The Scandinavian
Journal of Economics

Scand. J. of Economics 110(4), 663–693, 2008
DOI: 10.1111/j.1467-9442.2008.00556.x

Quasi-experimental Estimates of the Effect of Class Size on Achievement in Norway*

Edwin Leuven

ENSAE-CREST, F-92245 Malakoff Cedex, France
leuven@ensae.fr

Hessel Oosterbeek

Universiteit van Amsterdam, NL-1018 WB Amsterdam, Netherlands
h.oosterbeek@uva.nl

Marte Rønning

Statistics Norway, N-0033, Oslo, Norway
mro@ssb.no

Abstract

Using a comprehensive Norwegian administrative database, we exploit independent quasi-experimental methods to estimate the effect of class size on student achievement at the end of lower-secondary school. Identification based on maximum class-size rules and population variation (and variations on these methods) give very similar estimates. We cannot reject that the class-size effect is equal to zero, and can rule out effects as small as 1.5 percent of a standard deviation for a one-student change in class size during three consecutive years.

Keywords: Class size; educational production

JEL classification: *I*2

I. Introduction

One of the still unresolved issues in education research concerns the effects of class size on students' achievement. By now, it is well understood that endogeneity problems may severely bias naïve OLS estimates of the class-size effect, and that exogenous sources of variation in class size are key for a credible identification of the class-size effect. Various recent studies acknowledge this and apply convincing identification methods. However, this has not led to a definite conclusion about the magnitude or even the sign of the class-size effect.

* Leuven gratefully acknowledges financial support from an NWO-VENI grant. Rønning was a Ph.D. student in the Department of Economics, NTNU Norwegian University of Science and Technology, while working on this paper.

Most of the (quasi-) experimental studies report that a reduction in class size boosts achievement; cf. Angrist and Lavy (1999), Boozer and Rouse (2001), Browning and Heinesen (2007), Krueger (1999) and Urquiola (2006). Some of the studies that attempt to correct for endogeneity do, however, report opposite results; cf. e.g. Hoxby (2000). Moreover, in studies that report negative effects of bigger classes, the size of this effect varies considerably, thereby limiting the relevance of these results for policy conclusions.

Of course, there need not be a universal effect of class-size reduction on achievement. Effects may vary with characteristics of the students affected by the policy, or by contextual factors such as remedial instruction for low-performing students or the quality of teachers' education; cf. Wößmann and West (2006). This would imply that for policy purposes, studies have to be conducted for separate levels of education and for separate countries (or perhaps for groups of very similar countries).

This paper provides evidence about the effect of class size on achievement in Norwegian lower secondary schools. This evidence is obtained by means of two different and independent approaches. The first approach uses exogenous variation due to maximum class-size rules in Norwegian lower secondary education. This approach was first used by Angrist and Lavy (1999). The second approach exploits variation in actual class size that is attributable to demographic variation, as applied by Hoxby (2000).

Some features of our study are worth emphasizing. First, we have access to an extraordinarily rich dataset. The dataset covers two entire cohorts of students who participated in nationwide tests in the school years 2001/2002 and 2002/2003. The data are administrative, thereby giving rise to no or only slight measurement error in actual class size and enrollment. Achievement is measured as high-stake test scores, thereby also reducing measurement error in the dependent variables. Together, these characteristics of the dataset enable us to produce very precise estimates of class-size effects. Second, for all students in our sample we know actual class sizes during the three years they spent in lower secondary school. This allows a clear-cut interpretation of the effects that we estimate.

There are no previous studies with precise estimates of the effects of class size on achievement for any of the Nordic countries (Denmark, Finland, Norway and Sweden). Earlier studies have either considered other outcomes than achievement such as years of education, as in Bingley, Jense and Walker (2005) and Browning and Heinesen (2007), or used a very small sample, thereby lacking precision, as in Bonesrønning (2003) and Lindahl (2005). Moreover, most of the evidence on class-size effects pertains to primary education; our study is among the few dealing with class-size effects in lower secondary education.[1,2]

The results reported in this paper consistently point to a lack of any impact of class size on achievement. Effects as small as 1.5 percent of a standard deviation for a one-student change in class size during three consecutive years can be ruled out. This holds irrespective of identification method (maximum class-size rule or demographic variation), subject tested (math, languages) or the control variables included in the regressions. The finding also holds across various sub-groups of the population and is independent of teacher characteristics.

The remainder of this paper is organized as follows. The next section gives a brief summary of related studies. In Section III we describe the relevant institutional features of the Norwegian educational system. Section IV gives a description of the data used in the empirical analysis. Section V continues with an exposition of the empirical approaches applied in this paper and their limitations. In Section VI we report and discuss the main findings. Section VII investigates the possibility of heterogeneous class-size effects, and Section VIII concludes.

II. Related Literature

Over more than a decade, the common wisdom among economists was that extra resources for education—measured as the teacher–pupil ratio or as expenditures per pupil—have no systematic relation with students' achievement. This view was mainly based on Hanushek's (1986) influential review of the literature. Only recently has this received wisdom been challenged by a series of studies that use experimental and quasi-experimental approaches to identify the causal impact of class size on achievement.

Krueger (1999) analyzes data from a large-scale field experiment conducted in Tennessee. Students and their teachers were randomly assigned to a group of regular size (22–25 students), to a group of regular size including a teaching assistant, or to a small group (13–17), during their first four years in school. Krueger's findings are in line with what others have reported about this project, namely that students in smaller classes perform better on standardized achievement tests. Scores increase by four percentile points for the first year that a student is exposed to a small class and by one percentile point for each subsequent year. In a follow-up study, Krueger and Whitmore (2001) demonstrate that reduced class sizes in early school

[1] For a sub-sample of our schools we could also conduct analyses using class size in primary school as a class-size measure. We have chosen not to present these results since they are very similar to those for lower secondary schools.

[2] Häkkinen, Kirjavainen and Uusitalo (2003) study the lower secondary level for Finland, but they look at a broader measure of school resources.

years can have long-lasting effects. Students who attended small classes in this experiment were more likely to take a college-entrance exam and have somewhat higher test scores. The effects on taking exams are mainly concentrated among minority students.

Angrist and Lavy (1999) were the first to exploit the exogenous variation generated by maximum class-size rules to obtain quasi-experimental estimates of the class-size effect on achievement. They exploit the fact that according to official guidelines for Israeli public schools, maximum class size equals 40. If the size of an enrollment cohort in a school exceeds (a multiple of) 40, an extra class should be created. This rule creates discontinuities in the relation between cohort enrollment size and class size, which Angrist and Lavy then use in a regression discontinuity framework to identify the effect of class size on achievement.[3] When they do not correct for endogeneity bias, their estimates point to a positive relation between class size and achievement. In contrast, estimates based on the discontinuities in grade enrollment point to a negative effect of class size on achievement.

Other papers which exploit maximum class-size rules include Bonesrønning (2003) for Norway, Urquiola (2006) for Bolivia, Piketty (2004) and Gary-Bobo and Mahjoub (2006) for France, Browning and Heinesen (2007) and Bingley *et al.* (2005) for Denmark, and Wößmann (2005) for Denmark, France, Germany, Greece, Iceland, Ireland, Norway, Spain, Sweden and Switzerland. The papers by Bonesrønning and Wößmann are of particular interest for our analysis because they also deal with class-size effects in Norway. Bonesrønning uses data from a small-scale self-collected dataset. Class-size effects are estimated using the discontinuity at 30 as an instrument for actual class size. The reported effects depend on the exact specification but vary between 0.13 and 0.26 of a standard deviation for a 10-student reduction in class size, and are significantly different from zero but not very precisely estimated. The main differences between this study and ours are the following: (i) we apply two different methods rather than only one; (ii) we use a much larger dataset, so that our estimates have much more precision; (iii) we use high-stake exam scores as our achievement measure; (iv) in our regression discontinuity specifications we include controls for enrollment in a grade, whereas Bonesrønning does not. Without controls for enrollment, the estimate of the class-size effect will also pick up effects of enrollment.

Wößmann uses data from 38 Norwegian schools with a total of 1,351 pupils who participated in the Trends in International Mathematics and Science Study (TIMSS). Without controls for enrollment, a small but significantly negative estimate of the class-size effect is found, which

[3] In Section V we discuss their identification strategy in more detail because we apply the same method in this paper.

then vanishes once controls for enrollment are included in the specification. Wößmann uses a cut-off of 28 rather than 30, arguing that this gives the best fit to his data, so that it seems as if this implicit rule is actually used in most Norwegian schools. This “letting the data decide” approach is at odds with the basic philosophy underlying the regression discontinuity approach because assignment to treatment and control will no longer be based on an exogenous rule but on schools’ choices.

Hoxby uses demographic variation to identify the class-size effect. She exploits the idea that—after correcting for a trend—cohort sizes within school districts can be larger or smaller in some years than in others.[4] Using data on elementary school pupils in the state of Connecticut, she does not find any statistically significant effect of class size on student achievement, and her estimates are precise enough to rule out even modest effects.

Two recent papers apply identification strategies in the same spirit as Hoxby’s approach, although they do not control for a trend in cohort size. Urquiola (2006) identifies the class-size effect using variation in population size between schools in rural areas that are so small that fewer than 30 students (a number that fits into one class) are enrolled. These schools are more likely to be in small communities where class size is mainly determined by cohort size. Urquiola uses data from third graders in Bolivia and finds significantly negative effects of class size on achievement.

Wößmann and West (2006) exploit within-school differences in average class size between adjacent grades. An attractive feature of their study is that it uses data from students in 11 different countries, although this comes at a price, i.e., the numbers of (identifying) observations are in some cases rather small and precision therefore low. Sizable positive effects of smaller classes are reported for Greece and Iceland, even small class-size effects can be ruled out for four countries, and large beneficial effects can be ruled out for another four countries. As an explanation for the differences in class-size effects across countries, Wößmann and West (2006) advance the hypothesis that smaller classes are only beneficial where/when the average capability of the teaching force appears to be low.

Lindahl (2005) implements a value-added approach to estimate the effect of class size using Swedish data. He fails to find statistically significant class-size effects using standard value-added methods. When he identifies the class-size effect by taking the difference between school- and summer-period test-score changes, he finds that class size does matters. We are not able to implement this approach because we do not have repeated achievement measures for the students in our sample.

[4] This method is dealt with in more detail in Section V.

The variation in the findings reviewed here is reflected in the controversy between Krueger (2003) and Hanushek (2003). In an attempt to reconcile both views, Todd and Wolpin (2003) stress the difference between policy effects and parameters of the education production function. According to these authors, estimates of the class-size effect obtained from experimental and quasi-experimental research designs should be interpreted as policy effects, whereas estimates obtained from non-experimental research designs are aimed at identifying the education production function. To learn about the production technology, one needs exogenous variation in class size *while holding other inputs constant*. To learn about the policy impact, one needs exogenous variation in class size *while not holding other inputs constant*. Therefore the estimates in our paper are probably best understood as policy effects. Another important input not controlled for in experimental and quasi-experimental studies is that of parents. For instance, parents may respond to a reduction in class size by spending less time teaching their children at home. In that case, school and parental inputs are substitutes and the policy effect will be smaller than the technological effect. In principle, school and parental inputs can also work as complements, in which case the policy effect would exceed the technological effect. A similar line of reasoning holds with respect to, for example, teacher effort. Hægeland, Raaum and Salvanes (2007) argue that the maximum class-size rule is accompanied by a similar input substitution in terms of school resources. We will show that this is not a concern for our findings. When we replace class size by a commonly used measure for school resources in Norway, we find significant differences around the discontinuities and our results for achievement are unchanged.

This brief review of related studies only includes those by economists. For a recent review of the class-size literature from a non-economic perspective (although references to most of the studies cited above are included), see Hattie (2005). His reading of the literature is that class-size effects are often very small. As a candidate explanation, he proposes that teachers tend to use the same teaching methods independent of class size.

III. Institutional Settings

Compulsory schools in Norway are owned and operated by the 435 municipalities.[5,6] Municipalities receive funding to run their various activities (including schools) through a combination of a local income tax, property

[5] In Norway, the terms local school district, local government and municipality are interchangeable.

[6] In addition to compulsory public schooling, local governments are responsible for elderly care, preschool education and infrastructure. Spending on education amounts to about 30 percent of total spending of the available budget.

taxes and transfers from the central government; see e.g. Hægeland *et al.* (2007) for more details.

For the students in our sample, compulsory schooling consisted of nine years: grades 1–6 in primary school and grades 7–9 in lower secondary school.[7] Less than 3 percent of Norwegian students are enrolled in private schools. Thus private schools do not provide a realistic alternative to public schools and were therefore dropped from the analysis. Schools have catchment areas, i.e., parental choice among schools for given residence is not allowed. Most students attend separate primary and lower secondary schools, but due to the rural settlement pattern in Norway, about 23 percent are enrolled in so-called "combined" schools that offer both primary and lower secondary education. Students who begin their primary education in a combined school typically continue their lower secondary education there owing to a lack of school choice for given residence and moving costs. Combined schools are often situated in relatively rural areas. If total enrollment is less than 40, schools often mix grades in the classroom. These schools account for 3.2 percent of the student population and were excluded from the analysis.

Although we measure class-size effects in lower secondary school (instead of primary school, as most studies do), it is important to note that students in lower secondary schools in Norway are assigned to the same class during the entire day and year. Each class is taught by different teachers depending on the subject; the teachers rotate among classes.

Another feature of the Norwegian compulsory school system is that grade retention almost never occurs. Strøm (2004) attributes this to "the strong integration and equalizing policy that all students within a cohort should be treated equal, and be given education in their ordinary classes". As a result, at the end of compulsory schooling all Norwegian students have attended school for an identical length of time. This is important for our analysis since we have student-level data for the nationwide tests conducted in 2001/2002 and 2002/2003 as well as class-size data from these and previous years. Since students do not repeat grades, we know the class-size history of individual students during lower secondary school, provided that they did not change school. We are not aware of any study which documents the extent of students' school mobility in Norway. Hægeland *et al.* (2007), who use the same student data as we do, report that 95.3 percent of the students lived in their graduation municipality throughout all three years in lower secondary schools. While this does not prove

[7] From the school year 1997/1998 onwards it became compulsory to start schooling at the age of 6 (instead of 7), and from then on 10 years of schooling were implemented in Norway. The reform was implemented in such a way that the length of primary school was extended by one year.

low mobility across schools, it is certainly not inconsistent with it. Note furthermore that even selective movements are not problematical when we instrument average class during three years in lower secondary school by predicted class size in seventh grade.

IV. Data

Test Scores

We used administrative enrollment data from Statistics Norway that cover all students who were in the final grade of lower secondary school (ninth grade) for the school years 2001/2002 and 2002/2003. We merged this dataset with test-score data from centralized exit exams (also from Statistics Norway). Students in Norway are required to take such exams at the end of their final year in lower secondary school.[8] Their results on this exam are important for further schooling, i.e., upper secondary education. The exam is regarded as a high-stake test by all parties involved: students, their parents, teachers and school administrators.[9] Although the curriculum includes many different subjects, a written centralized exam is only undertaken in three subjects: mathematics, English and Norwegian.[10] To reduce the administrative burden, each student takes the exam in only one subject, determined by a random device shortly before the tests take place. Students are notified of the subject only three school days in advance, so that they have minimal scope to prepare for the specific subject. All students in a class take their exam in the same subject, but students in different classes at the same school may be tested in different subjects.

Each test is graded by two independent external examiners, plus a third in case of disagreement between the first two. Examiners receive detailed guidelines for grading the exams from the Ministry of Education. Each examiner grades 100–120 tests, in most cases from different schools. Examiners do not know the names of the pupils, but do know the name of the school(s). The exams are graded on a scale of 1–6, where 1 is fail, 2 the lowest pass and 6 the top score. Examiners are explicitly instructed not to normalize exam grades either within or between schools. The distributions of test scores form a bell-shaped curve and there are no signs of floor or

[8] Although this exit exam had already existed for many years, its results did not become available for research purposes until the school year 2001/2002.

[9] Although all students have the right to continue at upper secondary schools, and over 95 percent do so, their choice set among different schools and different study tracks depends on their achievement in lower secondary schools.

[10] Norway has two official written languages, main Norwegian (Hovedmål) and a second Norwegian language (Sidemål). When students are examined in Norwegian, they get a score on both languages.

ceiling effects (low frequencies for 1 and 6). Average scores for each of the subjects are around 3.5, with standard deviations almost equal to 1.

Class Size

Our class-size information is from the Norwegian Ministry of Education which registers, for all schools in Norway, the number of classes per grade. For each student, we combined this information with the enrollment data mentioned above to calculate average class-size at the grade level as

$$\text{class size} = \text{enrollment}/\text{no. of classes}.$$

Note that we have data on average class size per grade and not actual class size (except when schools have exactly one class in a grade). It is important to keep in mind that this eliminates biases resulting from within-school sorting, while the associated measurement error is removed by our 2SLS approach.

An attractive feature of the data is that we not only have information on contemporaneous class size (the school year of the exam), but also on class size during the previous years in lower secondary school. Unfortunately, class sizes across years within the same school are too highly correlated to examine their separate impacts (the correlations are always higher than 0.9). We therefore chose to estimate the effect of class size as the average class size during the three years of lower secondary school. In this way we avoid confounding the impact of class size in grade 9 with the impact of class size in earlier grades. Moreover, impact estimates of class size as defined here are relevant from a policy point of view. Hoxby (2000) also focuses on the average class size experienced by a cohort up until the time it takes the test.

Figure 1 shows the distribution of average class size in lower secondary schools in Norway. It shows that the majority of average class sizes per grade was between 20 and 30 students, although a substantial fraction (25 percent) of the classes had 21 students or less. The (unweighted) average class size is equal to 23.3 (SD = 4.1). The graph also suggests that the maximum class size of 30 is enforced since no class was larger than the threshold of 30. At the grade level, average class size exceeds 30 in only 22 out of 5,032 cases.

Control Variables

We used other administrative databases to merge information on students' age, gender, ethnic minority background, household income, whether parents live together or not, and years of education of both parents. This information pertains to the same year as the exam scores (2001/2002 and

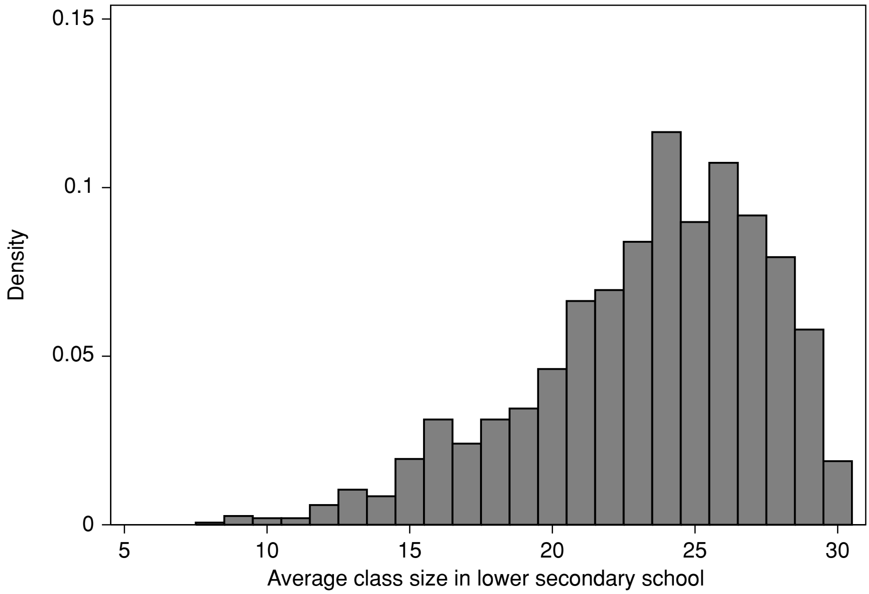

Fig. 1. Distribution of class size

2002/2003). Comparable information is available and exploited in most (but not all) other studies which examine the effect of class size on achievement.

We also control for a number of teacher characteristics based on employer register data from the Ministry of Labor and Government Administration. Since we are not able to link teachers and students, we aggregated the teacher data up to the school level and weighted with the "workload per teacher". Teacher controls included experience, gender, temporary contract and years of schooling. The (log) size of the school district is also included, measured in terms of inhabitants and the number of people in a school district who live in rural areas. We also control for combined schools.

Table 1 lists descriptive statistics for the control variables. Half of the students are female, the average age is 14.5 years, the average levels of fathers' and mothers' education are almost equal, only 5 percent of the sample consists of students with an immigrant background, and 31 percent of the students live in a one-parent household. Almost a quarter of the students attend a combined primary and lower secondary school, average class size (weighted by numbers of students) is 24.8 and enrollment in seventh grade is 87.4. Teachers have on average 4.6 years of teacher training, 18.9 years of work experience as a teacher, 56 percent of the teachers are

Table 1. *Sample summary statistics*

	Mean (1)	SD (2)
Individual characteristics		
Girl	0.49	(0.50)
Age	14.53	(0.31)
ln(family income)	13.24	(0.76)
Education mother (years)	11.94	(2.90)
Education father (years)	12.23	(3.15)
First- or second-generation immigrants	0.05	(0.22)
Parents non-cohabiting	0.31	(0.46)
ln(pop. size school district)	10.11	(1.44)
ln(rural pop. size school district)	7.97	(0.74)
School characteristics		
Combined school	0.18	(0.38)
Average class-size grades 7–9	24.80	(3.24)
Enrollment grade 7	87.43	(40.13)
Teacher characteristics (*year t*)		
Average teacher education (years)	4.62	(0.19)
Average teacher experience	18.86	(3.24)
Fraction of female teachers	0.56	(0.11)
Fraction of teachers with a temp. contract	0.17	(0.11)
Year = 2002	0.52	0.50
N	111,463	
N schools	781	

women and 17 percent of the teachers are on a temporary contract (again all weighted by numbers of students).[11]

Class size is not distributed randomly in the population due to sorting of students and teachers, and the targeting of educational resources. This is illustrated by Table 2 which reports the results of regressions of actual average class size in lower secondary school on individual, teacher and school characteristics. The first column only includes students' characteristics and shows that actual class size increases with family income, parents' levels of education, immigrant background and non-cohabiting parents. When characteristics of the district are added to this specification, the positive effects of family income and parents' education remain significant but become smaller. The effect of having an immigrant background is reversed and the effect of living in a two-parent household is no longer significant (and also changes sign). These latter results indicate that immigrant families

[11] Student characteristics are measured at the time of testing, whereas school and teacher information refers to October 1 of the school year.

Table 2. *Regression of actual average class size in lower secondary school on observables, school years 2001/2002 and 2002/2003*

	(1)	(2)
Girl	0.001	0.004
	(0.024)	(0.022)
Age	−0.039	0.065
	(0.041)	(0.037)*
ln(family income)	0.296	0.112
	(0.036)***	(0.022)***
Education mother (years)	0.039	0.014
	(0.007)***	(0.006)**
Education father (years)	0.064	0.018
	(0.008)***	(0.005)***
First- or second-generation immigrants	0.859	−0.258
	(0.145)***	(0.122)**
Parents non-cohabiting	0.185	−0.007
	(0.037)***	(0.032)
ln(pop. size school district)		0.893
		(0.064)***
ln(rural pop. size school district)		0.202
		(0.131)
Year = 2002		0.105
		(0.130)
Adj. R-squared	0.015	0.173
N	111,463	111,463
N schools	781	781

Note: Standard errors are heteroskedasticity robust and corrected for school-level clustering. */**/*** statistically significant at the 10/5/1 percent level.

and separated parents are concentrated in more densely populated districts. The results in Table 2 indicate that class size is not randomly distributed across students. Students from more affluent families typically attend larger classes, thereby suggesting that class-size reduction is used as a compensatory policy. If we ignored selective placement in small and large classes, the true effect of class-size reduction on achievement would most likely be underestimated. Since these problems call for a strategy, we now turn to our approaches for addressing them.

V. Empirical Approaches

We follow the literature and assume that achievement of student i (y_i) is generated by the following equation:

$$y_i = x_i'\beta + w_{s(i)}'\alpha + \delta \cdot \overline{cs}_{s(i)} + \eta_{s(i)} + \psi_{t(i)} + \epsilon_i, \qquad (1)$$

where x_i is a vector of observable attributes of the student and his parents, $w_{s(i)}$ is a vector of observable school and teacher characteristics and $s(i)$

identifies the school of pupil i, $\overline{cs}_{s(i)}$ is the average class size that student i attended during her school career in school s, $\eta_{s(i)}$ is school effect, $\psi_{t(i)}$ is an effect for the year in which student i is in her final year of lower secondary school (2001/2002 or 2002/2003), and ϵ_i denotes all other determinants of achievement such as unobserved attributes of the student, parents and community.[12] The coefficient of interest is δ, the class-size effect. Note that a value-added specification is not feasible because achievement is only measured at the end of lower secondary school.

Conditioning on Observables

Table 3 shows the relation between student achievement and class size estimated using OLS. Results are presented from various specifications and separately for mathematics and languages, where we have pooled English and the two Norwegian subjects (and included subject dummies).[13] Columns (1) and (4) are obtained from a specification without covariates; both estimates are positive and significant, indicating that pupils in larger classes perform better than pupils in smaller classes. The results in columns (2) and (5) are obtained from a specification that includes individual characteristics as controls, while columns (3) and (6) report the results from specifications that also include school and teacher characteristics. We control for school-district characteristics in specifications (2), (3), (5) and (6). Including controls produces small negative but statistically insignificant estimates on class size in all specifications. If conditioning on a rather rich set of observables is sufficient to correct for biases of selective placement into (schools with) different class sizes, we can say that a one-pupil reduction in class size improves test scores by no more than 0.8 percent of a standard deviation with 95 percent probability.

However, not all selection into classes of different size needs to be on observables. To address any remaining endogeneity of class size, we need variation in actual average class size that is arguably not subject to the choices of parents and schools' principals or teachers. We exploit two sources of such exogenous variation, one induced by a maximum class-size rule and one based on population variation. In the remainder of this section we describe what these methods entail and how they can be implemented in the Norwegian context.

[12] In the estimations we allowed for clustering of this error term at the school level.

[13] We pooled the languages to reduce the number of outcomes to be repeated and discussed. Our findings are very similar, however, for the various languages separately.

Table 3. *OLS; dependent variable: exam results in ninth grade; unit of observations: individual student's exam result*

	Mathematics			Languages (pooled)		
	(1)	(2)	(3)	(4)	(5)	(6)
Average class-size grades 7–9	0.010	−0.002	−0.002	0.009	−0.001	−0.001
	(0.004)**	(0.003)	(0.003)	(0.003)***	(0.002)	(0.002)
Individual characteristics						
Girl		0.015	0.015		0.506	0.505
		(0.012)	(0.012)		(0.010)***	(0.010)***
Age (years)		0.069	0.070		0.031	0.030
		(0.020)***	(0.020)***		(0.015)**	(0.015)**
ln(family income)		0.179	0.180		0.143	0.143
		(0.009)***	(0.009)***		(0.008)***	(0.008)***
Education mother (years)		0.075	0.075		0.057	0.057
		(0.002)***	(0.002)***		(0.002)***	(0.002)***
Education father (years)		0.066	0.066		0.049	0.049
		(0.002)***	(0.002)***		(0.002)***	(0.002)***
First- or second-generation immigrants		−0.442	−0.434		−0.229	−0.224
		(0.032)***	(0.032)***		(0.025)***	(0.024)***
Parents have different address		−0.347	−0.345		−0.186	−0.185
		(0.013)***	(0.013)***		(0.009)***	(0.009)***

ln(pop. size school district)		0.015	0.025		0.009	0.013
		(0.008)*	(0.008)***		(0.005)	(0.006)**
ln(rural pop. size school district)		−0.019	−0.025		−0.014	−0.017
		(0.016)	(0.016)		(0.010)	(0.010)*
School characteristics						
Average teacher education (years)			0.035			0.061
			(0.057)			(0.043)
Average teacher experience			0.015			0.010
			(0.003)***			(0.002)***
Fraction of female teachers			−0.095			0.049
			(0.116)			(0.080)
Fraction of teachers with a temp. contract			0.072			0.022
			(0.090)			(0.067)
Combined school			0.049			0.024
			(0.028)*			(0.023)
R-squared	0.001	0.165	0.167	0.012	0.173	0.174
N pupils	36,915	36,915	36,915	74,548	74,548	74,548
N schools	608	608	608	752	752	752

Notes: Standard errors are heteroskedasticity robust and corrected for school-level clustering. */**/*** statistically significant at the 10/5/1 percent level. All regressions include a control for year of observation; the language regressions also include dummies for the language in which the student took the exam.

Maximum Class-size Rules

Lower secondary schools in Norway are subject to maximum class-size rules of 30 students. This rule creates a discontinuous relation between enrollment and class size. Just above multiples of 30, class size drops substantially. Following Angrist and Lavy (1999), we exploit this maximum class-size rule in a regression discontinuity design. For this approach to work, schools need to be located randomly around the thresholds and no other discontinuities that may affect outcomes should exist.

Identification in the regression discontinuity design is ultimately local, as in e.g. Hahn, Todd and Van der Klaauw (2001). However, Angrist and Lavy also proposed instrumenting actual class size, with predicted class size, while conditioning on a smooth function of enrollment which is supposed to capture any direct effect of this variable on achievement. This identifying assumption essentially boils down to an exclusion restriction with respect to the discontinuities. In the analyses we control for a cubic function of enrollment. An alternative to controlling for a smooth function of cohort enrollment is to restrict the sample to the regions around the kinks. We report separate results from analyses for a sample which is restricted to schools with cohort enrollment in grade 7 at most five students away from the kinks and refer to this regression discontinuity sample as $\text{DS} \pm 5$.

Our analysis differs from previous work in that we use predicted class size in grade 7, the first year of lower secondary school, to instrument average class size during the three years of lower secondary school. The underlying reason is that cohort enrollment in grades 8 and 9, and thereby predicted class size in these grades, may depend on actual class size in grade 7, and is therefore potentially endogenous. Such dependence could, for instance, result from parents' decisions to move from schools with large classes in seventh grade to schools where they observed small classes in seventh grade.[14] Although we reported evidence above which suggests that student mobility during lower secondary education is limited, our estimates will not be affected by selective mobility if it is orthogonal to our instrument.

Figure 2 plots predicted class size and average actual class size against cohort enrollment. Average actual class size closely tracks predicted class size especially around the first kink. Table 4 reports the results from the corresponding first-stage regressions, for specifications with different sets of control variables, including polynomials of enrollment. The first column presents these results for the entire sample of schools. The estimates in

[14] Students in combined schools enroll in lower secondary school in grade 7. This may depend on actual class size during the previous primary school period. But since the maximum class-size rule changes from 28 in grade 6 to 30 in grade 7 causes, students in grade 7 in the combined schools are also confronted with a new class size.

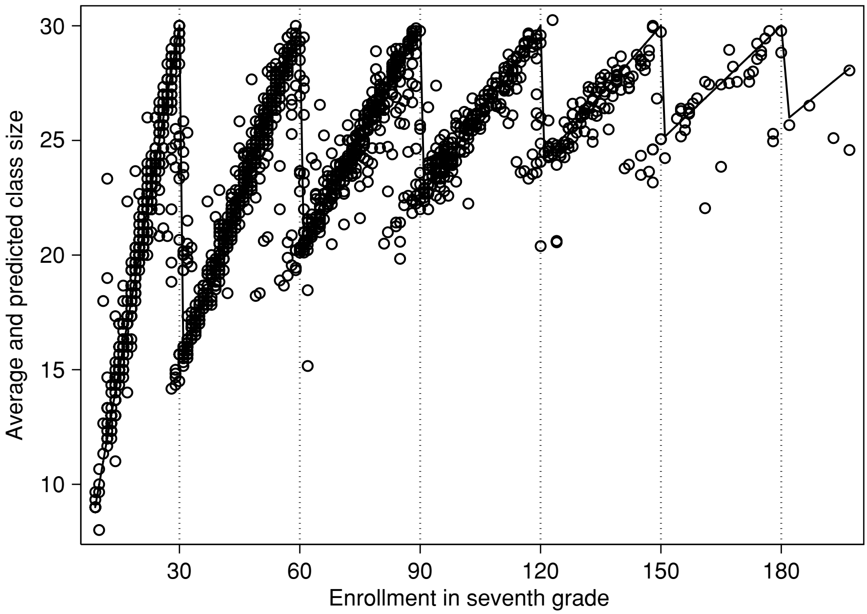

Fig. 2. Average and predicted class size, lower secondary school

the second column are only based on the restricted sample of schools with enrollment levels at most five students away from a multiple of 30 (hence, between 25 and 35, between 55 and 65, etc.). These results show that the predicted class size is a strong instrument, the F-values are equal to 454 and 193. Note, however, that the coefficients are between 0.52 and 0.63 and are thus smaller than 1, which would result if schools would perfectly adhere to the rule. But they are substantially higher than the first-stage estimates in Angrist and Lavy (1999) which are between 0.346 and 0.542, when they control linearly for enrollment.

As in any regression discontinuity design, one needs to ensure that the exclusion restriction is not violated. One way of testing this is to check that schools and/or parents do not sort around the cut-offs. We do not observe bunching after the cut-offs in Norway. To go one step further, we also compared the characteristics of students and schools around the kinks. A proper regression discontinuity design is like a local randomized experiment and observed characteristics should therefore be balanced. Although this is a necessary and not a sufficient condition (which also requires balancing of unobserved characteristics), evidence for balancing sends strong support since it seems difficult to imagine unobservables that matter for outcomes but that are orthogonal to observables that affect outcomes.

Table 4. *The impact of predicted class size according to the maximum class-size rule on average actual class size*

	Full sample		DS ± 5	
	Coef.	SE	Coef.	SE
Predicted class size in grade 7	0.627	(0.029)***	0.532	(0.038)***
Enrollment/10	0.802	(0.181)***	0.416	(0.540)
$(\text{Enrollment}/10)^2$	−0.064	(0.020)***	−0.040	(0.056)
$(\text{Enrollment}/10)^3$	0.002	(0.001)**	0.001	(0.002)
Individual characteristics				
Girl	−0.032	(0.015)**	−0.079	(0.038)**
Age	0.031	(0.024)	0.073	(0.054)
ln(family income)	0.021	(0.014)	0.020	(0.031)
Education mother (years)	0.002	(0.004)	0.015	(0.008)*
Education father (years)	−0.001	(0.003)	−0.005	(0.007)
First- or second-generation immigrants	−0.003	(0.056)	0.031	(0.140)
Parents non-cohabiting	−0.041	(0.023)*	−0.100	(0.054)*
ln(pop. size school district)	0.147	(0.056)***	0.552	(0.141)***
ln(rural pop. size school district)	0.093	(0.082)	0.206	(0.160)
Teacher/School characteristics				
Average teacher education (years)	0.091	(0.360)	−0.211	(0.798)
Average teacher experience	0.050	(0.022)**	0.129	(0.052)**
Fraction of female teachers	−0.385	(0.638)	−0.515	(1.677)
Fraction of teachers with a temp. contract	0.817	(0.495)*	3.091	(1.329)**
Combined school	0.316	(0.184)*	0.353	(0.530)
R-squared	0.653		0.513	
N	111,463		34,780	
N schools	781		378	
F-statistic	454.4		192.5	

Notes: Standard errors are heteroskedasticity robust and corrected for school-level clustering. */**/*** statistically significant at the 10/5/1 percent level.

To test this, we restricted the sample to students in classes at most five students away from the cut-offs and regressed the indicator for being above the cut-off (versus being below it) on various sets of observable characteristics (including enrollment). In the specifications that regress the above/below indicator on student characteristics and school characteristics, the p-values for joint significance of these characteristics are 0.443 and 0.495, respectively. The p-value for joint significance of student and school characteristics together is 0.518. The only separate variable that comes in marginally significant (10 percent level) is teachers' education (lower above the kinks).

Population Variation

The second approach exploits demographic variation and was first proposed by Hoxby (2000). Instead of using the variation in enrollment that—in the presence of a maximum class-size rule—triggers changes in the number of classes (and therefore class size), Hoxby exploits the population variation conditional on the number of classes.

Variation in enrollment may, however, correlate with other determinants of student achievement. This occurs, for example, when more educated parents avoid schools where enrollment is large, or when better schools face increased demand if parents selectively choose schools based on school quality. Where part of the variation in enrollment depends on (variation in) population characteristics, there is also a random component (u) that arises from random fluctuations in timing and number of births. Since it seems natural that the share of the random component in total births does not depend on population size, it is assumed that u affects e proportionally, whereby enrollment for a given school (and grade level) can be expressed by the following equation:

$$\log(e_{st}) = \log(\bar{e}_{st}) + \log(u_{st}), \tag{2}$$

where $\bar{e}_{st}$ is the deterministic part of enrollment, and $\log(u_{st})$ the i.i.d. part which captures the random variation in enrollment caused by idiosyncratic factors such as biology.

If $\log(u_{st})$ is not correlated with any of the determinants for student achievement (x_{ist}, w_{st} and ϵ_{ist}) in equation (1), a consistent estimate of $\log(u_{st})$ would be a valid instrument for class size since $\log(u_{st})$ is correlated with $\log(e_{st})$. Hoxby assumes that $\log(\bar{e}_{st})$ changes smoothly over time and can be approximated by a grade-school-specific intercept and a school-specific polynomial in time. In this case equation (2) can be written as:

$$\log(e_{st}) = \sum_{k=0}^{K} \alpha_{sk} t^k + \log(u_{st}). \tag{3}$$

To investigate the dynamics of enrollment we calculated autocorrelations of $\log(e_{st})$. These were always higher than 0.95 over the period for which we have enrollment data (1992–2002). Such high persistence suggests that there is indeed at least a school-specific intercept α_{s0}. As a next step we calculated the autocorrelation matrix for relative enrollment growth $\Delta\log(e_{st})$, which is reported in Table 5. It is clear from this table that after first-differencing all persistence is gone. The first off-diagonal elements are approximately -0.5 and the others are close to zero. Note that if there would be a school-specific linear trend, first-differencing would leave a school-specific effect and the high persistence would remain. Table 5 is therefore consistent with the following data-generating process:

$$\log(e_{st}) = \alpha_{s0} + \log(u_{st}),$$

which implies that $K = 0$ in equation (3).

Table 5. *Autocorrelation matrix of relative changes in school enrollment* ($\Delta\log(e_{st})$)

	Correlation($\Delta\log e_{s,n}$, $\Delta\log e_{s,m}$)							
$n=$	t	$t-1$	$t-2$	$t-3$	$t-4$	$t-5$	$t-6$	$t-7$
$m=$								
t	1.000							
$t-1$	−0.459	1.000						
$t-2$	−0.007	−0.464	1.000					
$t-3$	0.007	−0.011	−0.462	1.000				
$t-4$	0.040	0.013	−0.004	−0.483	1.000			
$t-5$	−0.053	0.018	0.028	0.021	−0.505	1.000		
$t-6$	0.042	−0.028	0.001	0.005	0.035	−0.504	1.000	
$t-7$	−0.031	0.039	−0.015	0.003	−0.014	0.038	−0.496	1.000

Table 6. *First-stage estimates in population variation approach: alternative instruments*

	$K=4$			$K=0$		
	Coef.	SE	F-stat.	Coef.	SE	F-stat.
Instrument						
$t-5$	−0.028	(0.030)	0.8	−0.041	(0.027)	2.3
$t-4$	0.011	(0.027)	0.2	−0.002	(0.027)	<0.1
$t-3$	0.026	(0.028)	0.9	0.022	(0.026)	0.7
$t-2$	0.003	(0.032)	<0.1	0.012	(0.031)	0.2
$t-1$	−0.321	(0.041)	61.1	−0.252	(0.039)	41.1
t	0.714	(0.036)	385.8	0.760	(0.033)	516.2
$t+1$	−0.265	(0.035)	58.3	−0.222	(0.033)	44.7
$t+2$	0.047	(0.042)	1.3	0.021	(0.031)	<0.1

Following Hoxby, we estimated equation (3) for each school separately to obtain the estimated residuals for $\widehat{\log(u_{st})}$ for both $K=0$ and $K=4$, which can serve as the instrument for class size in (1). As in the maximum class-size approach, we base our instruments on enrollment in grade 7. Table 6 reports the coefficients from first-stage regressions where we instrumented cohort's t enrollment with its own residual (t), but also with the enrollment residuals from previous ($t-1$, $t-2$, ...) and subsequent cohorts ($t+1$, $t+2$). For both specifications of the enrollment process, the contemporaneous residual is highly significant and has the expected sign; positive enrollment shocks increase class size. The residuals from the adjacent cohorts ($t \pm 1$) are also significant, but to a much smaller extent. The sign on these residuals is negative. This suggests that $\log(u_{st})$ is a first-order moving average. This is consistent with the biological variation motivation behind this approach: if women give birth in year t they are less likely to have another birth in year $t+1$.[15]

Even if $\log(e_{st})$ is correctly modeled, the validity of the population variation approach depends on the exclusion restriction

$$E[\epsilon_{ist} \cdot \widehat{\log(u_{st})}|\eta_s] = 0,$$

which is implemented in a fixed-effects 2SLS procedure which eliminates school-fixed effects. As a tentative test of the validity of the exclusion restriction, we regressed the estimated residuals on the individual and school observables in Table 3. There is no indication that our instruments correlate with observed determinants of students' achievements when we calculate the F-test of joint significance on these regressors for the models $K=4$ and $K=0$. The F-statistic for the former is 1.145 with a p-value of 0.315, and for the latter 1.167 with a $p\text{-value}=0.296$. The results are based on the model with the quartic trend ($K=4$) for the sake of comparability with previous studies. There is substantial variation in the instrument. It is on average −0.0045 with a standard deviation of 0.19, symmetric around zero and the 1st, 5th, 10th and 25th percentiles are −0.6, −0.3, −0.2 and −0.09, respectively.

Note that this approach is valid as long as the variation in enrollment does not trigger a change in the number of classes from 2001/2002 to 2002/2003 since this would violate monotonicity of the instrument. For this reason we only included schools where the predicted number of classes are the same in 2001/2002 and 2002/2003. Moreover, to take into account any remaining endogeneity—which would occur when parents transfer their children to other schools in response to the class size their children are experiencing—we follow Hoxby and carry out the analysis at the district level in order to cancel out within-school-district transfers. Other sources of bias such as transfers to private schools or selective grade retention or advancement are not relevant for Norway due to the negligible presence of private schools and the absence of grade retention/advancement practices.

Variations of Population Variation

A number of authors have used approaches that build on or combine the methods discussed above. We also report results based on these approaches and therefore describe them below.

Population Variation and Maximum Class-size Rules. In addition to exploiting population variation while conditioning on the number of classes,

[15] If $\log(u_{st})$ is a first-order moving average, i.e., $\log(u_{st}) = \vartheta_{st} - \rho\vartheta_{s,t-1}$, then this does not invalidate the population variation approach as long as $E[\epsilon_{ist}\vartheta_{s,t-1}|\eta_s]=0$. We make this assumption in addition to the contemporaneous exclusion restriction $E[\epsilon_{ist}\vartheta_{s,t}|\eta_s]=0$.

Hoxby (2000) also uses the population variation that triggers changes in the number of classes (and therefore class size) because of maximum class-size rules. Hoxby identifies all events where the number of classes changed due to maximum class-size rules. For all cases where the enrollment change was not more than 20 percent, she then estimates a first-differenced version of equation (1).

Small Schools. Urquiola (2006) focuses on small schools with only one predicted class (per grade). The idea here is that the endogeneity of class size is less of a problem in small rural schools since they are local monopolies and parents cannot enroll their children elsewhere, thereby ruling out between-school sorting. In addition, between-class sorting is not an issue since these small schools have only one class. As noted by Urquiola, class size may still correlate with unobserved factors that affect achievement since class size depends on community size, fertility, etc. All class-size variation is therefore generated by differences in enrollment. This approach is like Hoxby's, but does not control for school-fixed effects since it exploits cross-sectional variation, nor does it control for trends. The implementation is straightforward and amounts to estimating equation (1) by OLS, where schools with more than one predicted class per grade are excluded. Furthermore, the sample is restricted to districts with only one school. This strategy is suitable for the Norwegian context with rural settlement patterns and many small schools.

Within-school Between-grade Comparison. Wößmann and West (2006) use within-school and between-grade variation to estimate class-size effects. By comparing adjacent grades, they account for between-school sorting since they eliminate school-fixed effects (to the extent to which they are uniform across grades). To eliminate within-school sorting problems, actual class size is instrumented with average class size at the grade level. The reduced-form estimates are therefore equivalent to the population variation approach of Hoxby, but without correcting for a trend, and without discarding schools around the kinks. We implement a version of this approach where we exploit variation in enrollment between cohorts in the same school instead of variation between grades. In the next section we refer to this approach as "population variation without trend".[16]

[16] The reason for this deviation from Wößmann and West's approach is that we only have outcomes measured in the final grade of lower secondary school, whereas they have outcome data for two adjacent grades. Hence, their possible bias due to differences between grades is replaced by a possible bias due to differences across cohorts.

VI. Results

Estimates of class-size effects based on the different approaches are shown in Table 7.[17] These results are all based on specifications with the full set of control variables (e.g. the same control variables as those included in the OLS regression, and—for the estimates based on the maximum class-size rule—a third-order polynomial in enrollment). Less elaborate specifications produce very similar results (and are available from the authors on request). Moreover, for students in combined schools, we also know their class size during grades 1–6, assuming that students do not switch between combined schools and separate schools. We also conducted analyses using average class size during the six years in primary school as a class size measure. As mentioned above, we do not present these results here since they are very similar to those for average class size in lower secondary schools.[18]

The first row in Table 7 repeats the OLS results from the analysis of selection on observables in Section V. These estimates are negative, but small and not significant at conventional levels. At face value, these estimates would imply a 1–2 percent standard deviation improvement in achievement for a reduction in class size by 10 pupils. Out of the other 14 effect estimates in Table 7, only the two estimates based on the population variation approach at the district level have the expected negative sign. None of the effect estimates is significantly different from zero although almost all effects are estimated quite precisely, so that substantial negative effects of class size on achievement can be ruled out with high probability. In square brackets below each effect estimate we report the largest (most negative) effect that falls into the 95 percent confidence interval (point estimate −1.96* SE). With the exception of the population variation approach the estimates (including those in the first row) imply maximum effects of 1 percent of a standard deviation or less.

The results in Table 7 reveal no clear pattern in the estimated effect sizes across the different methods, apart from the fact that those which exploit population variation (rows 4 and 5) have substantially larger standard errors than the other methods, especially when applied to math achievement. The pooled estimates for the population variation approach shown in column (1) in Table 8 are more precise and equal −0.002 (0.006) for the pooled sample; at the district level (not reported here) it is −0.008 (0.009). These estimates imply lower bounds of −0.013 and −0.026, respectively.

---

[17] The first stages are always quite strong, with values from F-tests for significance of the instrument at least equal to 102.8.

[18] There may be some concern that the results for average class size in lower secondary schools pick up effects of average class size in primary schools. This is muted, however, by the fact that the maximum class-size rule in primary schools is different from that in lower secondary schools (28 versus 30).

Table 7. *Estimated class-size effects using different methods*

	Mathematics			Language		
	Effect	SE	N (N_S)	Effect	SE	N (N_S)
(1) Conditional on observables	−0.002 [−0.008]	(0.003)	36,915 (608)	−0.001 [−0.005]	(0.002)	74,548 (752)
(2) Maximum class size	0.002 [−0.009]	(0.006)	36,915 (608)	0.008 [−0.0004]	(0.004)*	74,548 (752)
(3) Maximum class-size discontinuity sample	0.006 [−0.011]	(0.009)	10,810 (190)	0.007 [−0.005]	(0.006)	23,970 (280)
(4) Population variation	0.016 [−0.031]	(0.024)	26,130 (424)	0.0004 [−0.015]	(0.008)	51,272 (549)
(5) Population variation at district level	−0.020 [−0.127]	(0.055)	240 (120)	−0.009 [−0.036]	(0.014)	420 (210)
(6) Maximum class size + population variation	0.095 [0.074]	(0.011)***	4,962 (67)	0.013 [−0.002]	(0.008)*	11,890 (84)
(7) Small monopoly schools (enrollment <25)	0.026 [0.0004]	(0.013)*	717 (42)	0.005 [−0.018]	(0.011)	1,636 (56)
(8) Population variation without trend	0.010 [−0.009]	(0.009)	36,915 (608)	0.0004 [−0.005]	(0.003)	74,548 (752)

Notes: Standard errors are heteroskedasticity robust and corrected for school-level clustering. */**/*** statistically significant at the 10/5/1 percent level. Square brackets indicate the upper bound of the 95 percent confidence interval. All coefficients come from different regressions. All regressions include controls for gender, age, ln(family income), education mother, education father, a dummy for immigrants, a dummy for parents living at different addresses, ln(population size of the school district), ln(rural population size in school district), average teacher education, average teacher experience, fraction of female teachers, fraction of teachers with a temporary contract, a dummy for combined schools, a dummy for year of observation. The language regressions also include dummies for the language in which the student took the exam. The regressions using the maximum class-size rule also include a third-degree polynomial of enrollment as control. The results in rows (1) and (7) are obtained using OLS; (8) with a fixed-effects regression; the other results are obtained using IV.

Table 8. *Class-size estimates based on population variation: alternative instrument sets and pooled estimates*

		Instrument				Hansen's J-test	
		t (1)	$t-1$ (2)	$t+1$ (3)	All (4)	(5)	
Math	b	0.016	0.042	−1.147	0.021	2.184	$p=0.336$
	SE	(0.024)	(0.034)	(13.453)	(0.023)		
	F	17.5	14.4	<0.1	10.3		
Language	b	0.0004	0.0002	0.008	0.001	0.255	$p=0.880$
	SE	(0.008)	(0.013)	(0.017)	(0.008)		
	F	85.1	29.8	13.8	29.8		
Pooled	b	−0.002	0.003	−0.022	−0.003	3.232	$p=0.198$
	SE	(0.006)	(0.010)	(0.013)*	(0.006)		
	F	385.8	61.1	58.3	129.8		

Note: See note to Table 7.

Columns (2)–(4) of Table 8 also show class-size estimates based on alternative instrument sets. Column (2) uses the predicted residuals from the previous (older) cohort (the first-stage regression coefficient was reported in Table 6), column (3) uses the instrument based on the following (younger) cohort and column (4) uses all three instruments. In terms of the validity of this identification approach, it is reassuring to see that we get similar point estimates when we exploit our additional instruments. This is confirmed by the Hansen J-tests (reported in the last column) which never reject the overidentifying restrictions.

Figure 3 illustrates the findings for the method based on the maximum class-size rule. The solid line plots predicted class size as a function of enrollment (in intervals of 10 pupils), while the dashed lines show average math and language achievement as a function of enrollment. If smaller class size would benefit pupil achievement, we would observe a jump in achievement around the discontinuities where class size drops. But in no way does achievement follow (mirror) the pattern of class size, as it does in Angrist and Lavy (1999).

In addition to the sorting, for which we do not find evidence above, it has been argued that maximum class-size rules may be inappropriate for estimating the impact of class size on achievement owing to potential input substitution around the discontinuities. To investigate whether this could be an issue in Norway, we replaced our class-size measure by teacher hours per student, measured as the total number of teacher hours in a grade (including extra education for students with specific needs) divided by the number of students in that grade. Although this does not cover resources such as school supplies, computers, teacher assistants, etc., it is the most important measure of resource use in Norway; cf. e.g. Hægeland, Raaum

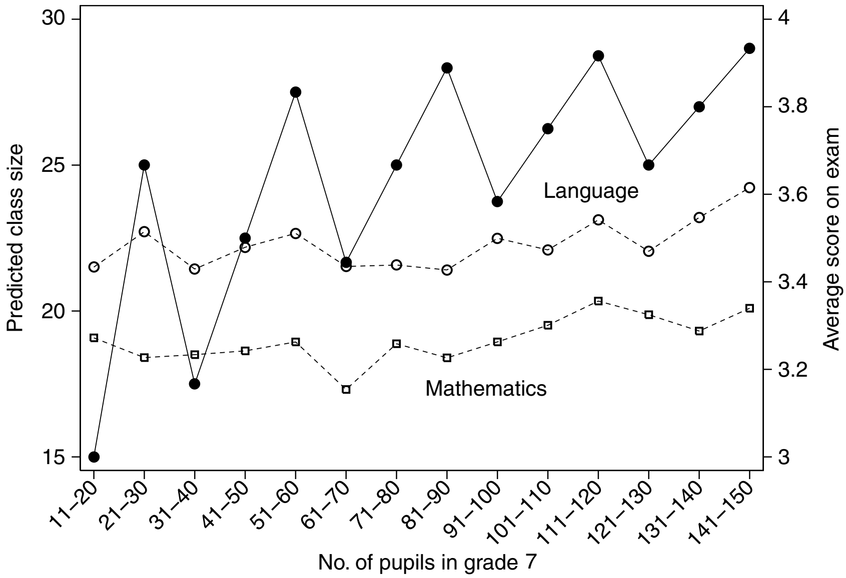

Fig. 3. Predicted class size and student achievement

Table 9. *The impact of school resources (teacher hours/pupil) on achievement: 2SLS estimates based on maximum class-size rules*

		Second stage	
	First stage (1)	Mathematics (2)	Language (3)
Full sample	−0.414	−0.004	−0.012
	(0.084)***	(0.013)	(0.006)**
	111,051 (777)	36,698 (605)	74,353 (749)
DS ± 5	−0.410	−0.012	−0.009
	(0.106)***	(0.020)	(0.007)
	34,669 (377)	10,701 (189)	23,968 (279)

Note: See note to Table 7.

and Salvanes (2005, 2007). The correlation between teacher hours per pupil and our class-size measure is −0.8.

Table 9 shows results from 2SLS estimations based on maximum class-size rules as in Table 7. Note that if input substitution is complete, our instrument would not affect teacher hours per pupil. As can be seen from the first column in Table 9, when predicted class size increases by one, teacher hours per pupil decrease by 0.41. Although there might be some input substitution, it is clear from these results that it is far from being complete. Columns (2) and (3) in Table 9 report impact estimates of teacher hours on mathematics and language achievement. The effects are in line with those in Table 7; they are highly insignificant and tend to have the wrong sign.

VII. Heterogeneous Class-size Effects

We now investigate whether the zero class-size effects reported above mask effects of class-size reduction on achievement for specific sub-groups. The reason for exploring this is that various studies report more pronounced effects of class-size reduction for disadvantaged groups, as in Angrist and Lavy (1999) and Krueger (1999), or dependence on teacher characteristics, as in Wößmann and West (2006).

Table 10 presents results for specific groups based on (i) OLS (conditioning on observables), (ii) the method that exploits maximum class-size rules (for the full sample and around kinks) and (iii) the method that exploits population variation (and includes a trend). The first sub-group is that of pupils with low educated mothers (less than 10 years of education). The OLS are small and negative. Results from the maximum class-size rule using the full sample show positive point estimates of the effect of class size on achievement in both language and math. The sign of the estimates obtained from the discontinuity sample and from the population variation method do, however, differ by subject matter. None of these estimates is statistically significant and there is therefore no evidence in favor of beneficial class-size effects for students with low educated mothers.

As a second sub-group, we consider children from immigrant families (their share in the population is only 5 percent). All estimates are positive and two are significantly different from zero at the 5 percent level.

The results for girls—reported in the third block—are all very small, and do not differ substantially or significantly from the results for girls and boys together. The fourth part of the table presents results for pupils in the lowest quartile of the age distribution. Here again, the emerging pattern is far from consistent. For math achievement, the population variation approach gives in a significantly positive estimate, whereas the point estimate for language is basically zero.

The final two parts of the table show effects for pupils in schools where the teaching staff belongs to the lowest quartile in terms of education level or experience. Again the different methods fail to reveal a consistent pattern. Point estimates are all negative with respect to mathematics achievement for pupils in schools with low educated teachers. This result concurs with the hypothesis advanced by Wößmann and West (2006). However, this finding is not confirmed by the estimates on language achievement. We do not see any indication of negative class-size effects in schools with less experienced teachers. Point estimates are almost all positive and rather small.

Table 10. *Class-size effects for sub-groups using different methods*

	Math			Language		
	Effect	SE	N (N_S)	Effect	SE	N (N_S)
Low educated mothers						
Conditional on observables	−0.007	(0.006)	3,833 (537)	−0.005	(0.005)	7,936 (715)
Maximum class size	0.008	(0.010)	3,833 (537)	0.0005	(0.008)	7,936 (715)
Maximum class-size discontinuity sample	0.023	(0.015)	1,112 (165)	−0.006	(0.011)	2,474 (267)
Population variation	0.092	(0.052)	2,711 (356)	−0.020	(0.018)	5,516 (493)
Immigrants						
Conditional on observables	0.001	(0.011)	2,000 (375)	0.017	(0.008)**	3,628 (493)
Maximum class size	0.027	(0.020)	2,000 (375)	0.012	(0.014)	3,628 (493)
Maximum class-size discontinuity sample	0.077	(0.040)*	642 (121)	0.016	(0.017)	1,168 (198)
Population variation	0.214	(0.202)	1,323 (190)	0.006	(0.038)	2,425 (259)
Girls						
Conditional on observables	−0.003	(0.004)	17,983 (582)	0.001	(0.003)	36,864 (741)
Maximum class size	0.005	(0.007)	17,983 (582)	0.008	(0.005)*	36,864 (741)
Maximum class-size discontinuity sample	0.015	(0.011)	5,283 (175)	0.006	(0.007)	11,784 (272)
Population variation	0.029	(0.027)	12,719 (420)	0.012	(0.010)	25,299 (541)

Youngest quartile						
Conditional on observables	−0.006	(0.005)	9,262 (576)	−0.003	(0.004)	18,602 (737)
Maximum class size	−0.003	(0.008)	9,262 (576)	0.007	(0.006)	18,602 (737)
Maximum class-size discontinuity sample	0.004	(0.014)	2,722 (171)	0.006	(0.009)	6,082 (270)
Population variation	0.101	(0.039)	6,519 (410)	−0.002	(0.013)	12,746 (534)
Teacher education in lowest quartile						
Conditional on observables	−0.003	(0.007)	6,020 (151)	−0.005	(0.004)	13,636 (202)
Maximum class size	−0.005	(0.010)	6,020 (151)	0.005	(0.007)	13,636 (202)
Maximum class-size discontinuity sample	−0.008	(0.083)	1,458 (37)	0.022	(0.011)*	4,382 (76)
Population variation	−0.123	(0.088)	4,498 (108)	−0.016	(0.020)	10,475 (156)
Teacher experience in lowest quartile						
Conditional on observables	0.001	(0.007)	9,416 (170)	−0.002	(0.004)	20,282 (206)
Maximum class size	0.005	(0.012)	9,416 (170)	0.010	(0.008)	20,282 (206)
Maximum class-size discontinuity sample	0.014	(0.024)	2,577 (52)	0.009	(0.010)	6,779 (84)
Population variation	0.158	(0.182)	7,028 (119)	0.014	(0.013)	14,077 (146)

Note: See note to Table 7.

We are inclined to conclude that there is no evidence in favor of heterogeneous class-size effects. If anything, pupils from low educated families and pupils in schools with low educated teachers benefit the most from a reduction in class size.

VIII. Conclusion

Based on estimation results that exploit arguably exogenous variation in class size, we find no significant effect of class size during lower secondary school on achievement in grade 9 in Norway. Depending on the identification approach used, we can exclude effects as small as 1–1.5 percent of a standard deviation for a one-student reduction in average class size during three consecutive years. Effects are rather similar for different social-background groups and for schools with different teaching staffs.

Our findings contrast sharply with most of the recent studies that apply experimental and quasi-experimental methods to estimate the class-size effect. Interestingly, while we applied the same identification strategy as Angrist and Lavy (1999) did in their study for Israel, the findings are very different. We interpret this as evidence that there is no such thing as a universal class-size effect.

Potential explanations for the negligible class-size effect in Norway are substitution of parental inputs and uniform teaching styles. Substitution of parental inputs occurs if the parents of pupils who are placed in small classes reduce their own inputs in the education production function; cf. Todd and Wolpin (2003). Uniform teaching styles annihilate potentially beneficial class-size effects if teachers are unable to take advantage of the extra time they could devote per student. However, the fact that we find some indication that pupils of low educated teachers benefit more from class-size reduction than pupils of high educated teachers seems to contradict this explanation. Further research is required to differentiate between these various explanations. This is important because the policy implications are quite different. If the zero effects are due to substitution of parental inputs, there is not much hope that the policy effects can be improved, although the reduction of parental inputs should be included in a cost–benefit analysis. If the zero effects are due to uniformity in teaching styles, there remains scope for improvement by educating teachers to take advantage of smaller classes.

References

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

Bingley, P., Jense, V. M. and Walker, I. (2005), The Effects of Class Size on Educational Attainment: Danish Quasi-experimental Evidence and Evidence that Controls for Family, School and Neighbourhood Effects, manuscript, Aarhus School of Business.

Bonesrønning, H. (2003), Class Size Effects on Student Achievement in Norway: Patterns and Explanations, *Southern Economic Journal 69*, 952–965.

Boozer, M. and Rouse, C. (2001), Intraschool Variation in Class Size: Patterns and Implications, *Journal of Urban Economics 50*, 163–189.

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

Gary-Bobo, R. J. and Mahjoub, M. B. (2006), Estimation of Class-size Effects, Using "Maimonides' Rule": The Case of French Junior High Schools, CEPR Discussion Paper no. 5754, London.

Hægeland, T., Raaum, O. and Salvanes, K. G. (2005), Pupil Achievement, School Resources and Family Background, IZA Discussion Paper no. 1459, Bonn.

Hægeland, T., Raaum, O. and Salvanes, K. G. (2007), Pennies from Heaven: Using Exogenous Tax Variation to Identify Effects of School Resources on Pupil Achievement, IZA Discussion Paper no. 3561, Bonn.

Hahn, J., Todd, P. and Van Der Klaauw, W. (2001), Identification and Estimation of Treatment Effects with a Regression-discontinuity Design, *Econometrica 69*, 201–209.

Häkkinen, I., Kirjavainen, T. and Uusitalo, R. (2003), School Resources and Student Achievement Revisited: New Evidence from Panel Data, *Economics of Education Review 22*, 329–335.

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

Hanushek, E. A. (2003), The Failure of Input-based Schooling Policies, *Economic Journal 113*, F64–F98.

Hattie, J. (2005), The Paradox of Reducing Class Size and Improving Learning Outcomes, *International Journal of Educational Research 43*, 387–425.

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

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

Krueger, A. B. (2003), Economic Considerations and Class Size, *Economic Journal 113*, F34–F63.

Krueger, A. B. and Whitmore, D. M. (2001), The Effect of Attending a Small Class in the Early Grades on College-test Taking and Middle School Test Results: Evidence from Project Star, *Economic Journal 111*, 1–28.

Lindahl, M. (2005), Home versus School Learning: A New Approach to Estimating the Effect of Class Size on Achievement, *Scandinavian Journal of Economics 107*, 375–394.

Piketty, T. (2004), L'impact de la taille des classes et de la ségrégation sociale sur la réussite scolaire dans les écoles françaises: une estimation à partir du panel primaire 1997, manuscript, EHESS, Paris.

Strøm, B. (2004), Student Achievement and Birthday Effects, manuscript, Norwegian University of Science and Technology, Trondheim.

Todd, P. and Wolpin, K. I. (2003), Towards a Unified Approach for Modeling the Production Function for Cognitive Achievement, *Economic Journal*, Features: F3–F33.

Urquiola, M. (2006), Identifying Class Size Effects in Developing Countries: Evidence from Rural Bolivia, *Review of Economics and Statistics 88*, 171–176.

Wößmann, L. (2005), Educational Production in Europe, *Economic Policy 20*, 445–504.

Wößmann, L. and West, M. (2006), Class-size Effects in School Systems around the World: Evidence from Between-grade Variation in TIMSS, *European Economic Review 50*, 695–736.