

WORKING PAPER SERIES

No. 3/2015

Do smaller classes always improve students' long run outcomes?

Torberg Falch and Bjarne Strøm
Department of Economics
Norwegian University of Science and Technology

Astrid Marie Jorde Sandsør
Department of Economics
University of Oslo

Department of Economics

 Norwegian University of Science and Technology

N-7491 Trondheim, Norway

<http://www.ntnu.edu/econ/working-papers>

Do smaller classes always improve students' long run outcomes?

Torberg Falch

Department of Economics
Norwegian University of
Science and Technology

Astrid Marie Jorde Sandsør

Department of Economics
University of Oslo

Bjarne Strøm

Department of Economics
Norwegian University of
Science and Technology

This version: July, 2015

Abstract:

We exploit the strict class size rule in Norway and matched individual and school register information for 1982–2011 to estimate long run causal effects on income and educational attainment. Contrary to recent evidence from the US and Sweden, we do not find any significant average effect on long run outcomes of reduced class size. We further use the large register data set and quasi-experimental strategy to estimate whether the class size effect depends on external conditions facing students and schools, such as teacher quality, extent of upper secondary school choice, school district size, local fiscal constraints and labor market conditions. Overall, we find that the class size effect does not depend on school district characteristics. The absence of class size effects on long run outcomes in Norway is consistent with earlier findings for short run outcomes using comparable data and empirical strategies.

Keywords: class size, school district, quasi-experiment, educational attainment, income

JEL-codes: I2, H7

1. Introduction

The impact of school resources on student performance has been disputed since the publication of the Coleman (1966) report. Although availability of data and empirical strategies to uncover causal effects have increased substantially in recent years, the evidence on the effect of resources on education outcomes is still inconclusive.¹ The literature is not conclusive even for more narrow and popular policy tools as class size. Although the results from the large famous randomized experiment in Tennessee (STAR) suggest that smaller classes are beneficial in terms of test scores,² studies using quasi-experimental approaches to identify causal effects differ substantially in their conclusions.³ One interpretation is that extra resources and reduced class size are effective tools in some contexts, while ineffective in other contexts.

This motivates the present study using rich register data from Norway from a long time period combined with a quasi-experimental empirical strategy to estimate both the average effect of class size and to which extent the effect varies with the external conditions facing schools and students. Test scores only measure cognitive skills, while class size may to some extent also affect non-cognitive skills. In addition, evidence based on test scores may be biased in settings where teachers systematically manipulate test scores as recently demonstrated in Angrist et al. (2015).⁴ Both arguments suggest that analyses of long run outcomes in terms of educational attainment and income in adulthood as used in our empirical study would provide the most credible evidence of the effect of school resources. Such studies will embed all short-run effects, including effects on non-cognitive skills that are difficult to measure directly.

¹ Summaries of the literature on the relationship between school resources and student achievement include Hanushek (1986, 2003, 2006), Krueger (2003) and Webbink (2005).

² See Krueger and Withmore (2001) and Chetty et. al. (2011) on evidence from the STAR experiment. In contrast to the STAR experiment, field experiments on class size conducted before WW II provided little evidence in support of the hypothesis that smaller classes increase student achievement, see Rockoff (2009) for an interesting review of these early field experiments.

³ The seminal paper by Angrist and Lavy (1999) initiated a literature exploiting class size rules in a regression discontinuity framework, Hoxby (2000) uses idiosyncratic variation in cohort size, and West and Woessmann (2006) employ a within-school across-classes strategy. While Angrist and Lavy (1999) find the expected negative effect of class size on student achievement for Israel, Hoxby (2000) and West and Woessmann (2006) find zero effects in Connecticut and for most OECD countries, respectively. In a recent paper, Denny and Oppedisano (2013) even find positive effects for the US and the UK. They use the same empirical strategy as West and Woessmann (2006) in addition to an approach based on restrictions on higher order moments.

⁴ Angrist et al (2015) exploit a class size rule in Italy and find a strong negative relationship between test scores and class size in Southern Italy. This relationship is, however, entirely driven by manipulation of the test scores by the teachers.

Three recent papers analyze long run effects of class size. Chetty et al. (2011) and Dynarski et al. (2013) study long run outcomes for participants in the STAR experiment, while Fredriksson et al. (2013) exploit a class size rule in Sweden to estimate both short run and long run outcomes. These papers all find positive long run effects of smaller classes, consistent with evidence in these contexts for short run outcomes using test scores. Of particular interest is Fredriksson et al. (2013) who find a positive short run effect on non-cognitive ability, which is an outcome rarely available for researchers. The similarity between short and long run effects may suggest that mixed results for short run outcomes in the literature can be due to imperfect measurement of students' skills.⁵ The results motivate further research on long run outcomes from contexts where only evidence on short run outcomes exist.

In this paper we estimate long run effects of class size for Norway where previous research has not been able to provide robust evidence of short run gains from smaller classes in terms of student achievement.⁶ In particular, the findings for short run outcomes differ substantially between the Scandinavian countries Sweden, Denmark and Norway with apparently similar educational and labor market institutions. All countries have small income differences, generous welfare state arrangements, and comprehensive public school systems seeking to equalize opportunities across families and students. Nevertheless, closer inspection reveals that important institutional differences prevail with regard to e.g. school district size and teacher shortages.⁷

⁵ In addition to Fredriksson et al. (2013), several studies from Sweden find that increased school resources increase student performance in the short run, including Björklund et al. (2004, ch. 4), Lindahl (2005) and Fredriksson and Öckert (2008). Browning and Heinesen (2007) and Heinesen (2010) find that lower class size in Danish compulsory education increases student performance in terms of both student test scores and educational attainment.

⁶ The Norwegian studies exploiting the class size rule are Bonesrønning (2003), Leuven et al. (2008), and Iversen and Bonesrønning (2013). They find small or zero average effects of class size. Hægeland et al. (2012) exploit variation in school resources across school districts with different income from local taxes on hydropower plants in Norway. They find that higher resources increase student achievement.

⁷ The institutional differences increased after the major reforms in Sweden in the mid-1990s. Our focus here is on institutional differences that have prevailed for several decades since several of the Swedish studies, including Fredriksson et al. (2013), use data on individuals graduating compulsory education before these reforms. See Björklund et al. (2004, ch. 4) for a description of the Swedish reforms in the 1990s and OECD (2011) and Bonesrønning (2013) for a description of recent Norwegian reforms.

We first exploit the strict class size rule in Norway and match individual and school register information from 1982 through 2011 to estimate causal effects on educational attainment and income.⁸ While experimental studies are often viewed as the “gold standard” in empirical research, exploiting the class size rule in a quasi-experimental approach makes it possible to circumvent the potential Hawthorn effect that might plague experimental studies (Ehrenberg et al., 2001). In contrast to Fredriksson et al. (2013), we are able to use register data for the whole population of schools for cohorts born 1966-1984 representing almost 1 million students and 1150 schools with separate catchment areas.⁹

Secondly, information on the whole population of schools and students for a long time period offers a unique possibility to use the quasi-experimental strategy to study whether the class size effect depends on characteristics of the environment in which the schools and students operate. We focus on dimensions that mirror differences in external conditions indicated by previous studies to be important for school efficiency and student performance, such as teacher quality, extent of upper secondary school choice, school district size, local fiscal constraints and labor market conditions.

We find insignificant effects of class size in grade 8-10 on educational attainment and income. While this is in contrast to the previous papers on long run effects, it is in accordance with the findings in the short run for Norway. Moreover, we find only weak evidence that class size effects vary with school district characteristics.

The paper is organized as follows. In section 2 we present arguments why the effect of resources may depend on characteristics of the external environment in which schools and students operate. Section 3 describes the institutions and the data, while the identification approach and model specification are presented in Section 4. Section 5 presents results from models estimating the causal average effect of class size on income and years of education,

⁸ During the work with the present paper, we became aware of Leuven and Løkken (2015) who includes similar analyses as we do using similar data on a slightly different period. Leuven and Løkken (2015) focus on the effect of class size both in primary and lower secondary education utilizing that some schools include grade 1 to 10, assuming that the students stayed in the same school during all school years. Their findings for the average effect of class size are qualitatively similar to ours.

⁹ Fredriksson et al. (2013) use data for a roughly 10 % sample of the cohorts born 1967, 1972, 1982 and 5% sample of the cohort born 1977. In addition, to ensure exogenous catchment areas for schools, they only include school districts (“rektorsområder”) with one school in their main analysis, implying that they are left with a sample of about 6000 students and 191 schools.

while Section 6 estimates interaction models where we investigate whether the class size effect depends on school district characteristics. Section 7 includes a discussion of the findings in relation to the present literature, and some concluding comments are provided in Section 8.

2. Why might class size effects vary?

Class size may change student outcomes through a number of mechanisms affecting both student and teacher behavior. Smaller classes may be beneficial for students by reducing crowding effects through student disruption (Lazear, 2001), increasing student attention or increasing the time teachers can use separately on each student. On the other hand, larger classes may be beneficial if a larger number of students increases the possibility that a student can find another student he/she can benefit from being in a class with, i.e. students with similar competencies, see Dobbelsteen et al. (2002). The literature in economics of education has also emphasized the impact of teachers, school district size and school district financing systems on student performance. In the following we discuss how these channels may affect class size effects.

2.1. Teacher quality

The class size effect might depend on teacher quality as argued by educationalists (Hattie, 2005) and economists (West and Woessmann, 2006). Hattie (2005) notes that “Without changing the teaching and ensuring rigor in the curriculum delivery then the effects of this most expensive policy is likely to be close to zero” (p. 417). This indicates that smaller classes are only productive with high-quality teachers. Mueller (2013) uses data from the STAR experiment and finds that being assigned to a small class increases test scores when the teacher is experienced.

On the other hand, West and Woessmann (2006) conclude that “smaller classes have an observable beneficial effect on student achievement only in countries where the average capability of the teaching force appears to be low” (p. 727). This finding is supported by evidence in Altinok and Kingdon (2012), who also use an international comparable data base. They exploit subject specific class sizes in a student fixed effects strategy. We extend this line of research to an RDD framework and analyze whether the class size effect depends on teacher supply conditions.

2.2. Student incentives

The simple human capital investment model assumes that students are forward looking and make optimal educational decisions given their preferences and information on private gains and costs of education. When making educational choices students trade off short run costs in terms of effort in school and foregone income against future utility benefits in terms of future income.¹⁰ Lavecchia et al (2014) extend this framework to incorporate elements from behavioral economics and discuss recent empirical evidence on the relationship between student achievement and incentives provided by schools and society in the context of deviations from long run rationality. One important element in this literature is that students are myopic and put too much weight on present effort relative to future gains. Under such circumstances external conditions affecting only short run educational costs can be very important for future educational outcomes.

While the literature has emphasized the direct effect of student incentives, we investigate whether a gain in student achievement from increased inputs in terms of lower class size only occurs if the schools and society in general provide sufficient incentives for students to exert effort. Evidence on this issue is very limited, but Bonesrønning (2003) finds some weak evidence that class size reduction has a larger positive effect on test results when teachers are able to install strong student effort incentives in terms of hard grading practices. We extend the research on student incentives to investigate whether the effect of class size is related to post-compulsory school choice systems and external labor market conditions.

Post-compulsory school choice

A large and still growing literature analyzes school choice as an incentive device. Although the empirical evidence is mixed, most studies find a modest positive effect of school choice and vouchers (Figlio and Hart, 2014). While school choice effects might be transmitted via a variety of mechanisms, our focus is on the effect of choice mediated by student incentives. Choice related incentives may also exist in traditional public school systems. In some cases students compete for admission to different tracks within compulsory school at certain ages based on prior performance. In other cases, competition is introduced by free school choice in

¹⁰ Examples of studies incorporating student effort in human capital investment models through educational standards is Costrell (1994), Betts (1998), and Becker and Rosen (2002).

upper secondary education based on prior student performance. These types of competition change the incentives for students to perform well in early school years.

Koerselman (2012) finds that the change from a tracking system to comprehensive schools in England reduced test scores at early ages. Using a difference-in-differences strategy, Haraldsvik (2012) finds that the introduction of free school choice in publicly provided upper secondary education in Norway increased student performance in lower secondary education. We investigate whether the effect of class size in compulsory education is related to the extent of competition for admission into post-compulsory education.

External labor market conditions

Several studies find that student opportunity costs in terms of foregone income during school and returns to schooling are important determinants of educational attainment. Clark (2011) finds a positive effect of regional unemployment on high school enrollment in England and Wales, while Reiling and Strøm (2015) find a countercyclical pattern in high school completion in Norway. Lee (2013) finds that increased job opportunities generated by repeal of Sunday shopping restrictions in US states decrease high school graduation. While these studies document the importance of job opportunities when students make educational choices after compulsory education, labor market conditions may also affect the student's allocation of time and effort during compulsory education. If class size effects depend on student incentives, the effect of class size could potentially depend systematically on labor market conditions. The fact that our data set covers a rather long time period makes it possible to investigate this issue by interacting class size with the local unemployment rate that prevails during compulsory education.

2.3. Fiscal constraints

In a traditional production function framework, more input implies higher production. Whether public sector services are produced technically efficient is, however, a widely discussed issue. In the public sector there are multiple principal-agent relationships (Dixit, 1996, 2002). Teachers and school principals might have different objectives than parents and the school district politicians. Thus, the institutional setting in which these actors operate is likely to affect the potential impact of exogenous changes in resources available for the schools. If student performance has no consequences for the decision makers in schools, it is

less likely that smaller classes would increase student performance. Instead, school principals and teachers might exploit extra resources to decrease effort, to make school days more pleasant, or to increase other types of “slack”.

A literature finds evidence that decentralized decision making improves student performance (Glaeser, 1996, Barankay and Lockwood, 2007, Falch and Fischer, 2012). Hoxby (1999) argues that local funding by local property taxation can work as a discipline device on local governments and improve cost control and effort. For Norway, Borge and Rattsø (2008) provide evidence that local property taxation reduces unit costs in utility services, while Fiva and Rønning (2008) find that property taxation increases student achievement. Studies from the US suggest that local funding increases technical efficiency in schools (Adkins and Moomaw, 2003) and student performance (Mensah et al, 2013). Further, Loeb and Strunk (2007) find substantial nonlinearities in the effect of accountability policies; i.e. accountability is more effective in US states with stronger local control in terms of local funding and local autonomy in hiring and spending decisions. While most of the studies so far find positive effects of local funding on efficiency and student performance, we ask whether the effect of exogenous variation in class size differ between school districts with and without access to local property tax revenue.

2.4. Interest groups

Chubb and Moe (1988) and Moe (2001, 2011) argue that teacher unions reduce the power of politicians to implement reforms and to use resources efficiently. Others argue that teacher unions may enhance efficiency by increasing teachers’ job satisfaction and productivity, see Gunderson (2005) for a discussion of union voice effects in the public sector. Hoxby (1996) finds evidence that teacher unions are able to increase the teacher-student ratio, but also decrease the productivity to such an extent that student performance declines. Lovenheim (2009) finds that while unions increase teacher employment, there is no corresponding impact on student performance. Strunk and Grissom (2010) find that school districts with strong teacher unions have less flexibility in school policy than districts with weaker unions, while the evidence in Lott and Kenny (2013) indicates that students in US states with strong teacher unions perform substantially worse than students in other states.

Since the large majority of teachers in Norwegian schools are members of a teacher union, it is almost impossible to study the impact of teacher unions on student performance and the interaction with class size effects. However, the impact of unions and other interest groups depends on the political setting in which they operate, i.e., by their ability to build coalitions in the government or directly affect the behavior of the decisive voter. Using survey data from Norway, Rattsø and Sørensen (2004) find that public employees prefer less public sector reform than others. Similar results are obtained by Bonesrønning (2013) who finds that school districts with a high share of public employees were less reluctant to implement a major accountability education reform in Norway in the period 2004-2006.¹¹ These findings motivate studying to what extent the impact of class size differs between school districts with high and low shares of public employment.

2.5. School district size

The size of school districts varies a lot between countries. A common argument is that the competency of education governance is higher in large school districts than in small school districts. However, the evidence on scale effects in public sector production in general is mixed, and the small literature on the effect of district size on student performance is also inconclusive.¹² We investigate whether there is a larger return to small class size in large school districts, more similar to the typical school district size in Sweden and Denmark.

3. Institutions and data

3.1. Institutions

Compulsory education in Norway consists of primary school and lower secondary school, and ends the year the students turn 16 years of age.¹³ Most students continue on to upper

¹¹Anzia (2011) argues that members of interest groups have higher turnout in off-cycle elections than other voters and that the policy in jurisdictions with off-cycle elections consequently are more favorable to interest groups. Consistent with this hypothesis she finds that US school districts with off-cycle elections have higher teacher pay than other districts.

¹² Using data from California, Driscoll et al. (2003) find that test scores are negatively related to district size. Using Danish data, Heinesen (2005) concludes that educational attainment is higher for students from larger districts, i.e. districts with population above 15,000.

¹³ During the empirical period, the school starting age was 7 years, but the school starting age was reduced from 7 to 6 years in 1997 such that primary school today consists of grades 1-7 (ages 6-13) and lower secondary school consists of grades 8-10 (ages 14-16). We refer to grades 8-10 as lower secondary education throughout the paper.

secondary education, which is divided into a three-year long academic study track and different vocational study tracks. After a major reform in 1994, vocational study tracks typically last for four years (including two years of apprenticeship training). Acceptance to upper secondary school is based on the grades achieved in grade 10. However, all students have been guaranteed admission to upper secondary education since 1994.

There is no possibility to fail a class neither in primary nor in lower secondary education during the empirical period, which implies that everyone finishes compulsory education on-time.¹⁴ The education is comprehensive with no tracking and a common curriculum for all students. The cutoff between grades is birth at January 1.

Compulsory education is free of charge and is the responsibility of the municipalities. Norwegian municipalities are multipurpose institutions, providing a large number of services, such as day care and care for the elderly, in addition to education.¹⁵ In the following we refer to municipalities as school districts. There are usually several primary schools within each school district, but many small school districts only have one lower secondary school. Parental school choice between public schools for a given residence is not allowed, and private schools are quite rare and do not represent a realistic alternative to public schools.

The classes could not exceed 30 students in lower secondary education during the empirical period. The class uses the same classroom for most subjects. The teachers, who are specialized in specific subjects, move between classrooms. The classes are established at the start of lower secondary education such that all classes have about the same socioeconomic composition, and it is very uncommon to change the composition of classes unless the number of classes changes.

¹⁴ In some cases, students do not start primary education at the expected age, which implies that they finish lower secondary education at a higher age. If a child is not considered to be mature enough, the parents together with the school and psychologists can postpone enrollment one year. In addition, some older students return to improve their grades, and immigrants are often over-aged at graduation.

¹⁵ Spending on primary and lower secondary education accounts for about 30% of total local government spending while spending on care for the elderly, preschool education, cultural services, infrastructure services and administration accounts for the rest.

3.2. Data

In this paper we study the cohorts born 1966-1984 who leave lower secondary education during 1982-2000. We use register data provided by Statistics Norway for all individuals leaving lower secondary education in this period. The data contain unique individual and school identifiers which allow us to combine detailed information on individuals and the schools they attended.

Our two main outcome variables are years of education and income. We measure the outcomes in a given year, for which the individuals are of different age, and fully control for age effects in the model by including fixed effects for the year of graduation from lower secondary education. Our measure of educational attainment is years of education in 2011, measured by degrees obtained. In higher education that is bachelor degree, master degree, and PhD, with 16, 18, and 21 years of education, respectively. We use the log of average pension qualifying income for the years 2009 and 2010 as our income measure,¹⁶ such that the youngest individuals in the sample are 25-26 years of age when income is measured.

We restrict the sample to students graduating lower secondary education the year they turn 16, which excludes 5 % of the observations. Table A1 reports the number of observations lost due to missing information on class size, the age restriction, requiring at least 10 school observations throughout the time period, and having missing information on either log income or educational attainment. We are able to use 86 % and 81 % of the population in the analysis on educational attainment and log income, respectively. The cohort leaving secondary education in 1990 has missing information on the school identifier, and is thus not included in the analysis. The number of observations in the analyses is about 950,000, with cohort sizes of about 50,000 students.

The distributions of the dependent variables are presented in Figures 1 and 2. Table 1 shows that the average years of education is 14.0 with standard deviation of 2.5, while log of income has mean 12.7 with standard deviation of 0.8.

¹⁶ We use the pension-qualifying income as reported in the tax registry. This income measure is not top coded and includes labor income, taxable sick benefits, unemployment benefits, parental leave payments, and pensions, see Black et al. (2013) p. 132. Information for 2011 is not available in our data.

Our individual register data contains information on gender, birth month and immigration status, as well as detailed data on educational attainment and income for all years after the individual leaves lower secondary school and up to 2011. Immigration status is divided into first and second generation immigrant, where first generation immigrants are born abroad and have both parents born abroad, while second generation immigrants are born in Norway and have both parents born abroad.

In the analysis, we also include information on parental education and parental employment status the year the individual turns 16, which is the year the individual leaves lower secondary school. Parental education is categorized as the highest completed education by one of the parents. The categories included are upper secondary education (High school), Bachelor's degree, Master's degree or PhD, and unknown education, with less than High school being the reference category. Regarding parental employment, we include indicators for only mother working, only father working, and both parents working, with the reference category being no parent working. Descriptive statistics are presented in Table 1.¹⁷

Data on the number of classes and enrollment by year and grade is obtained from a national school register administered by The Norwegian Ministry of Education. Variables are measured on October 1 of each year which is near the beginning of the school year. The information is provided for the school rather than for the class, so we are only able to calculate the average class size for each year and grade rather than the actual class size for each class. However, a benefit of using this measure is that we do not have to worry about sorting into classes of different class sizes within schools.

Figure 3 displays the distribution of the average class size in grades 8-10 for our sample. The typical student is in a class of 23-29 students. There are extremely few observations above 30 students per class, which reflects that the class size rule is strictly followed (see also Leuven et al., 2008).

¹⁷ Descriptive statistics on the school district characteristics used in the heterogeneity analysis are also presented in Table 1. These variables are closer described in Section 6 below.

4. Identification and model specification

There are several reasons why standard OLS regressions treating actual class size as an exogenous variable might yield biased estimates. For example, disruptive students with negative peer group effects might be placed in smaller classes; small remote schools with small classes might have problems in recruiting and retaining high quality teachers; student mobility might be motivated by observed class sizes; peers might correlate with class size; etc. To tackle the identification problem and estimate causal effects, one ideally want to explore only the part of variation in actual class size that is due to exogenous forces. A maximum class size rule serves this purpose.

4.1. The class size rule

During the time period we study, a national rule was in place saying that class size could not surpass 30 in lower secondary education. The class size rule creates exogenous variation in predicted class size depending on the number of students enrolled in a school.

Since learning is cumulative, we estimate the effect of average class size during lower secondary education (grades 8 to 10) and not the class size in one specific school year. Each student is matched to their lower secondary school at graduation, and we use information from this school also for the two previous school years to calculate average class size. For each grade level the data contain the number of classes and the number students enrolled.¹⁸

We follow Leuven et al. (2008) and use predicted class size based on enrollment in grade 8, two years prior to graduation, as the instrument in the analysis in order to avoid biased estimates due to possible endogenous mobility of students across schools during the years in lower secondary education. The instrument is given by

$$CS_{t-2}^{rule} = \frac{E_{t-2}}{\text{int}(1 + (E_{t-2} - 1)/CS^{max})} \quad (1)$$

where E_{t-2} is enrollment in grade 8 and CS^{max} is the maximum class size according to the rule, which in our case is 30 students during the whole empirical period. Using this formula, the

¹⁸ We only include students in schools that have non-missing information on the number of classes and enrollment for all three years of lower secondary education in the analysis, see Table A1.

strict maximum class size rule predicts a class size of 30 when 30 students are enrolled and a class size of 15.5 when 31 students are enrolled. Such a kink appears at each multiple of 30 and creates a nonmonotonic relationship between enrollment and predicted class size. We follow Angrist and Lavy (1999) in instrumenting actual class size by predicted class size defined in equation (1), while controlling flexibly for enrollment.

Figure 4 plots the predicted class size calculated by equation (1) and average actual class size for grades 8-10 against enrollment in grade 8. Average class size closely tracks predicted class size for all enrollment levels.

One possible threat to the validity of the instrument is manipulation of enrollment around the thresholds. Urquiola and Verhoogen (2009) finds this to be the case in Chile. Fredriksson et al. (2013) also find that sorting take place within school districts because “it is likely that school catchment areas are adjusted” (p. 254). Thus, their analysis includes only school districts with one school.

In Norway, it has been uncommon to adjust school catchment areas. Panel A in Figure 5 plots the distribution of enrollment in grade 8, where the vertical lines represent the class size thresholds. There is no evidence of manipulation of the enrollment. The density of observations just below and above the thresholds is similar. In fact, the enrollment is higher just above the threshold in 5 out of the 8 class size thresholds in the data. In addition, the figure shows that it is mainly the thresholds at enrollment of 30, 60, 90, and 120 students that will contribute to the identification of the class size effect. While the density in Panel A in Figure 5 is presented at the individual level, the identification is at the school level. Panel B uses the school as the observational unit, and shows that few schools have enrollment above 150 students in grade 8. Most schools have enrollment around the first threshold, for which there is the largest difference in class size across the threshold.

A more direct way to assess whether the instrument is valid is to examine whether socioeconomic characteristics are equal across observations above and below the class size thresholds. Table 2 tests the balancing of the covariates both individually and jointly.

The first two columns in Table 2 show that the socioeconomic characteristics are strong predictors of income and education as expected. The correlation with parental education is particularly strong. Column (3) presents results for a regression on predicted class size, using the control function for enrollment described below. None of the socioeconomic characteristics are significant at the 5% level, and the test for joint significance has a p-value of 0.08. Column (4) presents p-values for individual correlations, which are significant at the 5% level for two measures of parental education. Overall, however, the socioeconomic characteristics in the data are reasonably unrelated to predicted class size.

4.2. Model specification

We present results from two approaches to the regression discontinuity design. The first approach uses all information available, and includes a flexible control for the effect of cohort size at the school. The second approach discards observations away from the thresholds and uses a simpler specification for cohort size, see, e.g., Lee and Lemieux (2010) and Gelman and Imbens (2014) for a discussion of these approaches. We denote the former a “global” approach and the latter a “local” approach. In both approaches it is important to control for age effects because income and education are measured in a specific year, and thus at different ages. Since the analysis only includes individuals graduating lower secondary education at age 16, including time fixed effects is identical to including cohort and age fixed effects in our application.

Both approaches imply that we estimate variants of the following model:

$$y_{ist} = \alpha \overline{CS}_{st} + f(E_{st-2}) + \beta X_i + \delta_t + \varepsilon_{ist} \quad (2)$$

y_{ist} denotes the outcome for individual i graduating from school s in year t and \overline{CS} is the predicted average class size for grades 8-10.¹⁹ In addition, the model includes a flexible functional form of enrollment E in grade 8 at the school two year prior to graduation, individual characteristics, X_i , and cohort fixed effects (δ_t). The error term (ε_{ist}) is clustered at the school level. The first stage is simply

¹⁹ The average class size is calculated using information on grade 10, 9 and 8 in year t , $t-1$ and $t-2$, respectively, i.e., when the student was enrolled in the relevant grades.

$$\overline{CS}_{st} = \alpha' CS_{st-2}^{rule} + f(E_{st-2}) + \beta' X_i + \delta'_t + \varepsilon'_{ist} \quad (3)$$

It is important to ensure that the discontinuity generated by the class size rule is not confounded with a possible nonlinear relationship between the outcome variable and enrollment. Thus, a flexible modelling of enrollment effects in terms of the function $f(E_{st-2})$ is necessary when using the “global” approach. Define the thresholds for the class size rule in grade 8 as $\tilde{E}_{st-2} = \{30, 60, 90, \dots, 270\}$, and the segments of the class size rule as $S_{st-2} = I(\tilde{E}_{st} \pm 15)$. The following specification for the global approach seems to capture both the underlying functional form and to provide reasonable precision of the estimates

$$f(E_{st-2}) = \alpha_1 E_{st-2} + \alpha_2 E_{st-2}^2 + \alpha_3 E_{st-2}^3 + \alpha_4 E_{st-2}^4 + \alpha_5 S_{st-2} + \delta_s \quad (4)$$

where δ_s is school fixed effects.

The global approach essentially uses a bandwidth of ± 15 students. The local approach uses a substantially smaller bandwidth. In the case with the smallest possible bandwidth and only one discontinuity, $[\tilde{E}_{st-2}, \tilde{E}_{st-2} + 1]$, it is not possible to control for enrollment. The identifying assumption is that the outcome at these two enrollment levels would be equal in the absence of the discontinuity. Since we have several threshold levels in the data, we estimate local effects with the following model specification of enrollment.

$$f(E_{st-2}) = \alpha'_1 E_{st-2} + \alpha'_5 S_{st-2} \quad (5)$$

Figure 6 present average values of the outcomes for different levels of enrollment. The figure shows that the outcomes are positively related to enrollment. Since average class size is higher in larger schools than in small schools, this implies that class size and the outcomes are positively related, in contrast to the hypothesized class size effect. Our identification is the differences across the individual thresholds generated by the class size rule. The local polynomial regressions presented in the figure do not indicate any systematic changes in the outcomes related to the thresholds. For income, there seems to a difference for the threshold of 60 students, but in the opposite direction as expected. For educational attainment, there seems to be differences both for the thresholds 30 and 120 students, in the expected direction.

5. Average class size effects

For the global approach, in which all observations in the data are used, the results for different model specifications are presented in columns (1) - (8) in Table 3. Column (1) presents a simple OLS regression with cohort fixed effects and a linear enrollment control. With this specification, children in larger classes have higher income (t-value of -0.72) and complete more years of schooling (t-value of 5.60) than children in smaller classes, contrary to the expectations. However, when average class size is instrumented in this very simplistic model formulation (column (2)), the class size effect on income gets the expected sign, but is still insignificant. Predicted class size is a strong instrument. The F-value for the first stage is almost 5,000.

Columns (3)-(8) include various specifications of the enrollment control function. Regarding income, the point estimate is negative and clearly insignificant in all specifications. The result for educational attainment is more sensitive to the specification of the enrollment control function. The effect is positive and significant at 5% level in the models only including segment fixed effects (column 3) and enrollment to the fourth polynomial (column 4). When school fixed effects are introduced (column 5), the effect drops and becomes insignificant.

Column (6) additionally includes socioeconomic characteristics. This does not affect the class size effect, as expected from the balancing tests in Table 2. In column (7) and (8), enrollment is interacted with segment fixed effects. While the interaction is linearly in column (6), column (7) also includes interaction with enrollment up to the fourth polynomial. Although the strength of the instrument declines as the enrollment control function becomes more flexible, the F-value for the first stage is above 900 in each specification.

Column (6) is the model specification in equations (2) – (4) above. Taken at face value, the 95 % confidence interval of reduced class size of 10 students is [-0.0018, 0.0008] log points for income and [-0.0029, 0.0052] for years of education. Both intervals are very narrow. We can rule out even very small effects of class size.

The full results for the models in column (6) are presented in Appendix Table A2 columns (1) and (3). The effects of socioeconomic characteristics are as expected. Females have longer

education than males, but lower income. In addition, the appendix table shows results for the first stage. The first stage coefficient is 0.56, which is very close to the result in Leuven et al. (2008) despite that they only include students graduating lower secondary education in 2002 and 2003.

Figure 7 presents estimates for the local approach with 95% confidence intervals, shrinking the bandwidth from ± 10 students to ± 1 student. In the latter case, only observations just below and just above the thresholds are included (30 and 31 students, 60 and 61 students, etc.). The model formulation is equal to equations (2) and (5) above, and the results for bandwidth of ± 3 students, $[\tilde{E}_{st-2} - 2, \tilde{E}_{st-2} + 3]$ are presented in column (9) in Table 3.²⁰

For educational attainment, the estimated effects are insignificant for all bandwidths, and the point estimate is positive in all cases except one. Increased years of education for larger classes is in contrast with the intuitive hypothesis. For income, the point estimate is negative for all bandwidths except the most narrow. For large bandwidths, the effect is close to -0.002 and statistically significant at conventional levels. This is a stronger effect than for the global approach, but the enrollment control function is rather inflexible. It is simplistic in these models because it is specified for a model with a narrower bandwidth. For bandwidths of ± 6 students or smaller, the estimated effect is smaller and insignificant. Column (10) in Table 3 presents results for a model with a more flexible enrollment control function, including enrollment interacted with the segment fixed effects, for a bandwidth of ± 3 students. This changes the sign of the effect on both income and educational attainment, but the effects are still clearly insignificant. The strength of the instrument is reasonable also in this case with F-value for the first stage above 100.

To shed some light on what can be driving the insignificant results, Figure 8 presents cohort specific estimates using the model specification in column (6) in Table 3. The oldest cohort is born in 1966, graduated from lower secondary education in 1982, and years of education is measured at age 45 while income is measured as the average income at age 43 and 44. The youngest cohort is born in 1984. The estimate is not significant at the 5% level for any cohort

²⁰ For a full specification of the models, see columns (2) and (4) in Appendix Table A2.

and any outcome. For income, the point estimate is positive for four of the 19 cohorts, while for educational attainment, the estimate is positive for 12 cohorts.²¹

Figure 9 presents separate analysis for the different thresholds. The regressions are equivalent to column (9) in Table 3, with the segment fixed effects absorbed by the constant term. The regression denoted threshold 5 includes all thresholds from 5 (150 students) and upwards. As expected, confidence intervals at the 95% level rise with each threshold. In all cases, the effect of class size is insignificant at 5 percent level and close to zero.²²

6. Heterogeneous class size effects

In this section we investigate whether the class size effect depends on the external environment in which schools and students operate as discussed in Section 2 above. We focus on measures of teacher quality, fiscal constraints facing school districts, variables affecting student effort incentives, variables affecting interest group pressure and school district size. All variables are measured at the school district level. The small average treatment effect of class size in the long run might hide differences across school districts, and specific characteristics in some Norwegian school districts might explain the different average results compared to Chetty et al. (2001), Dynarski et al. (2013) and Fredriksson et al. (2013).

For each school district characteristic Z of interest, we estimate the following model,

$$y_{isdt} = \alpha \overline{CS}_{sdt} + \gamma Z_{dt} + \phi \overline{CS}_{sdt} * Z_{dt} + f(E_{sdt-2}) + g(E_{sdt-2}) * Z_{dt} + \beta X_i + \delta_t + \varepsilon_{isdt} \quad (6)$$

where subscript d indicates school district. This is equivalent to estimating equations (2) and (3), adding Z and the interaction terms with average class size and the control function for enrollment. The control functions $f(\cdot)$ and $g(\cdot)$ include the same elements, presented in equations (4) and (5) for the global and local approach, respectively. \overline{CS}_{st} and $\overline{CS}_{st} \cdot Z$ are instrumented using the class size rule and its interaction with Z . Since we use average class

²¹ Log income has a wide distribution, see Figure 1. However, this does not drive the results. In regressions including only observations with log of income between 10 and 15 (reduces the sample by 1.4%), the estimate for average class size is -0.00011 (0.0005) using the global approach specification in column (6) and -0.00098 (0.0016) using the local approach in column (9).

²² We have also run regressions using a binary variable for whether the student achieves a degree from higher education (completes more than 13 years of schooling) as an outcome variable. The effect is insignificant also for this measure of educational attainment. The estimate of average class size is 0.0006 (0.0004) using the global approach specification in column (6) and -0.0003 (0.0014) using the local approach in column (9).

size during grades 8-10 in the analyses, we measure the school district characteristics by the average value during the same time period. In order to facilitate interpretation, the interaction variables are standardized to have mean zero and standard deviation equal to unity, except when indicated. The level effect of Z is not reported since the interaction term with $g(\cdot)$ is included in the model.

6.1. Teacher quality

The evidence in the literature on the relationship between a class size effect and teacher quality is mixed. One empirical challenge is that teacher quality is not directly observed. Our approach is that teacher quality is related to the attractiveness of the school. According to the Norwegian school law, schools can only employ persons without a teaching certification if no certified teachers apply to a vacant teacher position, and non-certified teachers can only be employed for up to one school year. Teacher shortages measured by non-certified teachers thus reflect the state of the teacher labor market in a particular year. If the use of non-certified teachers increases, it reflects low interest for vacant positions, lack of options in the schools hiring processes, and thus low teacher quality. The share of certified teachers is thus a reasonable indicator of teacher quality, and is previously used by Bonesrønning et al. (2005) and Falch et al. (2009).

The first part of Table 4 presents the results.²³ Columns (1) and (3) use the global approach, while columns (2) and (4) use the local approach. The level effects of average class size are close to the findings in Table 3 as expected since the measure of teacher quality is standardized.²⁴ The joint strength of the instruments is tested by the Kleibergen-Paap F-statistic taking. The test value above 100 implies that the instruments are not weak

The interaction effect with our measure of teacher quality is negative or close to zero. The sign of the coefficient indicates that class size might have the expected negative effect when teacher quality is high. The best teachers might to be able to exploit the possibilities inherent

²³ Data for our measure of teacher quality is available from 1981. However, since we use 3 year averages in the estimations, the samples used in the analyses are from 1983 and onwards.

²⁴ Notice that since we have rescaled the variable for teacher quality to have mean zero, there are only two reasons why the level effect of class size could differ from the similar model in Table 3. First, the model includes an additional variable (teacher quality), and second, the sample size is about 5 % smaller. If we re-estimate the corresponding models in Table 3 using the same sample as in Table 4, we get the same coefficients for the level effect of class size.

in small classes. For the income-equation using the local approach, the interaction effect is significant at 5% level. The results imply that decreasing class size by 10 students in school districts with teacher quality 2 standard deviations above the average, increases the income by 0.057 log points, i.e., about 7% of a standard deviation in income.

6.2. Student incentives

Without student incentives, more resources can hardly improve student achievement. We investigate the effect of two different student incentives that are external to the school district authorities. First, upper secondary education is non-compulsory and is the responsibility of the 19 counties. Some counties have free school choice, while other counties use school catchment areas. With free school choice, the students rank schools in their applications, and admission to oversubscribed schools is solely based on grade point average from lower secondary education (GPA).²⁵ Thus, there are stronger incentives for study effort in lower secondary education in some counties than in others.²⁶ We use the classification developed by Haraldsvik (2004),²⁷ previously exploited by Falch and Naper (2013). Indeed, Haraldsvik (2012) finds that school choice in upper secondary education in Norway increases student achievement in lower secondary education. Our hypothesis is that since school choice increases student incentives, the effect of class size is larger than without school choice.

The results are presented in the second part of Table 4. The effect of the interaction between class size and the dummy variable for free school choice is negative as expected when using the local approach, but insignificant at conventional levels in all models. Taken at face value, the point estimate in the case of school choice of a reduction in class size of 10 students is 0.04 log points on income and 0.22 years of education.

²⁵ A closer description of one system of free school choice is given in Machin and Salvanes (2010). They study the effect on house prices of increased school choice from 1997 in the Oslo county.

²⁶ In addition, the students have to rank three different study tracks in their application to upper secondary education. They have a legal right to be enrolled into one of these three tracks, but whether they are enrolled in the first, second, or third preferred track depends on their GPA.

²⁷ Haraldsvik (2012) distinguishes between school districts where the students have (i) free school choice between at least five schools or (ii) with some limitations, (iii) free school choice but between less than five schools, (iv) no choice at all, and (v) some marginal school choice. We classify the former three school districts as free school choice and the two latter school districts as without school choice. School districts were in 2003 asked about their school choice rules for the past 10 years. The regression sample is therefore from 1993 and onwards.

Our second measure of student incentives is the unemployment rate in the school district. The interaction effects are negative as expected, but small and insignificant. Again the estimated class size effect is largest on income in the case with local identification, and of comparable size as in the model for teacher quality. But taken together, the results indicate that student incentives have no impact on how efficient schools use their resources.

6.3. Fiscal constraints

Local funding by local property taxation can work as a discipline device on local governments and lead to better cost control (Glaeser, 1996, and Hoxby, 1999). In Norway, some school districts have property taxes while others do not. We exploit this variation in order to investigate whether class size has the expected effect with stronger fiscal constraints, i.e., there is a stronger incentive for cost control and effort.

Local governments decide both on the valuation of houses, the tax-free allowance, and the tax rate, but data on these properties of the local tax systems are not available. In our analysis we follow Borge and Rattsø (2008) and use an indicator for whether the school district has property taxation or not, for which comparable data are available in the period 1997-1999. Introduction or abolishing of property taxation are political decisions with strong local interest, and does not happen often. The share of school districts with property taxation is 14.0 – 15.6 percent in this period, and is most common in the large school districts. Since we use three-year averages of the variables in the analyses, we extrapolate the information on property taxation in both ends, assuming that the values are the same in 1995 and 1996 as for 1997, and the same in 2000 as for 1999. The estimation period is therefore 1995-2000.

The results in Table 4 are again insignificant at the conventional level, and the sign of the interaction effect varies across the model specifications. The class size effect seems to be un related to local fiscal constraints.

6.4. Interest groups

Interest groups prefer increased resource use and reduced pressure on efficiency. As discussed in Section 2, there is some evidence in the literature indicating that public sector employees are more prone to interest groups than others. We use the share of public sector employment as an indicator of interest group influence, including employees both in local governments

and the central government, and test the hypothesis that the class size has a larger negative effect when this share is low.²⁸

Table 4 shows that also this interaction term is insignificantly related to the class size effect. The point estimates are small, and the sign of the effects varies across the specifications.

6.5. School district size

Are the resources used more efficiently in school districts with presumably more competent management of the schools? In the Norwegian setting it is usually argued that small school districts have challenges recruiting quality leadership and implementing efficient governance systems, which also was the main argument for the major school district consolidation in Denmark in 2007. There is a positive relationship between student achievement and school district size in Norwegian data.

We investigate the interaction between the class size effect and school district size in two different ways. Firstly we include interaction effects with the number of inhabitants in the school district. In this case the interaction effect is mainly negative as expected, but clearly insignificant. The F-value of the test of weak instruments is smaller in these models than in the models above, most likely because the schools are larger in the cities. Population size and predicted class size are positively correlated.

In general, the interaction effect with class size in this case might reflect unobserved characteristics of the school district. In addition, since the model using the whole sample includes school fixed effects and population changes only to a small extent from one year to another, little variation in school district size is used for identification in this case. Our second approach exploits that some school districts have merged during the empirical period, while the schools catchment areas did not change.

We combine a difference-in-differences approach with regard to school districts merging and the regression discontinuity approach with regard to class size. The model includes an

²⁸ Information on the share of public sector employment in the school district is available from 1984, which implies that the regression samples are from 1986 and onwards.

indicator variable for whether or not the school districts ever experiences a merger ($Treat$) and an indicator variable for the period after the merger in the treated school district ($Post$), in addition to the population size (Pop).

$$y_{isdt} = \alpha \overline{CS}_{sdt} + \gamma_1 Treat_d + \gamma_2 Treat_d * Post_t + \phi \overline{CS}_{sdt} * Treat_d * Post_t + f(E_{sdt-2}) + g(E_{sdt-2}) * Treat_d * Post_{dt} + \beta_1 Pop_{dt} + \beta X_i + \delta_t + \varepsilon_{isdt} \quad (0.1)$$

The term $\gamma_2 + g(E_{sdt-2})$ is the difference-in-differences estimator. Both terms including class size are instrumented in the same way as above.

Results are reported towards the end of Table 4. The results for the local approach can hardly be interpreted in this case because the instruments are weak. For the global approach, the interaction effects are relatively large, but insignificant and with opposite sign for income and education.

7. Discussion

Contrary to the results for the US, Sweden and Denmark, we find no long run effect of reduced class size. However, our study confirms that the long run effect of class size seems to be qualitatively similar to the short run effect on student achievement. While there appears to be positive effects of smaller class size both in term of student achievement, educational attainment, and income in contexts analyzed in the US, Sweden and Denmark, there appears to be no effect on neither student achievement, educational attainment nor income within the institutional setting in Norway.

The difference between our results and the other Scandinavian countries is of special interest since these countries are viewed as very similar. One potential explanation for the different results is that school districts are generally much smaller in Norway than in Sweden and Denmark.²⁹ However, our finding that the class size effect in Norway does not depend on school district size speaks against this explanation. Another possibility might be that teacher

²⁹ Both in Sweden, Denmark and Norway, the municipalities (school districts) are multi-purpose local governments with the major responsibility for local welfare services. A major consolidation reform in Sweden in 1974 reduced the number of municipalities to about 280, while Denmark in 2007 implemented a consolidation of municipalities from 271 to 98. In contrast, Norway has about 440 municipalities even though the population in 1990 (4.2 mill) was half of that in Sweden and roughly 20% lower than in Denmark. Average municipality size in 1990 was around 30,000, 19,000 and 10,000 in Sweden, Denmark and Norway, respectively

quality differs systematically between countries.³⁰ The absence of robust significant interaction effects between class size and our indicator for teacher quality does not support this explanation either.

At a general level, the class size effects might obviously also depend on characteristics of the students, although such characteristics vary to a smaller degree across countries. First, there is some evidence that the class size effect is largest at young ages. Ehrenberg et al. (2001) hypothesize that small classes during the elementary grades develop working habits that enable students to take advantage of learning opportunities in later grades. The STAR experiment was targeted towards students up to third grade. Fredriksson et al. (2013) investigate class size effects at ages 11-13. However, several papers find a positive effect of resources also in higher grades. Fredriksson and Öckert (2008) find for Sweden a positive effect of the teacher/student ratio on student performance at age 16 in a difference-in-differences framework. For Denmark, Browning and Heinesen (2007) find that lower class size in grade 8 increases the probability of completing high school and years of education, and Heinesen (2010) finds a positive effect of subject-specific class size in lower secondary education in a student fixed effects framework. This evidence suggests that our use of class size in lower secondary education (grades 8-10) cannot explain the different results between Norway and the other Scandinavian countries.

A final issue is that class size effects may differ across students with different socio-economic characteristics. First, there is evidence of gender differences in competitiveness (Buser et al., 2014), which might give gender differences in the class size effects. Larger classes arguably have a more competitive environment. However, also for gender differences, the evidence is mixed for class size reductions. In separate analyses reported in Appendix table A3, we do not find different class size effects for males and females in the Norwegian data.

Second, small classes might be most beneficial for students with disadvantaged backgrounds, who do not have the same resources in the home to support their education as other students. This is the typical finding from the STAR experiment (Dynarski et al., 2013) and other

³⁰ The share of teachers certified for teacher jobs varies substantially between Norway and Sweden. According to Andersson et al. (2011), above 15 percent of the Swedish teachers were non-certified on average in 2000, while Bonesrønning et al. (2005) shows that the corresponding number for Norway is about 6 percent.

studies (Bonesrønning and Iversen, 2013, Bosworth, 2014). On the other hand, Fredriksson et al. (2013) find strongest class size effects for students with high parental income. Appendix Table A3 shows that we do not find evidence of such heterogeneity.

8. Conclusion

The lack of conclusive evidence on the effect of school resources on student test scores calls for systematic studies of possible heterogeneous effects using credible identification strategies. This paper uses rich register data from Norway from a long time period combined with a quasi-experimental empirical strategy to estimate both the average effect of class size and to which extent the effect varies with a range of external conditions facing schools and students. Using a strict class size rule in an RDD framework, we first show that on average there is no evidence that lower class size increases long run outcomes as earnings and educational attainment. This is in accordance with the previous Norwegian results for short run outcomes.

Second, we investigate heterogeneity in class size effects by interacting class size with indicators of teacher quality, extent of upper secondary school choice, school district size, local fiscal constraints and labor market conditions within the same quasi-experimental framework. Overall, we find that class size effects do not depend on such external conditions.

Our results stand in sharp contrast to experimental evidence from the US and quasi-experimental evidence from Sweden and Denmark finding significant and numerically important positive effects of reduced class size on both short run and long run outcomes. The absence of interaction effects with measured external conditions indicate that between country differences in teaching practices and educational culture are relevant explanations for the different results.

References

- Adkins, L. C., and R. L. Moomaw (2003). The impact of local funding on the technical efficiency of Oklahoma schools, *Economics Letters* 81, 31-37.
- Altinok, N., and G. Kingdon (2012). New evidence on class size effects: A pupil fixed effects approach. *Oxford Bulletin of Economics and Statistics* 74, 203-234.
- Andersson, C., P. Johansson and N. Aldenström (2011). Do you want your child to have a certified teacher? *Economics of Education Review* 30, 65-78
- Angrist, J. D., E. Battistin, and D. Vuri (2015). In a small moment: Class size and moral hazard in the Mezzogiorno. IZA DP No. 8959.
- Angrist, J. D., and V. Lavy (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* 114, 533-575.
- Anzia, S. F. (2011). Election timing and the electoral influence of interest groups. *Journal of Politics* 73, 412-427.
- Barankay, I., and B. Lockwood (2007). Decentralization and the productive efficiency of government: Evidence from Swiss cantons. *Journal of Public Economics* 91, 1197-1218.
- Becker, W. E., and S. Rosen (1992). The learning effect of assessment and evaluation in high school. *Economics of Education Review* 11, 107-118.
- Betts, J. R. (1998). The impact of educational standards on the level and distribution of earnings. *American Economic Review* 88, 266-275.
- Björklund, A., P.-A. Edin, P. Fredriksson and A. Krueger (2004). Education, equality and efficiency – An analysis of Swedish school reforms during the 1990s. IFAU Report 2004:1.
- Black, S. E., P. J. Devereux and K. G. Salvanes (2013). Under pressure? The effect of peers on outcomes of young adults. *Journal of Labor Economics* 31, 119-153.
- Bonesrønning, H. (2003). Class size effects on student achievement in Norway: Patterns and explanations. *Southern Economic Journal* 69, 952-965.
- Bonesrønning, H. (2013). Public employees and public sector reform implementation. *Public Choice* 156, 309-327
- Bonesrønning, H., T. Falch and B. Strøm (2005). Teacher sorting, teacher quality, and student composition. *European Economic Review* 49, 457-483.
- Borge, L-E., and J. Rattsø (2008). Property taxation as incentive for cost control: Empirical evidence for utility services in Norway. *European Economic Review* 52, 1035-1054.
- Bosworth, R. (2014). Class size, class composition, and the distribution of student achievement. *Education Economics* 22, 141-165.

- Browning, M., and E. Heinesen (2007). Class size, teacher hours and educational attainment. *Scandinavian Journal of Economics* 109, 415-438.
- Buser, T., M. Niederle and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *Quarterly Journal of Economics* 129, 1409-1447.
- Chetty, R., J. N. Friedman, N. Hilger, E. Saez, E. W. Schanzenbach and D. Yagan (2011). How does your kindergarten classroom affect your earnings? Evidence from project STAR. *Quarterly Journal of Economics* 126, 1593-1660.
- Chubb, J. E., and T. M. Moe (1988): Politics, markets and the organization of public schools. *American Political Science Review* 82, 1065-1087.
- Clark, D. (2011). Do recessions keep students in school? The impact of youth unemployment on enrolment in post-compulsory education in England. *Economica* 78, 523-545.
- Coleman, J. S., et al (1966). *Equality of educational opportunity*. Washington. DC: US Government Printing Office.
- Costrell, R. M. (1994). A simple model of educational standards. *American Economic Review* 84, 956-971.
- Denny, K., and V. Oppedisano (2013). The surprising effect of larger class sizes: Evidence using two identification strategies. *Labour Economics* 23, 57-65.
- Dixit, A. K. (1996). *The Making of Economic Policy: A Transaction-Cost Politics Perspective*. The MIT Press.
- Dixit, A. K. (2002). Incentives and organizations in the public sector: An interpretative review, *Journal of Human Resources* 37, 696-727.
- Dobbelsteen, S., J. Levin and H. Oosterbeek (2002). The causal effect of class size on scholastic achievement: distinguishing between the pure class size effect from the effect of changes in class composition. *Oxford Bulletin of Economics and Statistics* 64, 17-38.
- Driscoll, D., D. Halcoussis and S. Svorny (2003). School district size and student performance, *Economics of Education Review* 22, 193-201.
- Dynarski, S., J. Hyman and D. W. Schanzenbach (2013). Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. *Journal of Policy Analysis and Management* 32, 692-717.
- Ehrenberg, R. G., D. J. Brewer, A. Gamoran and J. D. Willms (2001). Class size and student achievement. *Psychological Science in the Public Interest* 2, 1-30
- Falch, T., and J. A. V. Fischer (2012). Public sector decentralization and school performance: International evidence. *Economics Letters* 114, 276-279.
- Falch, T., K. Johansen and B. Strøm (2009). Teacher shortages and the business cycle. *Labour Economics* 16, 548-658.
- Falch, T., and L. R. Naper (2013). Educational evaluation and gender gaps in student achievement. *Economics of Education Review* 36, 12-25.

- Figlio, D., and C. D. Hart (2014). Competitive effects of means-tested school vouchers. *American Economic Journal: Applied Economics* 6, 133-156.
- Fiva, J., and M. Rønning (2008). The incentive effects of property taxation: Evidence from Norwegian school districts. *Regional Science and Urban Economics* 38, 49-62.
- Fredriksson, P., and B. Öckert (2008). Resources and student achievement - evidence from a Swedish policy reform. *Scandinavian Journal of Economics* 110, 277-296.
- Fredriksson, P., H. Oosterbeek and B. Öckert (2013). Long term effects of class size. *Quarterly Journal of Economics* 128, 249-285.
- Gelman, A., and G. Imbens (2014). Why high-order polynomials should not be used in regression discontinuity designs, NBER Working Paper No. 20405
- Glaeser, E. (1996). The incentive effects of property taxes on local government. *Public Choice* 89, 93-111.
- Gunderson, M. (2005): Two faces of union voice in the public sector. *Journal of Labor Research* 26, 393-413.
- Hanushek, E. A. (1996). School resources and student performance, in G. Burtless (ed.), *Does money matter? The effect of school resources on student achievement and adult success*, Brookings Institution Press, Washington, DC.
- Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal* 113, F64-F98.
- Hanushek, E. A. (2006). School resources, in Hanushek, E. A. and F. Welch (eds.). *Handbook of the Economics of Education, Volume 2*, Elsevier B.V.
- Haraldsvik, M. (2012). Does performance based school choice affect student achievement? In M. Haraldsvik: Influences on Educational Outcomes. *Doctoral theses at NTNU* 2012:346.
- Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research* 43, 387-425.
- Heinesen, E. (2005). School district size and student educational attainment: Evidence from Denmark. *Economics of Education Review* 24, 677-689.
- Heinesen, E. (2010). Estimating Class-size Effects using Within-school Variation in Subject-specific Classes. *Economic Journal* 120, 737-760
- Hoxby, C. M. (1996). How teachers' unions affect education production. *Quarterly Journal of Economics* 111, 671-718.
- Hoxby, C. M. (1999). The productivity of schools and other local public goods producers. *Journal of Public Economics* 74, 1-30.
- Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics* 115, 1239-1285.

- Hægeland, T., O. Raaum and K. G. Salvanes (2012). Pennies from heaven? Using exogenous tax variation to identify effects of school resources on pupil achievement. *Economics of Education Review* 31, 601-614.
- Iversen, J. M. V., and H. Bonesrønning (2013). Disadvantage students in the early grades: Will smaller classes help them? *Education Economics* 21, 305-324.
- Jeynes, W. (2012). A meta-analysis of the efficacy of different types of parental involvement programs for urban students. *Urban Education* 47, 706-742
- Koerselman, K. (2013). Incentives from curriculum tracking. *Economics of Education Review* 32, 140-150.
- Krueger, A. (2003). Economic Considerations and Class Size. *The Economic Journal* 113, 34-63.
- Krueger, A. B., and D. M. Whitmore (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project STAR. *Economic Journal* 111, 1-28.
- Lavecchia, A. M., H. Liu and P. Oreopoulos (2014). Behavioral economics of education: Progress and possibilities. NBER Working Paper 20609.
- Lazear, E. P. (2001). Educational production. *Quarterly Journal of Economics* 115, 777-803.
- Lee, D. N (2013). The impact of repealing Sunday closing laws on educational attainment. *Journal of Human Resources* 48, 287-310.
- Lee, D. S., and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature* 48, 281-355
- Leuven, E., and S. A. Løkken (2015). Long term impacts of class size in compulsory schooling. Mimeo.
- Leuven, E., H. Oosterbeek and M. Rønning (2008). Quasi-experimental estimates of the effect of class size on achievement in Norway. *Scandinavian Journal of Economics* 110, 663-693.
- 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.
- Loeb, S., and K. Strunk (2007). Accountability and local control: Response to incentives with and without authority over resource generation and allocation. *Education Finance and Policy* 2, 10-39.
- Lovenheim, M. F. (2009). The effect of teachers' unions on education production: evidence from union election certifications in three midwestern states. *Journal of Labor Economics* 27, 525-587
- Lott, J., and L. W. Kenny (2013). State teacher union strength and student achievement. *Economics of Education Review* 35, 93-103.

- Mensah, Y. W., M. P. Schoderbek and S. P. Sahay (2013). The effect of administrative pay and local property taxes on student achievement scores: Evidence from New Jersey public schools. *Economics of Education Review* 34, 1-16.
- Moe, T. M. (2001). A union by any other name. *Education Next* 1, 40-45.
- Moe, T. M. (2011). *Special interest: Teachers unions and America's public schools*. Washington DC: Brookings Institution Press.
- Mueller, S. (2013). Teacher experience and the class size effect - experimental evidence. *Journal of Public Economics* 98, 44-52.
- Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives* 20, 111-132.
- OECD (2011). OECD Reviews of evaluation and assessment in education: Norway. <http://www.oecd.org/norway/48632032.pdf>
- Rattsø, J., and R. J. Sørensen (2004). Public employees as swing voters: Empirical evidence on opposition to public reform. *Public Choice* 119, 281-310.
- Reiling, R. B., and B. Strøm (2015). Upper secondary school completion and the business cycle. *Scandinavian Journal of Economics* 117, 195-219.
- Rockoff, J. (2009). Field experiments in class size from the early twentieth century. *Journal of Economic Perspectives* 23, 211-230.
- Strunk, K. O., and J. A. Grissom (2010). Do strong unions shape district policies? Collective bargaining, teacher contract restrictiveness, and the political power of teachers unions. *Educational Evaluation and Policy Analysis* 32, 389-406.
- Urquiola, M., and E. Verhoogen (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review* 99, 179-215.
- Webbink, D. (2005). Causal effects in education. *Journal of Economic Surveys* 19, 535-560.
- West, M. R., and L. Wößmann (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review* 50, 695-736.

Table 1. Descriptive statistics

	Observations	Mean	Standard deviation
Outcome variables			
Log of income	903,828	12.715	0.765
Years of education	952,514	13.986	2.536
Class size variables			
Average class size grade 8-10	952,514	24.41	3.79
Predicted class size	952,514	24.93	3.98
Enrollment grade 8	952,514	87.49	43.98
Background variables			
Girl	952,514	0.490	0.500
Parental education: High School	952,514	0.546	0.498
Parental education: Bachelor	952,514	0.202	0.401
Parental education: Masters +	952,514	0.077	0.267
Parental education: Unknown	952,514	0.031	0.172
First generation immigrant	952,514	0.013	0.111
Second generation immigrant	952,514	0.006	0.076
Only mother working	952,514	0.172	0.378
Only father working	952,514	0.152	0.359
Both parents working	952,514	0.348	0.476
Birth month	952,514	6.342	3.335
School district variables			
Share of teachers with teacher certification (teacher quality)	893,546	0.960	0.039
Have school choice in upper secondary education	379,691	0.494	0.500
Unemployment rate	952,218	0.025	0.013
Have property tax	283,322	0.379	0.485
Share of the labor force employed in the public sector	563,570	0.221	0.067
Population size	952,218	60,496	114,704
District merger: Treatment school district	952,218	0.065	0.247
District merger: Treatment school district * post-merger	952,218	0.023	0.149

Note: Descriptive statistics corresponding to the estimation sample for years of education.

Table 2. Balancing sample

	(1)	(2)	(3)	(5)
	Log income	Years of education	Predicted class size	p-value
Girl	-0.336*** (0.0033)	0.545*** (0.0081)	-0.0013 (0.0046)	0.768
Parental education: High School	0.114*** (0.0025)	1.049*** (0.0098)	-0.0127 (0.0080)	0.010
Parental education: Bachelor	0.163*** (0.0033)	2.391*** (0.0121)	0.0039 (0.0090)	0.017
Parental education: Masters +	0.176*** (0.0051)	3.349*** (0.0158)	-0.0069 (0.0112)	0.986
Parental education: Unknown	0.009 (0.0062)	0.750*** (0.0250)	-0.0002 (0.0158)	0.685
First generation immigrant	-0.058*** (0.0104)	-0.032 (0.0366)	0.0012 (0.0339)	0.973
Second generation immigrant	0.030** (0.0147)	0.455*** (0.0438)	0.0955 (0.0682)	0.157
Only mother working	0.042*** (0.0026)	0.129*** (0.0092)	-0.0035 (0.0069)	0.185
Only father working	0.039*** (0.0026)	0.054*** (0.0094)	-0.0014 (0.0071)	0.400
Both parents working	0.108*** (0.0024)	0.458*** (0.0092)	0.0099 (0.0069)	0.081
Birth month	0.0007*** (0.0002)	0.006*** (0.0007)	0.0001 (0.0007)	0.876
Observations	903,828	952,514	952,514	
R-squared	0.107	0.151	0.305	
Number of grunn_id	1,156	1,156	1,156	
p-value of F-test	0	0	0.0809	

Note. Columns (1)-(3) report results of OLS regressions on the variables listed in the rows, where predicted class size is our class size instrument. These regressions also include the following control variables: fixed effects for enrollment segment, enrollment to the fourth polynomial, and time/age fixed effects. Independent variables are pre-determined parent and student characteristics. The p-value reported at the bottom of columns (1)-(3) is for an F-test of the joint significance of the variables listed in the table. Each row of column (4) reports a p-value from separate OLS regressions of the pre-determined variable (listed in the corresponding row) on the instrument, and the same set of control variables as in columns (1)-(3). Estimates in column (3) and (4) correspond to the sample used for educational attainment. The p-value is for a t-test of the significance of the class size instrument. Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors are clustered at the school level.

Table 3. Specification analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable: Log income										
Average class size grade 8-10	0.00035 (0.0005)	-0.00033 (0.0005)	-0.00077 (0.0007)	-0.00087 (0.0007)	-0.00029 (0.0007)	-0.00048 (0.0007)	-0.00046 (0.0007)	-0.00099 (0.0011)	-0.00002 (0.0021)	0.00025 (0.0030)
F-value first stage	-	4,843	2,391	2,327	1,975	1,976	1,864	911.4	295.9	146.1
Observations	903,828	903,828	903,828	903,828	903,828	903,828	903,828	903,828	170,604	170,604
Dependent variable: Years of education										
Average class size grade 8-10	0.0158*** (0.0028)	0.0117*** (0.0032)	0.0118*** (0.0038)	0.0106*** (0.0038)	0.0017 (0.0023)	0.0007 (0.0021)	0.0003 (0.0021)	-0.0001 (0.0034)	-0.0012 (0.0069)	0.0089 (0.0103)
F-value first stage	-	4,827	2,386	2,322	1,974	1,975	1,863	909.0	294.7	145.4
Observations	952,514	952,514	952,514	952,514	952,514	952,514	952,514	952,514	179,799	179,799
Estimation method	OLS	IV	IV	IV	IV	IV	IV	IV	IV	IV
Enrollment controls	Linear	Linear	Linear and segment FE	Polynomial and segment FE	Polynomial and segment FE	Polynomial and segment FE	Linear and interacted with segment FE	Polynomial and interacted with segment FE	Linear with segment FE	Linear and interacted with segment FE
Time/age fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	No	No	No	No	Yes	Yes	Yes	Yes	No	No
Socioeconomic characteristics	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Sample	All	All	All	All	All	All	All	All	\tilde{E}_τ +/- 3 students	\tilde{E}_τ +/- 3 students

Note: Standard errors clustered at the school level in parentheses, * p<0.05, ** p<0.01, *** p<0.001. Socioeconomic characteristics include birth month, gender, immigration status, parental education, and parental employment status. Full model results for columns (6) and (9) are presented in Appendix Table A3

Table 4. Heterogeneous effects of class size

Outcome	(1)	(2)	(3)	(4)
	Log income		Years of education	
<u>Teacher quality</u>				
Interaction effect with class size	0.00005 (0.0006)	-0.00270* (0.0016)	-0.0022 (0.0018)	0.0042 (0.0050)
Average class size grade 8-10	-0.00041 (0.0007)	-0.00028 (0.0023)	0.00003 (0.0022)	0.0018 (0.0076)
F-value, first stage	598.0	100.00	597.1	101.0
Observations	849,163	159,830	893,546	168,182
<u>School choice upper secondary education</u>				
Interaction effect with class size	0.00048 (0.0022)	-0.00668 (0.0078)	0.0089 (0.0067)	-0.0186 (0.0257)
Average class size grade 8-10	-0.00008 (0.0013)	0.00244 (0.0041)	-0.0016 (0.0039)	-0.0031 (0.0143)
F-value, first stage	204.3	19.62	203.4	19.82
Observations	364,670	69,101	379,619	71,876
<u>Local unemployment rate</u>				
Interaction effect with class size	-0.00016 (0.0007)	-0.00244 (0.0022)	-0.0019 (0.0021)	-0.0119 (0.0082)
Average class size grade 8-10	-0.00031 (0.0007)	-0.00054 (0.0020)	0.00075 (0.0021)	0.0008 (0.0068)
F-value, first stage	287.9	17.26	292.0	17.25
Observations	903,572	170,601	952,218	179,796
<u>Property tax</u>				
Interaction effect with class size	0.00122 (0.0028)	-0.01050 (0.0089)	-0.0046 (0.0087)	0.0269 (0.0317)
Average class size grade 8-10	-0.00109 (0.0015)	0.00851* (0.0052)	0.0033 (0.0041)	-0.0194 (0.0155)
F-value, first stage	128.4	13.71	127.7	13.54
Observations	272,724	52,087	283,322	54,053
<u>Share of public sector employment</u>				
Interaction effect with class size	0.00054 (0.0008)	-0.00099 (0.0030)	-0.0032 (0.0026)	0.0121 (0.0096)
Average class size grade 8-10	-0.00012 (0.0009)	-0.00061 (0.0029)	0.0016 (0.0028)	0.0012 (0.0094)
F-value, first stage	375.8	14.03	370.6	13.75
Observations	539,693	101,333	563,569	105,745
<u>School district size; population</u>				
Interaction with class size	0.00125 (0.0009)	-0.00184 (0.0039)	-0.0006 (0.0022)	-0.0017 (0.0103)
Average class size grade 8-10	0.00002 (0.0007)	0.00012 (0.0023)	0.0001 (0.0021)	0.0007 (0.0076)
F-value, first stage	61.83	8.062	60.72	8.152
Observations	903,572	170,601	952,218	179,796
<u>School district size; merger</u>				
Interaction with class size (treatment school district * Post-merger * average class size)	0.00780 (0.0082)	0.02860 (0.0278)	-0.0136 (0.0222)	0.0780 (0.0676)
Average class size grade 8-10	-0.00024 (0.0007)	-0.00011 (0.0021)	0.0004 (0.0021)	-0.0018 (0.0070)
Average population in the school district during grade 8-10, standardized	-0.246*** (0.0473)	-0.0123*** (0.0030)	0.456*** (0.0828)	-0.0325 (0.0217)
F-value, first stage	9.456	1.687	9.562	1.659
Observations	903,572	170,601	952,218	179,796
Enrollment controls	Polynomial and segment FE	Linear and segment FE	Polynomial and segment FE	Linear and segment FE
School FE	Yes	No	Yes	No
Subsample +/- 3 students	No	Yes	No	Yes

Note. Standard errors clustered at the school level in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The model specifications are equal to the model specifications in column (6) and (9) in Table 3, except as indicated. Instruments for average class size in grade 8-10 and the interaction effect with class size is the class size rule in grade 8 and the interaction with the class size rule in grade 8.

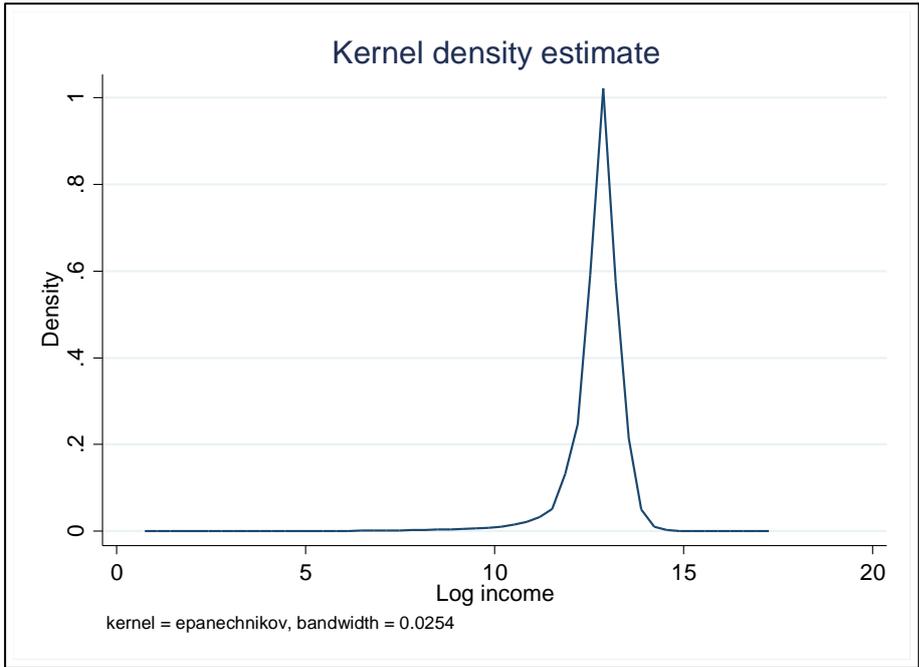


Figure 1: Distribution of log income conditional on cohort specific effects

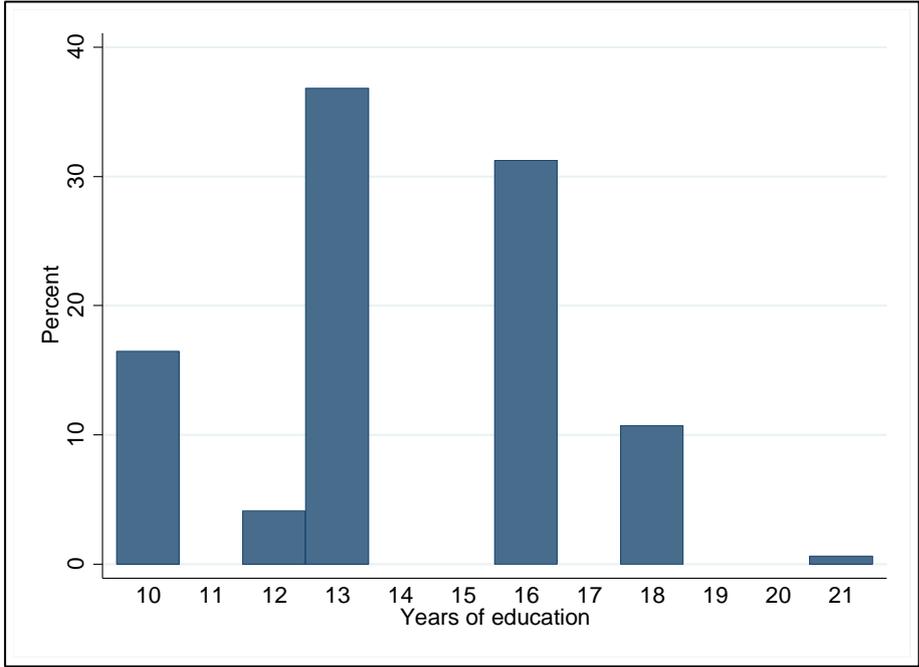


Figure 2: Distribution of years of education

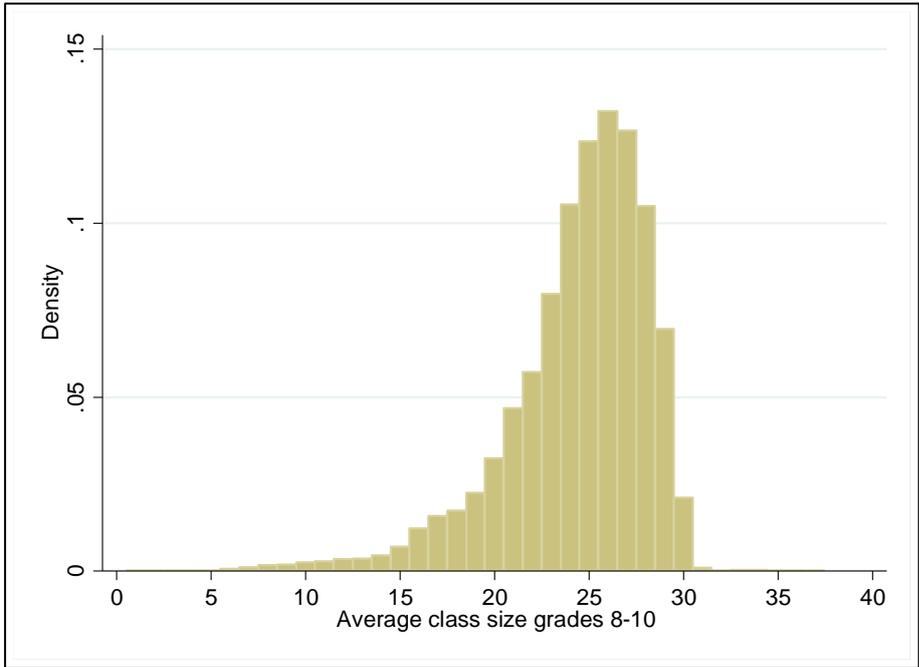


Figure 3. Average class in the empirical sample

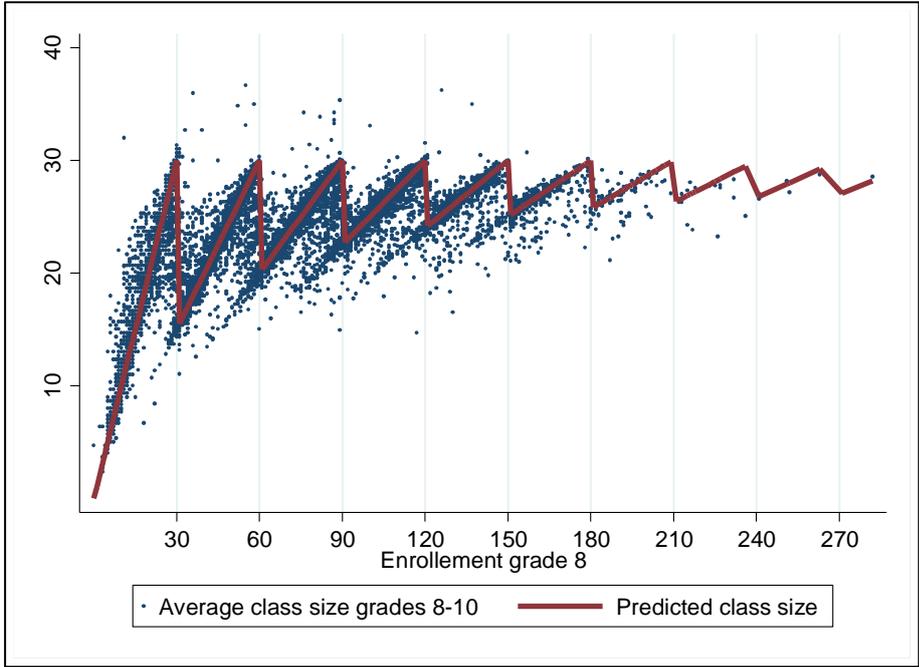
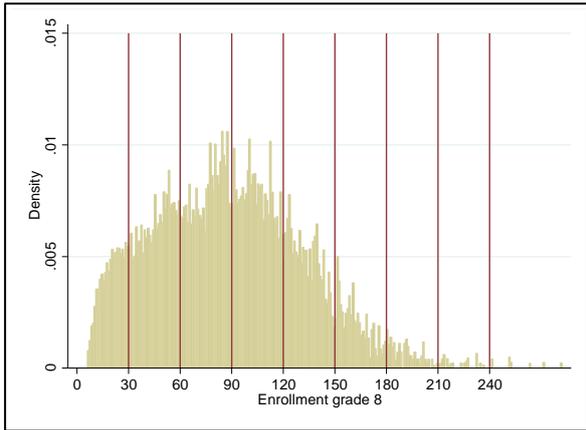
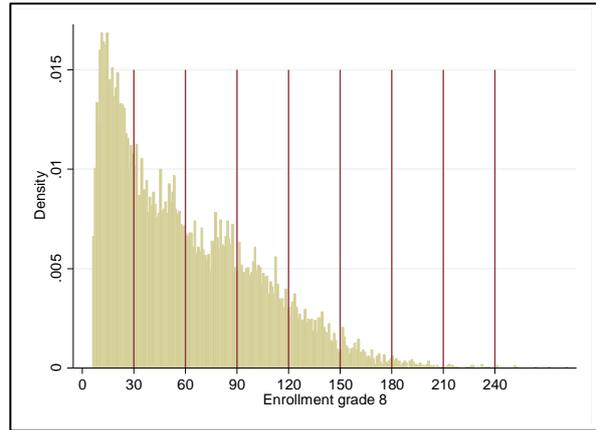


Figure 4. The first stage

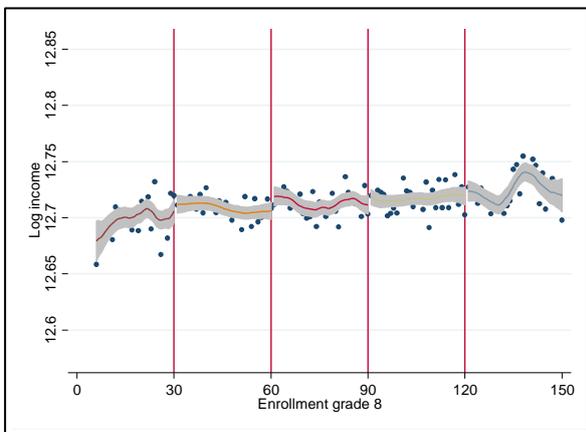


A. Individual level

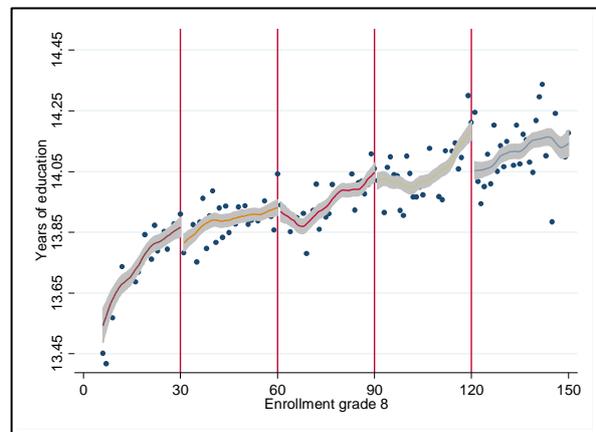


B. School level

Figure 5. Distribution of enrollment in grade 8 in the empirical sample

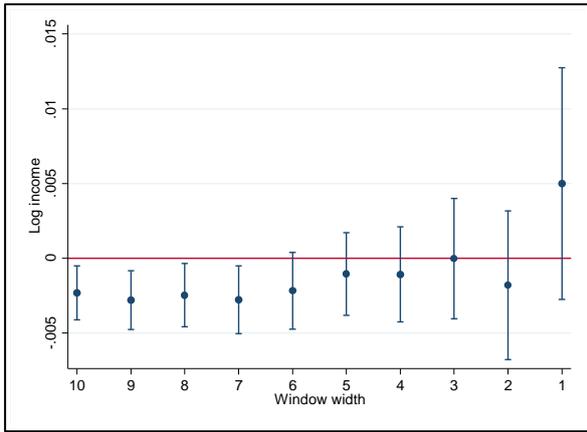


A. Log income

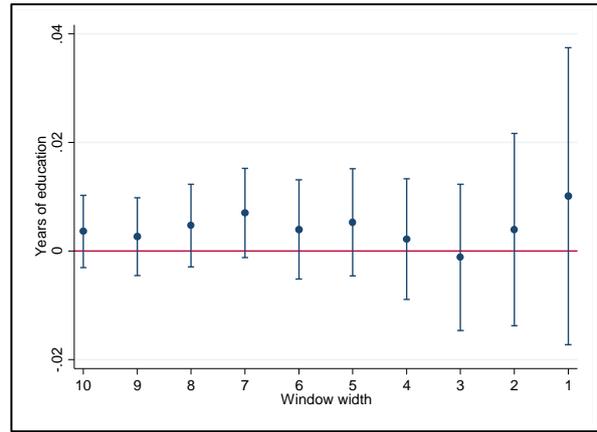


B. Years of education

Figure 6: Local polynomial regressions of enrollment in grade 8 on outcome variables for each segment. Log income and educational attainment are conditional on cohort specific effects. The markers indicate average outcome for each enrollment value. The y-axis is the mean value of the outcome variable +/- 0.2 standard deviations.

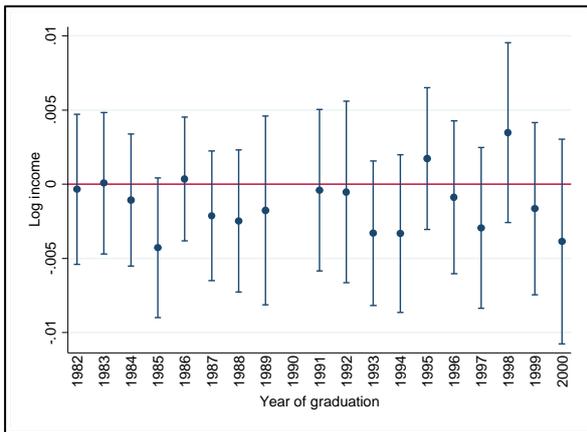


A. Log income

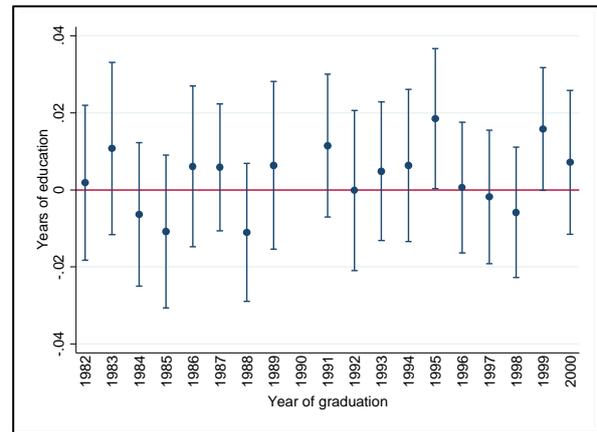


B. Years of education

Figure 7. Effect of class size with 95% confidence interval when reducing bandwidth from 10 to 1

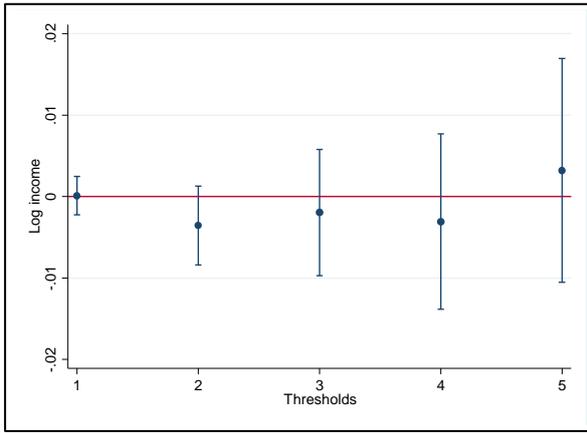


A. Log income

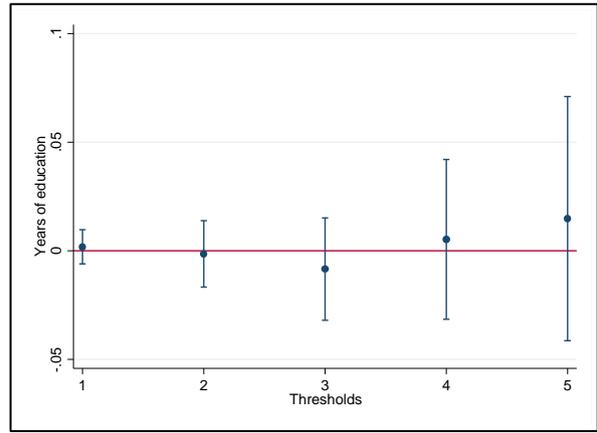


B. Years of education

Figure 8. Year specific estimates using the global approach with 95% confidence intervals



A. Log income



B. Years of education

Figure 9: Effect of class size with 95% confidence interval when running separate regressions for each threshold. Threshold 5 includes all thresholds from 5 and up.

Table A1. Data reduction

	Observations	Reduction	% Reduction
1. Sample 1982-2000 (without 1990) ¹	1,040,840		
2. Non-missing class size	1,003,149	37691	3,62 %
3. 16 years old when graduating from lower secondary school	953,512	49637	4,95 %
4. At least 10 school observations	953,183	329	0,03 %
5. Non missing years of education	952,514	669	0,07 %
5. Non missing log of income ²	903,828	49355	5,18 %

¹ Data on the school identifier is missing in 1990.

² 49,355 observations have zero income, which are excluded from the analysis because we use the logarithmic value of income.

Table A2. Main results with socioeconomic characteristics and enrollment controls

	Log income		Years of education	
	(1)	(2)	(3)	(4)
Average class size grade 8-10	-0.000396 (0.0007)	-2.35e-05 (0.0021)	0.000677 (0.0021)	-0.00117 (0.0069)
Girl	-0.336*** (0.0033)	-0.337*** (0.0052)	0.545*** (0.0081)	0.526*** (0.0148)
Parental education: High School	0.113*** (0.0025)	0.119*** (0.0054)	1.049*** (0.0098)	1.092*** (0.0182)
Parental education: Bachelor	0.163*** (0.0033)	0.161*** (0.0070)	2.391*** (0.0121)	2.495*** (0.0226)
Parental education: Masters +	0.176*** (0.0051)	0.181*** (0.0107)	3.349*** (0.0158)	3.485*** (0.0272)
Parental education: Unknown	0.00863 (0.0062)	0.0266** (0.0133)	0.750*** (0.0250)	0.831*** (0.0581)
First generation immigrant	-0.0575*** (0.0104)	-0.0642*** (0.0196)	-0.0318 (0.0365)	-0.184*** (0.0673)
Second generation immigrant	0.0305** (0.0146)	0.00807 (0.0318)	0.454*** (0.0438)	0.164 (0.1008)
Only mother working	0.0423*** (0.0026)	0.0349*** (0.0056)	0.129*** (0.0092)	0.107*** (0.0188)
Only father working	0.0393*** (0.0026)	0.0345*** (0.0055)	0.0535*** (0.0094)	0.0588*** (0.0189)
Both parents working	0.108*** (0.0024)	0.112*** (0.0050)	0.458*** (0.0091)	0.476*** (0.0172)
Birth month	0.000681*** (0.0002)	0.000235 (0.0005)	0.00602*** (0.0007)	0.00643*** (0.0017)
Enrollment	-0.00103 (0.0010)	0.00242 (0.0025)	0.00942*** (0.0033)	-0.0103 (0.0087)
Enrollment ²	1.59e-05 (0.0000)		-0.000117** (0.0001)	
Enrollment ³	-9.78e-08 (0.0000)		6.51e-07* (0.0000)	
Enrollment ⁴	2.09e-10 (0.0000)		-1.25e-09* (0.0000)	
Segment 1	-0.000288 (0.0094)	0.600 (0.6228)	0.0166 (0.0307)	-2.441 (2.1299)
Segment 2	0.00225 (0.0127)	0.530 (0.5421)	-0.0181 (0.0409)	-2.142 (1.8566)
Segment 3	0.00136 (0.0151)	0.458 (0.4635)	0.00627 (0.0472)	-1.802 (1.5851)
Segment 4	0.00230 (0.0172)	0.384 (0.3852)	0.0105 (0.0532)	-1.463 (1.3167)
Segment 5	0.0162 (0.0195)	0.304 (0.3079)	-0.0220 (0.0607)	-1.197 (1.0556)
Segment 6	0.0407* (0.0241)	0.243 (0.2312)	-0.0478 (0.0747)	-0.934 (0.7955)
Segment 7	0.0383	0.199	-0.113	-0.505

	(0.0307)	(0.1515)	(0.0985)	(0.5504)
Segment 8	0.00253	0.0722	-0.0265	-0.0854
	(0.0467)	(0.0771)	(0.1530)	(0.2659)
Segment 9	-0.0374		0.323	
	(0.0861)		(0.2947)	
R-squared	0.107	0.110	0.151	0.172
Predicted class size (the instrument), first stage	0.56*** (0.013)	0.41*** (0.023)	0.56*** (0.013)	0.40*** (0.023)
F-value first stage	1,935	295.9	1,934	294.7
R-squared first stage	0.4906	0.4864	0.4893	0.4855
Observations	903,828	170,604	952,514	179,799
No. of schools	1,156		1,156	
Enrollment controls	Polynomial and segment FE	Linear and segment FE	Polynomial and segment FE	Linear and segment FE
Sample	All		All	
Socioeconomic characteristics	Yes	Yes	Yes	Yes
Time/age FE	Yes	Yes	Yes	Yes
School FE	Yes	No	Yes	No

Note: Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors are clustered at the school level. Socioeconomic characteristics are described in section 3.1.

Table A3. Subsample analysis

Outcome	(1)	(2)	(3)	(4)
	Log income		Years of education	
<u>Girls</u>				
Average class size grade 8-10	0.000188	0.00227	-0.000529	0.00411
	(0.0009)	(0.0027)	(0.0029)	(0.0089)
Observations	443,057	83,593	466,957	88,101
<u>Boys</u>				
Average class size grade 8-10	-0.000916	-0.00186	0.00279	-0.00481
	(0.0009)	(0.0027)	(0.0029)	(0.0091)
Observations	460,770	87,011	485,557	91,698
<u>Parental education more than high school</u>				
Average class size grade 8-10	-0.000730	-0.000319	-0.000983	-0.00162
	(0.0010)	(0.0029)	(0.0013)	(0.0042)
Observations	392,413	74,591	405,286	77,021
<u>Parental education less than high school</u>				
Average class size grade 8-10	-0.000376	0.000225	-0.00170	0.00330
	(0.0008)	(0.0024)	(0.0016)	(0.0048)
Observations	511,412	96,013	547,225	102,778
<u>Immigrant</u>				
Average class size grade 8-10	-0.00764	0.00465	-0.000724	-0.0357
	(0.0073)	(0.0165)	(0.0195)	(0.0443)
Observations	15,206	2,977	17,427	3,364
<u>Non immigrant</u>				
Average class size grade 8-10	-0.000421	-0.000142	0.00116	-0.000582
	(0.0007)	(0.0021)	(0.0021)	(0.0069)
Observations	888,502	167,627	934,968	176,435
Enrollment controls	Polynomial and segment FE	Linear and segment FE	Polynomial and segment FE	Linear and segment FE
School FE	Yes	No	Yes	No
Subsample 3+/-	No	Yes	No	Yes