Marte Rønning

Studies in Educational Production and Organization

Thesis for the degree philosophiae doctor

Trondheim, November 2007

Norwegian University of Science and Technology
Faculty of Social Sciences and Technology Management
Department of Economics

NTNU

NTNU
Norwegian University of Science and Technology

Thesis for the degree philosophiae doctor

Faculty of Social Sciences and Technology Management
Department of Economics

ISBN 978-82-471-5174-7 (printed version)
ISBN 978-82-471-5188-4 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2007:233

Printed by NTNU-trykk

Studies in Educational Production and Organization

Marte Rønning

To my family

Preface

This thesis consists of four independent essays that deal with different topics within the field of economics of education. The essay in chapter 1 is joint work with Edwin Leuven and Hessel Oosterbeek. The essay in chapter 4 is joint work with Jon H. Fiva and is forthcoming in the *Regional Science and Urban Economics*.

Leuven and Oosterbeek are both affiliated with the Department of Economics at the University of Amsterdam and the Tinbergen Institute. Leuven is also affiliated with the CEPR, CREST-INSEE and IZA. Fiva was affiliated with the Department of Economics at the Norwegian University of Science and Technology when we started to write Chapter 4. Fiva is currently affiliated with Department of Economics, University of Oslo.

Contents

Acknowledgments

Many people have contributed to the final product of this thesis. It is here that I would like to express my gratitude to those who have helped me in one or another way.

I will start with thanking my supervisor Torberg Falch. I am deeply grateful that he had enough confidence in me to let me start writing a thesis four years ago. Torberg was the one who introduced me to the “research world”, and I appreciate his guidance, support and encouragement as well as his effort to find time for me in a busy schedule. Especially chapter 3 has improved a lot from suggestions from Torberg. I would also like to thank my co-supervisor Bjarne Strøm for his help and useful comments, particularly on Chapter 4. Co-author on Chapter 4 is Jon H. Fiva. In addition to being a good colleague and excellent team worker he has contributed with fruitful conversations.

Thanks to help from Jørn Rattsø and invitation from Eric Hanushek I had the privilege to visit Stanford University from September 2004 to February 2005. This stay was my first real meeting with international research of high quality, and triggered my curiosity for further adventures abroad.

The period from March 2005 until September 2006 I was given the opportunity to stay at the University of Amsterdam. The first year of this stay was funded by the EU/CEPR Network on "Economics of Education and Education Policy in Europe (EEEPE)". Financial support from EEEPE is therefore gratefully acknowledged. A special thanks goes to to Hessel Oosterbeek for admitting me to the SCHOLAR group. I appreciate Hessel for his kindness and also for sharing his knowledge with me. Hessel is one of my two co-authors on Chapter 1.

Cozy coffee breaks and small talks as well as discussions, assistance and support from my colleagues are also highly valuated. Besides the ones who are already mentioned, I particularly thank my good friends Linn and Silje. A thank goes also to Jørgen, Hildegunn, Per, Ragnar, Lars-Erik, Hans, Øystein, Gerd, Vigdis, Monique, Adam and Greg.

I would also like to thank my parents, my brother and his girlfriend Janne, aunt Laila and my cousins Mette Sofie and Stine for always being there for me. Vegard, Eirik, Marit and Arve deserve a thank for their thoughtful support in the beginning of my thesis.

Finally, my greatest thank goes to Edwin Leuven. This thesis would not have been the same without his valuable suggestions and discussions. I also appreciate Edwin for his frankness and affection. Edwin is my second co-author on chapter 1.

Marte Rønning

Introduction

Norwegian investment in education in terms of money and in terms of time is substantial. In 2003 Norway spent 4.8 percent of GDP on education compared to a 3.8 percent OECD average. There is however contradictory evidence whether resources alone is sufficient to guarantee educational success. During more than a decade, the common wisdom among economists was that extra resources for education - measured as the teacher-pupil ratio or as expenditure per pupil - have no systematic relation with student achievement. This view was mainly based on Hanushek's (1986) influential review of the literature. Recently this received wisdom has been challenged by a series of studies that use more elaborate methods to identify the causal impact of class size on achievement (Krueger, 1999; Angrist and Lavy, 1999).

The effect of class size on student achievement is still one of the unresolved issues in education research. It is by now well understood that endogeneity problems may severely bias naive OLS-estimates because of sorting of students and teachers, and the targeting of educational resources. Thus exogenous sources of variation in class size are key for a credible identification of the class size effect. Various recent studies acknowledge this and apply convincing identification methods. This has, however, not led to a definite conclusion about the magnitude or even the sign of the class size effect. For instance, the studies reporting negative effects of class size vary considerably in the size of this effect thereby limiting the relevance of these results for policy conclusions.

The first paper in this thesis (Chapter 1) provides evidence about the effect of class size on achievement in Norwegian lower secondary schools. This evidence is obtained from two different and independent approaches (Angrist and Lavy, 1999; Hoxby, 2000). The results consistently point to a lack of any impact of class size on achievement. Effects as small as 1.5 percent of a standard deviation for a one student change in class size during three consecutive years, can be ruled out. This holds irrespective of identification method, and indicate that decreasing class size is ineffective in raising average student achievement in lower secondary education in Norway.

The empirical literature on the impact of school inputs on educational performance increasingly recognizes that they cannot be well summarized by an average effect. The effects of different educational policies are very likely to vary across pupils from different family backgrounds. One important mechanism in this respect is substitution between the quality of the school environment and the quality of the home environment. This explains why such effects are typically bigger for children from disadvantage backgrounds (e.g. Guryan, 2001; Lindahl, 2001; Card and Payne, 2002; Papke, 2005).

The second paper in the thesis (Chapter 2) takes a slightly different approach and focus on complementarity between school inputs and the quality of the home environment. It does so by considering homework given to Dutch elementary school pupils. To the extent that parental input complements homework, we expect students from higher socioeconomic backgrounds to benefit more from homework assignments. The results indicate that the test score gap is indeed larger in classes where everybody gets homework than in classes where nobody gets homework. More precisely we find that pupils belonging to the upper part of the socioeconomic status scale benefit, whereas pupils from the lowest part are unaffected. Since we consider the effect of homework on within class differences, confounding effects of unobserved teacher characteristics drop out as long as they are homogeneous. Moreover, potential biases arising from unobserved school quality and pupil selection are accounted for by exploiting variation within schools.

The potential negative results with respect to the effectiveness of money based educational policies (school inputs) are also accompanied by i) a growing consensus that teachers are a crucial element to secure educational success of students (e.g. Rivkin et al. 2005), and ii) and increased interest in financial incentives as a way to motivate teachers, students and also other actors that play a role in the students learning (e.g. Lavy, 2002). Chapter 3 and 4 take a closer look at these two aspects.

In Norway, like in many other countries, the education sector faces high absenteeism rates. In upper secondary education 11 percent of the classes are canceled because of teacher absence (OECD, 2003). Accordingly there are reasons to believe that teacher absenteeism has a negative impact on student learning. It is therefore surprising that teacher absenteeism has not been devoted more space yet in the literature. Especially since observed measures on teacher quality like education and experience turn out to explain very little of the variation in student outcome (Hanushek, 2002). Studies on both the effect of teacher absenteeism on pupil learning as well as possible factors that explain teacher absenteeism are lacking in the literature.

The third paper in this thesis (chapter 3) analyzes the effects of workplace characteristics on absenteeism among teachers working in the Norwegian primary and lower secondary school. These characteristics are of interest because they may signal that some teachers are put under higher working pressure. For instance, teachers facing high pupil teacher ratios may have a more tiresome work day than teachers facing lower pupil teacher ratios. Moreover, teachers working in schools with many disadvantage students may be more often confronted with stressful situations. Teachers working hours and contract type are also considered in the model because the literature has found these two variables to be important determinants for teacher absenteeism (Leuven, 2006; Bradley et al., 2007). Exploiting intertemporal variation within teachers who have not changed schools, the findings indicate that the teachers lower their amount of absenteeism if the school's resource use increases. Increased work hours and a permanent work contract are also associated with higher absenteeism. When stratifying on the teachers age, increased workload appears to have a larger impact on old teachers.

The potential for financial incentives to successfully increase performance in the school sector has recently attracted attention. Both for teachers as in Lavy (2002), and for students as in Angrist and Lavy (2002) and Leuven et al. (2003). The last paper in this thesis (chapter 4) concerns financial incentives aimed at school bureaucrats and focus on organizational aspects within the school district and its effect on student outcome. More precisely it is an empirical investigation of theoretical contributions by Glaeser (1996) and Hoxby (1999) which highlight the favorable incentive effects of property taxation on public service providers. The mechanism is assumed to work through capitalization of school quality into housing prices. This involves that school bureaucrats in school districts that levy residential property taxation get incentives to deliver high quality on their services because it is associated with increased income. The institutional setting in Norway is well suited for analyzing the effects of property taxation because we can compare school districts with and without property taxation. To take into account potential endogeneity of the choice of implementing property taxation, we rely on instrumental variables techniques. The empirical results suggest that students in school districts that levy property taxation perform better on the national examination.

This thesis consists of four independent papers that all deal with different topics within the field of economics of education. Chapter 1 and 2 concern school inputs in educational production and focus on both homogeneous and heterogeneous effects. Chapter 3 and 4 are more related to organizational tasks both within schools and within school districts. The studies presented in this thesis contribute to the literature in the sense that they improve on measurement and address new questions. Yet it only covers a small fraction of the whole literature, and future research should therefore further expand the knowledge about how to implement successful and effective educational and teacher labor market policies.

1 Quasi-experimental estimates of the effect of class size on achievement in Norway

1.1 Introduction

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

Most of the (quasi-)experimental studies report that a reduction in class size boosts achievement (Angrist and Lavy, 1999; Krueger, 1999; Boozer and Rouse, 2001; Browning and Heinesen, 2006; Urquiola, 2006), other studies which also exploit credible identification strategies conclude the opposite (Hanushek et al., 1996; Hoxby, 2000). Moreover, the studies reporting negative class size effects vary considerably in the size of this effect thereby limiting the relevance of these results for policy conclusions.

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

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

Some features of our study are worth emphasizing. First, for this study we have access to an extraordinary rich dataset. The dataset covers two entire cohorts of students participating in nationwide tests in the schoolyears 2001/02 and 2002/03. The data are administrative thereby giving rise to no or only little measurement error in actual class size and enrollment. Achievement is measured as high stake test scores, thereby also reducing measurement error in the dependent variables. Together, these characteristics of the dataset enable us to produce very precise estimates of class size effects. Secondly, for all students in our sample we know actual class sizes during the three years they spent in lower secondary school. This allows a clear-cut interpretation of the effects that we estimate.

Furthermore, no previous study has presented precise estimates of the effects of class size on achievement for any of the Nordic countries (Denmark, Finland, Norway and Sweden). Previous studies have either considered other outcomes than achievement such as years of education (Bingley et al., 2005; Browning and Heinesen, 2006), or used a very small sample therefore lacking precision (Bonesrønning, 2003; Lindahl, 2005). Finally, most of the evidence on class size effects pertains to primary education, our study is among the few dealing with class size effects in lower secondary education.[1,2]

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

The remainder of this paper is organized as follows. The next section gives a brief summary of related studies. Section 1.3 describes the relevant institutional features of the Norwegian educational system. Section 2.2 gives a description of the data employed

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

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

in the empirical analysis. Section 1.5 continues with an exposition of the empirical approaches applied in this paper and their limitations. Section 1.6 presents and discusses the main findings. Section 1.7 investigates the possibility of heterogeneous class size effects, and Section 1.8 concludes.

1.2 Related literature

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

Krueger (1999) analyzes data from a large-scale field experiment conducted in the state of Tennessee. Students and their teachers were randomly assigned to a group of regular size (22-25 students), to a group of regular size including a teaching assistent, or to a small group (13-17), during their first four years in school. Krueger's findings are in line with what other people have reported about this project namely that students in smaller classes perform better on standardized achievement tests. Scores increase by four percentile points for the first year that a student is exposed to a small class and by one percentile point for each subsequent year. In a follow-up study, Krueger and Whitmore (2001) demonstrate that reduced class sizes in early school years can have long-lasting effects. Students who attended small classes in this experiment are more likely to take a college-entrance exam and have somewhat higher test scores. The effect on exam taking are mainly concentrated among minority students.

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

[3] In section 1.5 we will discuss their identification strategy in more detail because we will apply the same method in this paper.

estimates point to a positive relation between class size and achievement. In contrast to that, estimates based on the discontinuities in grade enrollment point to a negative effect of class size on achievement.

Other papers exploiting maximum class size rules include Bonesrønning (2003) for Norway, Urquiola (2006) for Bolivia, Piketty (2004) and Gary-Bobo and Mahjoub (2006) for France, Browning and Heinesen (2006) and Bingley et al. (2005) for Denmark, and Wößmann (2005) for Denmark, France, Germany, Greece, Iceland, Ireland, Norway, Spain, Sweden and Switzerland. Especially the papers by Bonesrønning and Wößmann are of interest for the analysis presented in this paper because they also deal with class size effects in Norway. Bonesrønning uses data from a small-scale self-collected dataset. Class size effects are estimated using the discontinuity at 30 as an instrument for actual class size. The reported effects depend on the exact specification but vary between 0.13 and 0.26 of a standard deviation for a 10-students reduction in class size, and are significantly different from zero but not very precisely estimated. The main differences between this study and ours are the following: (i) we apply two different methods rather than only one; (ii) we employ a much larger dataset, so that our estimates have much more precision; (iii) we use high-stake exam scores as our achievement measure (iv) in our regression discontinuity specifications we include controls for enrollment in a grade, something that Bonesrønning does not.

Wößmann uses data from 38 Norwegian schools (1,351 pupils) who participated in TIMSS. Not including controls for enrollment produces a small but significantly negative estimate of the class size effect. This effect vanishes once a controls for enrollment is included in the specification. Interestingly, Wößmann uses a cutoff of 28 rather than on 30 arguing that gives the best fit to his data so that it seems that this implicit rule is actually used in most Norwegian schools. This "letting the data decide" approach is at odds with the basic philosophy underlying the regression discontinuity approach because assignment to treatment and control is no longer based on an exogenous rule but on schools' choices.

Hoxby (2000) uses demographic variation to identify the class size effect. She exploits that - after correcting for a trend - cohort sizes within school districts can be larger or smaller in some years than in others.[4] Using data of elementary school pupils in the state of Conneciticut, she does not find any statistically significant effect of class size on student achievement, and her estimates are precise enough to rule out even modest effects.

[4]This method will also be discussed in more detail in section 1.5.

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

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

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

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

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

1.3 Institutional Settings

Compulsory schools in Norway are run and owned by multipurpose local governments.[5] Municipalities receive funding to run their various activities (including schools) through a combination of a local income tax, property taxes and transfers from the central government (see f.e Hægeland et al. (2007) for more details).

For the students in our sample compulsory schooling consisted of nine years; grades 1 to 6 in primary school; and grades 7 to 9 in lower secondary school.[6] Less than three percent of the students are enrolled in private schools. Thus private schools do not provide a realistic alternative to public schools and are therefore dropped from the analysis. Schools have catchment areas, implying that parental school choice between schools for given residence is not allowed. Most students go to separate primary and lower secondary schools, but because of the rural settlement pattern in Norway about 23 percent of the students are enrolled in so-called "combined" schools. These are schools

[5]In addition to compulsory public schooling local governments are responsible for elderly care, preschool education, and infrastructure. Spending on education consists of about 30% of total spending of the available budget. In Norway the names local school district, local government and municipality are interchangable.

[6]From the school year 1997/98 onwards it became compulsory to start schooling at the age of 6 (instead of 7), and from then on 10 years of schooling was implemented in Norway. The reform was implemented in such a way that the length of primary school was extended with one year.

that offer both primary and lower secondary education. Students who start taking their primary education in a combined school typically continue to take their lower secondary education there because of lack of school choice for given residence and moving costs. Combined schools are often situated in relatively rural areas.

Another feature of the Norwegian compulsory school system is that grade retention almost never happens. Strøm (2004) attributes this to "the strong integration and equalizing policy that all students within a cohort should be treated equal, and be given education in their ordinary classes." As a result, at the end of compulsory schooling all Norwegian students have identical length of schooling. This is important for our analysis since we have data at the student level of the nationwide tests conducted in 2001/2 and 2002/3 and of class size from these years as well as from previous years. Because students do not repeat grades we know the class size history of individual students during their stay in lower secondary school. We will use this information to construct our preferred class size measure (see below). According to anecdotal evidence, movements of students between schools are quite rare.[7]

1.4 Data

1.4.1 Test scores

We use administrative enrollment data from Statistics Norway that cover all students who were in the final grade of lower secondary school (9th grade) during the school years 2001/2002 or 2002/2003. We merged this dataset with test score data from centralized exit exams (also from Statistics Norway). The students in Norway have to sit such exams at the end of their final school year in lower secondary.[8] Their results on this exam are important for further schooling possibilities (which is upper secondary education), and the exam is considered to be a high stake test by all parties involved; students, their parents, teachers and school administrators.[9] Although the curriculum includes many different subjects a written centralized exam is only undertaken in three subjects: mathematics, English and Norwegian.[10] To reduce the administrative burden

[7] We are not aware of any study documenting the extent of students' school mobility in Norway.

[8] Although this exit exam existed already for many years, its results became available for research purposes for the first time in the schoolyear 2001/02.

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

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

each student takes the exam in one subject only. In which subject a student will take the exam, is determined by a random device shortly before the exams take place so that students have minimal scope to prepare for that specific subject only. Students are told only three school days in advance in which subject they are going to make the exam. All students in the same class do their exam in the same subject, but students in different classes of the same school may sit their exam in different subjects.

The grading of the exams is on a scale running from 1 to 6, where 1 is fail, 2 the lowest pass and 6 is the top score. The distributions of test scores have a bell curve shape, and there are no signs of floor or ceiling effects (low frequencies for 1 and 6). Average scores for each of the four subjects are around 3.5 with standard deviations almost equal to 1.

1.4.2 Class size

class size information comes from the Norwegian Ministry of Education (Grunnskolens Informasjonssystem) which registers, for all schools in Norway, the number of classes per grade. We combine this information with the enrollment data mentioned above to calculate for each student average class size at the grade level as

$$\text{class size} = \text{enrollment}/\#\ \text{classes}.$$

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

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

Figure 1.1 shows the distribution of class size in lower secondary schools. This figure

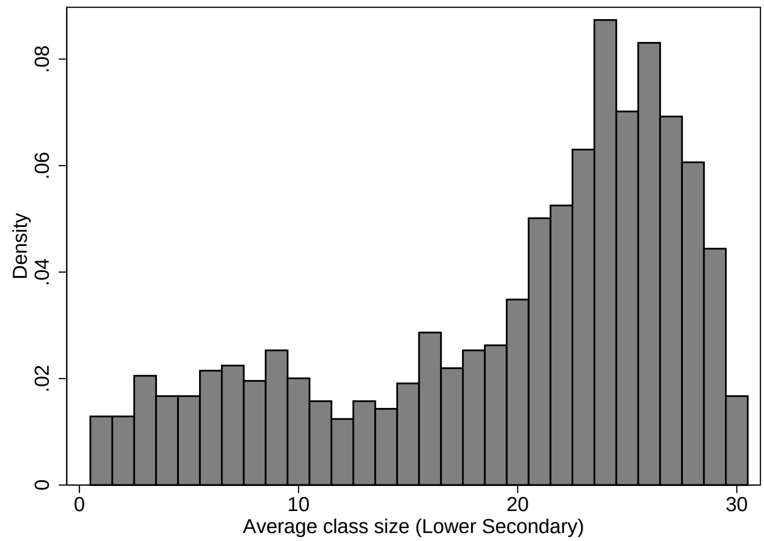

Figure 1.1 – Distribution of class size

shows that the majority of classes have between 20 and 30 students in it. A substantial fraction (35%) of the classes does, however, have fewer than 20 students. The (unweighted) average class size is equal to 20 (s.d. 7.4). The graph also suggests that the maximum class size of 30 is enforced; no class has a size exceeding the threshold of 30. And that the maximum class size is binding; without the restriction the distribution would have had some mass above 30.

1.4.3 Control variables

From other administrative databases we merged information on students' age, gender, ethnic minority background, household income, whether parents live together or not, and years of education of both parents. Comparable information is available and exploited in most (but not all) other studies examining the effect of class size on achievement.

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

Table 1.1 presents descriptive statistics of the control variables. Half of the students is female, the average age is 14.5 years, the average levels of fathers' and mothers' education are almost equal, only 3 percent of the sample consists of students with an immigrant background, and 30 percent of the students do not live in a household with both parents present. Almost a quarter of the students attends a school combining primary and lower secondary education, (weighted by numbers of students) average class size equals 24 and grade 7 enrollment equals 84.5. Teachers have on average 4.6 years of teacher training, 18.8 years of work experience as a teacher, 56 percent of the teachers is women and 17 percent of the teachers is on a temporary contract (again all weighted by numbers of students).[11]

Class size is not distributed randomly in the population because of sorting of students and teachers, and the targetting of educational resources. This is illustrated by Table 1.2 which reports the results of regressions of actual (average) class size on individual, teacher and school characteristics. The first column reports the results of a regression that only includes students' characteristics. This shows that actual class size increases with family income, with parents' levels of education, with parents being immigrants, and with parents not living together. When we add to this specification characteristics of the district, the positive effects of family income and parents' education remain significant but become smaller. The effect of having an immigrant background reverses (and is almost significantly negative) and the effect of living in a household with both parents present is no longer significant (and also changed sign). These latter results indicate that immigrant families and separated parents are concentrated in more densely populated districts. The results in Table 1.2 make clear that class size is not randomly distributed across students, and hence that some strategy is called for to identify the effects of class size on achievement. The next section discusses in detail the approaches that we will pursue.

1.5 Empirical approaches

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

$$y_i = x_i'\beta + w_{s(i)}'\alpha + \delta \cdot \overline{cs}_{s(i)} + \eta_{s(i)} + \psi_{t(i)} + \epsilon_i \tag{1.1}$$

[11] Student characteristics are measured at the moment of testing, whereas school and teacher information relates to October 1st of the schoolyear.

Table 1.1 – Sample summary statistics

	Mean (1)	s.d. (2)
Individual characteristics		
Girl	0.49	(0.50)
Age	14.53	(0.31)
ln(family income)	13.24	(0.75)
Education mother (years)	11.92	(2.89)
Education father (years)	12.20	(3.13)
1st or 2nd generation immigrants	0.05	(0.22)
Parents non-cohabiting	0.31	(0.46)
ln(pop. size school district)	10.05	(1.47)
ln(rural pop. size school district)	7.95	(0.75)
School characteristics:		
Combined school	0.20	(0.40)
Average class size grades 7-10	24.32	(4.20)
Enrollment grade 7	84.98	(41.79)
Teacher characteristics		
current (year t)		
Average teacher education (years)	4.61	(0.20)
Average teacher experience	18.86	(3.28)
Fraction of female teachers	0.56	(0.11)
Fraction of teachers with a temp. contract	0.17	(0.11)
Year=2002	0.52	0.50
N	115,084	
N schools	1,000	

Table 1.2 – Regression of actual class size on observables

	(1)	(2)
Girl	0.019	0.019
	(0.029)	(0.026)
Age	−0.064	0.071
	(0.051)	(0.046)
ln(family income)	0.430	0.168
	(0.042)∗∗∗	(0.025)∗∗∗
Education mother (years)	0.047	0.012
	(0.008)∗∗∗	(0.007)∗
Education father (years)	0.104	0.036
	(0.009)∗∗∗	(0.006)∗∗∗
1st or 2nd generation immigrants	1.313	−0.223
	(0.155)∗∗∗	(0.138)
Parents non-cohabiting	0.244	−0.017
	(0.045)∗∗∗	(0.040)
ln(pop. size school district)		1.211
		(0.068)∗∗∗
ln(rural pop. size school district)		0.434
		(0.148)∗∗∗
Year=2002		0.097
		(0.130)
adj. R-squared	0.0192	0.2103
N	115,084	115,084
N schools	1,000	1,000

x_i is a vector of observable attributes of the student and his parents, $w_{s(i)}$ a vector of observable school and teacher characteristics and $s(i)$ identifies the school of pupil i, $\overline{cs}_{s(i)}$ is the average class size that student i attended during her school career in school s, $\eta_{s(i)}$ a school effect, $\psi_{t(i)}$ is an effect for the year in which student i is in her final year of lower secondary school (2001/2 or 2002/3) and ϵ_i is all other determinants of achievement such as unobserved attributes of the student, parents and community. The coefficient of interest is δ, the class size effect. Note that a value-added specification is not feasible because achievement is only measured at the end of lower secondary school.

Table 1.3 shows the relation between student achievement and class size estimated using OLS. Results are presented from various specifications and separately for mathematics and languages, where we have pooled English and the two Norwegian subjects (and included subject dummies). Columns (1) and (4) are obtained from a specification without covariates; both estimates are positive and significant, indicating that pupils in larger classes perform better than pupils in smaller classes. The results in columns (2) and (5) are obtained from a specification that includes individual characteristics as controls, while columns (3) and (6) report the results from specifications that also include school and teacher characteristics. We control for school district characteristics in specifications (2), (3), (5) and (6). Including controls produces negative and highly statistically significant estimates on class size in all specifications. If we were to give these estimates credence, we would say that a one pupil reduction in class size improves test scores by between 0.4 and 0.6 percent of a standard deviation. The estimates exclude, however, class size effects larger than 1 percent of a standard deviation with 95% probability.

To address the potential endogeneity of class size we need variation in actual average class size that is arguably not subject to the choices of parents and schools' principals or teachers. We exploit two sources of such exogenous variation, one induced by a maximum class size rule and one based on population variation. The next two subsections describe briefly what these methods entail and how these can be implemented in the Norwegian context.

1.5.1 Maximum class size rules

Lower secondary schools in Norway are subject to maximum class size rules of 30 students per class. This rule creates a discontinuous relation between enrollment and class size. Just above multiples of 30 class size drops substantially. Following Angrist and Lavy (1999) we exploit this maximum class size rule in a regression discontinuity design. For

Table 1.3 – OLS. Dependent variable is exam results in 9th grade, unit of observations is the individual student's exam result

	Mathematics			Languages (pooled)		
	(1)	(2)	(3)	(4)	(5)	(6)
Average class size grade 7-9	0.005	−0.006	−0.005	0.004	−0.006	−0.005
	(0.003)*	(0.002)**	(0.002)**	(0.002)**	(0.002)***	(0.002)***
Individual characteristics						
Girl		0.018	0.018		0.506	−0.506
		(0.012)	(0.012)		(0.010)***	(0.0010)***
Age (years)		0.071	0.072		0.037	0.036
		(0.019)***	(0.019)***		(0.014)**	(0.014)**
ln(family income)		0.180	0.181		0.144	0.144
		(0.009)***	(0.009)***		(0.008)***	(0.008)***
Education mother (years)		0.075	0.075		0.057	0.057
		(0.002)***	(0.002)***		(0.002)***	(0.002)***
Education father (years)		0.066	0.066		0.049	0.049
		(0.002)***	(0.002)***		(0.002)***	(0.002)***
1st or 2nd generation immigrants		−0.446	−0.437		−0.232	−0.225
		(0.032)***	(0.032)***		(0.025)***	(0.025)***
Parents have different address		−0.345	−0.342		−0.189	−0.188
		(0.013)***	(0.013)***		(0.009)***	(0.009)***
ln(pop. size school district)		0.017	0.028		0.011	0.015
		(0.008)**	(0.008)***		(0.005)**	(0.006)**
ln(rural pop. size school district)		−0.018	−0.024		−0.011	−0.016
		(0.016)	(0.016)		(0.010)	(0.010)
School characteristics:						
Average teacher education (years)			0.042			0.070
			(0.055)			(0.042)*
Average teacher experience			0.016			0.011
			(0.003)***			(0.002)***
Fract. of female teachers			−0.097			0.059
			(0.110)			(0.077)
Fract. teachers with a temp. contract			0.067			0.022
			(0.088)			(0.065)
Combined school			0.052			0.028
			(0.027)**			(0.022)
R-squared	0.001	0.164	0.167	0.011	0.171	0.173
N pupils	38,045	38,045	38,045	77,039	77,039	77,039
N schools	755	755	755	958	958	958

Note: Standard errors are heteroscedasticity robust and corrected for school level clustering. All regressions include a control for year of observation; the laguage regressions also inlcude dummies for the language the student took the exam in.

this approach to work, schools need to be located randomly around the thresholds and no other discontinuities that may affect outcomes should exist.

Although identification in the regression discontinuity design is ultimately local (e.g. Hahn et al., 2001), Angrist and Lavy also proposed to instrument actual class size with predicted class size while conditioning on a smooth function of enrollment which is supposed to capture any direct effect of this variable on achievement. This identifying assumption essentially boils down to an exclusion restriction with respect to the discontinuities. In the analyses we will control for a cubic function of enrollment. An alternative for controlling for a smooth function of cohort enrollment is to restrict the sample to the regions around the kinks. We will present separate results from analyses for a sample which is restricted to schools with cohort enrollment in grade 7 at most 5 students away from the kinks.

Our analysis differs from previous analyses because we use predicted class size in grade 7, the first year in lower secondary school, to instrument average class size during the three years in lower secondary school. The reason behind this is that cohort enrollment in grades 8 and 9 and thereby predicted class size in these grades may depend on actual class size in grade 7, and is therefore potentially endogenous. Such dependence could for instance result from parents' decisions to move from schools where they experienced large classes in 7th grade to schools where they observed small classes in 7th grade.[12]

Figure 1.2 plots predicted class size and average actual class size against cohort enrollment. Average actual class size closely tracks predicted class size especially around the first kink.

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

[12] For students in combined schools enrollment in grade 7, the grade when they enroll in lower secondary school, may depend on actual class size during the previous primary school period. But the fact that the maximum class size rule changes from 28 in grade 6 to 30 in grade 7 causes that also in the combined schools students in grade 7 are confronted with a new class size.

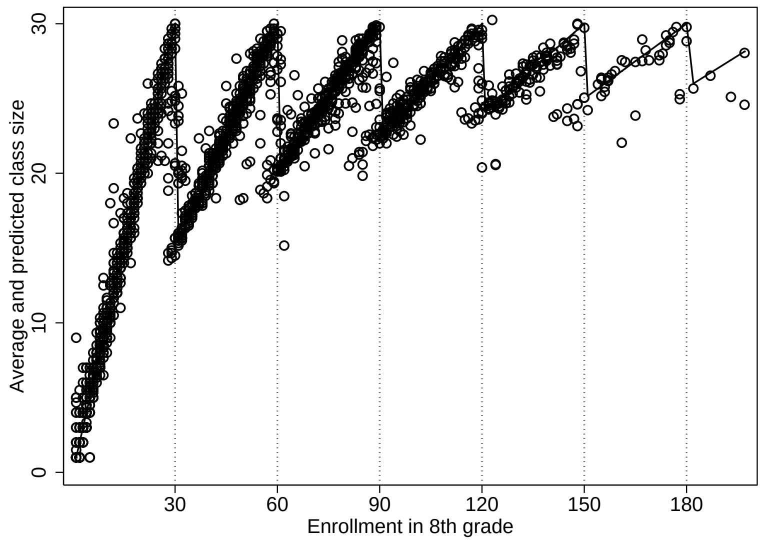

Figure 1.2 – Average and predicted class size, lower secondary school

To test this, we restricted the sample to students in classes close (at most 5 students away) to the cutoffs. For this sample, we regressed the indicator for being above the cutoff (versus being below it) on various sets of observable characteristics (including enrollment). In the specification that regresses the above/below indicator on student characteristics the p-value for joint significance of these characteristics equals 0.443. In the specification that regresses the above/below indicator on school characteristics the p-value for joint significance of these characteristics equals 0.495. And the p-value for joint significance of student and school characteristics together equals 0.518. The only separate variable that comes in marginally significant (10%-level) is teachers' education (lower above the kinks).

1.5.2 Population variation

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

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

$$\log(e_{st}) \quad = \quad \log(\bar{e}_{st}) + \log(u_{st}) \tag{1.2}$$

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

It is assumed that $\log(u_{st})$ is not correlated with any of the determinants for student achievement (x_{ist}, w_{st} and ϵ_{ist}) in equation (1.1), and since $\log(u_{st})$ is correlated with $\log(e_{st})$, a consistent estimate of $\log(u_{st})$ would be a valid instrument for class size. Hoxby assumes that $\log(\bar{e}_{st})$ changes smoothly over time and can be approximated by a grade-school-specific intercept and a school-specific polynomial in time. Equation (1.2) can therefore be written as

$$\log(e_{st}) = \sum_{k=0}^{K} \alpha_{sk} t^k + \log(u_{st}) \tag{1.3}$$

We estimate equation (1.3) for each school separately to obtain the estimated residuals $\widehat{\log(u_{st})}$ which serve as the instrument for class size in (1.1). As in the maximum class size approach we base our instruments on enrollment in grade 7. We have data on enrollment for ten years (from 1992 to 2002) and we show results that use up to a quartic in time because quartics appear to capture all of the smooth variation over time in enrollment. This approach therefore assumes

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

which is implemented in a fixed effects 2SLS procedure which eliminates school fixed effects.

Note that this approach is valid as long as the variation in enrollment does not trigger a change in the number of classes from 2001/02 to 2002/03 since this would violate monotonicity of the instrument. For this reason we only include schools where predicted number of classes are the same in 2001/02 and 2002/03. Finally, to take into account any remaining endogeneity – which would occur when parents transfer their children to other schools in response to the class size their children are experiencing – we follow Hoxby and also carry out the analysis at the district level in order to cancel out within school district transfers. Other sources of bias such as transfers to private schools or selective grade retention or advancement are not present in Norway because of the neglible presence of private schools and the absence of grade retention/advancement practices.

1.5.3 Variations of population variation

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

Population variation and maximum class size rules In addition to exploiting population variation while conditioning on number of classes, Hoxby (2000) also uses the population variation that triggers changes in the number of classes (and therefore class size) because of maximum class size rules. Hoxby identifies all events where the number of classes changed because of maximum class size rules. For all cases where the enrollment change was not more than 20 percent, she then estimates a first differenced version of equation (1.1).

Small schools Urquiola (2006) focuses on small schools with only one predicted class (per grade). The idea here is that the endogeneity of class size is less of a problem in small rural schools since they are local monopolies and parents cannot enroll their children elsewhere thereby ruling out between school sorting, in addition between class sorting is not an issue since these small schools have only one class. As noted by Urquiola, class size may still correlate with unobserved factors that affect achievement since class size depends on community size, fertility, etc. All class size variation is therefore generated by differences in enrollment and this approach is like Hoxby's, but does not control for school fixed effects since it exploits cross sectional variation, moreover it does not control for trends. The implementation is straightforward and amounts to estimating equation 1.1 by OLS, where schools with more than one predicted class per grade are excluded. Furthermore the sample is restricted to districts with only one school. This strategy is suitable for the Norwegian context with rural settlement pattern and many small schools.

Within school between grade comparison Wößmann and West (2006) use within school and between grade variation to estimate class size effects. By comparing adjacent grades they account for between school sorting since they eliminate school fixed effects (to the extent to which they are uniform across grades). To eliminate within school sorting problems, actual class size is instrumented with average class size at the grade level. The reduced form estimates are therefore equivalent to the population variation approach of Hoxby, but without correcting for a trend, and without discarding schools around the kinks. We will implement a version of this approach where we exploit variation in enrollment between cohorts in the same school instead of variation between grades.[13]

1.6 Results

Estimates of the class size effects based on the different approaches are presented in Table 1.4.[14] These results are all based on specifications with the full set of control variables (e.g. the same control variables as those included in the OLS-regression). Less elaborate specifications produce very similar results (and are available from the authors upon request). Moreover, for students in combined schools we also know their class size during grades 1 to 6, assuming that students do not switch between combined schools and separate schools. We have also conducted analyses using average class size during the six years in primary school as a class size measure. As mentioned above, we have chosen not to present these results since the results are very similar to the results for average class size in lower secondary schools.[15]

The first thing to note in Table 1.4 is that only one out of the twelve effect estimates presented in the table has the expected negative sign and that none of the effect estimates is significantly different from zero. With the true effect equal to zero, the likelihood of having 11 out of 12 estimates of the same sign is only 0.3%, thereby suggesting that the class size effect in Norwegian lower secondary schools is actually positive. Almost all effects are estimated quite precisely, so that substantial negative effects of class size of achievement can be ruled out with high probability. Below each effect estimate we report in square brackets the largest (most negative) effect that falls in the 95% confidence

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

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

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

interval (point estimate - 1.96*s.e.). In four cases, this maximum effect is in the range of 1 and 2 percent of a standard deviation for a one student reduction in class size. The other eight estimates imply maximum effects of 1 percent of a standard deviation or less.

The results in Table 1.4 reveal no clear pattern in the estimated effect sizes across the different methods, apart from the fact that those that exploit population variation including a trend (rows 3 and 4) have substantially larger standard errors than the other methods, especially when applied to math achievement.

Some have argued that maximum class size rules may be inappropriate to estimate the impact of class size on achievement because of potential input substitution around the discontinuities. To investigate whether this is an issue in Norway we replaced our class size measure by teacher hours per student (which is a common measure of resource use in Norway, e.g. Hægeland et al. (2005, 2007)). The correlation between teacher hours per pupil and our class size measure is 0.8.

Table 1.5 shows results from 2SLS estimations based on maximum class size rules as in Table 1.4. If input substitution is complete our instrument would not affect teacher hours per pupil. As can be seen from Table 1.5, when predicted class size increases by one, teacher hours per pupil decrease by 0.67. Around the discontinuity the coefficient equals to -0.41 but is still highly significant. Although there might be some input substitution, it is clear from these results that it is far from being complete. Moreover, there remains a statistically significant difference in terms of teacher hours around the discontinuities.

The two bottom panels of Table 1.5 report impact estimates of teacher hours on mathematics and language achievement. The effects are in line with those in Table 1.4; they are highly insignificant and tend to have the wrong sign.

1.7 Heterogeneous class size effects

In this section we investigate whether the zero class size effect reported in the previous section, hides positive effects of class size reduction for specific sub-groups. The reason to explore this is that various studies report that effects of class size reduction are more pronounced for disadvantaged groups (Angrist and Lavy, 1999; Krueger, 1999) or depend on teacher characteristics (Wößmann and West, 2006).

Table 1.6 presents results for specific groups based on the method that exploits maximum class size rules (for the full sample and around kinks) and on the method that exploits population variation (and includes a trend). The first sub-group is that of pupils with low educated mothers (less than 10 years of education). Results from the maximum

Table 1.4 – Estimated class size effects using different methods

	Math			Lang		
	Effect	s.e.	N (N_S)	Effect	s.e.	N (N_S)
Maximum class size	<.001	(0.005)	38,045 (755)	0.005	(0.004)	77,039 (958)
	[− 0.010]			[− 0.003]		
Maximum class size discontinuity sample	0.006	(0.009)	10,810 (190)	0.009	(0.006)	23,970 (280)
	[− 0.012]			[− 0.003]		
Population variation	0.018	(0.020)	26,799 (551)	0.002	(0.006)	52,941 (725)
	[− 0.020]			[− 0.010]		
Population variation at district level	0.025	(0.082)	16,215 (416)	0.035	(0.024)	30,207 (527)
	[− 0.136]			[− 0.012]		
Maximum class size+population variation	0.054	(0.026)	4,856 (66)	0.013	(0.009)	11,808 (83)
	[0.002]			[− 0.004]		
Small monopoly schools (enrollment<25)	0.016	(0.014)	672 (51)	0.005	(0.009)	1,680 (70)
	[− 0.011]			[− 0.013]		
Population variation without trend	0.011	(0.009)	38,045 (755)	−0.001	(0.003)	77,039 (958)
	[− 0.007]			[− 0.007]		

Table 1.5 – The impact of School Resources (Teacher Hours/Pupil) on Achievement - 2SLS estimates based on maximum class size rules

	All	DS±5
	First Stage[1]	
Predicted Class size	−0.671	−0.413
	(0.080)***	(0.106)***
N (N_S)	114,665 (994)	34,669 (377)
	Second Stage - Mathematics	
Teacher Hours/Pupil	0.005	−0.012
	(0.006)	(0.020)
N (N_S)	37,828 (752)	10,701 (189)
	Second Stage - Language	
Teacher Hours/Pupil	−0.002	−0.010
	(0.003)	(0.007)
N (N_S)	76,837 (953)	23,968 (279)

[1] The first stage results shown here are based on the full sample (i.e. both math and language scores).

class size rule using the full sample suggests that there is a significantly negative effect of class size on achievement in both language and math. The results obtained from the discontinuity sample and from the population variation method do, however, not point in the same direction. The evidence in favor of beneficial class size effects for students with low educated mothers is therefore limited.

As a second subgroup we consider children from immigrant families (their share in the population is only 5 percent). For this group we find no consistent pattern; one of the estimates is positive and significantly different from zero at the 5%-level. The five other estimates do, however, not confirm this result (none significant and one with a negative sign).

The results for girls - reported in the third block - are all very small, and do not differ substantially or significantly from the results from girls and boys together. The fourth part of the table presents results for pupils in the lowest quartile of the age distribution. Here again, the emerging pattern is far from consistent. For math achievement, the maximum class size rule applied to the full sample results in a significantly negative estimate, whereas the approach based on population variation shows the reverse (with a very large point estimate).

The final two parts of the table present effects for pupils in schools of which the teaching staff belongs to the lowest quartile in terms of education level or in terms of experience. Again the different methods fail to reveal a consistent pattern. Applying the maximum class size rule on the full sample indicates a negative effect of class size on language achievement for pupils in schools with low educated teachers. This result concurs with the hypothesis advanced by Wößmann and West (2006). This finding is, however, not confirmed by the other results in this block; three of the five other estimates even have the opposite sign.

We are inclined to conclude that there is only rather weak evidence for heterogeneous class size effects. If anything, it are pupils from low educated families and pupils in schools with low educated teachers that benefit the most from a reduction in class size.

1.8 Conclusion

Based on estimation results that exploit arguably exogenous variation in class size, we find no significant effect of class size during lower secondary school on achievement in grade 9 in Norway. We can even exclude effects as small as 1.5 percent of a standard deviation for a one student reduction in average class size during three years.

Effects are rather similar for different social backgrounds groups and for schools with different teaching staffs. There is only a rather weak indication that students with low educated mothers and students in schools with low educated teachers benefit somewhat from a reduction in class size.

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

Potential explanations for the zero class size effect in Norway are substitution of parental inputs and uniform teaching styles. Substitution of parental inputs occurs if the parents of pupils who are placed in small classes reduce their own inputs in the education production function (cf. Todd and Wolpin 2003). Uniform teaching styles annihilate potentially beneficial class size effects if teachers are unable to take advantage of the extra time they could have per student. Although the fact that we find some indication that pupils of low educated teachers benefit more from class size reduction than pupils of high educated teachers, seems to contradict this explanation. Further research is

Table 1.6 – Class size effects for sub-groups using different methods

	Math			Lang		
	Effect	s.e.	N (N_S)	Effect	s.e.	N (N_S)
Low educated mothers						
- Maximum class size	−0.010	(0.005)*	3986 (626)	−0.012	(0.004)***	8270 (843)
- Maximum class size discontinuity sample	0.020	(0.015)	1112 (165)	−0.008	(0.010)	2474 (267)
- Population variation	0.030	(0.041)	2778 (390)	−0.002	(0.016)	5732 (567)
Immigrants						
- Maximum class size	0.016	(0.013)	2015 (388)	0.004	(0.010)	3658 (517)
- Maximum class size discontinuity sample	0.086	(0.038)**	642 (121)	0.018	(0.015)	1168 (198)
- Population variation	0.211	(0.138)	1280 (189)	−0.010	(0.034)	2392 (261)
Girls						
- Maximum class size	−0.006	(0.004)	18533 (718)	−0.003	(0.003)	38041 (934)
- Maximum class size discontinuity sample	0.014	(0.011)	5283 (175)	0.008	(0.006)	11784 (272)
- Population variation	0.025	(0.025)	13038 (534)	0.005	(0.007)	26066 (698)
Youngest quartile						
- Maximum class size	−0.010	(0.005)**	9542 (689)	−0.004	(0.003)	19176 (905)
- Maximum class size discontinuity sample	0.004	(0.013)	2722 (171)	0.006	(0.008)	6082 (270)
- Population variation	0.091	(0.035)***	6669 (484)	0.004	(0.011)	13102 (654)
Teacher education in lowest quartile						
- Maximum class size	−0.003	(0.004)	8438 (269)	−0.009	(0.003)***	20448 (389)
- Maximum class size discontinuity sample	0.018	(0.016)	2284 (56)	0.010	(0.008)	6156 (96)
- Population variation	−0.056	(0.042)	6503 (211)	0.004	(0.010)	15326 (318)
Teacher experience in lowest quartile						
- Maximum class size	0.004	(0.006)	9271 (216)	−0.005	(0.004)	19430 (263)
- Maximum class size discontinuity sample	0.007	(0.024)	2363 (50)	0.016	(0.012)	6084 (74)
- Population variation	0.558	(1.137)	6627 (156)	0.008	(0.015)	13150 (190)

required to differentiate between these various explanations. This is important because the policy implications are quite different. If the zero-effects are due to substitution of parental inputs there is not much hope that the policy effects can be improved, although the reduction of parental inputs should be included in a cost-benefit analysis. If the zero-effects are due to uniformity in teaching styles, there remains scope for improvement by teaching teachers to take advantage of smaller classes.

2 Who benefits from homework?

2.1 Introduction

The empirical literature on the impact of school inputs on educational performance increasingly recognizes that such impacts cannot be well summarized by an average effect. The effects of different educational policies are very likely to vary across pupils from different family backgrounds because of substitution between the quality of the school environment and the quality of the home environment. This explains why the effects are typically bigger for disadvantage children (e.g. Guryan, 2001; Lindahl, 2001; Card and Payne, 2002; Papke, 2005). A good example is early schooling which is assumed to have a beneficial effect on pupils from disadvantage backgrounds since learning is substituted from the home environment to the class room at an early stage (see for example Leuven et al., 2006).

The current paper takes a slightly different approach and focus on complementarities between school inputs and the quality of the home environment. It does so by considering homework. Since homework takes learning out of the class room we expect pupils from higher socioeconomic backgrounds to benefit more from homework assignments.

Homework is commonly assigned to pupils in elementary education because it is believed to improve the performance. This belief is not confirmed by the education literature. Cooper (1989) conducted a review of nearly 120 empirical studies concerning the effect of homework on pupil outcome, and concluded that for elementary school pupils the effect of homework on achievement is negligible, if it exists at all.[1]

Younger pupils, especially those in elementary education, have less well-developed study habits (see for example Cooper, 2007). This suggests that they need parental help in order to complete their homework assignments. Since parental input is not equally distributed across pupils and is most likely a supplement to homework, children who are

[1]In the same study Cooper points out that the effect of homework on achievement is grade dependent. For high school students and also junior high school students homework has a positive effect. Other studies that find a positive effect of homework on student achievement in higher grades are; Betts (1996); Aksoy and Link (1999); Eren and Hendersen (2006).

most in need for extra work are the ones least likely to benefit from it. Consequently, the lack of a (clear) homogeneous effect on pupil outcome may be due to that homework exacerbate heterogeneous effects (Baker, LeTendre and Akiba, 2005 pp. 132).

This aspect of homework has received surprisingly little attention in the empirical literature. Using Dutch survey data on pupils and teachers in elementary school, this paper is the first study that empirically analyzes the heterogeneous effects of homework.

Suspect variation in homework may arise because of several reasons and the primary objective in the empirical analysis is to eliminate sources of bias that possibly contaminate our results. First of all, potential biases caused by unobserved school quality and student selection are taken out by exploiting variation within schools. And in order to distinguish the effect of homework from unobserved teacher and pupil effects, we proceed with comparing within class differences in test scores in classes where everybody gets homework to within class differences in test scores in classes where nobody gets homework. Thus the empirical strategy boils down to a difference in difference approach. The advantage with this latter procedure is that confounding effects of unobserved teacher characteristics drop out as long as they are homogeneous within class. And since everybody in the class either gets homework or does not get homework we can also rule out within class correlations between homework and unobserved individual pupil effects.

Our findings suggest that the test score gap is indeed larger in classes where everybody gets homework than in classes where nobody gets homework. More precisely, we find that pupils belonging to the upper part of the socioeconomic status scale gain, whereas pupils from the lowest part are unaffected.

If the intention with homework is to reinforce the children's learning process at home (and thereby benefit from homework) and families are unequal to the task, the students will not receive the same quality of education. Furthermore, it informs us about an early source of inequality and should be important in a society who cares about inequality in terms of extensive redistribution policy. The Netherlands is a country with a longstanding tradition in attempting to promote equality of opportunity in education (Leuven et al., 2007a)

The structure of the paper is as follows; Section 2.2 describes the data; in Section 2.3 we discuss complementarity between homework and the home environment; the empirical approach is lined out in Section 2.4; Section 4.4 presents the results; and Section 2.6 concludes.

2.2 Institutional settings and data

Elementary school in the Netherlands consists of eight grades. Children start school when they are 4/5 years old and finish when they are 11/12 years old. Every teacher covers all the subjects in the class. In the period under investigation, schools did not have catchment areas and there was free school choice.

The empirical analysis in this paper build on data from the four last waves of the Dutch PRIMA survey. This is a biannual survey which samples schools and contains data/information on about 10 percent of the Dutch pupils in grade 2, 4, 6 and 8. The first survey took place in the school year 1993/1994 and the last wave used in this paper is for the school year 2002/2003 (in total five school years). Several actors contributed to the collection of the data: The pupil's parents; the pupil's teachers; the school's principals; and the pupils themselves.

In the Netherlands homework is typically assigned on language related tasks such as reading and writing. To measure pupil outcome we therefore make use of test scores from a language-test.[2] This test is identical across schools and for all the four last PRIMA waves and graded externally. We standardize it by grade and year with zero mean and standard deviation equal to one in order to make it comparable.

Information on homework comes from the teacher questionnaires. The teachers in grades 4, 6 and 8 were asked how often they assign homework and could choose between four answers: i) hardly or never to anyone in my class; ii) only to the weak students in class; iii) only to the good students in class; and iv) to everybody in my class.

An overview of the teachers' homework practice in language are given in Table 2.1. In grade 4, a majority of the teachers hardly ever give homework to anyone in the class, whereas in grade 8 a majority of the teachers give homework to the whole class. In both grade 4 and 6, a substantial amount of teachers report that they give homework to only the weak pupils in their class. Basically nobody gives homework to the good pupils only. The data contains no information on how often the pupils get homework, but based on anecdotal evidence homework is typically given regularly, but not every day.

In the empirical analysis, we compare classes that get homework to classes that do not get homework. This involves that we exclude classes where only weak or good pupils get homework. This amounts to 1,677 classes and 31,572 pupils and does not lead to

[2]The test was taken halfway during the school years. There are some small differences regarding who was the responsible staff in the class room when the test was taken. In the second wave the test was monitored by an external examiner, while in the three reminding waves the teacher of the class was in charge during the test. Year dummies are added in the regression analysis.

Table 2.1 – Homework: Summary statistics, the teacher's homework practice in language.

	Grade		
	4	6	8
Hardly or never	1,222	926	353
Weak students only	767	659	232
Good students only	3	5	11
Everybody	527	813	1,698

major changes in the sample (we exclude row two and three in Table 2.1).

From the parent questionnaires we have information on the pupil's age, gender, education levels of parents and an indicator variable which equals one if the pupil comes from non-Western migrant backgrounds.[3] Parent's education level is divided into primary education, lower vocational, upper secondary/intermediate vocational and university/higher vocational. We also control for some class level characteristics such as the teacher's experience, gender and log of class size. These variables come from the teacher questionnaires. Since each teacher teaches all the subjects in the class, the observation unit is the same for teachers and classes. For simplicity reasons, the term "class characteristics" therefore denotes both characteristics with the class and the teacher in the remainder of the paper.

Table 2.2 gives a descriptive overview over the explanatory variables used in the empirical analysis. About half of the pupils are girls, the average age is 10 years and 24 percent come from a non-Western migrant background. Furthermore, 18 (14) percent of the mothers (fathers) have primary education, 32 (32) percent have lower vocational, 30 (24) percent have upper secondary/intermediate vocational and 12 (16) percent have higher vocational/university. Concerning the class characteristics, the average teacher has 18.4 years with experience and teaches a class consisting of 23 pupils. 54 percent of the teachers are females.

Homework is not distributed randomly in the population. This is illustrated by Table 2.3 which presents OLS estimates obtained from regressing the indicator variable for homework on observed pupil and class characteristics. The probability of getting homework is typically higher for older pupils and pupils from non-Western migrant backgrounds. With respect to parental education, pupils whose mother's education is

[3]This variable is derived from the funding scheme for Dutch primary schools that gives students with an ethnic minority background a weight equal to 1.9.

Table 2.2 – Sample summary statistics

	Mean	s.d.
Individual Characteristics (N = 96,723)		
Girl	0.47	0.50
Boy	0.47	0.50
Age	10.02	1.80
Non-Western migrant background	0.24	0.43
Mother's education		
- Primary	0.18	0.38
- Lower vocational	0.32	0.47
- Upper secondary/inetermediate vocational	0.30	0.46
- University/higher vocational (higher education)	0.12	0.33
- Missing	0.08	0.27
Father's education		
- Primary	0.14	0.35
- Lower vocational	0.32	0.47
- Upper secondary/intermediate vocational	0.24	0.43
- University/higher vocational (higher education)	0.16	0.36
- Missing	0.14	0.35
Class Characteristics (N = 5,539)		
Class size	23.10	5.90
Teacher's experience	18.40	10.60
Female teacher	0.54	0.50

Missing information: student's gender (5.42%); student's age (1.10%); female teacher (1.03%); class size (0.58%); teacher experience (0.31%)

higher than or equal to upper secondary are less likely to get homework than pupils with a lower vocational educated mother. And pupils with a primary (upper secondary) educated father are more (less) likely to get homework than pupils with a lower vocational educated father. The chances for homework assignments are also higher if the teacher is a woman. These findings indicate that homework is strongly correlated with observed characteristic. Moreover, they illustrate that there is a remedial aspect connected to giving homework in the sense that it is the weakest pupils who get homework. Hence potential endogeneity problems must be addressed.[4]

2.3 Complementarity between homework and home environment

As mentioned above, some home environments may be more conductive to learning than others, allowing heterogeneous effects of homework to arise. It is therefore interesting to shed some more light on the relationship between parental help with homework and parental background. The first wave of the PRIMA survey asked the parents of pupils in grade 4 how much they help their children with homework (conditional on that the children get homework). We have separate information on mothers and fathers, and the frequency of parental help with homework is divided into three categories; “almost never”; “sometimes”; and “often”. A descriptive overview over these answers are found in Table 2.4. More mothers than fathers “often” assist with homework, whereas more fathers than mothers “almost never” assist with homework. In order to check whether paternal help with homework is a supplement to maternal help with homework and vice versa, we have also cross tabulated maternal and paternal help. What is remarkable is that in 77 percent of the cases where the mother answers that she almost never helps with homework, also the father from the same household answers that he almost never helps with homework. This really indicate that some children do not get help with homework from neither parent. The next question we then should ask ourselves is why this is the case? Is it due to that these children are not in need for assistance, or is it because help with homework is totally lacking in some home environments?

To see which parents give help with homework, Table 2.5 proceeds with presenting the marginal effects obtained from regressing the frequency of maternal help with homework on parental background using a multinomial regression model. “Often” serves as the reference category.[5] This table shows that parental help with homework is strongly

[4]The results are unaltered if we apply a binomial model.

[5]If maternal help in Table 2.5 is substituted with paternal help the pattern remains unchanged, but the correlations are a bit less precisely estimated.

Table 2.3 – The determinants for assigning homework

	(1)
INDIVIDUAL CHARACTERISTICS	
Girl	0.0014
	(0.0030)
Age	0.0072
	(0.0030)∗∗
Non-Western migrant	0.0883
	(0.0095)∗∗∗
Mother's education (ref = Low. Voc.)	
- Primary	−0.0009
	(0.0077)
- Upper secondary	−0.0116
	(0.0053)∗∗
- Higher education	−0.0300
	(0.0075)∗∗∗
Father's education (ref = Low. Voc.)	
- Primary	0.0178
	(0.0080)∗∗
- Upper secondary	−0.0149
	(0.0052)∗∗∗
- Higher education	−0.0085
	(0.0076)
CLASS CHARACTERISTICS	
Log of class size	−0.0229
	(0.0244)
Teacher's experience	−0.0005
	(0.0006)
Female teacher	0.0320
	(0.0142)∗∗
N	96,723

Note: Reported estimates are from linear probability models where the dependent variable is a dummy variable which equals 1 if the class gets homework and 0 if the class does not get homework, unit of observation is the individual student. Standard errors are heteroscedastic robust and corrected for class level clustering. Included are also a constant term, year dummies, grade dummies, dummy variables for missing information on gender, age and parental education, missing information on class size, teacher's gender and experience. *, ** and *** denote statistical significance at the 10, 5 and 1 percent level respectively.

Table 2.4 – Parental help with homework

	Mother	Father
Almost never	8.7%	20.5%
Sometimes	44.2%	58.1%
Often	47.1%	21.4%

Note: 5,744 observations on maternal help with homework and 4,496 observations on paternal help with homework.

related to parental education and ethnic background. First, conditional on the education level, the probability that a mother with a non-Western migrant background "almost never" help with homework is 14 percent. And second, a pupil with a primary educated mother is 6 percent more likely to "almost never" get help with homework compared to a pupil with a lower vocational educated mother. Furthermore, if the mother's highest level of obtained education is primary school, the chances are also higher that she *only* sometimes help with homework than a mother whose education is lower vocational. Note that a higher educated educated mother is more likely to "almost never" assist with homework than a lower vocational educated mothers. This could be due to time constraint or because the children of these mothers need less help. Finally, girls get less help than boys.[6]

This analysis already illustrates that children from more disadvantaged backgrounds receive less help with their homework assignments. It implicitly assumes that the quality of the parental inputs is the same across groups. It seems however very likely that the unobserved quality of parental input correlates positively with the amount of human capital they posses. This suggests that, even keeping constant the time parents spend on homework with their children, those with less able parents learn less from homework. An example in place are children from ethnic minority families where the parents have limited Dutch language skills.

One could think of many possible reasons why children from ethnic minority and lower educated backgrounds get less help with homework. One explanation is parental priority. Families from lower socioeconomic backgrounds may prioritize other activities more than academic achievement. Another explanation is that some parents are not

[6]Note: Among mothers with primary level as highest level of obtained education, 60 percent are from migrant families. Also note that 24% of the migrant mothers have missing information on highest level of obtained education.

Table 2.5 – Relation between parental help with homework and parental background, estimates from a multinomial logit model. "Often" is the reference category.

	Frequency of parental help with homework	
	Almost never	Sometimes
Girl	0.0155	0.0515
	(0.0066)$**$	(0.0132)$***$
Non-Western migrant	0.1400	-0.0134
	(0.0170)$***$	0.0218
Mother's education (ref = Low. Voc.)		
- Primary	0.0581	0.0427
	(0.0159)$***$	(0.0257)$*$
- Upper secondary	-0.0051	-0.0032
	(0.0099)	(0.0171)
- Higher education	0.0369	-0.0290
	(0.0172)$**$	(0.0245)

Note: N = 5744. $\hat{p}_1 = 0.0725$ and $\hat{p}_2 = 0.4520$. Reported are marginal effects. Standard errors in (...). *, ** and *** denote statistical significance at 10, 5 and 1 percent level respectively. Included are also dummy variables for missing information on gender and mother's education.

able to help their children, because of lack of the relevant skills. Summarized, these findings indicate that the effects of homework on achievement differ across children from different socioeconomic backgrounds, and are the first indications that heterogeneous effects of homework may exist. The next section is concerned with how to empirically identify and test this.

2.4 Empirical approach

To estimate the impact of homework we start out with assuming that pupil i's achievement (y) in class j and school s can be explained by the following education production function:

$$y_{ijs} = x_i'\beta + \omega_{js}'\varphi + \delta_i hw_j + \varepsilon_{ijs} \tag{2.1}$$

where x_i is a vector of observed attributes of the pupils and his parents; ω_j a vector of observed class characteristics; and hw_j is an indicator variable taking the value 1 if the class receives homework and zero if the class does not get homework. The parameter of interest is δ_i, where subscript i indicates that the effect of homework may differ between pupils as already discussed above.

The central problem we face when estimating equation (2.1) by Ordinary Least squares (OLS) is that the estimate of the homework effect may be contaminated by omitted variables such as the influence of unmeasured class and teacher characteristics; unobserved school characteristics; as well as unobserved pupil and family characteristics. Consequently, we must be careful with giving this estimate of homework a causal interpretation. Note that the sign of the bias is not clear a priori. Good schools may give homework to do even better, or bad schools may give homework to make up for the poor environment. Similarly, poor teachers may use homework to compensate for the lack of teaching skills, while good teachers may use it to achieve ambitious goals. Correlation between homework and unobserved school and class characteristics may also arise because of pupil sorting. How these correlations net out is unclear.

Since one of the potential sources of bias is unobserved school characteristics, we include a school fixed effect, ψ_s, to equation (2.1):

$$y_{ijs} = x_i'\beta + \omega_{js}'\varphi + \delta_i hw_j + \psi_s + u_{ijs} \tag{2.2}$$

This identification rests on schools where there is variation in homework practices between classes within grades (i. e., ψ_s is actually a school-grade fixed effect). As a first attempt to check whether heterogeneous effects of homework exist we will estimate equation (2.2) for different sub samples of pupils stratified after mother's education.[7]

Although fixed effects estimation improves on OLS, u_{ijs} may still contain unobserved characteristics of the teachers and pupils, allowing within school differences in homework assignment to correlate with differences in teacher quality and attributes of the pupils. A standard way to solve this problem would be estimate a more elaborate fixed effects model and to include a teacher fixed effect (θ_j) and a pupil fixed effect (ϕ_i) and estimate the following equation:

$$y_{ijs} = x_i'\beta + \omega_{js}'\varphi + \delta_i hw_j + \theta_j + \phi_i + \psi_s + u_{ijs} \tag{2.3}$$

Unfortunately we cannot estimate equation (2.3) using our data. The remainder of this section is therefore concerned with how to purge our homework estimates from the confounding effects of these unobserved characteristics.

We will do so by looking at within class differences in test scores. The advantage of this procedure is that any unobserved teacher fixed effect, θ_j, will drop out as long as it is homogeneous across pupils within class. Moreover, since everybody in the class either gets homework or does not we also have that at the class level homework does not correlate with unobserved pupil ability, i.e. $E[\phi_i|hw_j] = 0$. Under these assumptions we will estimate the effect of homework on within class differences in test scores using the reduced form equation given by equation (2.4).

$$ineq(y_i)_{js} = \rho hw_j + \psi_s + \nu_{js} \tag{2.4}$$

where as a measure on $ineq(y_i)_j$ we use variance of the test scores as well as differences between various percentiles within the class. To simplify the analyses we base our inequality measures on residuals from OLS regressions that correct test scores for observed student and class characteristics. A descriptive overview of the inequality measures are given in Appendix Table 2.10.

Since we still condition on the school-grade fixed effect, ψ_s, this approach boils down to a difference-in-difference strategy. If the impact of homework on achievement is homogenous its effect on inequaltiy will be zero. The hypothesis we want to investigate is whether $\rho > 0$, that inequalities in classes where everybody gets homework is larger

[7]Since a crucial part of the identification strategy involves exploiting variation in homework practices across class rooms within schools and grades, a further restrictions in the sample is required. More precisely, we need to eliminate schools with no variation in homework within grades and schools. A descriptive overview over this reduced sample is shown in Appendix Table 2.9.

than inequalities in classes where nobody gets homework.

Although the assumptions made here are restrictive they improve substantially on the (individual) level equations. Another advantage of this approach is that possible problems connected to reverse causality weaken. When analyzing within class differences in test scores, reverse causality is only a problem if teachers respond to large inequalities by giving homework to the whole class. If the aim of homework is to smooth out inequalities, we would think that the rational way of reacting is to give homework to only the weak pupils in the class.[8]

2.5 Results

Although homework is not randomly assigned across pupils, it is still of interest to look at the relation between homework and pupil achievement estimated with a simple OLS. This is reported in Table 2.6. In column (1) which is obtained from a specification without covariates, the homework estimate is negative and highly significant. More precisely it indicates that pupils who get homework perform 12 percent of a standard deviation worse on average than pupils who do not get homework. As already discussed above, this effect cannot be given a causal interpretation since homework tends to be given to the weakest pupils. Column (2) confirms this. When controlling for individual characteristics, the effect decreases to 0.3 percent of a standard deviation and is clearly insignificant. In column (3) where we also control for class characteristics the effect is further reduced to 0.02 percent of a standard deviation. Hence, this confirms that homework appears to be highly correlated with both individual and class characteristics, and some more elaborated strategies are indeed essential in order to identify the effects of homework.

One such strategy is to compare pupils within schools and grades only. Table 2.7 presents result from estimating equation (2.2) with school-grade fixed effects. The left panel (column 1 to 3) reports results from various specifications including all pupils, whereas the right panel stratifies pupils by mothers' education and reports results from the most elaborate specification only.

The first thing to note in Table 2.7 is that the point estimates of homework are not sensitive to the inclusion of individual and class characteristics. This can be seen in column (1) to (3). This suggests that homework is not correlated with (observed)

[8]Also note that ruling out reverse causality is a bigger problem if the dependent variable is the average achievement level. A teacher may decide to give homework to the whole class in reaction to low average achievement level in that class.

Table 2.6 – The relation between homework and achievement, OLS

	(1)	(2)	(3)
Homework	−0.1175	−0.0031	−0.0002
	(0.0163)***	(0.0114)	(0.0114)
Individual characteristics			
Girl		0.0738	0.0738
		(0.0061)***	(0.0061)***
Age		−0.1669	−0.1663
		(0.0053)***	(0.0053)***
Non-Western migrant		−0.5344	−0.5223
		(0.0114)***	(0.0115)***
Mother's education (ref. = Low. Voc.)			
- Primary		−0.1339	−0.1303
		(0.0117)***	(0.0117)***
- Upper secondary		0.2100	0.2083
		(0.0085)***	(0.0085)***
- Higher educateion		0.3923	0.3911
		(0.0127)***	(0.0127)***
Father's education (ref. = Low. Voc.)			
- Primary		−0.0744	−0.0717
		(0.0119)***	(0.0119)***
- Upper secondary		0.1431	0.1422
		(0.0088)***	(0.0088)***
- Higher education		0.2981	0.2976
		(0.0115)***	(0.0115)***
Class characteristics			
Log of class size			0.0960
			(0.0211)***
Teacher's experience			0.0028
			(0.0005)***
Female teacher			−0.0139
			(0.0111)
N	96,723	96,723	96,723
R-squared	0.0028	0.2022	0.2039

Note: The unit of observation is the individual student. Standard errors are heteroscedastic robust and corrected for class level clustering. Included in all specifications are a constant term, grade and year dummies. The specifications which control for individual characteristics also include dummy variables for missing information on gender, age and mother's and father's education. The specification(s) which control for class characteristics also include(s) dummy variables for missing information on class size, teacher's gender and experience. *, ** and *** denote statistical significance at the 10, 5 and 1 percent level respectively.

individual and class characteristics when comparing pupils within the same school and grade. Moreover, the effect of homework on pupil achievement is positive, but only significant at the eighteen percent level. Turning to the right panel, column (4) shows that the point estimate is negative for pupils with a primary educated mother, but the point estimate is close to zero and insignificant. On the other hand, homework appears to have a positive effect on the achievement for pupils whose mothers have at least lower vocational (column 5). The effect is also significant at the 10 percent level. If we were to give the latter result credence we would say that homework increases pupil achievement with 7 percent of a standard deviation.[9]

Summarized, when ruling out suspect variation across schools and grades, the findings suggest that homework has a positive effect on average pupil achievement, but the estimates lack precision. However, when stratifying on mother's education the effect strengthens for pupils from the upper part of the socioeconomic status scale, but disappears for pupils from the lowest part of the socioeconomic status scale.

To the extent that measurable characteristics of the teachers explain little of the true variation in teacher quality, and that most of the variation in teacher quality occurs within schools (Rivkin et al., 2005), we must be careful with ruling out correlations between homework and unobserved teacher attributes. The next section is concerned with whether this pattern remains when analyzing the effect of homework on within class differences in test scores.

2.5.1 The effect of homework on within class differences

Table 2.8 presents results from estimating equation (2.4) which compares within class differences in test scores across class rooms within the same school and grade. Each row represents one regression. Out of totally six point estimates, four are significantly different from zero at the five and ten percent level. Thus the general picture is that homework increases the within class differences in the test scores.

In the upper panel of the table, both the variance as well as the distances between the 85th and the 15th and the 75th and the 25th percentiles are significantly larger in homework-classes than in non-homework classes. The point estimates being respectively 0.05 (21 percent of a standard deviation), 0.17 (15 percent of a standard deviation) and 0.09 (20 percent of a standard deviation). The distance between the 95th and the 5th

[9]The argument for separately looking at students from primary educated mothers versus the remaining students is because of the findings in section 2.3 that primary educated mothers help their children less with homework than the others.

Table 2.7 – The effect of homework on student achievement, fixed effect estimates

				Education Mother	
				Primary	> Primary
	(1)	(2)	(3)	(4)	(5)
Homework	0.0450	0.0471	0.0521	-0.0104	0.0723
	(0.0416)	(0.0389)	(0.0385)	(0.0625)	(0.0404)*
Controls					
- Individual	No	Yes	Yes	Yes	Yes
- Class	No	No	Yes	Yes	Yes
N	4,316	4,316	4,316	967	3,349
R-squared	0.0006	0.0951	0.0960	0.0533	0.0601

Note: See Table 2.6

percentile is also positive, but not significant.

Turning to the lowest panel, the results in Table 2.8 confirm the findings in Table 2.7 to a large degree. Giving homework to the whole class has a positive and significant impact on the deviation between the 85th percentile and the median, whereas the distance between the median and the 15th percentile seems to be the same in homework and non-homework classes. Thus pupils from the upper part of the distribution significantly benefit from homework.

A non-parametric way of analyzing the same problem, is to divide the pupils into two groups depending on whether they are in homework classes or not, and plot the density distribution of test score separately for these two groups. This is illustrated in Figure 2.1. Since the relevant approach is a within school within grade comparison, the test scores are standardized by grade, school and year. The lower part of the distribution coincide in homework and non-homework classes indicating that the weakest pupils are unaffected by homework. Concerning the the upper part of the distribution, it is skewed to the right in classes where pupils gets homework, and thereby confirming that it is the better pupils who are the ones who benefit from homework.

2.6 Concluding remarks and policy implications

Using Dutch data on children and their teachers in elementary school this papers shows that within class differences in test scores are larger in classes where everybody gets homework than in classes where nobody gets homework. More precisely we find that

Table 2.8 – The effect of homework on spread in test score, fixed effect estimates

Dependent variable	Effect	s.e.
Variance	0.0485	(0.0271)∗
Percentile ranges		
75th - 25th	0.0860	(0.0411)∗∗
85th - 15th	0.1665	(0.0643)∗∗
95th - 5th	0.1292	(0.1023)
85th - 50th	0.1392	(0.0514)∗∗
50th - 15th	0.0273	(0.0397)
N	258	

Note: The unit of observation is the class. Included are also a constant term and year dummies. The standard errors are robust and corrected for grade level clustering. *, ** and *** denote statistical significance at the 10, 5 and 1 percent level respectively.

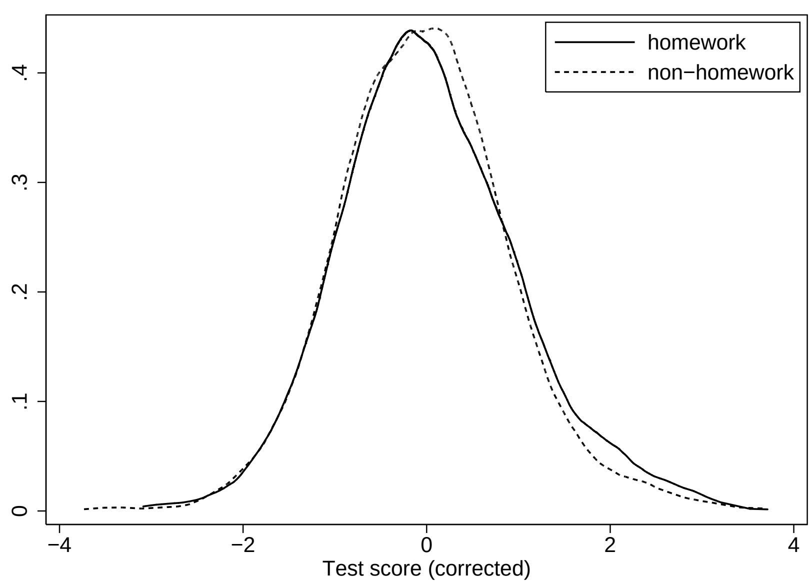

Figure 2.1 – The distribution of test scores in homework classes and non-homework classes

pupils in the upper part of the socioeconomic statues scale gain, whereas pupils in the lowest part of the scale are unaffected. Under the assumption that unobserved teacher attributes only affect average achievement, we argue that we manage to separate the effect of homework from unobserved class and teacher characteristics. Moreover, potential biases arising from unobserved school quality and pupil selection are accounted form by exploiting variation within schools.

These findings are important because they informs us about an early source of inequality, and should be of interest for both policy makers and teachers and school principals. It is well documented that pupils from disadvantage backgrounds fall behind at a very early age (even before they start school), and many education subsidies are motivated as an attempt to reduce these inequalities. It is therefore both essential and necessary to learn about potential sources that generate or increase (already existing) inequalities. If giving homework to children in elementary school contributes to growing inequalities at an early stage in the schooling career, we might want to reconsider current homework practices. One solution is the introduction of homework-help during or after school time for those pupils who do not get sufficient help at home. Another solution is that parents with less education should receive guidance from the school in order to assist their children in productive ways (Balli, Demo and Wedman 1998).

A. Appendix

Table 2.9 – Sample summary statistics, reduced sample

	Mean	s.d.
Individual Characteristics (N = 4,316)		
Girl	0.47	0.50
Boy	0.47	0.50
Age	9.53	1.70
Non-Western migrant background	0.31	0.46
Mother's education		
- Primary	0.22	0.42
- Lower vocational	0.31	0.46
- Upper secondary/intermediate vocational	0.29	0.45
- University/higher vocational (higher education)	0.12	0.33
- Missing	0.06	0.24
Father's education		
- Primary	0.17	0.38
- Lower vocational	0.30	0.46
- Upper secondary/intermediate vocational	0.23	0.42
- University/higher vocational (higher education)	0.16	0.36
- Missing	0.15	0.35
Class Characteristics (N = 258)		
Class size	22.45	5.35
Teacher's experience	16.80	10.80
Female teacher	0.60	0.49

Table 2.10 – Descriptive statistics, different inequality measures on class level

	Mean	s.d.
Variance	0.828	0.229
Percentile Ranges		
75th - 25th	1.039	0.354
85th - 15th	1.650	0.560
95th - 5th	2.696	0.831
85th - 50th	0.871	0.428
50th - 15th	0.778	0.346
N	258	

3 The effect of workplace characteristics on teacher sickness absenteeism

3.1 Introduction

Sickness absenteeism is expensive for several reasons. The society must pay sickness benefits and replacement; the sick individuals themselves loose work experience and acquired human capital; and having sick colleagues may be destructive for the overall moral on the workplace. In the school sector, sickness absenteeism may have an additional cost. Teacher absenteeism may be costly to the extent that the pupils learn less during their absence. This hypothesis is supported by Miller, Murname and Willett (2006) who find evidence that teacher absenteeism has a negative impact on the mathematics achievement of elementary school students in the Ormondale School district in the US. Potential channels this may work through are; i) ordinary teachers are substituted by less qualified replacement teachers; ii) the change of the teacher itself may have a disruptive effect on the student's learning environment; iii) in the worse scenario classes may be canceled when a replacement teacher cannot be found.

A common agreement in the literature is that teacher quality is a crucial input in explaining student outcome, but that observed measures on teacher quality like education and experience turn out to explain very little of the variation in student outcome (see Hanushek, 2002). It is therefore surprising that teacher absenteeism has not been devoted more space. There is however a small empirical literature showing that different sick-leave policies and financial incentives may be effective devices to lower teacher absenteeism. These studies are typically older and based on US data (Winkler, 1980, Jacobsen, 1989 and Ehrenberg et al., 1991).[1]

[1]The literature on the effect of financial incentives on sickness absenteeism in other sectors is more elaborated. See for instance Johansson and Palme (2005) and references therein.

One may argue that financial incentives are implemented in order to reduce absenteeism caused by abuse of sick-leave privileges by healthy employees. It is however difficult to predict whether this type of absenteeism is due to pure shirking only, or whether it also occurs because some teachers value their health over a raise in pecuniary and non-pecuniary rewards. One example in place is: to obtain goodwill from the principal, teachers that are marginally sick may show up on their work. Both Winkler (1980) and Jacobsen (1989) look at short term effects only. In the long run, financial incentives may have weaker effects simply because marginally sick teachers cannot continue to yield effort in order to obtain "goodwill". If that is the case other types of policy devices may have the potential to be equally good or even more successful. It is however important to point out that the optimal sick leave policy may be a combination of financial incentives and money spent on preventive systems. On the other hand, deriving this optimal strategy is difficult since it requires information on shirking behavior versus involuntary absenteeism. Notwithstanding, in order to get closer to an efficient policy device, further knowledge of sources that influence absenteeism are essential.

A couple of studies focus on different factors that can explain variation in teacher absenteeism. The first one is Leuven (2006) who investigates the effects on teacher absenteeism of a Dutch policy that allows employees older than 52 to reduce their working hours by 10 percent at the cost of a 3.5 percent reduction in salary. The aim of this policy is to lower absenteeism among older workers in the Netherlands through subsidized work time reduction. Exploiting longitudinal micro data on all teachers in primary and secondary education in the Netherlands, he finds that male teachers lower their absenteeism in response to this policy. Bradley et al. (2007) analyze so-called social multiplier effects (Manski, 1993). More precisely they investigate whether absenteeism has spillover effects on other teachers absenteeism. Using a database of matched teachers and schools obtained from the Queensland Government of Australia they find evidence that teachers absenteeism depends on the absenteeism of their co-workers.

Another interesting result in both Leuven (2006) and Bradley et al. (2007) is that teachers on temporary contracts have lower absenteeism rates than teachers on permanent contracts (see also Ichino and Riphahn, 2005). This suggest that less certain employment contracts may have a disciplinary effect in the sense that low sick rate improves the chances of renewed contract and permanent position. These findings are in line with the so-called shirking theory, that says that workers will shirk less and work harder when for instance the punishment is hardened (Barmby et al., 1994; Shapiro, and Stiglitz, 1984).

Using longitudinal register data on teacher absenteeism linked to school (workplace) as well as individual teacher characteristics, the current paper's contribution to the literature is to analyze the effect of workplace characteristics on teacher sickness absenteeism. These characteristics are also of interest because they may signal that some teachers are put under higher working pressure. For instance, teachers facing high pupil teacher ratios may have a more tiresome work day than teachers facing lower pupil teacher ratios. Similarly, teachers working in schools with many disadvantage students may be more often confronted with stressful situations. Teachers' working hours and contract type are also considered in the model because the literature has found these two variables to be important determinants for teacher absenteeism (Leuven, 2006; Bradley, 2007).[2]

The literature concerning absenteeism in other sectors have found working conditions to be important predictors for absenteeism (e.g. Hemningway et al., 1997; Brown et al., 1999; Ose, 2004). Yet, no earlier studies have particularly looked at how teacher absenteeism relates to workplace characteristics.

To identify the effects we exploit intertemporal variation within teachers who do not move to another schools. In addition to controlling for variation across teachers, this strategy also rules out that the results are driven by teachers who change schools. The results will be reported jointly and separately for male and female teachers which is a common practice in the sickness absenteeism literature. Moreover, we also stratify on the teacher's age.

The structure of the paper is as follows; section 2 describes the institutional settings; section 3 presents the data; section 4 outlines the empirical approach; the results are presented in section 5; we perform some robustness checks in section 6; and finally section 7 concludes.

3.2 Institutional settings

The Norwegian Public Service Pension Fund ("Statens Pensjonkasse") guarantees teachers 100 percent (monetary) replacement from the first day of incapacity up to one year ("sickness benefit year"). The first 16 days of the absence period is paid by the employer, which for the teachers is the municipality. The Norwegian National Insurance company ("Folketrygden") reimburses the remaining days (maximum 248 working days) up to an

[2]Analyzing social multiplier effects as in Bradley (2007) are more complicated and is therefore not a part of this analysis.

Table 3.1 – Summary statistics: Teacher sickness absenteeism measured in percent per teacher per year

	Mean	s.d.	N obs	N teachers
All teachers	5.99	14.38	258,903	84,001
Female teachers	6.74	15.03	178,859	58,991
Male teachers	4.31	12.65	80,044	25,010
Teachers younger than 50 (young)	5.28	12.92	172,570	60,828
Teachers older than 50 (old)	7.39	16.84	86,333	12,196

Note: The table shows statistics for the whole population covering the school years 2000/2001 to 2004/2005.

upper limit of 6G's.[3] Average teacher salary in Norway is approximately close to this cap. Employees are allowed to use self-certification for the first three days of absence, but need a medical certificate for all absence exceeding three days.

After one year, the insured employee either goes back to work again or is transferred to the rehabilitation scheme. Rehabilitation is the step before disability, and its intention is to avoid inflow into disability, either in forms of physical treatment or vocational occupational rehabilitation. The maximum length of stay on the rehabilitation scheme is one year, and the monetary compensation payed by "Folketrygden" is about 67 percent. Because the probability is low that they will go back to work again, old employees may be transferred directly to the disability scheme at the end of the "sickness benefit year" if they do not go back to work again. This is common for employees beyond 60 years.

3.3 Data

Longitudinal register data on medical certificated sickness absenteeism from Statistics Norway covering all teachers in Norwegian primary and lower secondary school from 2000/01 to 2004/05 (in total five school years) are linked to school (workplace) characteristics from the Norwegian Ministry of Education as well as individual teacher characteristics also from Statistics Norway. Since a medical certificate is necessary only for absence longer than three days, we do not have information on absenteeism shorter than

[3]In the period under investigation 1G increased from 48 377 NOK in 2000 to 58 139 NOK in 2004. 1 Euro $\simeq$8 NOK.

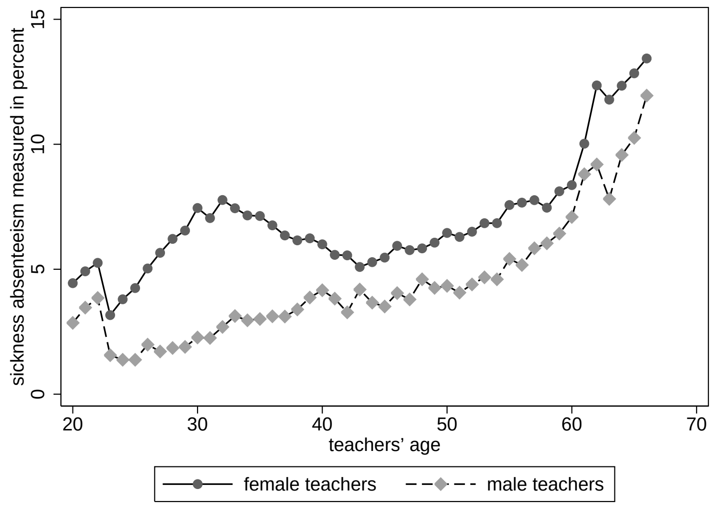

Figure 3.1 – The relationship between teacher's age and absenteeism

four days.

A few teachers are continuously sick for more than one year. They constitute a large part of of the sickness absenteeism and may not be representative for the average teacher. We therefore exclude them from the sample. Moreover, the national retirement age in Norway is 67, thus teachers older than 67 years are dropped. In order to avoid mixing sickness absenteeism with being on leave we also drop all teachers on leave.[4] Because of the rural settlement pattern in Norway, some very small schools exist. These schools are neither representative and are therefore excluded in the analysis (more precisely we exclude schools with less than 10 pupils).

3.3.1 Sickness absenteeism

For each teacher we know how many "day's work" per school year she is supposed to work according to the work-contract and how many of these "day's work" she was absent. The yearly average amount of medical certificated sickness absenteeism for the time period under investigation is 6 percent which can be seen in the first row of Table 3.1.[5] A general finding in the sickness absenteeism literature is that females are more

[4] All female employees in Norway are entitled generous maternity leave. 3.6 percent of the female teachers in our data that are not registered as being on leave have children younger than or equal to 1 year. In order not to mix sickness absenteeism with being on maternity leave, these teachers are dropped.

[5] Average number of "day's work" and "day's work lost due to medical certificated absenteeism" are respectively 195 and 12 per year. Since we do not have data on short term absenteeism, the total absenteeism exceeds 6 percent.

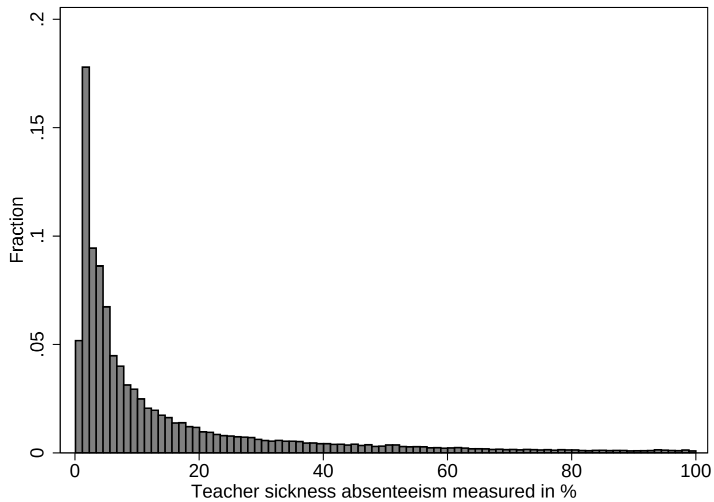

Figure 3.2 – The distribution of average sickness absenteeism per teacher per year conditional on that the teacher was absent (N = 105,322).

sick than males, and that absenteeism is increasing in age (e.g. Barmby et al., 2002). Norwegian teachers are no exception. In Norway, average sickness absenteeism among female teachers is about 2.5 percentage points higher than average sickness absenteeism among male teachers. And teachers born before 1950 (50 years old in 2000) are on average 2.1 percentage points more absent than teachers born after 1950. Figure 3.1 illustrates the relationship between age and sickness absenteeism separately for both genders. For male teachers the relationship seems to be fairly linear until the age of 55, whereas female teachers between 30 and 35 years are more absent than female teachers between 40 and 45 years. A likely explanation for the latter pattern is the age of their own children. Sickness absenteeism for both genders seem to increases rapidly from about 55 years and onwards. Figure 3.2 shows the distribution of teacher sickness absenteeism conditional on that the teachers have been absent. This figure shows that a majority of these teachers lose less than 20 percent "day's work" due to medical reported sickness absenteeism. Moreover, average absenteeism among teachers in this group is about 15 percent.

3.3.2 Workplace characteristics

As already mentioned, our key indicators for workplace pressure are the schools resource use and student composition. Two measures on resource use are available; teachers hours and teacher man-year per pupil (the number of working hours per week multiplied with the number of workweeks per year*)*. Teacher hours measures how much resources are allocated to teachers' interaction with students, either in the class-room or as extra education to specific students. It is not perfectly correlated with teacher man-year because the work load varies between subjects and depends on the amount of non-teaching tasks. Teachers hours is regarded as a very precise measure of the teacher resource use in schools. The relationship between teacher hours per pupil and teacher man-year per pupil is strong. The correlation coefficient is equal to *0.85* and is also illustrated in figure 1 in Falch et al. (2006). Both resource measures are also strongly related to school size as further shown in Falch et al. (2006), thus it is important to separately control for school size in the model. An alternative measure on the school's resource use would be class size. However, the maximum class rule terminated the school year 2002/03. As a consequence data on the number of classes per grade stopped being collected.

To measure the student composition in the school we include students with special needs and minority students. These two variables are derived from calculating the fraction of students within a school who are entitled to extra lessons or instructions with education personnel and the fraction of students who are given extra lessons in Norwegian. It is important to emphasize that conditional on schools' resource use (which also capture extra resources used on special education to both Norwegian students with special needs and minority students), minority and disadvantage students are not confounded with extra money allocated to the school to compensate a unfavorable student composition.

We argued above that a unfavorable student composition and a high pupil teacher ratio are likely to be associated with a more stressful and tiresome work environment. Accordingly we expect sickness absenteeism to raise if the percent of minority and special need students increases and to decline of the school's resource use increases.

Teachers in Norway usually start their teaching career on a temporary contract. A permanent contract is obtained in subsequent year(s) depending on availability of vacant positions and also teaching performance. The latter contract type involves that the teacher is granted a job in the school district. And in accordance with previous findings (Bradley et al., 2007; Leuven, 2006; Ichino and Riphahn, 2005) we expect absenteeism to raise if a permanent work contract is given to the teacher. Being on temporary job contract is assumed to have a disciplinary effect because it is an element of uncertainty connected to whether the contract is extended the next school year.

Concerning the teacher's individual workload, the number of weekly work hours are regulated by the work contract. However no strict guidance is imposed, involving that before the start of each school year the teachers can negotiate their own workload with the principal to a certain degree. In line with the finding in Leuven (2006), decreased number of work hours is predicted to lower absenteeism. A reduced work week involves more leisure time which can be used on other recreational activities that are good for the general well being.

3.3.3 Individual control variables

In addition to gender and age, it is a common practice in the absenteeism literature to control for education and income (e.g. Røed and Fevang, 2006; Askildsen et al., 2005; Leuven, 2006; Ose, 2004; Winkler, 1980).[6] Askildsen et al. (2005) also control for marital status and own children. They find that absenteeism is higher for separated and divorced employees, whereas no strong significant effects relating to the number of small children are found. The latter variable is meant to capture that small children cannot stay home alone when they are sick. The fact that it is insignificant may be due to that their absenteeism data lack information on spells shorter than 14 days. Nevertheless, we acknowledge that the family situation may be an important determinant for absenteeism and control for both marital status, the age of own children (indicator variable that equals one if the teacher has own children younger than 12 years) as well as number of own children. Finally, to take into account that the relationship between age and sickness absenteeism is non linear we choose to control for a quartic age function which is assumed to capture all smooth variation.

3.3.4 Descriptive statistics

Table 3.2 provides descriptive statistics for all the explanatory variables used in this paper covering the whole period under investigation. The average Norwegian teacher is 44 years old, has a workload of 90 percent and earns about 26 000 NOK per month. She teaches in a school consisting of 284 pupils where 6 and 5 percent of the students are minority or special need students. Average teacher hours per pupil and average teacher man-year per pupil is 75 and 8.8. Furthermore, 70 percent of the teachers are females, 65 percent have a permanent work contract, less than 5 percent do not have sufficient

[6]Teachers salary is to a large extend centrally decided and equal across teachers who share the same education and experience. Moreover, since it is a function of experience and education in a nonlinear and interacting way its effect is difficult to interpret. From the school year 2001/02 some local discretion in the teacher's wage bargaining were also introduced.

Table 3.2 – Summary statistics: Explanatory variables

	Mean	s.d.
WORKPLACE CHARACTERISTICS		
Minority students (measured in percent)	5.91	10.17
Students with special needs (measured in percent)	4.57	3.85
Man-year per pupil	8.78	2.60
Teacher hours per pupil	75.04	20.15
School size (number of pupils)	283.73	142.64
Primary school	0.55	0.51
Lower secondary school	0.24	0.43
Combined school	0.21	0.41
Workload (measured in percent)	90.14	18.79
Contract type		
.-Permanent position	0.65	0.48
- Temporary position	0.15	0.36
- Missing	0.20	0.40
INDIVIDUAL CONTROL VARIABLES		
Female	0.70	0.46
Male	0.30	0.46
Age (years)	44.80	11.04
Salary (NOK)/1000	25.96	3.61
Education		
- Unqualified (<= 12 years)	0.04	0.19
- Bachelor degree (12+3/4 years)	0.92	0.29
- Master or PhD degree (12+5/6 -9/10 years)	0.04	0.20
Marital status		
- Unmarried	0.24	0.43
- Married	0.64	0.48
- Widow/widower	0.01	0.12
- Divorced/separated	0.10	0.30
- Missing	0.01	0.04
Children		
- Number of children children	1.86	1.19
- Child(ren) <= 12 year(s)	0.32	0.47

Note. Number of observations: 258,903. All the variables are reported Oct. 1st every school year by the principal or other school leaders.

education to be certified, 65 percent are married and have on average 1.9 children and 32 percent have own children younger than 12 years. 55 percent of the Norwegian teachers work in pure primary schools (1-7 grade), 24 percent work in pure lower secondary schools (8-10 grade), whereas the remaining 21 percent work in combined schools (1-10 grade).

3.4 Empirical approach

Regarding sickness absenteeism, many aspects are of interest. The current paper focuses on one of them; the amount of absenteeism and how it is affected by observed characteristics. This section is concerned with how to empirically estimate this.

Assume that the amount of sickness absenteeism of teacher i in school s and year t (a_{ist}) is generated by the following equation:

$$a_{ist} \quad = \alpha + x'_{ist}\gamma + \eta_i + \theta_s + \delta_t + u_{ist} \tag{3.1}$$

where x'_{ist} is a vector consisting of all observable attributes of the teachers and schools presented in table 3.2. Clearly, a large part of the sickness variation is explained by the teacher's general health condition which is unobservable. Under the assumption that the teacher's health state is constant over the period under investigation, it is captured by the teacher fixed effect, η_i. A school fixed effect θ_s is included to control for all other unobserved attributes with the schools that potentially is correlated with the sickness absenteeism, δ_t are time indicator variables and u_{ist} is a random error term.[7]

In principle, all types of fixed effects can be handled by including indicator variables in the model. This is however a high dimensionality problem given the size of our sample. One way to handle this is to exploit that with fixed individual effects included, school specific effects are solely identified by teachers moving between different schools during the sample period. This implies that we escape possible biases caused by teacher movements between schools if we treat a teacher who moves from one school to another as two different individuals. Technically this involves that each teacher-school combination is given a unique identifier. In the empirical implementation we then only use intertemporal variation within teacher-school matches to identify the effects of observables on absenteeism. This is equivalent to including a full set of individual and school effects and the interaction between these effects (Falch and Strøm, 2005; see also Goux and

[7] The model in equation (3.1) is on reduced form since we do not know the accurate reason why the teachers are absent. Any underlying structural model is therefore not formalized.

Maurin, 1999). In addition to controlling for variation across teachers, this strategy also rules out that the results are driven by teachers who change schools. The exact model we estimate is give by equation (3.2).

$$a_{ist} \quad = \alpha + x'_{ist}\gamma + \eta_i \cdot \theta_s + \delta_t + u_{ist} \tag{3.2}$$

3.5 Results

This section presents results from estimating different variations of equation (3.2). Because of the high correlation between the two variables that measure the school's resource use, separate regressions are run for each of them, and only those which include teacher hours per student are reported in full tables. The teacher man-year coefficients are reported in Table 3.5.[8] In all the tables we only report the estimation results for the workplace characteristics, while the effects of the individual control variables are reported in appendix Table 3.7 and 3.8.

To shed some light on how the amount of sickness absenteeism is related to observed characteristics, we start out with presenting estimates based on a simple OLS regression. The results are reported in column (1) of Table 3.3. Although it is difficult to give estimation results obtained from a cross sectional analysis causal interpretations, it is of interest to compare these findings with the results obtained from the fixed effect analysis (FE) where intertemporal variation within teachers who have not changed schools are used to identify the effects. The results obtained from the latter approach are reported in column (2) of the same table.[9]

The OLS analysis produces a zero correlation between the school's resource use and teacher absenteeism, suggesting that the school's resource use have no impact on absenteeism. It is difficult to explain why this occurs, but it is important to point out that schools in Norway differ substantially in terms of size, location, student body and also teaching staff. Larger schools in big urban cities are typically faced with relatively few resources per pupil, whereas small schools situated in remote areas have relatively more resources per pupil.[10]

[8] In a "horse-race" competition between teacher man-year per student and teacher hours per student, the latter is significant at the five percent level (point estimate = 0.112), whereas the first one is clearly insignificant.

[9] Since many teachers have zero absenteeism, the arguably most correct model would be a Tobit specification. However, teachers who are non-absent over the whole sample period will drop out from the fixed effect approach.

[10] Winkler (1980) also find zero effect of the pupil-teacher ratio on absenteeism

Table 3.3 – The effect of workplace characteristics on teacher sickness absenteeism.

	OLS (1)	FE (2)
Workplace characteristics		
Teacher hours per pupil	−0.0021	−0.0139
	(0.0024)	(0.0056)**
Minority students	0.0148	−0.0184
	(0.0039)***	(0.0126)
Special need students	0.0280	0.0247
	(0.0108)***	(0.0178)
Pupil	0.0027	0.0097
	(0.0010)***	(0.0058)*
Pupil^2/1000	−0.0032	−0.0155
	(0.0014)**	(0.0067)**
School type (ref = Primary)		
- Lower secondary	−0.4241	
	(0.0906)***	
- Combined	−0.0510	
	(0.0943)	
Workload	−0.0198	0.2102
	(0.0022)***	(0.0058)***
Permanent position	0.9187	0.9458
	(0.0921)***	(0.1781)***
N observations	258,903	176,499
N teachers	84,001	44,988

Note: The dependent variable is teacher *i's* sickness absenteeism in year t measured in percent. Standard errors are heteroscedasticity robust and corrected for individual level clustering. Year dummies; dummy variables for missing information on the teacher's contract type, length of education and marital status are included. *. ** and *** denote statistical significance at the 10, 5 and 1 percent level respectively

However, when condition on fixed effects we find evidence that increased use of resources has a significant negative effect on sickness absenteeism. More precisely, when ruling out variation across teachers and schools we find that sickness absenteeism declines with 2 percent of a standard deviation if teacher hours per student increases with one standard deviation. In terms of percentage points this effect amounts to 0.3, and is not negligible given that average sickness absenteeism is 6 percent. The effect of teacher man-year per pupil is similar and reported in column 1 of Table 3.5.

The relationships between sickness absenteeism and both variables that measure student composition are positive and highly significant in the OLS analysis. The point estimates suggest that a one standard deviation increase in minority and special need students will increase the amount of sickness absenteeism with 1 and 0.8 percent of a standard deviation respectively. Although the effects are rather small, they indicate that teachers working in schools that face a large share of minority and special needs students have more sick-leave. The literature looking at teacher quit behavior typically find that teacher mobility is strongly related to characteristics of the students (e.g. Hanushek et al. 2004). The true effect of student composition on sickness absenteeism is then overestimated if teachers with good health sort themselves to schools with few minority and special needs students. With respect to minority students this seems to be confirmed in column (2). It is however somewhat puzzling that the effect is negative and almost significant at the ten percent level. The point estimate of the special need students is on the other hand fairly unchanged in the fixed effect specification, but note that the standard error has increased (t-value equals 1.4 in column 2).

Furthermore, absenteeism is highest in schools with 313 pupils which is slightly larger than average school size. Teachers working in pure lower secondary schools are less absent than teachers working in schools with students at the primary level only.

The predicted positive effect on absenteeism of both workload and being on a permanent contract may be underestimated in an OLS analysis if teachers with good health more often ask for a higher workload and are more easily given a permanent position. Regarding workload this seem to be the case. The correlation between workload and sickness absenteeism is negative and significant in the OLS analysis, but obtains the expected significant positive sign in the fixed effect analysis. The latter implies that a ten percentage point increase in workload will increase sickness absenteeism by 2 percentage points. This corresponds to 14 percent of a standard deviation. The point estimate of having a permanent position is similar in the OLS and fixed effect analysis. Teachers on a permanent job contract are almost 1 percentage point (about 7 percent of a standard deviation) more absent per year than teachers who have a temporary position which is consistent with earlier evidence and the shirking theory. It is however important to emphasize that being on a temporary contract may not be a “good” incentive if marginally sick teachers force themselves to work to hard in fear of not getting the contract renewed the next school year.

Although the fixed effect approach handles a large part of the potential selection bias, a small route for bias in the fixed effect estimates may remain. Biases arise if workload and contract type are correlated with unobserved changes in the teacher’s health condition over time. However, since the most likely scenario is a positive correlation (higher workload and permanent positions are given to teachers with improved health), the potential unobserved time varying health effect will indeed bias the fixed effect estimates downwards, *ceteris paribus.*

The effects of the remaining individual control variables that correspond to Table 3.3 are reported in appendix Table 3.7. As expected divorced teachers and teachers with small children are more absent. Note also that married teachers seem to be more absent than unmarried teachers. It is difficult to interpret why the latter result is obtained, but it may be due to that being married is correlated with other conditions that we do not control for in this model. Moreover, increased salary seems to be negatively correlated with absenteeism (both in the OLS and FE specification), but as already emphasized this coefficient may also capture omitted variables (see footnote 6).[11]

3.5.1 Gender differences

Since female employees turn out to have a higher sickness absenteeism than male employees, a common practice in the absenteeism literature is to separately look at at the genders (e.g. Johansson and Palme, 2005; Leuven, 2006; Askildsen et al., 2005). We follow this tradition and estimate equation 3.2 separately for female and male teachers.

Results obtained from fixed effect estimations are presented in the first two columns of Table 3.4. An increase in teacher hours per pupil has strongest effect on male teachers who reduce their absenteeism with 4.5 percent of a standard deviation if teacher hours per pupil increases with one standard deviation. This may indicate that only male teachers react on the school’s resource use. However, female teachers are more sensitive to an increase in teacher man-year per pupil as is shown in column 2 of Table 3.5. If teacher man-year per pupil increases with one standard deviation, female teachers lower

[11] With respect to the fixed effect specification(s) in Appendix A, it is important to point out that the results are driven by variation within very few teachers.

their absenteeism with 1.2 percent of a standard deviation. A general conclusion about resource use and gender differences is then difficult to derive. Rather we are inclined to say that both female and male teachers respond to the school's resource use. The point estimate of minority students is similar to the one in the pooled sample, whereas only female teachers appear to increase their absenteeism if the fraction of special need students increases. Regarding changes in workload and the contract type, the impact on both genders resembles the findings in Table 3.3 to a large extent.

Columns (1) and (2) in appendix Table 3.8 report the effects of the remaining individual control variables that correspond to column (1) and (2) of Table 3.4. Only female teachers seem to respond to marital status and small children. The fact that only married women have higher sick-leave is in accordance with Barmby et al. (2002). Salary has still a negative effect on absenteeism and the size of the coefficients are similar to the ones in appendix Table 3.7.

3.5.2 Age differences

Another dimension for which differences in sickness absenteeism are large is age. Although sickness absenteeism is increasing in age, we still do not know the exact reason why this is observed. One explanation is that health is negatively affected by external factors such as work pressure, allowing health problems to arise over time. On the other hand, we cannot rule out that health itself simply is a declining function of time, implying that poorer health is inevitable when we become older. Separating these two effects are difficult. However if only the latter explanation is relevant, we expect no systematic differences to exist between young and old teachers with respect to the effect of observed characteristics. As an attempt to check this we stratify on the teachers age and estimate equation 3.2 separately for old (born latest 1950) and young teachers (born after 1950).

The fixed effects estimates are presented in columns (3) and (4) of Table 3.4. The point estimates of teacher hours per student is larger for old teachers than for young teachers. However, if we interpret the effects in terms of percent of a standard deviation, both effects are close to 2 if teacher hours per pupil increases with one standard deviation.[12] Furthermore, both young and older teachers increase their absenteeism with about 7 percent of a standard deviation if they are given a fixed work contract. Concerning the student composition, the fraction of special need students has still the expected positive sign for both age-groups, but as emphasized earlier, this effect is very small. The effect of minority students is unchanged compared to the findings in previous tables. Thus, with

[12]Once again, the effects of teacher man-year per pupil is fairly similar.

Table 3.4 – The effect of workplace characteristics on teacher sickness absenteeism, fixed effect estimates for different sub groups

	Female (1)	Male (2)	Young (3)	Old (4)
WORKPLACE CHARACTERISTICS				
Teacher hours per pupil	−0.0087	−0.0277	−0.0104	−0.0171
	(0.0066)	(0.0103)***	(0.0061)*	(0.0109)
Minority students	−0.0240	−0.0062	−0.0101	−0.0381
	(0.0157)	(0.0196)	(0.0140)	(0.0254)
Special need students	0.0339	0.0028	0.0137	0.0431
	(0.0218)	(0.0308)	(0.0199)	(0.0346)
Pupil	0.0103	0.0079	0.0060	0.0166
	(0.0070)	(0.0103)	(0.0066)	(0.0108)
$Pupil^2/1000$	−0.0149	−0.0168	−0.0107	−0.0252
	(0.0082)*	(0.0118)	(0.0078)	(0.0125)**
Workload	0.2153	0.1845	0.1499	0.3210
	(0.0065)***	(0.0129)***	(0.0064)***	(0.0112)***
Permanent position	1.0197	0.7257	0.9677	1.1820
	(0.2161)***	(0.2950)**	(0.1819)***	(0.6459)*
N observations	127,983	48,516	112,672	63,827
N teachers	33,075	11,913	29,855	15,133

Note: See Table 3.3.

Table 3.5 – The effect of teacher man-year per pupil on teacher sickness absenteeism, fixed effect estimates for all teachers and different sub groups of teachers

		Sub groups			
	All	Female	Male	Young	Old
	(1)	(2)	(3)	(4)	(5)
Teacher man-year per pupil	−0.0536	−0.0731	0.0161	−0.0322	−0.0954
	(0.0271)**	(0.0303)**	(0.0620)	(0.0283)	(0.0626)

Note: The dependent variable is teacher i's sickness absenteeism in year t measured in percent. Standard errors are heteroscedasticity robust and corrected for individual level clustering. Each row represent one regression. Included control variables are the same as in corresponding specifications in table 3.3 to 3.4.

respect to resource use, student composition and contract type the effects are similar to the ones in Table 3.3.[13] Increased workload turns now out to have larger impact on old teachers. The effect amounts to 19 percent of a standard deviation if workload increases with 10 percentage points and is highly significant. This is a 35 percent increase compared to the effect in the pooled sample in column (2) of Table 3.3. The effect for young teachers is 11.6 percent of a standard deviation and is also significant.

Summarized, older teachers appear to be more sensitive to changes in workload than young teachers. This finding therefore supports that subsidized work time reduction for older workers may be a good policy to lower sickness absenteeism among older workers (Leuven, 2006).

Concerning the effects of the remaining individual control variables that correspond to the two last columns of Table 3.4, no big changes compared to the overall findings are revealed. The only exception is that the dummy variable for small children only affects young teachers' absenteeism (which is not surprising since having small children is less frequent among older teachers). This is reported in the two last columns of appendix Table 3.8.

3.6 Robustness checks

A relevant question that arises when studying sickness absenteeism is to what extent the effects are driven by the composition of the labor force (Askildsen et al., 2005; Araj and Thoursie, 2005). In periods with excess demand for labor, marginal workers will be employed to a larger degree, and these workers are expected to have higher sickness rates on average due to poorer health. Similarly, the same type of workers may also be the first ones to leave when the labor market gets slacker again.

With respect to teachers, the teaching staff is a function of the school size and also the student composition. The demand for teachers automatically increases if school size increases. Schools are also in need for extra teachers if the student composition worsen. We can neither rule out that teachers leave the school sector for a job in other sectors, especially in periods when the labor market is tight. In order to check whether our results are driven by variation in the composition of the teaching staff from year to year, we restrict the sample to teachers who are present in the whole sample period (stable worker sample).[14] The results are presented in Table 3.6. The effects of all variables are

[13] With respect to old teachers, it is important to point out that most of them have a fixed position.

[14] Average sickness absenteeism for the stable worker sample is 4.76. This is 1.23 percentage points lower compared to the whole sample. The standard deviation is 12.03.

Table 3.6 – The effect of workplace characteristics on teacher sickness absenteeism, fixed effect estimates for the stable worker sample

		Sub groups			
	All	Female	Male	Young	Old
	(1)	(2)	(3)	(4)	(5)
WORKPLACE CHARACTERISTICS					
Teacher hours per pupil	−0.0175	−0.0132	−0.0283	−0.0135	−0.0211
	(0.0062)***	(0.0074)*	(0.0114)**	(0.0073)*	(0.0109)*
Minority students	−0.0132	−0.0182	−0.0007	−0.0186	−0.0045
	(0.0130)	(0.0162)	(0.0206)	(0.0145)	(0.0246)
Special need students	0.0176	0.0281	−0.0016	0.0206	0.0088
	(0.0193)	(0.0240)	(0.0322)	(0.0229)	(0.0339)
Pupil	0.0120	0.0149	0.0069	0.0116	0.0113
	(0.0063)*	(0.0077)*	(0.0107)	(0.0074)	(0.0109)
$\text{Pupil}^2/1000$	−0.0191	−0.0204	−0.0175	−0.0211	−0.0150
Workload	0.2099	0.2197	0.1784	0.1421	0.3045
	(0.0084)***	(0.0094)***	(0.0183)***	(0.0101)***	(0.0139)***
Permanent position	0.4442	0.5428	0.2621	0.5153	0.7545
	(0.2367)*	(0.2986)*	(0.3744)	(0.2424)**	(0.7543)
N	106,644	74,084	32,560	63,010	43,634
N teachers	22,234	15,411	6,823	13,364	8,870

Note: see Table 3.3. The effects of the remaining individual control variables are not reported in any tables.

basically unchanged compared to the whole sample, but note that the effect of teacher hours becomes clearer. The only exception is being on a permanent work contract which has now a smaller effect, although it is still positive and mainly significant at the 10 percent level. This suggest that teachers opting in and out of the education sector account for a large part of the positive effect of this variable derived in earlier tables.

3.7 Discussion and conclusion

In order to fight sickness absenteeism it is necessary to accumulate knowledge about possible sources that explain variation in sickness absenteeism.

Using longitudinal register data on teachers sickness absenteeism linked to school and individual characteristics, this paper studies the effect of workplace characteristics on the amount of medical reported sickness absenteeism among teachers in public primary and lower secondary school in Norway. Our findings suggest that teachers lower their amount of sickness absenteeism when the school's resource use increases. We also find evidence that increased workload and employment protection are associated with higher absenteeism. Moreover, old teachers appear to be more sensitive to changes in workload than young teachers. All results are derived from specifications where variation within teachers who have not changed schools is used to identify the effects.

To what extent we should be concerned about teacher absenteeism depends on its influence on students outcome. This is a question which so far has been (almost) neglected in the literature and more research on his field is therefore highly demanded.

A. Appendix

Table 3.7 – Table 3.3 cont'd

	OLS (1)	FE (2)
INDIVIDUAL CONTROL VARIABLES		
Male teacher	−2.3459	
	(0.0734)∗∗∗	
Age	6.4526	
	(0.7519)∗∗∗	
Age^2	−0.2098	0.4674
	(0.0277)∗∗∗	(0.0838)∗∗∗
Age^3	0.0029	−0.0091
	(0.0004)∗∗∗	(0.0013)∗∗∗
$Age^4/1000$	−0.0136	0.0622
	(0.0026)∗∗∗	(0.0075)∗∗∗
Number of children	−0.3394	−2.7057
	(0.0450)∗∗∗	(0.2326)∗∗∗
Child(ren) $<=$ 12 year(s)	0.7036	0.3210
	(0.0954)∗∗∗	(0.2048)
Salary (NOK)/1000	−0.1433	−0.1697
	(0.0155)∗∗∗	(0.0556)∗∗∗
Education (ref = Unqualified)		
- Bachelor	0.1840	−0.2720
	(0.1922)	(0.8150)
- Master or PhD	−0.0417	−2.8499
	(0.2563)	(1.3715)∗∗
Marital status (ref = Unmarried)		
- Married	0.4067	2.4455
	(0.1061)∗∗∗	(0.3669)∗∗∗
- Widow/widower	0.3032	−10.2182
	(0.3577)	(1.2490)∗∗∗
- Divorced/separated	2.6415	1.2082
	(0.1604)∗∗∗	(0.5322)∗∗

Note: See Table 3.3.

Table 3.8 – Table 3.4 cont'd

	Female (1)	Male (2)	Young (3)	Old (4)
INDIVIDUAL CONTROL VARIABLES				
Age^2	0.3669	0.8791	−0.0083	−19.0954
	(0.1003)∗∗∗	(0.1540)∗∗∗	(0.1829)	(6.0842)∗∗∗
Age^3	−0.0079	−0.0144	−0.0014	0.2103
	(0.0016)∗∗∗	(0.0024)∗∗∗	(0.0032)	(0.0709)∗∗∗
$Age^4/1000$	0.0568	0.0874	0.0163	−0.8558
	(0.0091)∗∗∗	(0.0137)∗∗∗	(0.0208)	(0.3092)∗∗∗
Number of children	−4.1323	0.0600	−2.9359	−3.7072
	(0.3385)∗∗∗	(0.2564)	(0.2357)∗∗∗	(1.5431)∗∗
Child(ren) $<=$ 12 year(s)	0.6982	−0.3896	0.5080	0.3269
	(0.2608)∗∗∗	(0.3160)	(0.2186)∗∗	(0.6277)
Salary (NOK)/1000	−0.1609	−0.1969	−0.1847	−0.1894
	(0.0692)∗∗	(0.0929)∗∗	(0.0592)∗∗∗	(0.1173)
Education (ref = Unqualified)				
- Bachelor	−0.6071	0.4665	0.0719	0.2517
	(0.9485)	(1.4571)	(0.8327)	(1.6077)
- Master or PhD	−4.3810	0.3639	−2.4591	−2.1009
	(1.6847)∗∗∗	(2.2811)	(1.4247)∗	(3.3410)
Marital status (ref = Unmarried)				
- Married	3.4397	0.4233	2.3442	5.5363
	(0.4848)∗∗∗	(0.4864)	(0.3692)∗∗∗	(2.5855)∗∗
- Widow/widower	−9.8520	−8.7715	−11.5517	−6.7296
	(1.4229)∗∗∗	(2.5753)∗∗∗	(2.6127)∗∗∗	(2.9027)∗∗
- Divorced/separated	1.5454	0.9771	1.1929	4.0576
	(0.6612)∗∗	(0.8887)	(0.5527)∗∗	(2.7598)

Note: See Table 3.3.

4 The incentive effects of property taxation: Evidence from Norwegian school districts

4.1 Introduction

Brennan and Buchanan (1980) emphasize that if governments are not entirely benevolent, it is in the interest of voters to design tax systems to control a 'Leviathan' government. Glaeser (1996) applies the ideas of Brennan and Buchanan to local property taxation. He shows that when local government bureaucrats act as revenue-maximizing agents, a tax base sensitive to changes in the quality of local goods and services produced is favorably for the voters. When housing demand is inelastic, property taxation reduces waste compared to lump sum taxation because local bureaucrats take into account the income feedback via increased property values. High quality of local public goods and services is associated with increased housing prices, yielding additional property tax revenue, which in turn benefits the local public sector bureaucrats. Wilson and Gordon (2003) develop the argument of Glaeser (1996) in a richer model that accentuates competition between jurisdictions. In their model, interjurisdictional competition reduces waste, raises the utility of the residents, and increases the residents' desired quality of public goods. The argument is analogous to the property tax mechanism, as public officials benefit from "taking a smaller slice out of a larger pie" through the Tiebout process (Wilson and Gordon, 2003, p. 401). Hoxby (1999) also provides a framework to analyze local costs and effort in which property taxation works as a discipline device. In line with Glaeser she also argues that property taxation links school quality to school finance and helps to control costs and improve effort in schools.

Empirically, the incentive effect of property taxation is hard to identify since most countries have property taxation for all local governments. The Norwegian case, however, is well suited for empirical analyses because one can compare local governments with and without property taxation. Moreover, since many local governments have property taxation for exogenous reasons, the data allows taking into account potential endogeneity problems. A credible test of the hypotheses put forward by Glaeser (1996) and Hoxby (1999) can be conducted.

Borge and Rattsø (2006) present a first investigation of the incentive effects of property taxation relying on data from Norwegian local governments. Using regression analysis and matching on propensity scores, they find that property taxation is negatively associated with the unit costs in the production of sewage services. The current paper takes a different approach, and focus on public sector quality rather than costs. Theoretically it is possible to look at the quality on all services provided by the local government. Empirically we are restricted to focus on the sectors where data is available. We have access to administrative test score data for two entire cohorts of students from their last year in compulsory schooling. A natural choice is therefore to concentrate on the school district's quality of public schooling. This is also a highly conspicuous sector since the largest part of the local budget (about 30 percent on average) is allocated to this sector. As each school district corresponds to one local government, we use the two terms interchangeably.

A large related empirical literature has focused on the relationship between school quality and housing prices. The key challenge faced by these studies is to distinguish the impact of school quality from the impact of other neighbourhood characteristics. Most studies, typically based on US data, find that school quality is capitalized into housing prices, e.g. Black (1999) and Kane et al. (2006), both applying a boundary fixed effects approach and Figlio and Lucas (2004) evaluating whether the publication of 'school report cards' contributed to changes in housing values. Others studies find weaker effects, in particular with respect to value added measures, e.g. Brasington and Haurin (2006) applying spatial econometrics as an identification strategy, and Clapp et al. (2007) using panel data. The only study applying Norwegian data is Machin and Salvanes (2007). Exploiting a policy experiment they seek to identify the relationship between housing prices and school performance on Norwegian high school data. Their findings indicate that parents are willing to pay more to live in catchment areas of better performing schools.

The empirical strategy of this paper follows the spirit of Hanushek et al. (1996). First, we regress district-level fixed effects on student achievement, controlling for a rich set of individual student demographics. Second, we use comprehensive district level characteristics, including the use of property taxation, to explain the variation in the district fixed effects. The key methodological challenge is that districts can decide whether or not to levy property taxation. It follows that differences in achievement cannot be causally attributed to differences in the presence of property taxation. To handle this endogeneity problem we rely on instrumental variable techniques. The instrumental variables are derived from characteristics of the tax laws of 1911 and the property tax law of 1975.

The results are consistent with the theoretical framework put forward by Glaeser (1996) and Hoxby (1999). Conditional on resource use, property taxation improves school quality measured as students' result on the national examination. With low local flexibility in teacher salaries and work hours, we believe the established beneficial effect of property taxation to be related to organizational tasks within the school district executed by school leaders and bureaucrats. Even though school districts face central regulations on several aspects, such as maximum class size, they still have considerable discretion on how to organize the school sector. However, it is important to point out that in this paper we do not aim to establish a structural model where we explicitly derive the channels the incentive mechanisms of property taxation work through. Our contribution is to estimate a reduced form model where we investigate whether such beneficial effects exist.

The structure of the paper is as follows; Section 4.2 presents the institutional setting; Section 4.3 lines out the empirical strategy; Section 4.4 presents the results; Section 4.5 discusses the robustness of the main result and section 4.6 concludes the paper.

4.2 Institutional setting

4.2.1 Schooling

Compulsory schooling in Norway consists of primary education (1.-7. grade) and lower secondary education (8.–10. grade). Public schools operated at the local government level account for 98 percent of the students. Private schools are either religious schools or schools that use alternative education methods and are not a part of this analysis. All students are allocated to public schools by catchment areas within local governments. There is no school choice for given residence. In addition to compulsory schooling, local governments in Norway are also responsible for preschool education, care for the elderly and some other services, such as infrastructure.

Our measure of school quality is based on a national written external examination that all Norwegian students have to undertake at the end of tenth grade (the last year in lower secondary school). Although the curriculum includes many different subjects, both elective and compulsory, written examination is undertaken only in mathematics, English, and Norwegian.[1] Each student is examined in only one of the three subjects, which is decided centrally shortly before the exam. In addition, students are graded by their teachers. Both the exam results and the teacher grades matter for entrance to upper secondary school. Since exam grades are externally determined, this measure provides the most accurate measure of student achievement. Teacher grading may be severely biased due to grade inflation and relative grading within schools and districts. The student achievement data used in the present analysis are exam results for 118,178 students who finished lower secondary school in the school years 2001/2002 and 2002/2003.[2] The grading goes from one to six, where six is the top score and one is fail. The exam grades are fairly normally distributed, with a mean of 3.44 and a standard deviation of 1.07.

4.2.2 Local government financing

Unlike local governments in the US, Norwegian local governments are largely financed through block grants and regulated income tax sharing. Close to 90 percent of the local governments' revenues are generated from central government grants and regulated income taxes. The grants are distributed as block grants and are based on objective criteria, partly as income tax equalization and partly as spending equalization. The local governments set the income tax rate within a band determined by the central government. Since all local governments apply the maximum income tax rate (and have done so since 1977), both grants and income tax revenue consequently appear as given from above. The remaining 10 percent of the local governments' revenue comes from user charges and property taxes which are the main revenue decisions made by the local governments.[3]

Two types of property taxes exist: commercial and residential property taxation. The former kind of taxation is not interesting in our setting because it is basically a tax on power stations and primary industries. The current choice to have residential property taxation is regulated by the property tax law of June 6th, 1975. This law restricts residential property taxation to urban areas (§ 3). Urban areas can choose to levy property taxation. Non-urban areas are prohibited from levying residential property taxation.[4]

[1]Norway have two official written languages (Norwegian and New Norwegian), and an exam has to be undertaken in both of them.

[2]The total number of students finishing in 2001/2002 and 2002/2003 is 122,281 students. Because of missing data on some explanatory variables we are left with 118,178 observations in the analysis

[3]User charges are primarily levied for infrastructure services to residents (water supply etc.) and are limited by national regulations saying that user charges cannot exceed total production costs.

[4]Property taxation can also be levied in non-urban areas from 2007.

The definition of an urban area is not clear cut and there are many court cases where property owners have argued that the area under taxation is not urban. However, residential property taxation in 2001 seems to be a choice among school districts with more than 2500 inhabitants. Only three local governments with less than 2500 inhabitants (out of 130) use residential property taxation.

Local governments take spending decisions regarding 15-20 percent of GDP, but face regulations concerning both coverage and standards of welfare services. The regulations imply a sort of centralized spending that matches the centralized financing (Rattsø, 2003). In this setting, user charge financing and property taxation are important revenue sources at the margin. Earlier Norwegian studies illustrate that decisions on user charges and property taxation are important political decisions. Borge (2000) analyzes the determinants of user charges, and finds strong evidence of revenue substitution. User charges increase when other sources of revenue become more restricted. The choice of having property taxation is investigated by Fiva and Rattsø (2007) in a model allowing for yardstick competition. Borge and Rattsø (2004) find that an equal income distribution affects the mix of revenues from property taxation and user charges.

4.3 Empirical strategy

The object of the current paper is to investigate whether property taxation works as a disciplining device on local school leaders and bureaucrats. When aiming to isolate the incentive effect of property taxation on school quality, it is clearly important to control for all other features of the school districts that affect student achievement (and might be correlated with property taxation). In particular, it is well known that family background characteristics are important determinants of student achievement. In our analysis we are able to control for a large number of family characteristics at the individual level.

Our estimation strategy contains two steps and draws on Hanushek et al. (1996).[5] In the first step we estimate the impact of each student's family background characteristics (F_{ij}) on individual exam performance (A_{ij}) and a separate indicator variable for each school district (E_{ij}):[6]

$$A_{ij} = \delta + F_{ij}\zeta + \sum_j \alpha_j E_{ij} + \mu_{ij} \tag{4.1}$$

[5]Borjas (1987) applied a similar strategy to study the impact of political and economic conditions in US immigrants home countries on US earnings (see Borjas' footnote 21 for his two step approach).

[6]Because achievement is only measured by the end in grade 10, a value added specification is not feasible.

where (E_{ij}) are dummy variables that equals 1 if student i lives in school district j and 0 otherwise, and μ_{ij} is an i.i.d. error term. Because of the discrete nature of the student achievement variable, there is a rationale for utilizing the ordered probit model. However, when we are interested in the school district fixed effects we need to rely on standard Ordinary Least Squares (OLS). The fixed effects ordered probit is inconsistent and suffers from the so-called incidental parameters problem (Wooldridge, 2002, p. 484).

Included in F_{ij} is the student's gender, age and immigrant status, parent's education and marital status and the household's income.[7] The school district fixed effects, α_j, can be interpreted as student performance conditional on family background. In the second step we utilize the school district fixed effects to investigate whether school districts that levy residential property taxation have a higher level of student achievement. The following education production function is estimated at the school district level:

$$\alpha_j = \lambda + \beta DPTAX_j + \gamma_1 R_j + \theta_1 C_j + \varepsilon_j \tag{4.2}$$

where the regressor of primary interest, $DPTAX_j$ is a dummy for residential property taxation, and R_j is a variable capturing resource use in the school sector. It is crucial not to mix up the effects of increased resources spent and the incentive mechanism caused by property taxation. As a measure on resource use we include teacher hours per student. This is a very precise measure on how much time the teacher spends in the classroom.[8] In addition we control for other characteristics at the district level C_j. ε_j is an error term. Included in C_j are several school district characteristics that capture scale effects in the production of education: enrollment, enrollment squared, (the log of) population size and traveling distance within zone. The latter variable explicitly captures time used on traveling within the school district.[9] Since other studies have found teacher experience to be an important determinant of student achievement in Norway we also include average teacher experience as an explanatory variable (Leuven et al., 2007b). Finally, we acknowledge that the controls for family background in the first step, given

[7] In other settings one would look to ethnic differences to explain differences in educational performance. In the Norwegian setting however, immigrant status is sufficient to capture differences in educational performance related to ethnic differences, since the general population is ethnically very homogenous.

[8] An alternative measure is expenditures per student which we will use as a robustness check. The raw correlation between expenditures per student and teacher education hours per student is 0.76.

[9] Travelling distance is measured as the average number of kilometres from the centre of each 'neighbourhood' (weighted by population size) to the centre of the zone. Each zone is a contiguous geographical area of 'neighborhoods' and is intended to reflect a proper organization of the school district. The centre of the zone is defined as the most populous 'neighborhood'. A densely populated school district will typically consist of many zones. If fewer than 2000 inhabitants live in the school district then the school district constitute one zone.

by Equation (4.1), may not completely capture the effect of demographic composition in the district. Unobserved differences in family background across districts may still remain in the error term. Therefore, we include several family background characteristics at the school district level (median private income, educational level in the population, the unemployment rate and the share of the population that is divorced or separated). Descriptive statistics of all variables are provided in the appendix, Table 4.7.

Because the actual school district fixed effects are not observed, we need to use the predicted values from the first step estimation. Using the first step predicted values introduces an additional error to the second step regression because of sampling error of the form $\alpha_j = \alpha_j^* + \eta_j$, which implies that the second step becomes a random components model:

$$\alpha_j = \lambda + \beta DPTAX_j + \gamma_1 R_j + \theta_1 C_j + \omega_j \tag{4.3}$$

where $\omega_j = \varepsilon_j + \eta_j$. Because the precision in the estimates of α_j will differ across school districts (in particular between small and large school districts), $\eta_j$will be heteroskedastic. A possible strategy to correct for this is to assume that the variance of η_j is proportional to the first stage sampling variance of α_j and use weighted least squares in the second stage regression, weighted with the inverse of the sampling variance of α_j. Such a strategy implies however that the other component of the error term ε_j has a variance that is either 0 or proportional to η_j (Hanushek et al., 1996). To allow a more flexible structure we use a specialized form of generalized least squares (GLS) in the second step. First we estimate Equation (4.2) with ordinary least squares (OLS). Second, we run an auxiliary regression where we regress the square of the residuals on the sampling variance of α_j and a constant term. Finally, we use the inverse of the predicted square of residuals from this regression as the weight in the GLS estimation of (4.3).

4.3.1 Endogeneity

Since residential property taxation is voluntary among urban school districts differences in achievement cannot be causally attributed to differences in the presence of property taxation. Even with the extensive list of controls included on both the individual and on the school district level, a GLS estimate of β may still capture the effect of some omitted school district characteristic or reverse causality. Note that the sign of the bias is not clear a priori. To address this potential endogeneity issue we need variation in property taxation that is arguably not subject to the choice of school district. We argue that characteristics of the historic and current property tax legislation provide such variation. The tax laws of 1911 governed the use of property taxation prior to 1975, while the property tax law of 1975 are the regulations in force in 2001. In what follows we elaborate this argument.

4.3.2 Instruments for property taxation

Historically Norwegian local governments were separated into *bykommuner* (towns) and *herredskommuner* (countryside local governments). The formal division between *bykommuner* and *herredskommuner* was established in 1837 when local self-government was introduced. Until 1975 two different tax laws existed for the two different types of local governments. Both tax laws, passed August 18th, 1911, included regulations regarding the use of residential property taxation. While residential property taxation was mandatory in towns, school districts on the countryside could choose to levy residential property taxation.

Until 1996 town status was historically solely assigned by the central government. And the decision was very stable over time. Based on the school district structure of 2001, 56 (out of a total number of 434) school districts had town status in 1911. And 45 out of these 56 remained towns from 1911 to 1975. Moreover 34 out of these 45 school districts had town status since the establishment of the local self-government in 1837. This historical township status in combination with the historical tax laws of 1911 creates exogenous variation in the school districts' use of property taxation. As a first instrument we therefore suggest an indicator variable which equals one if the school district continuously had town status from 1911 to 1975 and zero otherwise (town). 15 school districts had town status for some, but not all years between 1911 and 1975. However all of these had town status uninterruptedly for at least 33 years.[10] The main results from this analysis do not change if we include these school districts as towns in our instrumental variable strategy.

Although residential property taxation was no longer mandatory for towns after the new property tax law of 1975, 77 percent of the school districts in our sample that were

[10] In 1911, 661 school districts existed, but several have merged since 1911, in particular due to a reform taking place in the 1960s. This reform consolidated around 750 school districts into around 450. Two school districts lost their town status in 1944 and another nine lost their town status in the 1960s. Four school districts were given town status between 1911 and 1924. Three of these lost their status in the 1960s. No new towns were established between 1924 and 1975 (or between 1975 and 1995). The Local Government Act of 1992 removed the last formal division between towns and other local governments. There were no longer any need for the central government to assign town status, and from 1996 on, the local governments could choose to define themselves as towns. From 1995 to 2005 the number of Norwegian towns increased from 46 to 93.

uninterruptedly classified as towns from 1911 till 1975 still levied residential property taxation in 2001. The lower panel of Table 4.1 illustrates that to a great extent the school districts' current use of property taxation is associated with the historical township status of the school district.[11]

Our second instrument variable relates to the property tax law of 1975. This law restricts residential property taxation to urban areas. Consequently, variables that capture the current settlement pattern in the school district also have the potential to work as instruments. We pursue this strategy and suggest 'the share of the population in the school district living in rural areas (rural)' as a second instrumental variable.[12] The upper panel of Table 4.1 shows that there is a strong relationship between residential property taxation and settlement pattern.

Our first stage regression in the two stage least squares approach (2SLS), is given by:[13]

$$DPTAX_j = \lambda_2 + \delta rural_j + \varpi town_j + \gamma_2 R_j + \theta_2 C_j + u_j \tag{4.4}$$

We expect 'rural' to have a negative impact on implementation of property taxation ($\delta < 0$) and 'town' to have a positive impact ($\varpi > 0$). Note that since our potential endogenous regressor, DPTAX, is binary, it may seem appropriate to use a probit or a logit to estimate Equation (4.4) and obtain predicted values. However, this is not necessary and may even produce inconsistent estimates in the second stage. Consistency of the second stage estimates does not depended on getting the first stage functional form right (Kelejian, 1971; Angrist and Krueger, 2001). Hence, we rely on standard OLS in the first stage regression.

The exclusion restriction $E[\omega_j|town_j] = E[\omega_j|rural_j] = 0$ requires that the only effect of 'rural' and 'town' on achievement is via the presence of property taxation. This may seem as a strong assumption to make for two reasons. First, student and family characteristics are likely to be correlated with the rural-urban dimension while simultaneously being important in determining student achievement. Second, economics of scale in the production of education is likely to be correlated with the rural-urban dimension. We argue, however, that with the broad set of controls that we have available we are able

[11] Table 4.1 is based on the 425 school districts that are used in the empirical analysis.

[12] Statistics Norway has generated the "share of population in the school district living in rural areas" in the following way: Every household in a school district are assigned to a cluster of houses where each cluster is classified to be either rural or urban. For a cluster to be classified as urban it must fulfil two criteria: 1) the cluster of houses must contain at least 200 people, 2) The distance between the houses within the cluster cannot exceed 50 metres (except where houses cannot be built because of parks, rivers etc).

[13] It is important not to mix this first stage with the "first step" represented by Equation (4.1).

Table 4.1 – Current settlement pattern (2001), historical town status (1911-1975) and residential property tax.

	School districts	
	With property taxation	Total
% Living in rural areas (2001)		
0-10%	50%($N = 17$)	34
10-20%	50%($N = 20$)	40
20-30%	39%($N = 14$)	36
30-40%	36%($N = 21$)	58
40-50%	29%($N = 14$)	49
50-60%	24%($N = 12$)	50
60-70%	13%($N = 7$)	52
70-80%	6%($N = 3$)	54
80-90%	6%($N = 1$)	16
90-100%	0%($N = 0$)	36
Town status (1911-1975)		
Town	77%($N = 34$)	44
No town	20%($N = 75$)	381
Overall	26%($N = 109$)	425

Note: The definition of 'rural areas' is provided in footnote 13. The most rural school district that levies residential property taxation has 81% of the population living in rural areas. The raw correlation between 'town status' and 'the share of the population living in rural areas' is -0.37.

to control properly for family and school district characteristics that may affect student achievement. 'Traveling distance within zone' (defined in footnote 9) is in particular an important control variable. It captures differences along the rural-urban dimension that may be important for the outcome of interest (due to scale effects), and is at the same time uncorrelated with the decision to have residential property taxation, given our instruments.[14] Notwithstanding, the exclusion restriction is further investigated in section 4.5.2.

4.4 Results

In the first step in our estimation strategy we find the expected effects of family background characteristics on student achievement. In particular, we find that the achievement level increases with the parents' income and educational levels. Table 4.2 presents the results from Equation (4.1).

The school district fixed effects, which will serve as our dependent variable in the remaining analysis, come out as highly jointly statistically significant. The F-statistic for joint significance equals 4.72, statistically significant at the 1 percent level. There is substantial variation in the school district fixed effects.[15] Compared to Oslo, the 'worst' and 'best' school districts have average achievement levels of around one grade lower and one grade higher, respectively.

The second step of the analysis is to estimate the effect of residential property taxation on student achievement, represented by the school district fixed effects. However, as discussed above we need to rely on instrumental variable techniques to obtain reliable inference on the causal effect of property taxation on student achievement. Table 4.3 presents results from the first stage regression given by Equation (4.4). Specification (a) and (b) report estimated coefficients when the two instrumental variables, 'town' and 'rural', are included separately, while both are included in specification (c). In line with the descriptive statistics presented in Table 4.1, the first stage shows that the two instruments are strong predictors of residential property taxation. The F-test for joint

[14]The raw correlation between 'traveling distance within zone' and the dummy for residential property taxation is -0.16.

[15]The school district fixed effects is based on data from all 434 school districts. In the further analysis we exclude the capital Oslo because the organization of the local council differs from the rest of the school districts. We also drop one school district which is an extreme outlier from the first step to ensure that all estimated variances in the second stage are positive and another seven school districts due to missing observations, leaving us with 425 observations. Estimating Equation (1) where we restrain the sample to include only those 425 school districts we utilize in the second step yields identical results.

Table 4.2 – Computation of school district fixed effects. Dependent variable is the individual exam result.

Girl	0.340
	(0.006)***
Age (years)	0.062
	(0.009)***
ln (family income)	0.108
	(0.005)***
Education mother (years)	0.076
	(0.001)***
Education father (years)	0.067
	(0.001)***
1st or 2nd generation immigrant	−0.095
	(0.015)***
Parents non-cohabiting	−0.124
	(0.014)***
R-squared	0.1947
N	118,178
School district fixed effects	Yes
Estimation method	OLS

Note: The unit of observation is the individual student. Included in the regression, but not reported are a constant term, a year dummy, dummies for different subjects, and school districts fixed effects. The capital, Oslo, is the reference school district, which is not included in the second step. Standard errors are in parentheses. *** denotes statistical significance at the 1% level.

significance of the excluded instrument(s) in the linear probability model obtains values between 19.35 and 21.39. We conclude that 'town' and 'rural' do not suffer from the problems related to weak instruments. As expected 'traveling distance within zone' does not come out statistically significant.

Table 4.4 presents the second stage results obtained from Equation (4.3). As a benchmark for evaluating the results we provide the standard GLS ignoring the potential endogeneity problem related to our key explanatory variable. The effect of property taxation is estimated to be 0.052 and is statistically significant at the 1 percent level. This result is the first indication that property taxation has a favorable incentive effect on school officials, stimulating them to provide effective and high quality schooling in line with arguments of Glaeser (1996) and Hoxby (1999). The positive and statistically significant relationship between property taxation and student achievement is confirmed when we use 'rural' and 'town' as instruments for the dummy for residential property taxation. The point estimate increases from 0.052 in standard GLS to 0.163 in the 2SLS estimation using both instruments simultaneously. The naive GLS estimate seems to be downward biased which suggests a negative correlation between property taxation and the disturbance term in Equation (4.3). This suggests that public sector quality is negatively correlated with the decision to implement property taxation. The estimated coefficient for property taxation is considerably higher than the benchmark GLS estimation also when 'town' and 'rural' are used separately as instruments and the effect is statically significant at the 10 and 5 percent level, respectively. We interpret this as evidence of property taxation having favorable incentive effects on local school leaders and bureaucrats. An alternative interpretation is that property taxation increases popular control of the bureaucrats because property taxation represents a direct and visible influence on the private economy of the inhabitants.

Although we are primarily interested in testing whether there are beneficial incentive effects of property taxation, it is interesting to evaluate how large this effect is. An increase in average student achievement of 0.163 grading points, as suggested by our main specification (g), corresponds to (slightly above) one standard deviation increase in the school district fixed effects. The disciplining effect of property taxation on school leaders leads to a one grading point increase for around one out of six students. Although the standard error related to the estimated effect of property taxation on student achievement is admittedly quite large, the incentive effect seems to be substantial.

Table 4.3 – First stage estimations: Historical town status (1911-1975), current settlement pattern and the probability of levying residential property tax. Dependent variable is a dummy for residential property tax.

Specification	(a)	(b)	(c)
Rural		−0.543	0.319
		(0.119)∗∗∗	(0.078)∗∗∗
Town	0.361		−0.479
	(0.079)∗∗∗		(0.118)∗∗
Teacher hours per student	0.004	0.004	0.004
	(0.002)∗∗∗	(0.002)∗∗	(0.001)∗∗
Traveling distance within zone	−0.003	−0.001	0.000
	(0.003)	(0.003)	(0.003)
Enrollment	0.004	0.002	0.004
	(0.003)	(0.002)	(0.002)
$\text{Enrollment}^2/1000$	−0.028	−0.017	−0.027
	(0.014)∗∗	(0.014)	(0.014)
ln (population)	0.158	0.163	0.116
	(0.039)∗∗∗	(0.039)∗∗∗	(0.040)∗∗∗
Teacher experience	0.013	0.013	0.014
	(0.007)∗∗	(0.007)∗∗	(0.007)∗∗
Low. sec.	−0.056	−0.572	−0.091
	(0.586)	(0.574)	(0.575)
High. sec.	−0.330	−0.494	−0.246
	(0.699)	(0.697)	(0.686)
Unemployment rate	1.869	0.439	0.291
	(1.282)	(1.343)	(1.318)
Divorce rate	−0.884	−0.040	−1.208
	(1.495)	(1.470)	(1.470)
Median income (1000 NOK)	−0.002	−0.006	−0.005
	(0.001)	(0.002)∗∗∗	(0.002)∗∗∗
R-squared	0.247	0.246	0.276
N	425	425	425
Joint significance of instruments,			
F-test (p-value)	21.39 (0.00)	20.81 (0.00)	19.35 (0.00)
Estimation method	OLS	OLS	OLS

Note: The unit of observation is the school district. Included in the regressions, but not reported, is a constant term. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4.4 – The effect of residential property tax on student achievement. Dependent variable is the school district fixed effects.

Specification	(d)	(e)	(f)	(g)
DPTAX	0.052	0.190	0.138	0.163
	(0.019)***	(0.090)**	(0.083)*	(0.065)**
Teacher hours per student	−0.001	-0.002	-0.001	−0.001
	(0.001)	(0.001)*	(0.001)*	(0.001)*
Traveling distance within zone	0.004	0.005	0.005	0.005
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Enrollment	−0.003	-0.003	-0.003	−0.003
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Enrollment^2/1000	0.011	0.012	0.012	0.012
	(0.006)**	(0.006)**	(0.006)*	(0.006)**
ln (population)	0.000	-0.032	-0.020	−0.026
	(0.016)	(0.027)	(0.025)	(0.022)
Teacher experience	0.008	0.006	0.007	0.006
	(0.003)***	(0.003)*	(0.003)**	(0.003)**
Low. sec.	−1.074	-0.969	-1.009	−0.990
	(0.242)***	(0.265)***	(0.255)***	(0.256)***
High. sec.	−0.711	-0.629	-0.660	−0.645
	(0.290)**	(0.312)**	(0.301)**	(0.304)**
Unemployment rate	−0.656	-1.058	-0.906	−0.979
	(0.572)	(0.659)	(0.632)	(0.622)
Divorce rate	−1.476	-1.590	-1.547	−1.567
	(0.615)**	(0.657)**	(0.634)**	(0.642)**
Median income (1000 NOK)	−0.001	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
R-squared	0.189			
N	425	425	425	425
Instrument(s) for property tax		town	rural	town, rural
Sargan's test (p-value)				0.21 (0.64)
Estimation method	GLS	GLS-IV	GLS-IV	GLS-2SLS

Note: The unit of observation is the school district. Included in the regressions, but not reported, is a constant term. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

4.5 Sensitivity analysis

4.5.1 Resource use and the disciplining effect of property taxation

A potential critique of our central result might be that revenue from property taxation yields increased spending in the education sector, which our resource use variable fails to capture. This may be important if increased spending increases student performance. To further investigate whether this is a relevant critique, we have run an additional regression where 'expenditures per student' is included as an alternative resource measure.

The property tax effect is basically unaltered when 'teacher hours per student' is substituted with 'expenditures per student' (specification (h) in Table 4.5) and when both measures are included simultaneously (not reported). In line with a large literature, we find that resource use do not seem to have any impact on student performance (e.g. Hanushek, 2002). However, since resource use is likely to be endogenous we are not necessarily obtaining the causal effect of resource on student achievement. The important point in our setting is that the estimated effect of property taxation is very robust to different definitions of resource use and also that the effect is unaltered if we exclude measures of resource use (specification (i) in Table 4.5).

Finally we investigate whether the effect of resource use is dependent on the cost structure in the school district, by interacting our measure of resource use with 'traveling distance', 'enrollment' and 'enrollment^2'. These measures are likely to capture potential differing cost effects across districts in the production of education. We find some support in favor a differential effect of resource use. In particular, the positive sign of the interaction effect of enrollment^2 and 'teacher hours per student' suggest that the resource use effect is increasing non-linearly in enrollment. As an example, consider two school districts where one has an average (student weighted) school size ('enrollment') of 28 and the other has an average (student weighted) school size of 78. These numbers correspond to the first and third quartile in 'enrollment'. Based on specification (k), the effect of increasing 'teacher hours per student' with one standard deviation (18 hours) is estimated to be -0.007 for the district with small schools and 0.081 for the district with large schools. But, most importantly in our setting, allowing for a differential effect of resource use based on traveling distance or enrollment leave the estimated impact of property taxation on student achievement basically unaltered.

Table 4.5 – The effect of residential property tax on student achievement, further investigations of the role of resource use. Dependent variable is the school district fixed effects.

Specification	(h)	(i)	(j)	(k)
DPTAX	0.161	0.162	0.171	0.153
	(0.066)∗∗	(0.065)∗∗	(0.063)∗∗∗	(0.063)∗∗∗
Teacher hours per student (th)			−0.001	−0.001
			(0.001)	(0.001)
Expenditures per student	0.001			
	(0.001)			
Traveling distance within zone (tz)	0.004	0.004	0.015	0.005
	(0.002)∗∗	(0.002)∗∗∗	(0.008)∗	(0.002)∗∗∗
(tz*th)/100			−0.010	
			(0.008)	
Enrollment	−0.002	−0.002	−0.003	0.001
	(0.001)∗	(0.001)∗∗	(0.001)∗∗∗	(0.004)
Enrollment^2/1000	0.009	0.010	0.012	−0.053
	(0.006)	(0.006)∗	(0.006)∗∗	(0.004)
(Enrollment*th)/1000				−0.007
				(0.005)
(Enrollment^2*tz)/1000				0.001
				(0.000)∗∗
ln (population)	−0.017	−0.020	−0.026	−0.019
	(0.022)	(0.022)	(0.022)	(0.021)
Teacher experience	0.007	0.007	0.006	0.006
	(0.003)∗∗	(0.003)∗∗	(0.003)∗∗	(0.003)∗∗
Low. sec.	−0.929	−0.942	−0.968	−0.997
	(0.259)∗∗∗	(0.257)∗∗	(0.257)∗∗∗	(0.252)∗∗∗
High. sec.	−0.521	−0.574	−0.646	−0.618
	(0.309)∗	(0.304)∗	(0.306)∗∗	(0.301)∗∗
Unemployment rate	−1.108	−1.102	−0.937	−0.792
	(0.630)∗	(0.628)∗	(0.631)	(0.612)
Divorce rate	−1.799	−1.716	−1.495	−1.608
	(0.645)∗∗∗	(0.642)∗∗∗	(0.650)∗∗	(0.637)∗∗
Median income (1000 NOK)	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
N	423	425	425	425
Sargan's test (p-value)	0.37 (0.54)	0.33 (0.58)	0.18 (0.67)	0.18 (0.67)

Note: The unit of observation is the school district. Both 'town' and 'rural' are used as instruments in all specifications (h) -(k) and all specifications (h) - (k) are estimated with GLS - 2SLS. Included in the regressions, but not reported, is a constant term. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

4.5.2 Empirical investigation of the exclusion restriction

The validity of the identification strategy pursued in this analysis rests on the assumption that 'rural' and 'town' are excludable from our structural equation in the sense that they relate to student achievement only through their influence on residential property taxation. The key challenge for obtaining reliable inference is that differences along the rural-urban dimension still exist even after conditioning on the broad set of controls that we have available. If this is the case, 'rural' and 'town' are invalid as instruments and should be included in the second stage regression, rather than in the first stage regression. According to the Sargan instrument validity test, this does not seem to be a relevant critique. The test fails to reject the null that our instruments are uncorrelated with the error term in all specifications (with p-values above 0.5 in all specifications). However, since it is widely acknowledged that overidentification tests have low power in certain settings (e.g. Newey, 1985), we also cut the sample according to property tax status and run the following 'second step' regressions:

$$\alpha_j = \lambda + \varphi rural_j + \phi town_j + \gamma_2 R_j + \theta_2 C_j + \omega_j \tag{4.5}$$

If our exclusion restriction is valid then 'rural' and 'town' should not have any impact on student achievement when we condition on the decision to have property taxation (and the other control variables). The effects of 'rural' and 'town' on student achievement are statistically insignificant both for the sample of school districts with property taxation and for school districts without property taxation (results are reported in the appendix, Table 4.8. We conclude that unobserved differences across the rural-urban dimension do not seem to be driving our main result.

4.5.3 Spatial dependence

In an extension of our baseline model we focus on interdistrict school competition. School district competition is likely to appear as a spatial in pattern in school quality. Interdistrict school district competition can arise either because school districts face outmigration (in-migration) if they provide low (high) quality schooling or because of some form of yardstick competition. An observed spatial pattern in school district quality is not necessarily due to competition among school districts. Common shocks and unobserved correlates will also appear as spatial dependence. Whatever the causes of spatial dependence, ignoring it may give biased and inconsistent estimates. Spatial dependence is however not biasing the effect of property taxation on student achievement, given that our instruments are valid. Nonetheless, incorporating spatial dependence provide a relevant robustness test of the key finding in this analysis.

To test for spatial dependence we follow the empirical literature on interjurisdictional competition and relate the value of the dependent variable at a given school district to its value at neighboring locations in our regression analysis.[16] We define neighbors as school districts with a common border. The spatial weight matrix that we apply is row-standardized, ensuring that the spatially lagged dependent variable can be interpreted as a weighted average of neighboring school districts' school quality. A non-zero coefficient is consistent with interdistrict school competition.

The spatial econometric literature recommends either instrumental variable or maximum likelihood methods to handle the simultaneity problem inherent in spatial lag models. For simplicity we have decided to ignore the simultaneity problem, since we are introducing spatial dependence primarily as a robustness check. To provide a method that unify the econometric approach based on Hanushek et al. (1996) and take into account the endogeneity of the spatially lagged dependent variable is beyond the scope of the current paper. The simultaneity problem yields upward biased estimates of the spatial lag parameter.

Table 4.6 reproduces the results from Table 4.4, but includes a spatially lagged dependent variable, assumed to be exogenous. The positive parameter estimate with upward bias of around 0.3 is consistent with the school district competition hypothesis.[17] Blair and Staley (1995), which also ignore simultaneity, find for comparison an effect of approximately 0.4 on US data. In another recent US study, Millimet and Rangaprasad (2007) find that school districts compete with neighboring school districts in the choice of educational inputs.

Controlling for spatial dependence, the estimated effects of property taxation on school quality are similar, but somewhat reduced to the corresponding non-spatial specifications. The property tax effect is statistically significant at least at the 10 percent level in all specifications, except in specification (n) where we only utilize 'rural' as an instrument. We conclude that the property tax dummy does not seem to capture school district competition or other spatially correlated omitted variables.

[16] Revelli (2005) presents an overview of different studies utilizing this approach and discusses econometric issues.

[17] The Moran test on the residuals from the specifications reported in Table 4.4 also shows spatial dependence (statistically significant at the 5 percent level).

Table 4.6 – The effect of residential property tax on student achievement controlling for spatial dependence. Dependent variable is the school district fixed effects.

Specification	(l)	(m)	(n)	(o)
DPTAX	0.039	0.163	0.070	0.120
	(0.019)∗∗	(0.093)∗	(0.094)	(0.069)∗
Teacher hours per student	−0.001	−0.001	−0.001	−0.001
	(0.001)	(0.001)∗	(0.001)	(0.001)∗
Traveling distance within zone	0.004	0.005	0.004	0.004
	(0.002)∗∗	(0.002)∗∗∗	(0.002)∗∗	(0.002)∗∗∗
Enrollment	0.009	0.010	0.009	0.010
	(0.006)	(0.006)∗	(0.006)∗	(0.006)∗
Enrollment^2/1000	−0.002	−0.003	−0.002	−0.003
	(0.001)∗∗	(0.001)∗∗	(0.001)∗∗	(0.001)∗∗
ln (population)	0.000	−0.029	−0.007	−0.019
	(0.016)	(0.026)	(0.026)	(0.022)
Teacher experience	0.009	0.007	0.008	0.007
	(0.003)∗∗∗	(0.003)∗∗	(0.003)∗∗	(0.003)∗∗
Low. sec.	−0.948	−0.893	−0.934	−0.912
	(0.237)∗∗∗	(0.253)∗∗∗	(0.241)∗∗∗	(0.244)∗∗∗
High. sec.	−0.713	−0.638	−0.694	−0.665
	(0.281)∗∗	(0.301)∗∗	(0.287)∗∗	(0.290)∗∗
Unemployment rate	−0.591	−0.971	−0.685	−0.838
	(0.560)	(0.651)	(0.628)	(0.608)
Divorce rate	−0.722	−1.051	−0.804	−0.936
	(0.626)	(0.701)	(0.673)	(0.664)
Median income (1000 NOK)	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Spatial lag	0.348	0.246	0.323	0.282
	(0.086)∗∗∗	(0.117)∗∗	(0.115)∗∗∗	(0.103)∗∗∗
R-squared	0.225			
N	425	425	425	425
Instrument(s) for property tax		town	rural	town,rural
Sargan's test (p-value)				0.60 (0.44)
Estimation method	GLS	GLS – IV	GLS - IV	GLS – 2SLS

Note: The unit of observation is the school district. Included in the regressions, but not reported, is a constant term. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively

4.6 Conclusion

Several researchers have asserted that inefficiency in the public sector arises from a lack of incentives to behave efficiently. When output is not well defined, as in the education sector, it may be hard to detect inefficient production. Monitoring may simply be too difficult. Residential property taxation may help overcome this problem by establishing a direct relationship between the tax level and the benefits received by taxpayers through capitalization of public sector quality into housing prices.

Utilizing a well-suited data set from the educational sector in Norway, we find evidence that students living in school districts levying residential property taxation perform better at the nationally decided external examination than other students. The main result holds when we utilize variation in property taxation driven by the tax laws from 1911 and the property tax law from 1975 in a two stage least squares framework. All the results presented in this analysis corroborate the theoretical framework put forward by Glaeser (1996) and Hoxby (1999) well. These studies emphasize that local property taxation gives bureaucrats incentives to maintain an efficient and well functioning public sector by stronger linking public sector quality to public sector financing. The current analysis shows that these theoretically consistent mechanisms also seem to matter in practice.

A. Appendix

Table 4.7 – Summary statistics and data description.

Variable	Mean	s.d.
Individual characteristics (N=118,178)		
- Girl	0.49	0.50
- Age (measured in years)	14.52	0.30
- ln (family income)	13.27	0.66
- Education mother (measured in years)	11.82	2.71
- Education father (measured in years)	11.96	2.85
- 1st or 2nd generation immigrant	0.04	0.20
- Parents non-cohabiting	0.30	0.46
- Parents married	0.65	0.48
School district characteristics (N=425)		
- School district fixed effects[a]	-0.02	0.14
- Property taxation, dymmy = 1 if school district levy property tax and zero otherwise (DPTAX)	0.26	0.44
- Teacher hours per student	85.39	17.83
- Expenditures per student (1000 NOK)	64.63	13.73
- Traveling distance within zone[b]	8.14	6.96
- Enrollment	54.87	34.77
- Enrollment^2/1000[c]	4.42	4.99
- Teacher experience (measured in years)	19.27	2.98
- ln (population)	8.48	1.07
- The share of the population (16–74 years) with completed lower secondary education as highest educational level ("Low. sec.")	0.21	0.06
- The share of the population (16–74 years) with completed upper secondary education as highest educational level ("High. sec.")	0.62	0.04
- Unemployment rate, yearly average	0.03	0.02
- Divorce rate, yearly average	0.06	0.02
- Median income for persons 17 years and older (1000 NOK)	195.10	20.98

Note:[a] weighted with the inverse of the sampling variance from the first step, [b] defined in footnote 9, [c] student weighted average. The data on residential property taxation is collected in a survey carried out in 2001. Data on student achievement are from the Norwegian Board of Education. The other data are provided by the Norwegian Social Science Data Services and Statistics Norway. Neither of these institutions is responsible for the analysis conducted or for the conclusions drawn.

Table 4.8 – Empirical investigation of the exclusion restriction. Dependent variable is the school district fixed effects.

Specification	(p)	(q)	(r)	(s)	(t)	(u)
Town		0.019	0.017		0.071	0.068
		(0.041)	(0.042)		(0.058)	(0.059)
Rural	−0.058		-0.051	-0.035		−0.025
	(0.140)		(0.142)	(0.065)		(0.065)
Teacher hours per student	−0.003	-0.003	-0.003	0.000	0.000	0.000
	(0.001)**	(0.001)**	(0.001)**	(0.001)	(0.001)	(0.001)
Traveling distance within zone	0.007	0.007	0.007	0.004	0.003	0.004
	(0.004)*	(0.004)*	(0.004)*	(0.002)*	(0.002)*	(0.002)*
Enrollment	−0.001	-0.001	-0.001	-0.004	-0.003	−0.003
	(0.002)	(0.002)	(0.002)	(0.002)**	(0.002)**	(0.002)**
Enrollment^2/1000	−0.002	-0.001	-0.001	0.017	0.015	0.015
	(0.009)	(0.009)	(0.009)	(0.009)*	(0.010)	(0.010)
ln (population)	−0.007	-0.008	-0.010	0.006	0.002	0.000
	(0.026)	(0.026)	(0.027)	(0.023)	(0.023)	(0.024)
Teacher experience	0.003	0.003	0.003	0.011	0.011	0.011
	(0.005)	(0.005)	(0.005)	(0.004)***	(0.004)***	(0.004)***
Low. sec.	−1.206	-1.182	-1.172	-0.907	-0.861	−0.856
	(0.478)**	(0.485)**	(0.487)**	(0.298)***	(0.300)*	(0.302)***
High. sec.	−0.239	-0.153	-0.219	-0.707	-0.762	−0.745
	(0.500)	(0.467)	(0.505)	(0.381)**	(0.377)**	(0.382)**
Unemployment rate	−0.074	0.136	-0.073	-1.332	-1.296	−1.361
	(1.005)	(0.818)	(1.009)	(0.800)	(0.779)*	(0.800)*
Divorce rate	−2.058	-2.044	-2.175	-1.317	-1.551	−1.561
	(1.230)*	(1.209)*	(1.269)*	(0.748)*	(0.774)**	(0.778)**
Median income (1000 NOK)	0.000	0.001	0.000	-0.001	-0.001	−0.001
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
R-squared	0.244	0.244	0.245	0.197	0.202	0.201
N	109	109	109	316	316	316

Note: The unit of observation is the school district. Included in the regressions, but not reported, is a constant term. All specifications from (p) to (u) are estimated with GLS and the sample in all specifications from (p) to (u) consists of those school district with DPTAX = 1. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Bibliography

[1] Aksoy T. and Link C. R. (1999). A Panel analysis of student mathematics achievement in the US in the 1990s: does increasing the amount of time in learning activities affect math achievement? *Economics of Education Review*, 19: 261-277.

[2] Angrist, J. D. and Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114(2): 533-575.

[3] Angrist, J. D. and Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives,* 15: 69-86.

[4] Angrist, J. D. and Lavy, V. (2002). The effect of high school matriculation awards: Evidence from randomized trials. NBER Working Paper No. 9389

[5] Arai, M. and Thoursie, P. S. (2005). Incentives and Selection in Cyclical absenteeism. *Labor Economics*, 12: 269-280.

[6] Askildsen, J. E., Bratberg, E. and Nilsen, Ø. A. (2005). Unemployment, labor force composition and sickness absence: a panel data study. *Health Economics*, 14: 1087-1101.

[7] Baker D. P., LeTendre, G. K., and Akiba, M. (2005). Schoolwork at home? Low-quality schooling and homework. In Baker, D. P. and LeTendre, G. K. (Eds.), *National differences, global similarities: World culture and the future of schooling* (pp. 117-133). Stanford, CA: Stanford University Press.

[8] Balli S. J., Demo D. H. and Wedman J. F. (1998). Family Involvement with Children's Homework: An Intervention in the Middle Grades. *Family Relations*, 47: 149-157.

[9] Barmby, T., Sessions, J. G. and Treble, J. G. (1994): Absenteeism efficiency wages and shirking. *Scandinavian Journal of Economics,* 96: 561-566.

[10] Barmby, T. A., Ercolani, M. G. and Treble, J. G. (2002). Sickness Absence: An International Comparison. *The Economic Journal,* 112(480): F315-F331.

[11] Betts J. R. (1996). The Role of Homework in Improving School Quality. Unpublished manuscript, Department of Economics, University of California, San Diego.

[12] Bingley, P., Jensen, V. M. and Walker, I. (2005). The effect of class size on educational attainment: Danish quasi-experimental evidence and evidence that controls for family, school and neighbourhood effects. Unpublished manuscript.

[13] Black, S. (1999). Do better schools matter? Parental evaluation of elementary education. *Quarterly Journal of Economics,* 114: 577-600.

[14] Blair, J. P. and Staley, S. (1995). Quality competition and public schools: Further evidence. *Economics of Education Review,* 14: 193-198.

[15] Bonesrønning, H. (2003). Class size effects on student achievement in Norway: Patterns and explanations. *Southern Economic Journal*, 69(4): 952-965.

[16] Boozer, M. and Rouse, C. (2001). Intraschool variation in class size: Patterns and implications. *Journal of Urban Economics*, 50(1): 163-189.

[17] Borge, L.-E. and Rattsø, J. (2004). Income distribution and tax structure: Empirical test of the Meltzer-Richards hypothesis. *European Economic Review,* 48: 805-826.

[18] Borge, L.-E. and Rattsø, J. (2006). Property taxation as incentive for cost control: Empirical evidence for utility services in Norway. Discussion Paper No. 12/2006, Norwegian University of Science and Technology.

[19] Borge, L.-E. (2000). Charging for public services: The case of utilities in Norwegian local governments. *Regional Science and Urban Economics*, 30: 703-718.

[20] Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review,* 77: 531-553.

[21] Bradley, S., Green, C. and Leeves, G. (2007). Worker absence and shirking: Evidence from matched teacher-school data. *Labor Economics,* 14: 319-334.

[22] Brasington, D. M. and Haurin, D. R. (2006). Educational outcomes and house values: A test of the value-added approach. *Journal of Regional Science,* 46: 245-268.

[23] Brennan, G. and Buchanan, J. (1980). The power to tax. *Cambridge University Press*, New York.

[24] Brown, S., Fakhfakh, F. and Sessions, J. G. (1999). Absenteeism and employee sharing: An empirical analysis based on French panel data, 1981-1991. *Industrial and Labor Relation Review*, 52: 234-251.

[25] Browning, M. and Heinesen, E. (2006). Class size, teacher hours and educational attainment. *Scandinavian Journal of Economics*, forthcoming.

[26] Card, D. and Payne, A. A. (2002). School finance reform, the distribution of school spending, and the distribution of student test scores. *Journal of Public Economics*, 83: 49-82.

[27] Clapp, J. M., Nanda, A. and Ross, S. L. (2007). Which school attributes matter? The influence of school district performance and demographic composition of property values. *Journal of Urban Economics*, forthcoming.

[28] Cooper H. (1989). Homework. White Plains, N. Y.: Longman.

[29] Cooper H. (2007). The Battle Over Homework. Third edition, Corwin Press, California

[30] Ehrenberg, R. G., Ehrenberg, R. A., Rees, D. I. and Ehrenberg, E. L. (1991). School District Leave Policies, Teacher Absenteeism, and Student Achievement. *Journal of Human Resources*, 26: 72-105.

[31] Eren, O. and Hendersen D. J. (2006). The Impact of Homework on student achievement. Unpublished manuscript, Department of Economics, State University of New York, Binghamton, NY.

[32] Falch, T., Rønning, M. and Strøm, B. (2006): A cost model of schools: School size, school structure and student composition. In Sougel, N and Jaccard, P., editors, Governance and performance of education systems. Springer, Dordrecht.

[33] Falch, T. and Strøm, B. (2005). Wage Bargaining and Political Strength in the Public Sector. CESifo Discussion Paper No. 1629.

[34] Figlio, D. N. and Lucas, M. E. (2004). What's in a grade? School report cards and housing market. *American Economic Review,* 94: 591-604.

[35] Fiva, J. H. and Rattsø, J. (2007). Local choice of property taxation: Evidence from Norway. *Public Choice*, forthcoming.

[36] Gary-Bobo, R. J. and Mahjoub, M. B. (2006). Estimation of class size effects, using 'maimonides rule': The case of French junior high schools. CEPR Discussion Paper No. 5754.

[37] Glaeser, E. (1996). The incentive effects of property taxes on local governments. *Public Choice,* 89: 93-111.

[38] Goux, D. and Maurin, E. (1999). "Persistence of Interindustry Wage Differentials: A Reexamination using Matched Worker-Firm panel data". *Journal of Labor Economics,* 17(3): 492-533.

[39] Guryan, J. (2001). Does money matter? Regression-discontinuity estimates from education finance reform in Massachusetts. NBER Discussion Paper No. 8269.

[40] Hahn, J., Todd, P. and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1): 201-209.

[41] Häkkinen, I., Kirjavainen, T. and Uusitalo, R. (2003). School Resources and Student Achievement Revisited: New Evidence from Panel Data. *Economics of Education Review*, 22(3): 329-335.

[42] Hanushek, E. A., Rivkin, S. and Taylor, L. (1996). Aggregation and the Estimated Effects of School Resources. *The Review of Economics and Statistics*, 78(4): 611-627.

[43] Hanushek, E. A., Kain, J. F. and Rivkin, S. G. (2004). Why Public Schools Lose Teachers. *Journal of Human Resources,* 39: 326-354.

[44] Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24: 1141-1177.

[45] Hanushek, E. A. (2002). Publicly provided education. In Auerbach, A. and Feldstein, M., editors, *Handbook of Public Economics*, volume 4, chapter 30. Elsevier Science Publishers B. V., Amsterdam

[46] Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal*, 113: F64-F98.

[47] Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research*, 43: 387-425.

[48] Hemningway, H., Shipley, M. J., Stansfeld, S. and Marmot, M. (1997). Sickness absence from back pain, psychosocial work characteristics and employment grade among office workers. *Scandinavian Journal of Work Environment Health*, 23: 121-129.

[49] Hægeland, T., Raaum, O. and Salvanes, K. G. (2005). Pupil achievement, school resource use and family background. IZA Discussion Paper No. 1459, Bonn.

[50] Hægeland, T., Raaum, O. and Salvanes, K. G. (2007). Pennies from heaven: Using exogenous variation to identify effects of school resources on pupil achievement. Unpublished manuscript.

[51] Hoxby, C. M. (1999). The productivity of schools and other local public goods producers. *Journal of Public Economics,* 74: 1-30.

[52] Hoxby, C. M. (2000). The effect of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics*, 115(4): 1239-1285.

[53] Ichino, A. and Riphahn, R. T. (2005). The effect of employment protection on worker effort: a comparison of worker absenteeism during and after probation. *Journal of European Economic Association*, 3: 120-143.

[54] Jacobsen, S. L. (1989). The effects of pay incentives on teacher absenteeism. *Journal of Human Resources*, 24(2): 280-286.

[55] Johansson, P. and Palme, M. (2005). Moral Hazard and sickness insurance. *Journal of Public Economics*, 89: 1879-1890.

[56] Kane, T. J., Riegg, S. K. and Staiger, D. O. (2006). School quality, neighborhoods, and housing prices. *American Law and Economics Review,* 8: 183-212.

[57] Kelejian, H. H. (1971). Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variables. *Journal of the American Statistical Association,* 66: 373-374.

[58] Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, 114(2): 497-532.

[59] Krueger, A. B. (2003). Economic Considerations and Class Size. *Economic Journal*, 113: F34-F63.

[60] Krueger, A. B. and Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project star. *Economic Journal*, 111: 1-28.

[61] Lavy, V. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of Political Economy*, 110: 1286-1317.

[62] Leuven, E., Oosterbeek, H. and van der Klaauw, B. (2003). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. CEPR Discussion Paper No. 3921.

[63] Leuven E., Lindahl M., Oosterbeek H. and Webbink D. (2006). Expanding schooling opportunities for 4-year-olds. Unpublished manuscript. Department of Economics, University of Amsterdam.

[64] Leuven E., Lindahl M., Oosterbeek H. and Webbink D. (2007a). The effect of extra funding for disadvantaged pupils in achievement. *Review of Economics and Statistics,* forthcoming.

[65] Leuven, E., Oosterbeek, H. and Rønning, M. (2007b). Quasi-experimental estimates of the effect of class size on achievement in Norway. Unpublished manuscript, Norwegian University of Science and Technology.

[66] Leuven, E. (2006). Sick of Work? The effect of work time reduction on the sickness absenteeism of teachers. Unpublished manuscript, University of Amsterdam

[67] Lindahl, M. (2005). Home versus School Learning: A New Approach to Estimating the Effect of Class Size on Achievement. *Scandinavian Journal of Economics*, 107(2): 375-394.

[68] Lindahl, M. (2001). Summer Learning and the Effect of Schooling: Evidence from Sweden. IZA Discussion Paper No. 262, Bonn

[69] Machin, S. and Salvanes, K. G. (2007). School choice and the housing market: Valuation through an admission policy reform. Paper presented at the Royal economic society's annual conference, April 2007.

[70] Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60: 531-542.

[71] Miller, R. T., Murnane, R. J. and Willett J. B. (2006). The causal impact of teacher absence on student achievement. Unpublished Manuscript, Harvard University, Graduate School of Education.

[72] Millimet, D. L. and Rangaprasad, V. (2007). Strategic competition amongst public schools. *Regional Science and Urban Economics,* 37: 199-219.

[73] Newey, W. (1985). Generalized method of moments specification testing. *Journal of Econometrics,* 29: 229-256.

[74] OECD, Education At A Glance., Paris: OECD, 2003.

[75] Ose, Solveig O. (2004). Working conditions, compensation and absenteeism. *Journal of Health Economics*, 24: 161-188.

[76] Papke, L. E. (2005). The effect of spending on test pass rates: evidence from Michigan. *Journal of Public Economics*, 89: 821-839.

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

[78] Rattsø, J., (2003). Vertical Imbalance and fiscal behavior in a welfare state: Norway, in: Rodden J., G. Eskeland and J. Litvack, eds., *Fiscal decentralization and the challenge of hard budget constraints*, MiT Press, Cambridge, MA, 133-160.

[79] Revelli, F. (2005). On spatial public finance empirics. *International Tax and Public Finance,* 12: 475-492.

[80] Rivkin S. G., Hanushek, E. A. and Kain J. F. (2005). Teachers, schools and academic achievement. *Econometrica*, 73: 417-458.

[81] Røed, K. and Fevang, E. (2006). Organizational Change, Absenteeism, and Welfare Dependency. *The Journal of Human Resources*, 42: 156-193.

[82] Shapiro, C. and Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review,* 74: 433-444.

[83] Strøm, B. (2004). Student achievement and birthday effects. Unpublished manuscript. Norwegian University of Science and Technology.

[84] Todd, P. and Wolphin, K. I. (2003). Towards a unified approach for modeling the production function for cognitive achievement. *Economic Journal*, Features: F3-F33.

[85] Urquiola, M. (2006). Identifying class size effects in developing countries: Evidence from rural Bolivia. *Review of Economics and Statistics*, 88: 171-176.

[86] Wilson, J. D. and Gordon, R. (2003). Expenditure competition. *Journal of Public Economic Theory,* 5: 399-417.

[87] Winkler, D. R. (1980). The effects of sick-leave policy on teacher absenteeism. *Industrial and labor Relations review*, 3(2): 232-240.

[88] Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data. *MiT Press,* Cambridge, MA.

[89] Wössmann, L. (2005). Educational Production in Europe. *Economic Policy*, 20(43): 445-504.

[90] Wössmann, L. and West, M. (2006). Class size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review*, 50(3): 695-736.