✓	✓
	Sex	✓	✓	✓	✓	✓	✓
	Year	✓	✓	✓	✓	✓	✓

Source: Authors' calculations

Note: The tables show the results of the estimation of the effect of class size on non-cognitive skills. The unit of observations is a student. The sample is divided into either low SES student or not by the level of books at home. If a student doesn't have books at home, the student is classified into low SES students, otherwise. The estimation is made for each wave("grade6"-"grade9") and pooled sample("ALL"). In upper tables, The estimated coefficient of class size for each wave is reported. All estimated specification are two-stage least square estimates. In lower tables, The estimated coefficient of class size for the pooled sample are reported. All estimated specification are two-stage least square estimates. "Model3" and "Model4" represents estimated specification. "Model3" specification is two-stage least squares(2SLS), and "Model4" specification is also 2SLS which control school fixed effect. All dependent variable measures are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors in parentheses are clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

36

Table 12: HETEROGENOUS EFFECT OF SES ON COGNITIVE SKILLS (2)

| | | Dependent Variable | | | |
| | | Japanese | | Math | |
		No Private Tutoring	Private Tutoring	No Private Tutoring	Private Tutoring
Wave(grade in 2018)	grade6	-0.004*	0.001	-0.003	0.001
		(0.002)	(0.003)	(0.002)	(0.002)
	grade7	-0.004*	-0.002	-0.005***	-0.003
		(0.002)	(0.002)	(0.002)	(0.002)
	grade8	-0.003	-0.003	-0.005**	-0.006***
		(0.002)	(0.002)	(0.002)	(0.002)
	grade9	-0.005*	-0.001	-0.007***	-0.002
		(0.002)	(0.003)	(0.003)	(0.003)
Fixed Effect	School				
	Student				
Control	School Enrollment	✓	✓	✓	✓
	Sex	✓	✓	✓	✓
	Year	✓	✓	✓	✓
ALL	Model3 (IV)	-0.004***	-0.001	-0.005***	-0.002*
		(0.001)	(0.001)	(0.001)	(0.001)
		-0.001	0.001	-0.002	-0.001
	Model4 (IV + School Fixed Effect)	(0.001)	(0.001)	(0.001)	(0.001)
Control	School Enrollment	✓	✓	✓	✓
	Sex	✓	✓	✓	✓
	Year	✓	✓	✓	✓
	Grade	✓	✓	✓	✓

Source: Authors' calculations

Note: The tables show the results of the estimation of the effect of class size on cognitive skills. The unit of observations is a student. The sample is divided into either whether to go cram school or not. The estimation is made for each wave("grade6"-"grade9") and pooled sample("ALL"). In upper tables, The estimated coefficient of class size for each wave is reported. All estimated specification are two-stage least square estimates. In lower tables, The estimated coefficient of class size for the pooled sample are reported. All estimated specification are two-stage least square estimates. "Model3" and "Model4" represents estimated specification. "Model3" specification is two-stage least squares(2SLS), and "Model4" specification is also 2SLS which control school fixed effect. All dependent variable measures are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors in parentheses are clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

37

Table 13: HETEROGENOUS EFFECT OF SES ON NON-COGNITIVE SKILLS (2)

		Selfcontrol		Self-efficacy		Conscientiousness	
Dependent Variable		No Private Tutoring	Private Tutoring	No Private Tutoring	Private Tutoring	No Private Tutoring	Private Tutoring
Wave(grade in 2018)	grade6	-0.002 (0.002)	0.000 (0.002)				
	grade7			-0.001 (0.002)	-0.001 (0.002)		
	grade8	0.004 (0.002)				-0.003 (0.002)	-0.003 (0.002)
	grade9		0.007** (0.003)				
Fixed Effect	School						
	Student						
Control	School Enrollment	✓	✓	✓	✓	✓	✓
	Sex	✓	✓	✓	✓	✓	✓
	Year	✓	✓	✓	✓	✓	✓

Source: Authors' calculations

Note: The tables show the results of the estimation of the effect of class size on non-cognitive skills. The unit of observations is a student. The sample is divided into either whether to go cram school or not. The estimation is made for each wave("grade6"-"grade9") and pooled sample("ALL"). In upper tables, The estimated coefficient of class size for each wave is reported. All estimated specification are two-stage least square estimates. In lower tables, The estimated coefficient of class size for the pooled sample are reported. All estimated specification are two-stage least square estimates. "Model3" and "Model4" represents estimated specification. "Model3" specification is two-stage least squares(2SLS), and "Model4" specification is also 2SLS which control school fixed effect. All dependent variable measures are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors in parentheses are clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

38

5.5 Discussion

5.5.1 Comparison with previous estimates

Our preferred estimates based on the IV regressions without school fixed effects for Japanese and math test scores range from 0.01 to 0.07 standard deviations for a 10-student reduction. They are in the same ballpark of existing estimates from Japanese data, although slightly smaller than estimates by Ito et al. (2017) and the estimate for Japanese test scores for sixth graders by Akabayashi and Nakamura (2014).

Ito et al. (2017) argue that previous papers report small or insignificant effects of class size because the previous papers rely on between-school comparison, and hence, fail to control for unobserved school characteristics. While they are right in the sense that controlling for school fixed effects is important for their data from an anonymous city in Chubu region, controlling for school fixed effects does not seem to bias estimates substantially in our data from an anonymous prefecture and data from Yokohama.[4] We argue that, although small, the differences in the estimates are likely to reflect the fact that samples are taken from different locations, rather than differences in their statistical methods.

5.5.2 Can class-size effects on grade eight students be real?

Although the effects of the class-size reduction seems small, if any, the estimates for Japanese and math for grade eight students are larger than those for other grades. While this may simply be by a chance, this might be related to the fact that students are promoted from an elementary school to a junior high school.

Because students in our sample take the achievement test in April, which is the beginning of school year, we estimate the effects of class size in the

---

[4]In their robustness check, Akabayashi and Nakamura (2014) control for school fixed effects and find that estimates do not differ from their preferred estimates without school fixed effects.

previous school year on the outcomes in April. Therefore, effects on outcomes of eighth graders reflect the effects of class size in grade seven. Note that grade seven is the first year in junior high school, which is usually completely separated from elementary schools.

In informal conversations, teachers mentioned that the new junior high school students (i.e. seventh graders) often faced challenges caused by drastic changes in school environments. Perhaps, smaller class size may help the students adjust themselves to the new environment.

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 from four to nine in an anonymous prefecture. Our identification approach is based on Maimonides' rule for class size.

We find that the effects of class-size reduction on Japanese and math test scores are rather small. The estimates from our preferred specification range from 0.01 to 0.07 standard deviations for a 10-student reduction. We also find no effects of class-size reduction on non-cognitive skills. Our estimates are robust to the choice of statistical methods. Most importantly, controlling for school fixed effects does not change the estimates substantially.

The results from subsample analysis indicate that the effects of class-size reduction are stronger for students not going to a private tutoring school, although the effect size is still small even for these students.

Given that our estimates are robust to the choice of statistical methods and that class-size effects vary among students, the different estimates across existing studies on Japan are likely due to 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 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 a particular circumstance 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, data on teacher characteristics and their teaching environment are necessary.

References

Akabayashi, Hideo, and Ryosuke Nakamura. 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., and V. Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." The Quarterly Journal of Economics, 114(2): 533–575.

Bandura, Albert. 1986. "Social foundations of thought and action." Englewood Cliffs, NJ, 1986.

Barbaranelli, Claudio, Gian Vittorio Caprara, Annarita Rabasca, and Concetta Pastorelli. 2003. "A questionnaire for measuring the Big Five in late childhood." Personality and Individual Differences, 34(4): 645–664.

Borghans, Lex, Angela Lee Duckworth, James J Heckman, and Bas Ter Weel. 2008. "The economics and psychology of personality traits." Journal of human Resources, 43(4): 972–1059.

Carneiro, Pedro, Claire Crawford, and Alissa Goodman. 2007. "The impact of early cognitive and non-cognitive skills on later outcomes."

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

Dee, Thomas S., and Martin R. West. 2011. "The Non-Cognitive Returns to Class Size." Educational Evaluation and Policy Analysis, 33(1): 23–46.

Duckworth, Angela L, Eli Tsukayama, and Teri A Kirby. 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.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek. 2013. "Long-Term Effects of Class Size *." The Quarterly Journal of Economics, 128(1): 249–285.

Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." Journal of Labor economics, 24(3): 411–482.

Heineck, Guido, and Silke Anger. 2010. "The returns to cognitive abilities and personality traits in Germany." Labour economics, 17(3): 535–546.

Hojo, Masakazu. 2013. "Class-size effects in Japanese schools: A spline regression approach." Economics Letters, 120(3): 583–587.

Hoxby, Caroline 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, Hiroyuki, Megumi Hamada, Yasuo Murayama, Takayanagi Nobuya, Kazuo Nomura, Mitsunobu Myogan, and Masatsugu Tsuji. 2017. "Effects of Class Size on Academic Achievement and Emotional and Behavioral Problems:." The Japanese Journal of Educational Psychology, 65(4): 451–465.

Krueger, Alan B. 1999. "Experimental estimates of education production functions." Quarterly Journal of Economics, 114(2): 497–532.

Mueller, Gerrit, and Erik Plug. 2006. "Estimating the effect of personality on male and female earnings." ILR Review, 60(1): 3–22.

Niki, Minae. 2013. "The effects of class size on cognitive and non-cognitive abilities." Economic Analysis, 186: 30–49.

Pajares, Frank, and John Kranzler. 1995. "Self-efficacy beliefs and general mental ability in mathematical problem-solving." Contemporary educational psychology, 20(4): 426–443.

Pajares, Frank, and Laura Graham. 1999. "Self-efficacy, motivation constructs, and mathematics performance of entering middle school students." Contemporary educational psychology, 24(2): 124–139.

Pajares, Frank, and M David Miller. 1994. "Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis." Journal of educational psychology, 86(2): 193.

Pintrich, Paul R, and Elisabeth V De Groot. 1990. "Motivational and self-regulated learning components of classroom academic performance." Journal of educational psychology, 82(1): 33.

Pintrich, Paul R, et al. 1991. "A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)."

Poropat, Arthur E. 2009. "A meta-analysis of the five-factor model of personality and academic performance." Psychological bulletin, 135(2): 322.

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

Senoh, Wataru, Masakazu Hojo, Takehisa Shinozaki, and Shinpei Sano. 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, James H, Jonathan H Wright, and Motohiro Yogo. 2002. "A survey of weak instruments and weak identification in generalized method of moments." Journal of Business & Economic Statistics, 20(4): 518–529.

Tsukayama, Eli, Angela Lee Duckworth, and Betty Kim. 2013. "Domain-specific impulsivity in school-age children." Developmental Science, 16(6): 879–893.

Urquiola, Miguel. 2006. "Identifying class size effects in developing countries: Evidence from rural bolivia." Review of Economics and Statistics, 88(1): 171–177.

A Additional tables and figures

Table 14: ITEMS CONSISTING OF NON-COGNITIVE SKILLS

Class	Group	Question Items
Non Cognitive Ability	Self-control	I forgot something needed for school
		I interrupted other people.
		I said something rude.
		I could not find something because of mess.
		I lost temper.
		I did not remember what someone said to do.
		I mind wandered.
		I talked back when upset.
	Self-efficacy	I believe I will receive an excellent grade in this clas.s
		I'm certain I can understand the most difficult material presented in the readings for this course.
		I'm confident I can understand the basic concepts taught in this course.
		I'm confident I can understand the most complex material presented by the instructor in this course.
		I'm confident I can do an excellent job on the assignments and tests in this course.
		I expect to do well in this class.
		I'm certain I can master the skills being taught in this class.
		Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this class.
	Conscientiousness	I do my job without carelessness and inattention.
		I work hard and with pleasure.
		I engage myself in the things I do.
		During class-time I am concentrated on the things I do.
		When I finish my homework, I check it many times to see if I did it correctly.
		I respect the rules and the order.
		If I take an engagement I keep it.
		My room is in order.
		When I start to do something I have to finish it at all costs.
		I like to keep all my school things in a great order.
		I play only when I finished my homework.
		It is unlikely that I divert my attention.
		I do my own duty.

Table 15: DISCONTINUITY SAMPLES ±3

		Dependent Variable				
		Japanese	Math	Selfcontrol	Self-efficacy	Conscientiousness
Model Type		limited sample(+-3)	limited sample(+-3)	limited sample(+-3)	limited sample(+-3)	limited sample(+-3)
Wave(grade in 2018)	grade6					
	grade7	0.013	0.008		0.005	0.011
		(0.014)	(0.012)		(0.013)	(0.032)
	grade8	-0.019	-0.013			
		(0.028)	(0.031)			
	grade9					
Fixed Effect	School	✓	✓	✓	✓	✓
	Student					
Control	School Enrollment	✓	✓	✓	✓	✓
	Sex	✓	✓	✓	✓	✓
	Year	✓	✓	✓	✓	✓

Source: Authors' calculations

Note: The table shows the results of the estimation of the effect of class size on cognitive skills and non-cognitive skills in the discontinuity samples. The unit of observations is student. The coefficient of class size is reported in each cell in the rows of "Wave". The estimation is made for each wave("grade6"-"grade9") and pooled sample("ALL"). The label of the wave is named by grade in 2018, so "grade6" wave represents the sample of six grade students in 2018. All estimated specification are two-stage least square estimates which control school fixed effect. All dependent variable measures are standardized to have a mean of zero and a standard deviation of unity within each grade. Standard errors in parentheses are clustered by the school. "***", "**", and "*" represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

Table 16: HETEROGENOUS EFFECT OF SES ON COGNITIVE SKILLS: INTERACTION TERM EFFECT

	Dependent Variable			
	Japanese	Math	Japanese	Math
Interaction Factor	Class Size * LowSES	Class Size * LowSES	Class Size * Private Tutoring	Class Size * Private Tutoring
Wave(grade in 2018) grade6	-0.006**	-0.007**	0.006***	0.008***
	(0.003)	(0.003)	(0.002)	(0.002)
grade7	-0.003	-0.004*	0.002	0.004**
	(0.002)	(0.002)	(0.001)	(0.002)
grade8	-0.001	-0.003	0.004**	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
grade9	-0.003	-0.002	0.007***	0.011***
	(0.003)	(0.002)	(0.002)	(0.002)
Fixed Effect School				
Student				
Control School Enrollment	✓	✓	✓	✓
Sex	✓	✓	✓	✓
Year	✓	✓	✓	✓
ALL Model3 (IV)	-0.002	-0.002*	0.005***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)
Model4 (IV + School Fixed Effect)	-0.003**	-0.003***	0.005***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)
Control School Enrollment	✓	✓	✓	✓
Sex	✓	✓	✓	✓
Year	✓	✓	✓	✓
Grade	✓	✓	✓	✓

Source: Authors' calculations

Note: Standard errors in parentheses are adjusted for clustering on schools. All dependent measures are standardized to have a mean of zero and a standard deviation of unity within each grade. All regressions use the data at the school level. ***, **, and * represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.

47

Table 17: HETEROGENOUS EFFECT OF SES ON NON COGNITIVE SKILLS: INTERACTION TERM EFFECT

Dependent Variable	Selfcontrol	Self-efficacy	Conscientiousness
Interaction Factor	Class Size * LowSES	Class Size * LowSES	Class Size * LowSES
Wave(grade in 2018) grade6	-0.002 (0.003)	-0.005** (0.002)	
grade7			
grade8	0.002 (0.003)		0.004* (0.003)
grade9			

Dependent Variable	Selfcontrol	Self-efficacy	Conscientiousness
Interaction Factor	Class Size * Private Tutoring	Class Size * Private Tutoring	Class Size * Private Tutoring
Wave(grade in 2018) grade6	0.000 (0.002)	0.001 (0.002)	
grade7			
grade8	-0.002 (0.002)		0.000 (0.001)
grade9			

Fixed Effect			
School			
Student			
Control			
School Enrollment	✓	✓	✓
Sex	✓	✓	✓
Year	✓	✓	✓

Source: Authors' calculations

Note: Standard errors in parentheses are adjusted for clustering on schools. All dependent measures are standardized to have a mean of zero and a standard deviation of unity within each grade. All regressions use the data at the school level. ***, **, and * represent 0.1 percent, 1 percent, and 5 percent significance level, respectively.