OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 79, 5 (2017) 0305–9049
doi: 10.1111/obes.12161

Do Smaller Classes Always Improve Students' Long-run Outcomes?*

TORBERG FALCH†, ASTRID MARIE JORDE SANDSØR‡ and BJARNE STRØM†

†*Department of Economics, Norwegian University of Science and Technology Dragvoll, N-7491 Trondheim, Norway (e-mails: torberg.falch@svt.ntnu.no; bjarne.strom@svt.ntnu.no)*
‡*Department of Economics, University of Oslo, Postboks 1095 Blindern, N-0317 Oslo, Norway (e-mail: a.m.j.sandsor@econ.uio.no)*

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 labour market conditions. Overall, we find that the class size effect does not depend on these factors measured at the school district level. 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.

I. Introduction

The impact of school resources on student performance has been disputed since the publication of the Coleman *et al.* (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

JEL Classification numbers: I2, H7.
*We greatly acknowledge comments from two referees, Karl Ove Moene and participants at the Workshop on Applied Economics of Education in Catanzaro, the CESifo Area Conference on Economics of Education, and seminars at University of Lancaster and the Institute of Social Research in Oslo.

[1] Summaries of the literature on the relationship between school resources and student achievement include Hanushek (1996, 2003, 2006), Krueger (2003) and Webbink (2005).

[2] See Krueger and Whitmore (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

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.

Test scores only measure cognitive skills, while class size may 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, Battistin and Vuri (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.

Three recently published papers analyse long-run effects of class size. Chetty *et al.* (2011) and Dynarski, Hyman and Schanzenbach (2013) study long-run outcomes for participants in the STAR experiment, while Fredriksson, Oosterbeek and Öckert (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, which suggests that the mixed effects in the literature on short run effects are related to imperfect measurement of student skills.[5] However, the findings for the long run are also consistent with the findings in the short run, using test scores within the same contexts.[6] 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. These results motivate studies on long-run outcomes from contexts where the evidence indicates no class size effect on short-run outcomes.

In this paper, we estimate long-run effects of class size for Norway where previous research has not been able to provide evidence of short-run gains from smaller classes in terms of student achievement.[7] We investigate whether the class size effect in lower secondary education depends on characteristics of the environment in which the schools

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 Wößmann and West (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 Wößmann and West (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 Wößmann and West (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.

[5]Chetty *et al.* (2011) find positive effects of smaller classes on college attendance and earnings, but the latter is imprecisely estimated and insignificant.

[6]See for example Krueger and Whitmore (2001) for the STAR experiment. 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.

[7]The Norwegian studies exploiting the class size rule in short-run studies are Bonesrønning (2003), Woessmann (2005), Leuven, Oosterbeek and Rønning (2008), and Iversen and Bonesrønning (2013). They find small or zero average effects of class size. Hægeland, Raaum and Salvanes (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.

and students operate. Several hypotheses of heterogeneous effects across school districts are derived and tested. Leuven and Løkken (2015) explore similar Norwegian data, estimating the effect of class size both in primary and lower secondary education. Their analysis utilizes that some schools include grade 1–10 and estimates the effect of class size in primary education in addition to the effect in lower secondary education, assuming that the students stayed in the same school during all school years. We find qualitatively similar average effects of class size as they do.

The findings for short-run outcomes differ substantially between the Scandinavian countries Sweden, Denmark and Norway with apparently similar educational and labour 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.[8]

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. 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 Hawthorne 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 between 1966 and 1984, representing almost 1 million students and 1,150 schools with separate catchment areas.[9]

Secondly, information on the whole population of schools and students 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 labour 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 and the long-run effect in Leuven and Løkken (2015). Moreover, we find no evidence that class size effects vary with school district characteristics.

The paper is organized as follows. In section II, we present arguments as to why the effect of resources may depend on characteristics of the external environment in which schools and students operate. Section III describes the institutions and the data, while

[8] 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 Nusche *et al.* (2011) and Bonesrønning (2013) for a description of recent Norwegian reforms.

[9] Fredriksson *et al.* (2013) use data for a roughly 10% sample of the cohorts born 1967, 1972 and 1982, and a 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 6,000 students and 191 schools. Since we include all schools in our analysis, the statistical power in robustness and heterogeneity analyses should be higher.

the identification approach and model specification are presented in section IV. Section V presents results from models estimating the causal average effect of class size on income and years of education, while section VI estimates interaction models where we investigate whether the class size effect depends on school district characteristics. Section VII includes a discussion of the findings in relation to the present literature, and some concluding comments are provided in section VIII.

II. Why might class size effects vary?

Class size may change student outcomes through a number of mechanisms affecting both student and teacher behaviour. 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, Levin and Oosterbeek (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.

Teacher quality

The class size effect might depend on teacher quality as argued by educationalists (Hattie, 2005) and economists (Wößmann and West, 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, Wößmann and West (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), see also Woessmann (2005). 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 analyse whether the class size effect depends on teacher supply conditions.

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, Liu and Oreopoulos (2014) extend this framework to

[10] Examples of studies incorporating student effort in human capital investment models through educational standards is Costrell (1994), Betts (1998), and Becker and Rosen (1992).

incorporate elements from behavioural 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 positive effect on test results only 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 postcompulsory school choice systems and external labour market conditions.

Postcompulsory school choice

A large and still growing literature analyses 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 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 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 (2013) 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 postcompulsory education.

External labour market conditions

Several studies find that student opportunity costs in terms of foregone earnings during schooling and returns to schooling are important determinants of educational attainment. Clark (2011) finds a positive effect of regional unemployment on high school enrolment 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, labour market conditions may also affect the student's allocation of time and effort during compulsory education. If student effort and class size are complementary inputs in the educational production function, lower class size is less (more) likely to increase student performance when student effort is low (high). Moreover, if student effort varies countercyclically, this suggests that class size reductions can be more productive during economic downturns than during upturns.

Direct student incentives may be modified by cyclical variation in parental input during compulsory education. Parents' investment in children's education can vary with external labour market conditions. The income effect suggests that parental involvement is higher in an economic upturn, while the substitution effect goes in the opposite direction.

While the direction of the effect of external labour market conditions is in general ambiguous, the upshot of the discussion above is that class size effects can potentially depend systematically on labour 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.

Fiscal constraints

Property taxation can work as a discipline device on decision-makers in local governments as argued in the theoretical contributions by Glaeser (1996) and Hoxby (1999). The main mechanism in their papers is that lower slack improves the quality of local services, makes the local government more attractive and increases the property tax base through capitalization in the housing market. Fischel (2001) extended this literature and introduced the home-voter hypothesis. The argument is that property taxation mainly works as a discipline device through the direct effect of service quality on property values for homeowner voters in a local government. According to this view, property tax financing of schools and other public service providers increases the incentives for voters and hence elected politicians to put pressure on agents to produce public sector goods and services in an efficient way.

The prediction from these models is thus that services are produced more efficiently in terms of higher production and lower costs in local governments financed by local property taxes than in other local governments. There is some empirical support for this theoretical prediction. Separating between municipalities with and without property taxation, Fiva and Rønning (2008) find that student achievement measured by grades in lower secondary education is higher in Norwegian municipalities with local property taxation. Borge and Rattsø (2008) provide evidence that local property taxation reduces unit costs in utility services in Norwegian municipalities. Barankay and Lockwood (2007) show that educational attainment is higher in Swiss cantons with local financing of education, while Falch and Fischer (2012) use cross-country data from international tests and find that student performance is higher in countries with decentralized financing of the public sector. Studies from the US suggest that local funding increases technical efficiency in schools (Adkins and Moomaw, 2003) and student performance (Mensah, Schoderbek and Sahay, 2013).

Most empirical studies have tested the hypothesis that property tax financing increases outcomes and decreases costs due to the disciplining mechanism. A related prediction from theory is that schools located in districts with property taxation are better able to take advantage of a sudden exogenous decrease in class size to increase student performance. Our rich register data from Norway combined with the strict class size rule enables us to test this hypothesis. A related hypothesis of nonlinearity in the effect of accountability policies is investigated in Loeb and Strunk (2007). They find that accountability is most effective in the US states with strong local control in terms of local funding and local autonomy in hiring and spending decisions.

Interest groups

Chubb and Moe (1988) and Moe (2001, 2011) argue that teacher unions reduce the power of politicians to implement necessary educational reforms and to use school resources efficiently. Others argue that teacher unions enhance efficiency by increasing teachers' job satisfaction and productivity, see Gunderson (2005) for a general discussion of union voice effects in the public sector.

Hoxby (1996) provides an explicit theoretical discussion of the different approaches along with an empirical study of the effect of teacher unions on student performance and school resources. She distinguishes between two possible types of teacher unions: The efficiency enhancing and the rent maximizing union. The efficiency enhancing union has the same objective function as parents and voters and seeks to maximize student performance. However, the teacher union has superior information about the actual educational production function and can allocate resources in a more efficient way than non-unionized districts. In contrast, the rent maximizing union has an objective function that differs from parents and voters and seeks to maximize the personal wellbeing of the teachers by seeking high salary, low effort and pleasant working conditions. The theory predicts that an efficiency enhancing teacher union will increase student performance, while a rent maximizing union is likely to reduce student performance.

The evidence from US school districts in Hoxby (1996) suggests that teacher unions are able to increase the teacher-student ratio, but also decrease the productivity to such an extent that student performance declines, consistent with the rent-maximizing model. In a more recent study, 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.

While most empirical studies have focused on the effect of teacher unions on student performance, a related question raised here is to what extent schools located in unionized districts are more or less able to take advantage of an exogenous reduction in class size than schools located in non-unionized ones. The theory models above suggest that a class size reduction is more (less) likely to increase student performance if the union is of the efficiency enhancing (rent maximizing) type.

The ideal setting for an empirical assessment of this issue would be to utilize exogenous changes in class size across districts with different teacher unionization rates. Unfortunately, this is not possible in our empirical setting, since a large majority of teachers in Norwegian schools are members of a teacher union and the unionization rate of teachers does not vary much across schools and school districts.

While unionization rates do not vary much across districts, the power of teacher unions is likely to depend on the political setting in which they operate, i.e. by their ability to build coalitions in the government or directly affect the behaviour 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 more reluctant to implement a major accountability education reform in Norway in the period 2004–06.[11] If teacher unions are more (less) powerful in school districts with high (low) shares of public employment, it makes sense to investigate whether the impact of class size on student outcomes depends on the share of public employment in the school district.

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. For example, Driscoll, Halcoussis and Svorny (2003) find that test scores are negatively related to district size in California. Using Danish data, Heinesen (2005) concludes that educational attainment is higher for students from larger districts, i.e. districts with population above 15,000. Berry and West (2010) exploit variation in the timing of consolidation across US states and find that larger districts have some modest gains with respect to returns to education. We investigate whether there is a larger return to small class size in large school districts, which are more similar to the typical school district size in Sweden and Denmark.

III. Institutions and data

Institutions

Compulsory education in Norway consists of primary schools and lower secondary schools, and ends by grade 10 the year the students turn 16 years of age.[12] Most students continue on to upper secondary education, which is divided into a 3-year long academic study track and different vocational study tracks. After a major reform in 1994, vocational study tracks typically last for 4 years (including 2 years of apprenticeship training). Acceptance to an 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

[11] Anzia (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 favourable 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] 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 today primary education consists of grades 1–7 (ages 6–13) and lower secondary education consists of grades 8–10 (ages 14–16). We refer to grades 8–10 as lower secondary education throughout the paper.

on-time.[13] The education is comprehensive with no tracking and a common curriculum for all students. The cutoff between grades is birth at 1 January.

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.[14] 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 socio-economic composition, and it is very uncommon to change the composition of classes unless the number of classes changes.

Data

In this paper, we study the cohorts born 1966–84 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 with the school 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 empirical model. 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,[15] 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

[13] 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 enrolment 1 year. In addition, some older students return to improve their grades, and immigrants are often over-aged at graduation.

[14] 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.

[15] We use the pension-qualifying income as reported in the tax registry. This income measure is not top coded and includes labour income, taxable sick benefits, unemployment benefits, parental leave payments, and pensions, see Black, Devereux and Salvanes (2013, p. 132). Information after 2010 is not available in our data.

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.[16]

The distributions of the dependent variables are presented in Figures 1 and 2. The average years of education is 14.0 with SD of 2.5, while log of income has mean 12.7 with SD of 0.8 (Table 1).[17]

Data on the number of classes and enrolment by year and grade are obtained from a national school register administered by The Norwegian Ministry of Education. Variables are measured on 1 October 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, while Table 1 provides descriptive statistics. 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).

Our individual register data contains information on gender, birth month and immigration status, parental education and parental employment status the year the individual turns 16, as well as detailed data on educational attainment and income for all years after the individual leaves lower secondary education and up to 2011.[18] Descriptive statistics of the socio-economic characteristics are presented in Table 1.[19]

IV. 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

[16] These operationalizations differ somewhat from Leuven and Løkken (2015). While we measure the outcomes the last year with available data, creating a cross-section on individual data, Leuven and Løkken use yearly data, creating an individual panel-data. We include the age span 25–45 years, while Leuven and Løkken use the age span 27–43 years. In addition, while we only include individuals graduating compulsory education at the normal age, Leuven and Løkken include all students and control for age at graduation in the regression model.

[17] The specification of the two outcome variables is inspired by Fredriksson *et al.* (2013). They use years of education measured in 2009 and average earnings during 2007–09 for the age span 27–42 years. According to a working paper version of Fredriksson *et al.* (2013), the average value and standard deviation of years of education in their sample are almost exactly similar to ours. Regarding earnings, Fredriksson *et al.* (2013) use income statements made by employers, while we include welfare benefits. Consequently, there is a much lower number of individuals with zero income in our sample than in Fredriksson *et al.* (2013). They use both the level of earnings and the log of the wage for wage earners, where the sample is almost halved in the latter case. The standard deviation of our income measure is smaller than for the level of earnings in Fredriksson *et al.* (2013), most likely because we do not include individuals with zero income, but is three times larger than for the log of the wage in Fredriksson *et al.* (2013), most likely because we include non-employed. Since Fredriksson et al. find significant effects on all their outcomes related to education and income, different variability of the outcomes cannot explain the different results in our paper.

[18] Regarding immigration status, we distinguish between first and second generation immigrants, where the former are born abroad and have both parents born abroad, while the latter are born in Norway and have both parents born abroad.

[19] Descriptive statistics on the school district characteristics used in the heterogeneity analysis are also presented in Table 1. These variables are described in section VI below.

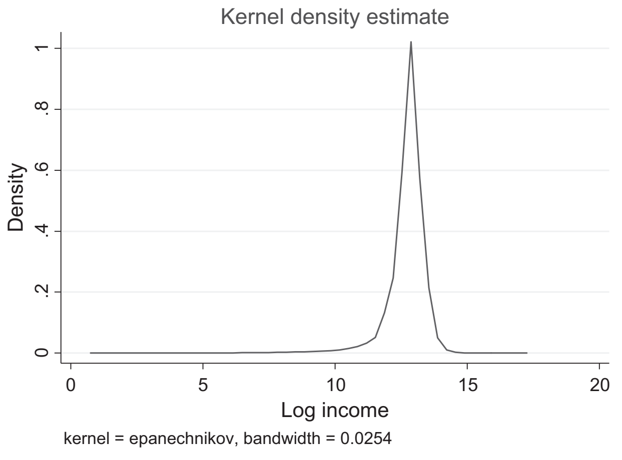

Figure 1. Distribution of log income conditional on cohort specific effects

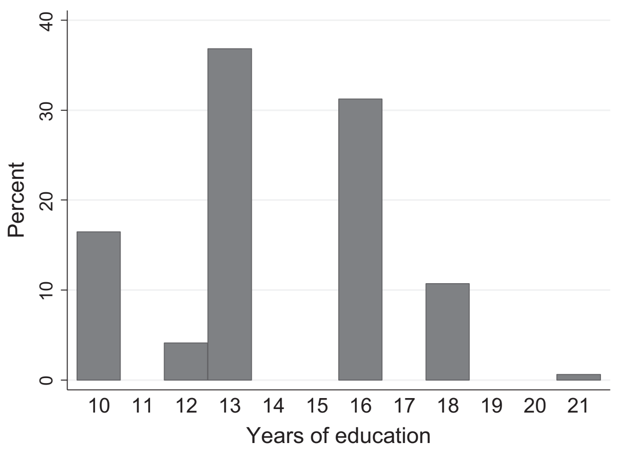

Figure 2. Distribution of years of education

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 wants 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.

The class size rule

During the time period we study, a national rule was in place saying that class size could not surpass 30 students 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 8–10) and not the class size in one specific school year. The

TABLE 1

Descriptive statistics

	Observations	*Mean*	*SD*
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
Enrolment grade 8	952,514	87.49	43.98
Socio-economic characteristics			
Girl	952,514	0.490	0.500
Mother's education: Less than high school	952,514	0.327	0.469
Mother's education: High School	952,514	0.485	0.500
Mother's education: Bachelor	952,514	0.163	0.369
Mother's education: Masters +	952,514	0.015	0.122
Mother's education: Unknown	952,514	0.010	0.098
Father's education: Less than high school	952,514	0.265	0.441
Father's education: High School	952,514	0.498	0.500
Father's education: Bachelor	952,514	0.141	0.348
Father's education: Masters +	952,514	0.074	0.262
Father's education: Unknown	952,514	0.022	0.147
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
None of parents working	952,514	0.328	0.475
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 taxation	283,322	0.379	0.485
Share of the labour 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 × postmerger	952,218	0.023	0.149

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

students are 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.[20] For each grade level the data contain the number of classes and the number students enrolled.

We follow Leuven *et al.* (2008) and use predicted class size based on enrolment in grade 8, 2 years prior to graduation, as the instrument in the analysis in order to avoid biased

[20]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.

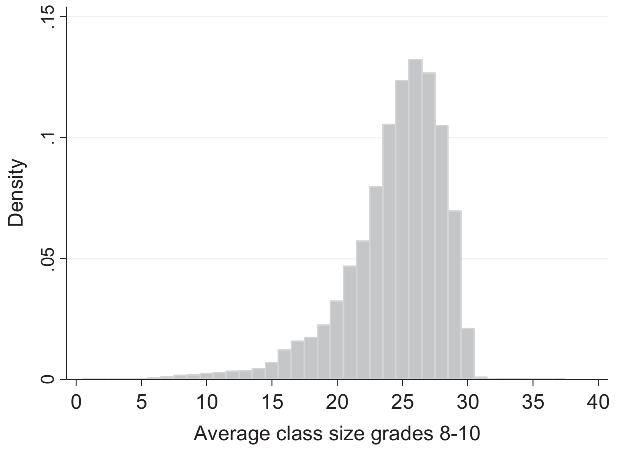

Figure 3. Average class in the empirical sample

estimates due to possible endogenous mobility of students across schools during the years in lower secondary education. The instrument is given by

$$\mathrm{CS}_{t-2}^{\mathrm{rule}} = \frac{E_{t-2}}{\mathrm{int}\left(1 + \left(E_{t-2} - 1\right)/\mathrm{CS}^{\max}\right)} \tag{1}$$

where E_{t-2} is enrolment in grade 8 and $\mathrm{CS}^{\max}$ is the maximum class size according to the rule. Using this formula, the 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 non-monotonic relationship between enrolment 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 enrolment.

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

One possible threat to the validity of the instrument is manipulation of enrolment around the thresholds. Urquiola and Verhoogen (2009) find 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 enrolment in grade 8, where the vertical lines represent the class size thresholds. There is no evidence of manipulation of the enrolment. The density of observations just below and above the thresholds is similar. In fact, the enrolment 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 enrolment of 30, 60, 90 and 120 students who 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 enrolment above 150 students in grade 8. Most schools have enrolment

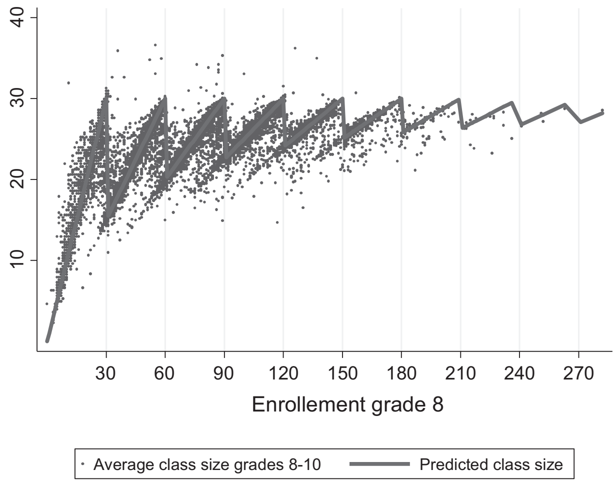

Figure 4. The first stage

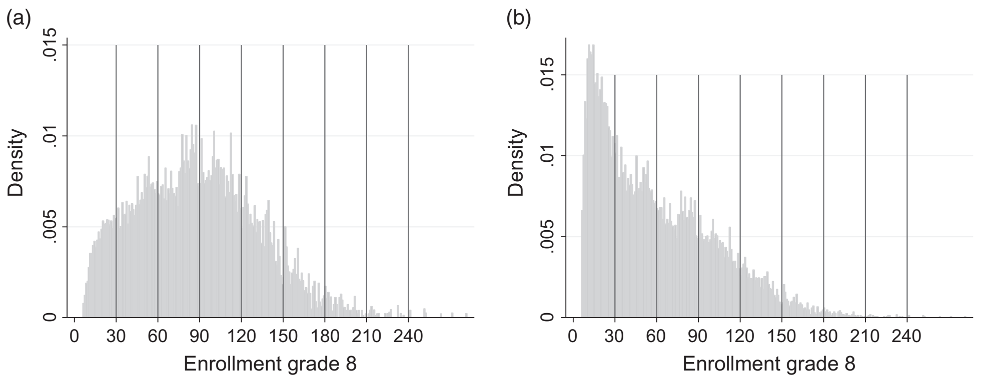

Figure 5. Distribution of enrolment in grade 8 in the empirical sample. (a) Individual level; (b) School level

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 class size rule, i.e. the instrument, is valid is to examine whether socio-economic characteristics are equal across observations above and below the class size thresholds. Table 2 tests whether the covariates are correlated with the class size rule, both individually and jointly.

The first column in Table 2 presents P-values for pairwise correlations between the class size rule and each socio-economic characteristic. None of the 15 correlations are significant at the 5% level, while two are significant at the 10% level (Father's education is bachelor degree and Both parents are working). Column (2) presents results for a regression on the class size rule, using the control function for enrolment described below and including time/age fixed effects. None of the socio-economic characteristics are significant at conventional levels, and the test for joint significance has a P-value of 0.36. The last two columns in Table 2 simply show that even though the socio-economic variables are unrelated to the class size rule, they are strongly related to the outcomes income and

TABLE 2

Balancing of socio-economic characteristics

	(1) P-value	(2) Predicted class size	(3) Log income	(4) Years of education
Girl	0.768	−0.00135	−0.336***	0.547***
		(0.0046)	(0.0033)	(0.0081)
Mother's education: High School	0.342	−0.00594	0.0783***	0.800***
		(0.0060)	(0.0019)	(0.0070)
Mother's education: Bachelor	0.329	−0.00161	0.0715***	1.589***
		(0.0083)	(0.0030)	(0.0103)
Mother's education: Masters+	0.568	−0.0166	0.0141	1.955***
		(0.0185)	(0.0101)	(0.0222)
Mother's education: Unknown	0.582	−0.0372	−0.0230**	0.0734**
		(0.0309)	(0.0106)	(0.0345)
Father's education: High School	0.541	0.00219	0.0722***	0.699***
		(0.0065)	(0.0021)	(0.0072)
Father's education: Bachelor	0.0837	0.0130	0.104***	1.563***
		(0.0086)	(0.0033)	(0.0103)
Father's education: Masters+	0.991	0.00434	0.121***	2.150***
		(0.0113)	(0.0046)	(0.0137)
Father's education: Unknown	0.556	0.0176	−0.0468***	0.129***
		(0.0175)	(0.0069)	(0.0219)
First generation immigrant	0.973	0.00823	−0.0478***	0.00458
		(0.0315)	(0.0104)	(0.0381)
Second generation immigrant	0.157	0.102	0.0431***	0.580***
		(0.0661)	(0.0146)	(0.0442)
Only mother working	0.185	−0.00401	0.0398***	0.0850***
		(0.0069)	(0.0026)	(0.0089)
Only father working	0.400	−0.00213	0.0361***	0.0266***
		(0.0071)	(0.0026)	(0.0091)
Both parents working	0.081	0.00930	0.101***	0.359***
		(0.0069)	(0.0024)	(0.0087)
Birth month	0.876	7.10e-05	0.000755***	0.00643***
		(0.0007)	(0.0002)	(0.0007)
Observations	–	952,514	903,828	952,514
R^2	–	0.305	0.107	0.151
P-value of F-test	–	0.364	0	0
Estimation method	–	OLS	OLS	OLS
Enrolment controls	–	Polynomial and segment FE	Polynomial and segment FE	Polynomial and segment FE
Time/age fixed effects	–	Yes	Yes	Yes
School fixed effects	–	Yes	Yes	Yes

Notes: All regressions include the following control variables: fixed effects for enrolment segment, enrolment to the fourth polynomial, time/age fixed effects and school fixed effects. Independent variables are pre-determined parent and student characteristics. Each row of column (1) reports a P-value from separate OLS regressions of the pre-determined variable (listed in the corresponding row) on the instrument. Column (2) reports the result of an OLS regressions on the variables listed in the rows, where predicted class size, our class size instrument, is the dependent variable. Columns (3) and (4) report results from a similar regression where the outcomes, log of income and educational attainment are the dependent variables. The P-value reported at the bottom of columns (2)–(4) is for an F-test of the joint significance of the variables listed in the table. Estimates in column (1) and (2) correspond to the sample used for educational attainment. Standard errors clustered at the school level in parentheses, $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$.

educational attainment as expected. For example, females have 0.34 log points lower income and 0.5 more years of education than males, in the age span 25–45 years that we have in the data. Overall, the socio-economic characteristics are unrelated to the class size rule despite being strong predictors of individuals' long-run outcomes.

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 discussions 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 cohort fixed effects is identical to including age fixed effects in our application.

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

$$y_{ist} = \alpha\widehat{\overline{\text{CS}}}_{st} + f\left(E_{st-2}\right) + \beta X_i + \delta_t + \varepsilon_{ist} \tag{2}$$

y_{ist} denotes the outcome for individual i graduating from school s in year t and $\widehat{\overline{\text{CS}}}$ is the predicted average class size for grades 8–10. In addition, the model includes a flexible functional form of enrolment E in grade 8, 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

$$\overline{\text{CS}}_{st} = \alpha' \text{CS}^{rule}_{st-2} + f\left(E_{st-2}\right) + \beta' X_i + \delta'_t + \varepsilon'_{ist}. \tag{3}$$

When using the 'global' approach, a flexible modelling of enrolment effects in terms of the function $f(E_{st-2})$ is necessary in order to avoid that the discontinuity generated by the class size rule is confounded with a possible nonlinear relationship between the outcome variable and enrolment. Define the thresholds for the class size rule in grade 8 as $\tilde{E}_{st-2} = \{30, 60, 90, \ldots, 270\}$, and the segments of the class size rule as $S_{st-2} = \text{I}(\tilde{E}_{st} \pm 15)$. The following specification for the global approach seems to capture both the underlying functional form and to provide reasonable precision of the estimates

$$f\left(E_{st-2}\right) = \alpha_1 E_{st-2} + \alpha_2 E^2_{st-2} + \alpha_3 E^3_{st-2} + \alpha_4 E^4_{st-2} + \alpha_5 S_{st-2} + \delta_s \tag{4}$$

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 enrolment. The identifying assumption is that the outcome at these two enrolment 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 enrolment.

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

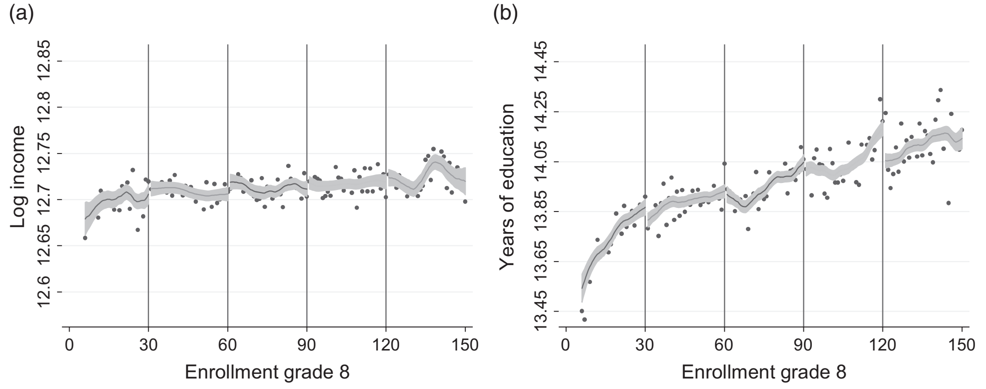

Figure 6. Local polynomial regressions. (a) Log income; (b) Years of education
Notes: Local polynomial regressions of enrolment in grade 8 on outcome variables for each segment. Log income and educational attainment are conditional on cohort specific effects. The markers indicate average outcome for each enrolment value. The y-axis is the mean value of the outcome variable ± 0.2 SDs.

Figure 6 presents average values of the outcomes for different levels of enrolment and shows that the outcomes are positively related to enrolment. 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. 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 be a difference for the threshold of 60 students in the expected direction. For educational attainment, there seems to be differences both for the thresholds 30 and 120 students, but in the opposite direction of what is expected.

V. 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 enrolment 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. The class size rule is a strong instrument. The F-value for the first stage is almost 5,000.

Columns (3)–(8) include various specifications of the enrolment 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 enrolment control function. The effect is positive and significant at 5% level in the models only including segment fixed effects (column 3) and enrolment to the fourth polynomial (column 4). When school fixed effects are introduced (column 5), the effect drops and becomes insignificant.

TABLE 3

Average class size effect and specification analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable: Log income										
Average class size grade 8–10	0.00035	−0.00033	−0.00077	−0.00087	−0.00029	−0.00041	−0.00047	−0.00045	−0.00006	0.00009
	(0.0005)	(0.0005)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0020)	(0.0030)
F-value first stage	–	4,843	2,391	2,327	1,975	1,936	1,864	2,000.0	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**	0.0117**	0.0118**	0.0106**	0.0017	0.0008	0.0005	−0.0005	−0.0024	0.0067
	(0.0028)	(0.0032)	(0.0038)	(0.0039)	(0.0023)	(0.0020)	(0.0021)	(0.0021)	(0.0065)	(0.0096)
F-value first stage	–	4,827	2,386	2,291	1,932	1,934	1,863	1,996.0	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
Enrolment 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
Socio-economic characteristics	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Sample	All	All	All	All	All	All	All	All	$\tilde{E}_\tau +/-3$ students	$\tilde{E}_\tau +/-3$ students

Notes: Standard errors clustered at the school level in parentheses, $*P < 0.05$ and $**P < 0.01$. Socio-economic 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.

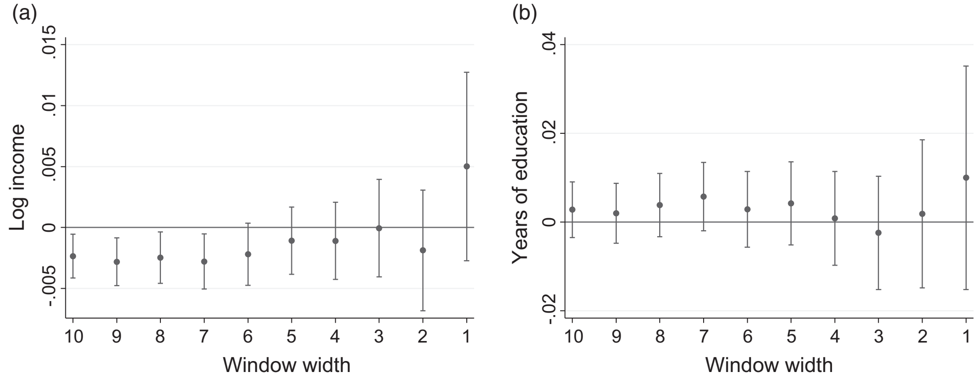

Figure 7. Effect of class size with 95% confidence interval for different bandwidths. (a) Log income; (b) Years of education

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

Column (6) in Table 3 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 [−0.018, 0.001] log points for income and [−0.032, 0.048] 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 socio-economic 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 are presented in column (9) in Table 3.[21]

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 enrolment control function is rather 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

[21] For a full specification of the models, se columns (2) and (4) in Appendix Table A2.

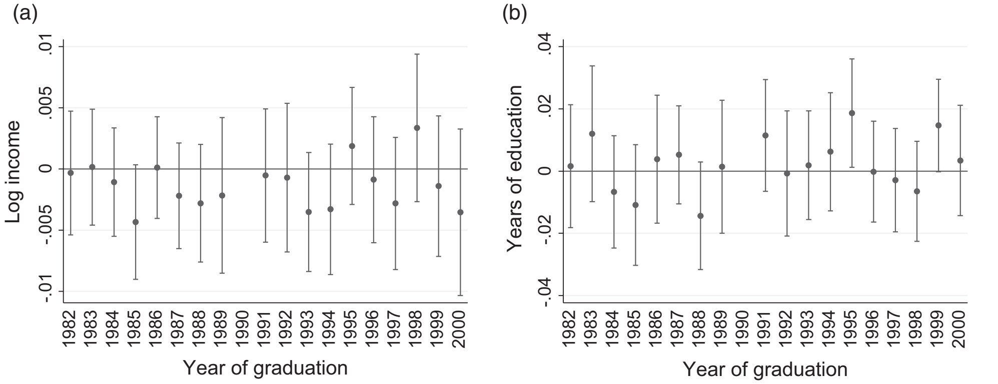

Figure 8. Cohort specific estimates using the global approach with 95% confidence intervals. (a) Log income; (b) Years of education

Table 3 presents results for a model with a more flexible enrolment control function, including enrolment 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.

Several of the individuals in the data are relatively young and in their early career. It might take some experience in the labour market before there is a class size effect on wages. There is some evidence of such delayed effects of individual characteristics that are not easy to observe for the employers (Altonji and Pierret, 2001; Light and McGee, 2015). 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, well beyond the labour market experience level with employer learning effects found in the literature. The estimate is not significant at the 5% level for any cohort and any outcome. For income, the point estimate is positive for four of the 19 cohorts,[22] while for educational attainment, the estimate is positive for 12 cohorts.[23]

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 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% level and close to zero.

[22] 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 −0.00002 (0.0005) using the global approach specification in column (6) in Table 3 and −0.0010 (0.0016) using the local approach in column (9).

[23] We have also done a similar analysis for income splitting the sample using potential experience instead of age. If class size affects schooling, despite our evidence it does not, then the class size would affect experience in the labour market at a given age. In the case of nonlinear effect of experience, this could bias the class size effect on income downwards. Potential experience is calculated using the educational attainment outcome and a maximum value of 28 in the data. As expected, the class size effect on income is mainly insignificant also at each level of potential experience, with the exception of potential experience around 15 years.

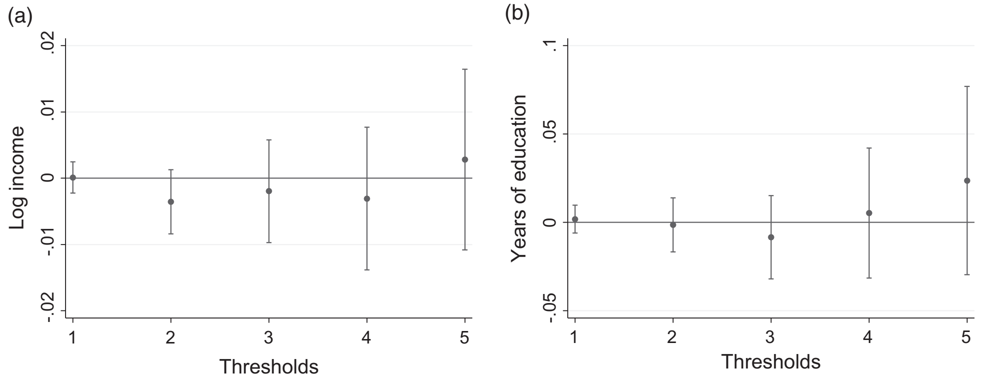

Figure 9. Effect of class size at different thresholds with 95% confidence interval. (a) Log income; (b) Years of education
Note: Threshold 1 is 30 students, threshold 2 is 60 students, etc. Threshold 5 includes all thresholds from 5 and up.

We have also investigated whether the class size affects binary variables of education and employment. However, there is no effect on the probability of graduating college or on the probability of being employed in 2011.[24]

One common argument for smaller classes is that it can improve the possibility to support students most in need of learning support. The evidence from, e.g. the STAR experiment suggests that students with a disadvantaged background benefit the most from smaller classes (Dynarski, Hyman and Schanzenbach, 2013), which suggests that smaller classes have the potential to reduce the variation in student outcomes. We investigate this issue in Table 4 where we use data collapsed to school-by-year observations, and the standard deviation in the outcomes are the dependent variables.[25] Columns (1) and (2) present results for the global approach, while columns (3) and (4) present results for the local approach. The measures of the socio-economic composition at the schools included in columns (2) and (4) are simply the average values over the relevant students for the individual characteristics presented above.

The effects of smaller classes in column (1) in Table 4 are negative as expected. Reduced class size of 10 students significantly decreases the variation in log income by 0.03, which is 12% of a standard deviation. The effect on the variation in years of education is about 6% of a standard deviation, but insignificant. However, including measures of the socio-economic composition at the school reduces the class size effect considerably.[26] In addition, the effects estimated by the local approach are negative and insignificant. Overall, it does not seem like smaller classes reduces the variation in student outcomes.

[24] For a model using whether the student achieves a degree from higher education (completes more than 13 years of schooling) as the dependent variable, the estimated effect of class size is 0.0006 (0.0004) using the global approach specification in column (6) in Table 3 and −0.0005 (0.0013) using the local approach in column (9), with standard errors in parentheses. For a model using whether the individual has been employed during 2011 as the dependent variable, the estimated effect of class size is −0.0002 (0.0003) and 0.0008 (0.0010) for the global and local approach specification, respectively.

[25] The average values (standard deviation) for the dependent variables are 0.69 (0.26) and 2.42 (0.32) for the standard deviation in log income and years of education, respectively.

[26] The effect disappears when controlling for parental education.

TABLE 4

Effect of class size on variation in outcomes, school level analysis

	(1)	(2)	(3)	(4)
Dependent variable: Standard deviation in Log income				
Average class size grade 8–10	0.00322**	0.00154	−0.00322	−0.00601*
	(0.0012)	(0.0011)	(0.0037)	(0.0035)
F-value first stage	2,620	2,593	215,1	220.3
Observations	16,731	16,731	2,713	2,713
Dependent variable: Standard deviation in Years of education				
Average class size grade 8–10	0.00204	−0.000267	−0.00126	−0.00455
	(0.0014)	(0.0014)	(0.0041)	(0.0040)
F-value first stage	2,623	2,595	215,1	220.3
Observations	16,734	16,734	2,713	2,713
Enrolment controls	Polynomial and segment FE	Polynomial and segment FE	Linear and segment FE	Linear and segment FE
School fixed effects	Yes	Yes	No	No
Time/age fixed effects	Yes	Yes	Yes	Yes
Socio-economic composition at school	No	Yes	No	Yes
Subsample ±3 students	No	No	Yes	Yes

Notes: Standard errors clustered at the school level in parentheses, $*P<0.05$ and $**P<0.01$. Socio-economic composition at school is measured as average values of the socio-economic characteristics included in Table 2.

VI. 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 II 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.* (2011), Dynarski *et al.* (2013) and Fredriksson *et al.* (2013).

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

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

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 enrolment. The control functions $f(.)$ and $g(.)$ include the same elements as above. $\overline{\text{CS}}_{st}$ and $\overline{\text{CS}}_{st} \times Z$ are instrumented using the class size rule and it's interaction with Z. Since we use average class size during grades 8–10 in the analysis, we measure the school district characteristics by the average value during the same time period. In order to facilitate interpretation, the continuous interaction variables are standardized to have mean zero and standard deviation equal to unity. The level effect of Z is not reported since the interaction term with $g(.)$ is included in the model.

An alternative to the interaction specification is to split the sample with respect to the school district characteristics. This would allow all coefficients in the model to vary across subsamples, and is a less restrictive IV-specification. However, the qualitative results are equal to the results presented below.[27]

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 teacher 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 labour 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 expected teacher quality. The share of certified teachers is thus a reasonable indicator of teacher quality, and is previously used with this interpretation by Bonesrønning, Falch and Strøm (2005) and Falch, Johansen and Strøm (2009).[28]

The first part of Table 5 presents the results.[29] 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.[30] The joint strength of the instruments is tested by the Kleibergen-Paap F-statistic, and 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 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 SDs above the average, increases the income by 0.053 log points, i.e. about 7% of a standard deviation in income.

[27] The results are available upon request.

[28] It is beyond the scope of the present paper to investigate why the share of certified teachers varies across schools. Bonesrønning *et al.* (2005) find that it depends on the student composition of the school, while Falch *et al.* (2009) find that it depends on the business cycle. There are other potential factors, such as features of the school leadership. Thus, this variable might capture broader factors such as overall school quality or local labour market conditions.

[29] 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.

[30] 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 5, we get the same coefficients on class size.

TABLE 5

Heterogeneous effects of class size

Outcome	(1) Log income	(2)	(3) Years of education	(4)
Teacher quality				
Interaction effect with class size	0.00003	−0.00264*	−0.00228	0.00366
	(0.0006)	(0.0016)	(0.0017)	(0.0049)
Average class size grade 8–10	−0.000421	−0.000268	0.000278	0.000926
	(0.0007)	(0.0022)	(0.0022)	(0.0072)
F-value, first stage	597.9	99.98	597.0	100.9
Observations	849,163	159,830	893,546	168,182
School choice upper secondary education				
Interaction effect with class size	0.000459	−0.00678	0.00945	−0.0115
	(0.0022)	(0.0077)	(0.0065)	(0.0241)
Average class size grade 8–10	0.00006	0.00258	−0.00101	−0.00001
	(0.0013)	(0.0041)	(0.0038)	(0.0136)
F-value, first stage	204.3	19.61	203.4	19.81
Observations	364,670	69,101	379,619	71,876
Local unemployment rate				
Interaction effect with class size	−0.000164	−0.00234	−0.00169	−0.00999
	(0.0007)	(0.0022)	(0.0020)	(0.0079)
Average class size grade 8–10	−0.000322	−0.000526	0.000898	−1.41e-05
	(0.0007)	(0.0020)	(0.0020)	(0.0064)
F-value, first stage	287.9	17.25	291.9	17.24
Observations	903,572	170,601	952,218	179,796
Property tax				
Interaction effect with class size	0.00118	−0.0103	−0.00497	0.0261
	(0.0028)	(0.0089)	(0.0085)	(0.0293)
Average class size grade 8–10	−0.00106	0.00860*	0.0033	−0.0181
	(0.0015)	(0.0051)	(0.0041)	(0.0144)
F-value, first stage	128.4	13.71	127.7	13.54
Observations	272,724	52,087	283,322	54,053
Share of public sector employment				
Interaction effect with class size	0.000564	−0.000888	−0.00245	0.0133
	(0.0008)	(0.0030)	(0.0026)	(0.0094)
Average class size grade 8–10	−0.000149	−0.000621	0.00182	−0.000249
	(0.0009)	(0.0029)	(0.0027)	(0.0089)
F-value, first stage	375.8	14.03	370.6	13.74
Observations	539,693	101,333	563,569	105,745
School district size; population				
Interaction with class size	0.00123	−0.00157	−0.00118	0.00142
	(0.0009)	(0.0038)	(0.0021)	(0.0089)
Average class size grade 8–10	0.000016	0.000136	0.000122	0.000175
	(0.0007)	(0.0023)	(0.0021)	(0.0070)
F-value, first stage	61.90	8.066	60.78	8.154
Observations	903,572	170,601	952,218	179,796

continued

TABLE 5

Continued

Outcome	(1) *Log income*	(2)	(3) *Years of education*	(4)
School district size; merger				
Interaction with class size (treatment school district × Postmerger × average class size)	0.00744 (0.0081)	0.0274 (0.0270)	−0.0158 (0.0203)	0.0723 (0.0594)
Average class size grade 8–10	−0.000257 (0.0007)	−0.000127 (0.0020)	0.000567 (0.0021)	−0.00284 (0.0066)
Average population in the school district during grade 8–10, standardized	−0.239*** (0.0465)	−0.0127*** (0.0030)	0.456*** (0.0823)	−0.0473*** (0.0179)
F-value, first stage	9.456	1.687	9.562	1.659
Observations	903,572	170,601	952,218	179,796
Enrolment 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

Notes: Standard errors clustered at the school level in parentheses, $^{*}P < 0.05$ and $^{**}P < 0.01$. 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.

Student incentives

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).[31] Thus, there are stronger incentives for study effort in lower secondary education in some counties than in others.[32] We use the classification developed by Haraldsvik (2004),[33] 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.

[31] A closer description of one system of free school choice is given in Machin and Salvanes (2016). They study the effect on house prices of increased school choice from 1997 in the Oslo county.

[32] 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.

[33] 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.

The results are presented in the second part of Table 5. 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 0.068 log points on income and 0.19 years of education.

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 do not have a robust impact on how efficient schools use their resources.

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; 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 1997–99. 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% in this period, and is most common in the large school districts. Since we use 3-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 5 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 unrelated to local fiscal constraints.

Interest groups

Interest groups prefer increased resource use and reduced pressure on efficiency. As discussed in section II, 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.[34]

[34] 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.

Table 5 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.

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 the 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 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 1 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 mergers and the regression discontinuity approach with regard to class size. The model includes an indicator variable for whether or not the school district 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).

$$\begin{aligned} y_{isdt} = {} & \alpha\widehat{\overline{\mathrm{CS}}}_{sdt} + \gamma_1 \mathrm{Treat}_d + \gamma_2 \mathrm{Treat}_d \times \mathrm{Post}_t + \phi\widehat{\overline{\mathrm{CS}}_{sdt} \times \mathrm{Treat}_d} \times \mathrm{Post}_t \\ & + f\left(E_{sdt-2}\right) + g\left(E_{sdt-2}\right) \times \mathrm{Treat}_d \times \mathrm{Post}_{dt} + \beta_1 \mathrm{Pop}_{dt} + \beta X_i + \delta_t + \varepsilon_{isdt} \end{aligned} \tag{7}$$

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 5. 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.

VII. 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 analysed 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.[35] 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 quality differs systematically between countries.[36] The absence of robust significant interaction effects between class size and our indicator for teacher quality does not support this explanation either.

If schools use compensatory policies and increase the use of other inputs in grades with larger classes, such as teacher assistants, the estimated class size effect will be biased towards zero. Unfortunately, other measures on education inputs than the class size are not available for the time period of the present paper. The potential for such policies used to be low, but has increased over time by increased school budgets, availability of computers, and due to school budgets being uncoupled with the class size rule to an increasing degree. Using data on the number of teachers for the school years 2002–03, Leuven *et al.* (2008) investigate whether input substitution can explain the absence of any class size effect in the short run. They find only week evidence of input substitution, and that such a substitution cannot drive the results. In addition, it is unlikely that any compensatory policies towards large classes should be different across the Scandinavian countries.

At a general level, the class size effects might obviously also depend on characteristics of the students, although such characteristics vary to a smaller extent 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. In addition, Leuven and Løkken (2015) find no long-run effect of class size in primary education in Norway. This evidence clearly

[35] Both in Sweden, Denmark and Norway, the municipalities (school districts) are multipurpose 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 million) 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.

[36] The share of teachers certified for teacher jobs varies substantially between Norway and Sweden. According to Andersson, Johansson and Aldenström (2011), above 15% 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%.

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, Niederle and Oosterbeek, 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 studies (Iversen and Bonesrønning, 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, which is consistent with the findings for variation in student outcomes in Table 4 above.

VIII. 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, the extent of upper secondary school choice, school district size, local fiscal constraints, and labour 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.

Appendix A. Data reduction and main results with complete list of explanatory variables

Tables A1–A3.

TABLE A1

Data reduction

	Observations	*Reduction*	*% Reduction*
1. Sample 1982–2000 (without 1990)*	1,040,840		
2. Non-missing class size	1,003,149	37,691	3.62
3. 16 years old when graduating from lower secondary education	953,512	49,637	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†	903,828	49,355	5.18

Notes: *Data on the school identifier is missing in 1990.†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 socio-economic characteristics and enrolment controls

	(1)	*(2)*	*(3)*	*(4)*
	Log income		*Years of education*	
Average class size grade 8–10	−0.000413	−5.67e-05	0.000789	−0.00244
	(0.0007)	(0.0020)	(0.0020)	(0.0065)
Girl	−0.336***	−0.337***	0.547***	0.528***
	(0.0033)	(0.0052)	(0.0081)	(0.0148)
Mother's education: High School	0.0783***	0.0803***	0.800***	0.822***
	(0.0019)	(0.0043)	(0.0070)	(0.0137)
Mother's education: Bachelor	0.0715***	0.0666***	1.589***	1.650***
	(0.0030)	(0.0072)	(0.0103)	(0.0209)
Mother's education: Masters+	0.0142	0.0102	1.955***	2.029***
	(0.0101)	(0.0219)	(0.0222)	(0.0475)
Mother's education: Unknown	−0.0230**	0.0142	0.0734**	−0.00895
	(0.0106)	(0.0234)	(0.0345)	(0.0751)
Father's education: High School	0.0722***	0.0789***	0.699***	0.730***
	(0.0021)	(0.0043)	(0.0072)	(0.0154)
Father's education: Bachelor	0.104***	0.101***	1.563***	1.607***
	(0.0033)	(0.0066)	(0.0103)	(0.0200)
Father's education: Masters+	0.121***	0.127***	2.150***	2.198***
	(0.0046)	(0.0093)	(0.0137)	(0.0260)
Father's education: Unknown	−0.0468***	−0.0279*	0.129***	0.211***
	(0.0069)	(0.0149)	(0.0219)	(0.0496)
First generation immigrant	−0.0477***	−0.0597***	0.00444	−0.119*
	(0.0104)	(0.0205)	(0.0380)	(0.0617)
Second generation immigrant	0.0434***	0.0166	0.579***	0.320***
	(0.0146)	(0.0312)	(0.0441)	(0.0878)
Only mother working	0.0398***	0.0323***	0.0850***	0.0594***
	(0.0026)	(0.0056)	(0.0089)	(0.0180)
Only father working	0.0361***	0.0308***	0.0266***	0.0287
	(0.0026)	(0.0055)	(0.0091)	(0.0185)
Both parents working	0.101***	0.105***	0.359***	0.370***
	(0.0024)	(0.0050)	(0.0087)	(0.0165)
Birth month	0.000755***	0.000302	0.00643***	0.00660***
	(0.0002)	(0.0005)	(0.0007)	(0.0017)

continued

TABLE A2

Continued

	(1) *Log income*	(2)	(3) *Years of education*	(4)
Enrolment	−0.000989	0.00244	0.00885***	−0.0108
	(0.0010)	(0.0025)	(0.0033)	(0.0083)
Enrolment^2	1.53e-05		−0.000110**	
	(0.0000)		(0.0001)	
Enrolment^3	−9.43e-08		6.16e-07*	
	(0.0000)		(0.0000)	
Enrolment^4	2.03e-10		−1.18e-09*	
	(0.0000)		(0.0000)	
Segment 1	−0.000543	0.599	0.00957	−2.569
	(0.0094)	(0.6157)	(0.0302)	(2.0234)
Segment 2	0.00243	0.527	−0.0211	−2.267
	(0.0126)	(0.5359)	(0.0400)	(1.7635)
Segment 3	0.00175	0.453	0.00167	−1.929
	(0.0150)	(0.4581)	(0.0461)	(1.5059)
Segment 4	0.00275	0.379	0.00447	−1.585
	(0.0171)	(0.3808)	(0.0519)	(1.2509)
Segment 5	0.0161	0.298	−0.0316	−1.305
	(0.0194)	(0.3043)	(0.0592)	(1.0024)
Segment 6	0.0401*	0.237	−0.0584	−1.011
	(0.0239)	(0.2284)	(0.0730)	(0.7534)
Segment 7	0.0367	0.191	−0.130	−0.599
	(0.0305)	(0.1498)	(0.0964)	(0.5249)
Segment 8	0.00106	0.0653	−0.0508	−0.153
	(0.0462)	(0.0758)	(0.1478)	(0.2484)
Segment 9	−0.0356	–	0.266	–
	(0.0765)		(0.3943)	
R^2	0.108	0.111	0.176	0.199
Predicted class size	0.56***	0.40***	0.56***	0.40***
(the instrument), first stage	(0.013)	(0.024)	(0.013)	(0.024)
F-value first stage	1,936	295.9	1,934	294.7
R^2 first stage	0.4909	0.4865	0.4896	0.4856
Observations	903,828	170,604	952,514	179,799
No. of schools	1,156		1,156	
Enrolment controls	Polynomial and segment FE	Linear and segment FE	Polynomial and segment FE	Linear and segment FE
Sample	All		All	
Socio-economic characteristics	Yes	Yes	Yes	Yes
Time/age FE	Yes	Yes	Yes	Yes
School FE	Yes	No	Yes	No

Notes: Standard errors in parentheses, *$P<0.05$, **$P<0.01$, ***$P<0.001$. Standard errors are clustered at the school level. Socio-economic characteristics are described in section ‘Institutions’.

TABLE A3

Subsample analysis

Outcome	*(1)* *Log income*	*(2)*	*(3)* *Years of education*	*(4)*
Girls				
Average class size grade 8–10	0.000193	0.00219	−0.000357	0.00285
	(0.0009)	(0.0027)	(0.0028)	(0.0087)
Observations	443,057	83,593	466,957	88,101
Boys				
Average class size grade 8–10	−0.000797	−0.00185	0.00207	−0.00627
	(0.0009)	(0.0027)	(0.0028)	(0.0087)
Observations	460,770	87,011	485,557	91,698
Parental education more than high school				
Average class size grade 8–10	−0.000719	−0.00271	−0.00318	−0.000336
	(0.0014)	(0.0040)	(0.0042)	(0.0140)
Observations	257,205	49,267	271,400	51,965
Parental education less than high school				
Average class size grade 8–10	−0.000229	0.000573	0.00120	−0.00410
	(0.0007)	(0.0020)	(0.0024)	(0.0079)
Observations	646,606	121,337	681,097	127,834
Immigrant				
Average class size grade 8–10	−0.00772	0.00417	−0.00144	−0.00725
	(0.0073)	(0.0166)	(0.0191)	(0.0440)
Observations	15,206	2,977	17,427	3,364
Non-immigrant				
Average class size grade 8–10	−0.000354	−0.000185	0.000833	−0.00247
	(0.0007)	(0.0020)	(0.0021)	(0.0066)
Observations	888,502	167,627	934,968	176,435
Enrolment 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

Notes: Standard errors in parentheses, $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$. Standard errors are clustered at the school level. Socio-economic characteristics are described in section 3.1.

Final Manuscript Received: February 2016

References

Adkins, L. C. and Moomaw, R. L. (2003). ‘The impact of local funding on the technical efficiency of Oklahoma schools’, *Economics Letters*, Vol. 81, pp. 31–37.

Altinok, N. and Kingdon, G. (2012). ‘New evidence on class size effects: a pupil fixed effects approach’, *Oxford Bulletin of Economics and Statistics*, Vol. 74, pp. 203–234.

Altonji, J. G. and Pierret, C. R. (2001). ‘Employer learning and statistical discrimination’, *Quarterly Journal of Economics*, Vol. 116, pp. 313–350.

Andersson, C., Johansson, P. and Aldenström, N. (2011). ‘Do you want your child to have a certified teacher?’, *Economics of Education Review*, Vol. 30, pp. 65–78

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

Angrist, J. D., Battistin, E. and Vuri, D. (2015). *In a Small Moment: Class Size and Moral Hazard in the Mezzogiorno*, IZA Discussion Paper No. 8959.

Anzia, S. F. (2011). 'Election timing and the electoral influence of interest groups', *Journal of Politics*, Vol. 73, pp. 412–427.

Barankay, I. and Lockwood, B. (2007). 'Decentralization and the productive efficiency of government: evidence from Swiss cantons', *Journal of Public Economics*, Vol. 91, pp. 1197–1218.

Becker, W. E. and Rosen, S. (1992). 'The learning effect of assessment and evaluation in high school', *Economics of Education Review*, Vol. 11, pp. 107–118.

Berry, C. R. and West, M. R. (2010). 'Growing pains: the school consolidation movement and student outcomes', *Journal of Law, Economics, and Organization,*, Vol. 26, pp. 1–29.

Betts, J. R. (1998). 'The impact of educational standards on the level and distribution of earnings', *American Economic Review*, Vol. 88, pp. 266–275.

Björklund, A., Edin, P.-A., Fredriksson, P. and Krueger, A. (2004). *Education, Equality and Efficiency–An Analysis of Swedish School Reforms During the 1990s*, IFAU Report No. 2004:1.

Black, S. E., Devereux, P. J. and Salvanes, K. G. (2013). 'Under pressure? The effect of peers on outcomes of young adults', *Journal of Labor Economics*, Vol. 31, pp. 119–153.

Bonesrønning, H. (2003). 'Class size effects on student achievement in Norway: patterns and explanations', *Southern Economic Journal*, Vol. 69, pp. 952–965.

Bonesrønning, H. (2013). 'Public employees and public sector reform implementation', *Public Choice*, Vol. 156, pp. 309–327

Bonesrønning, H., T, Falch and Strøm, B. (2005). 'Teacher sorting, teacher quality, and student composition', *European Economic Review*, Vol. 49, pp. 457–483.

Borge, L-E. and Rattsø, J. (2008). 'Property taxation as incentive for cost control: empirical evidence for utility services in Norway', *European Economic Review*, Vol. 52, pp. 1035–1054.

Bosworth, R. (2014). 'Class size, class composition, and the distribution of student achievement', *Education Economics*, Vol. 22, pp. 141–165.

Browning, M. and Heinesen, E. (2007). 'Class size, teacher hours and educational attainment', *Scandinavian Journal of Economics*, Vol. 109, pp. 415–438.

Buser, T., Niederle, M. and Oosterbeek, H. (2014). 'Gender, competitiveness, and career choices', *Quarterly Journal of Economics*, Vol. 129, pp. 1409–1447.

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

Chubb, J. E. and Moe, T. M. (1988): Politics, markets and the organization of public schools', *American Political Science Review*, Vol. 82, pp. 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*, Vol. 78, pp. 523–545.

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F. and York, R. (1966). *Equality of Educational Opportunity*, Washington, DC, US Government Printing Office.

Costrell, R. M. (1994). 'A simple model of educational standards', *American Economic Review*, Vol. 84, pp. 956–971.

Denny, K. and Oppedisano, V. (2013). 'The surprising effect of larger class sizes: evidence using two identification strategies', *Labour Economics*, Vol. 23, pp. 57–65.

Dobbelsteen, S., Levin, J. and Oosterbeek, H. (2002). 'The causal effect of class size on scholastic achievement: distuingishing between the pure class size effect from the effect of changes in class composition', *Oxford Bulletin of Economics and Statistics*, Vol. 64, pp. 17–38.

Driscoll, D., Halcoussis, D. and Svorny, S. (2003). 'School district size and student performance', *Economics of Education Review*, Vol. 22, pp. 193–201.

Dynarski, S., Hyman, J. and Schanzenbach, D. W. (2013). 'Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion', *Journal of Policy Analysis and Management*, Vol. 32, pp. 692–717.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A. and Willms, J. D. (2001). 'Class size and student achievement', *Psychological Science in the Public Interest*, Vol. 2, pp. 1–30

Falch, T. and Fischer, J. A. V. (2012). 'Public sector decentralization and school performance: international evidence', *Economics Letters*, Vol. 114, pp. 276–279.

Falch, T. and Naper, L. R. (2013). 'Educational evaluation and gender gaps in student achievement', *Economics of Education Review*, Vol. 36, pp. 12–25.

Falch, T., Johansen, K. and Strøm, B. (2009). 'Teacher shortages and the business cycle', *Labour Economics*, Vol. 16, pp. 548–658.

Figlio, D. and Hart, C. D. (2014). 'Competitive effects of means–tested school vouchers', *American Economic journal: Applied Economics*, Vol. 6, pp. 133–156.

Fischel, W. A. (2001): Homevoters, municipal corporate governance, and the benefit view of the property tax', *National Tax Journal*, Vol. 54, pp. 157–174.

Fiva, J. and Rønning, M. (2008). 'The incentive effects of property taxation: evidence from Norwegian school districts', *Regional Science and Urban Economics*, Vol. 38, pp. 49–62.

Fredriksson, P. and Öckert, B. (2008). 'Resources and student achievement – evidence from a Swedish policy reform', *Scandinavian Journal of Economics*, Vol. 110, pp. 277–296.

Fredriksson, P., Oosterbeek, H. and Öckert, B. (2013). 'Long term effects of class size', *Quarterly Journal of Economics*, Vol. 128, pp. 249–285.

Gelman, A. and Imbens, G. (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*, Vol. 89, pp. 93–111.

Gunderson, M. (2005). 'Two faces of union voice in the public sector', *Journal of Labor Research*, Vol. 26, pp. 393–413.

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*, Vol. 31, pp. 601–614.

Hanushek, E. A. (1996). 'School resources and student performance', in Burtless G. (ed.), *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Washington, DC: Brookings Institution Press, pp. 43–73.

Hanushek, E. A. (2003). 'The failure of input–based schooling policies', *The Economic Journal*, Vol. 113, pp. F64–F98.

Hanushek, E. A. (2006). 'School resources, in Hanushek E. A. and Welch F. (eds), *Handbook of the Economics of Education*, Vol. 2, Amsterdam: Elsevier B.V, pp. 865–908.

Haraldsvik, M. (2004). *Elevprestasjoner og konkurranse i den videregående skolen: Gir konkurranse om skoleplassene elevene insentiver til å jobbe hardere på ungdomsskolen?* Norwegian University of Science and Technology (Master thesis).

Haraldsvik, M. (2012). 'Does performance based school choice affect student achievement?', in Haraldsvik M. (ed), *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*, Vol. 43, pp. 387–425.

Heinesen, E. (2005). 'School district size and student educational attainment: evidence from Denmark', *Economics of Education Review*, Vol. 24, pp. 677–689.

Heinesen, E. (2010). 'Estimating class–size effects using within–school variation in subject–specific classes', *Economic Journal*, Vol. 120, pp. 737–760

Hoxby, C. M. (1996). 'How teachers' unions affect education production', *Quarterly Journal of Economics*, Vol. 111, pp. 671–718.

Hoxby, C. M. (1999). 'The productivity of schools and other local public goods producers', *Journal of Public Economics*, Vol. 74, pp. 1–30.

Hoxby, C. M. (2000). 'The effects of class size on student achievement: new evidence from population variation', *Quarterly Journal of Economics*, Vol. 115, pp. 1239–1285.

Iversen, J. M. V. and Bonesrønning, H. (2013). 'Disadvantage students in the early grades: will smaller classes help them?', *Education Economics*, Vol. 21, pp. 305–324.

Koerselman, K. (2013). 'Incentives from curriculum tracking', *Economics of Education Review*, Vol. 32, pp. 140–150.
Krueger, A. (2003). 'Economic considerations and class size', *The Economic Journal*, Vol. 113, pp. 34–63.
Krueger, A. B. and Whitmore, D. M. (2001). 'The effect of attending a small class in the early grades on college–test taking and middle school test results: evidence from project STAR', *Economic Journal*, Vol. 111, pp. 1–28.
Lavecchia, A. M., Liu, H. and Oreopoulos, P. (2014). *Behavioral Economics of Education: Progress and Possibilities*, NBER Working Paper No. 20609.
Lazear, E. P. (2001). 'Educational production', *Quarterly Journal of Economics*, Vol. 115, pp. 777–803.
Lee, D. N. (2013). 'The impact of repealing Sunday closing laws on educational attainment', *Journal of Human Resources*, Vol. 48, pp. 287–310.
Lee, D. S. and Lemieux, T. (2010). 'Regression discontinuity designs in economics', *Journal of Economic Literature*, Vol. 48, pp. 281–355
Leuven, E. and Løkken, S. A. (2015). *Long Term Impacts of Class Size in Compulsory Schooling*, Mimeo.
Leuven, E., Oosterbeek, H. and Rønning, M. (2008). 'Quasi–experimental estimates of the effect of class size on achievement in Norway', *Scandinavian Journal of Economics*, Vol. 110, pp. 663–693.
Light, A. and McGee, A. (2015). 'Employer learning and the importance of skills', *Journal of Human Resources*, Vol. 50, pp. 72–107.
Lindahl, M. (2005). 'Home versus school learning: a new approach to estimating the effect of class size on achievement', *Scandinavian Journal of Economics*, Vol. 107, pp. 375–394.
Loeb, S. and Strunk, K. (2007). 'Accountability and local control: response to incentives with and without authority over resource generation and allocation', *Education Finance and Policy*, Vol. 2, pp. 10–39.
Lott, J. and Kenny, L. W. (2013). 'State teacher union strength and student achievement', *Economics of Education Review*, Vol. 35, pp. 93–103.
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*, Vol. 27, pp. 525–587
Machin, S. and Salvanes, K. G. (2016). 'Valuing school quality via a school choice reform', *The Scandinavian Journal of Economics*, Vol. 118, pp. 3–24.
Mensah, Y. W., Schoderbek, M. P. and Sahay, S. P. (2013). 'The effect of administrative pay and local property taxes on student achievement scores: evidence from New Jersey public schools', *Economics of Education Review*, Vol. 34, pp. 1–16.
Moe, T. M. (2001). 'A union by any other name', *Education Next*, Vol. 1, pp. 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*, Vol. 98, pp. 44–52.
Nusche, D., Earl, L., Maxwel, W. and Shewbridge, C. (2011). *OECD Reviews of Evaluation and Assessment in Education:* Norway, OECD, Paris
Rattsø, J. and Sørensen, R. J. (2004). 'Public employees as swing voters: empirical evidence on opposition to public reform', *Public Choice*, Vol. 119, pp. 281–310.
Reiling, R. B. and Strøm, B. (2015). 'Upper secondary school completion and the business cycle', *Scandinavian Journal of Economics*, Vol. 117, pp. 195–219.
Rockoff, J. (2009). 'Field experiments in class size from the early twentieth century', *Journal of Economic Perspectives*, Vol. 23, pp. 211–230.
Strunk, K. O. and Grissom, J. A. (2010). 'Do strong unions shape district policies? Collective bargaining, teacher contract restrictiveness, and the political power of teachers unions', *Educational Evaluation and Policy Analysis*, Vol. 32, pp. 389–406.
Urquiola, M. and Verhoogen, E. (2009). 'Class–size caps, sorting, and the regression–discontinuity design', *American Economic Review*, Vol. 99, pp. 179–215.
Webbink, D. (2005). 'Causal effects in education', *Journal of Economic Surveys*, Vol. 19, pp. 535–560.
Wößmann, L. and West, M. R. (2006). 'Class–size effects in school systems around the world: evidence from between–grade variation in TIMSS', *European Economic Review*, Vol. 50, pp. 695–736.
Woessmann, L. (2005). 'Educational production in Europe', *Economic policy*, Vol. 20, pp. 446–504.