[Preliminary]

Do Class Size Effects Differ Across Grades?

Anne Brink Nandrup

Department of Economics and Business, Aarhus University, annebn@asb.dk

This version: January, 2014

Abstract

Class size effects on pupil outcomes have been intensely analyzed in economics of education. The contribution of this paper is to study whether short-run class size effects are constant across grade levels in compulsory school. Results are based on administrative data on all pupils enrolled in the Danish public school system. By exploiting exogenous variation in class sizes created by a government imposed maximum class size rule, I am able to evaluate the effects of class size separately across both lower and upper primary school as well as lower secondary school. Significant (albeit modest in size) positive effects of class size reductions are found for children in primary school. Also, the effects are statistically different across primary and secondary school.

Keywords: Class size, regression discontinuity, literacy, test scores
JEL codes: I21, I28, C31

Acknowledgments. I gratefully acknowledge comments from Nina Smith, Maria Knoth Humlum and Helena Skyt Nielsen.

1 Introduction

This paper evaluates the short-run effects of class sizes on pupil performance *across* different grade levels of the Danish public school system[1]. More specifically, I estimate class size effects on pupil abilities within both mathematics and reading for children attending lower and upper primary school as well as lower secondary school, separately.

[1] I have chosen to denote the public primary and lower secondary school (or in Danish terms 'folkeskolen') as the public school system. Of course there are many other components of the public school system in Denmark, e.g. upper secondary schools and technical colleges. These are self-governing institutions in contrast to the public primary and lower secondary schools owned by the municipalities.

While there exists a vast literature estimating short- and medium-run class size effects in primary and secondary school, previous studies are often concerned with only one or a few close grades in the same setup (e.g. Heinesen and Browning (2007) on 8^{th} graders, Finn and Achilles (1999) on kindergarten through grade 3, Angrist and Lavy (1999) on 4^{th} and 5^{th} graders, and Fredriksson et al. (2013) on $4^{th} - 6^{th}$ graders). Thus, there is little empirical insight into the mechanisms of class size effects across the years of schooling.

There is a number of reasons why class size effects potentially differ across grade levels[2]. Pupils in their early school years may be more dependent on adult supervision and help, for example, peer-tutoring or group work may be more effective in later grades (Blatchford and Mortimore 1994). Teaching methods may vary depending on the difficulty level of the material taught. The self-control of pupils may increase with age as well as a number of other psychological factors presumably change as pupils mature. Also, parents may be more qualified to assist their children, e.g. with homework, supplementary reading etc., in the early school years. This potentially has large implications for the optimal class composition across compulsory schooling. Particularly, for certain institutional settings, such as the current Danish one, where pupils typically attend the same class throughout primary and lower secondary school.

The quality of the public primary and secondary schools have increasingly been at the center of attention in many countries over the course of the last decades. This follows partly from the recognition that the formation of human capital has important implications for both the individual and the society as well. As such, early test score measures of pupils' academic achievement are possibly strongly related to measures of sustained success in adulthood such as wages and length of education (Todd and Wolpin, 2003). Additionally, public school programs in numerous countries are faced with substantial budget cuts or periods of no growth in consequence of the global financial crisis. Thus, school administrators and politicians on both national and local levels are preoccupied with creating the best school systems within a tight budget.

When considering schemes to improve the quality of school systems class and school sizes are a recurrent issue. They are readily measured and are in general considered easier to manipulate than other school inputs. Furthermore, class size reductions seem to be popular with (almost) every party interested in schools such as parents, teachers, school administrators etc. However, reducing class sizes are expensive; In Denmark, teachers' salaries alone constitute 80 pct. of compulsory schooling expenditure. Thus, class size increments comprises large potential monetary gains in terms of budget savings - if there are little or no adverse effects of increasing the number of pupils in a class room.

Exploiting test results from the unique Danish national test system in combination with detailed register-based data, I am able to identify the effects of changes in class size on contemporaneous cognitive ability estimates for at least three different levels of compulsory schooling. Following in the footsteps of An-

[2]The same arguments apply to possible differences in class size effects across subjects such as reading and math, as these are generally perceived to be based on very different mindsets.

grist and Lavy (1999), I employ a fuzzy regression discontinuity design arising from a mandatory maximum class size rule of 28 pupils. I apply this identification strategy to data covering pupils enrolled in the $2^{nd}/3^{rd}$, 6^{th} and 8^{th} grade in the Danish public school system between 2009/2010 and 2011/2012. Furthermore, in order to gain insight into the dynamics of class size effects, I estimate the effect of class size on reading and math test results, respectively.

I find that significant (albeit modest) positive effects of class size reductions are present in the Danish public school system where the average class size is 21 with a modal value of 23. Most effects of a class size increment in primary school are significantly negative, whereas none of the 8^{th} grade estimates are significant. More importantly, under certain circumstances I am able to reject that the results do not differ across grades levels. Furthermore, I employ a wide range of robustness checks to underpin the validity of my results.

The remainder of the paper is organized as follows. Section 2 reviews relevant literature and places the contribution of this study within that. Section 3 summarizes the institutional setting of the Danish public school system and the national test system while Section 4 presents the available data as well as the identifying variation of the IV estimates. The identification strategy is described in Section 5 where empirical results are presented and discussed in Section 6. Finally, Section 7 summarizes and concludes.

2 Literature Review

In accordance with the importance of the topic, the literature in the area of schooling inputs is extensive. In the following, I highlight selected contributions that are of particular relevance to this paper, thus, focusing on the estimation of causal effects of class size.

A primary goal of the education production function literature is to understand the technology of schooling inputs such as class size in the creation of cognitive achievement outcomes. The analogy between education production and firm production should be interpreted with caution, however (Hoxby 2000). The production function of a firm is the result of maximizing an objective function given a set of production possibilities, but such maximization objectives are not imposed on schools. Even though a reduction in class size necessarily increases the *opportunity* of investment in the ability formation of the child, it is not obvious that these opportunities are seized depending on the incentives faced by the respective schools. For example, teachers, who suddenly experience a class size reduction, may not adapt his/her teaching methods accordingly.

Lazear (2001) proposes the 'Disruption model' to explain the effect of class size on pupil outcome. Here, class room learning is considered a public good. Thus, congestion effects, where one pupil impedes the learning of all others, are increasing with class size assuming the probability of a pupil interrupting is constant across individuals. Because teaching is temporarily suspended during an interruption, the negative externality to everyone else in class is increasing with class size. Consequently, optimal class size is larger with more well-behaved pupils or if particularly difficult children are selected into smaller classes by school administrators. While not incorporating all (dis)advantages of class sizes, this model proposes a specific functional form for the educational production function. Using the framework of Lazear (2001), class size effects likely differ across grade levels if one is willing to entertain the notion that the probability of a given child behaving disruptively is not constant across age and grade level.

On the empirical front, class sizes are rather difficult to study because the majority of class size variations is likely the result of choices made by parents, school administrators, teachers and politicians on a local or national level. Thus, the variation of class size is potentially correlated with other determinants of pupil achievement (Hoxby 2000), and OLS estimates do not have causal interpretation. Parents focused on the education of their children may choose to locate close to schools with more resources per pupil (corresponding to a lower class size) or otherwise self-select into such schools. Furthermore, as proposed by Lazear (2001) school administrators may chose to allocate pupils with learning disabilities or from disadvantaged backgrounds to smaller classes, hoping that they will not be left (further) behind. And also, larger classes may be assigned extra teaching hours assuming that it is not the class size itself determining pupils' test results but rather teacher resources per pupil. All of which causes upward-biased estimates of class size effect on pupil achievement. On the other hand, parents may choose to invest less in their children when per pupil school resources are high, resulting in a downwards bias of the estimated direct effect of class size. Datar and Mason (2008) find evidence that class size and parental investment are complements for certain parental inputs while they are substitutes for others. Along this line, the estimated effects should in general be interpreted as 'total policy effects' because of limited data on other school inputs (Todd and Wolpin 2003). I.e. the total effect of an exogenous change in class size that include both the *ceteris paribus* effect of class size as well as an indirect effect through responses of other inputs to the formation of human capital. Although one is usually most interested in the total policy effect, these estimates provide little insight into the nature of the human capital production function.

Tennessee's STAR experiment, where pupils are randomly assigned to classes of varying size, has given rise to a number of studies focusing on the first four years of compulsory schooling, e.g. Krueger (1999), and Finn and Achilles (1990) find that attending a class of $13-17$ pupils improves test scores by $0.15-0.28$ standard deviations compared to a 'regular' class size of $22-25$ pupils in kindergarten trough grade 3. Equally important, this effect seem to be lessened for pupils entering the small classes at a later point. By comparing IV estimates of class size effects in the regular classes to those obtained between large and small classes, Krueger (1999) does not find evidence in favor of the Hawthorne phenomenon, where individuals, knowing they are being evaluated, act to increase their productivity. In order to be eligible for participation in the STAR project, schools had to be large enough to accommodate at least three classes in each grade implying that at least 57 pupils should be enrolled in each grade. This is somewhat larger than the average Danish public school (see Table 2).

Also the effects of class sizes in upper primary school have been heavily estimated. The identification method presented in this paper is similar to that of Angrist and Lavy (1999). They were the first to propose an instrument based on fuzzy discontinuities in class size triggered by administrative maximum class size rules. Applying this strategy to Israeli 3^{rd}, 4^{th} and 5^{th} graders they find evidence of significant negative effects of a one-pupil class size increment on the distribution of class test score means for Israeli 4^{th} and 5^{th} graders[3]. These results are not directly comparable to those of the STAR project, but after recalculating, the authors estimate them to be in the lower end of the findings from the Tennessee STAR project. On an individual level the class size effects of 3^{rd} graders are numerically smaller and all insignificant[4]. Unfortunately, the only control included in addition to functions of enrollment is an index of the fraction of pupils from disadvantaged backgrounds; based on a function of the father's education, continent of birth, and family size. Fredriksson et al. (2013)[5] slightly alter the identification method and find significant negative effects of increasing the average class size in the $4^{th}-6^{th}$ grade largely in line the Tennessee STAR findings. Moreover, the authors are able to identify significant beneficial long-term effects e.g. on adult wages. Thus, recent evidence suggest that short-run class size effects of at least upper primary school are potentially highly persistent.

Trying a different approach, Hoxby (2000) uses both an instrument based on cohort size arising from natural variation in the timing of births as well as one based on maximum class size rules on 4^{th} and 6^{th} graders from Connecticut. Because the two identification strategies are independent, she argues that the two methods can be used as checks on each other. Nevertheless, she finds only insignificant effects (of varying sign) of a class size reduction on pupil test scores for the two grade levels. Bingley et al. (2007) question the validity of the birth-rate instrument because parents very likely have better information than the researcher concerning the forecast of local cohort size, and may take action accordingly.

Using the same approach, Heinesen and Browning (2007) find modest and marginally significant negative effects of larger class size on years of education for Danish 8^{th} graders. While Heinesen (2010) finds significant and substantial positive effects on 9^{th} grade examination marks in French from reducing class sizes using within-school variation over time in the subject-specific French classes.

Summing up, there is more or less convincing evidence that class size reduc-

[3]The authors only has access to micro-level data on 3^{rd} graders.

[4]The authors argue that this is likely caused by different test conditions. 3^{rd} graders were tested the year after the other pupils and consequently teachers had the opportunity to prepare the pupils for the test.

[5]The authors use a representative sample of approximately $5-10$ pct. of Swedish cohorts born in 1967, 1972, 1977, and 1982 that is further reduced by restricting the sample to pupils residing in single-school school districts.

tions are beneficial for pupils across most grade levels in compulsory schooling. Unfortunately, comparison of these effects are complicated by the varying institutional settings across countries as well as the outcome measures seldom being comparable. In this paper, I will attempt to remedy this by employing the same identification method across grade levels in the Danish public school system while using directly comparable measures of pupil ability as the outcomes of interest.

3 Institutional Setting

In this section, I will provide an overview on the institutional setting pertaining to the grade levels and school years in the scope of this paper. In 2009 the Danish legislation was changed from nine to ten years of compulsory schooling. In effect this means that children must now receive teaching from the calendar year they turn six years old, beginning with preschool.[6] In contrast, preschool was optional before 2009 with compulsory schooling beginning with 1^{st} grade in the year of the pupil's seventh birthday. However, schools were still obliged to provide a preschooling scheme which was already chosen by the vast majority of parents (Nikolajsen and Molsgaard 2012). The amendment seems to only have motivated a more 'timely' commencement of schooling, in the sense that fewer pupils are being held back a year before entering the school system. The 2^{nd} graders of the school year 2011/2012 are the only ones in the sample influenced by the sparse effects of the amendment.

Public schools are free, run and financed by local municipalities through municipality income tax in combination with a complex between-municipality redistribution scheme subsidizing expenditures in low-income municipalities, and subject to a maximum class size rule of 28 pupils per class[7]. In reality, class sizes vary considerably across schools and cohorts. This is partially caused by some municipalities adopting an even lower suggested maximum class size rule, both to ensure a better learning environment for the pupils from the perspective of potential negative effects of larger class size as well as further minimizing the risks of having to split up already existing classes, with the increased costs that follow. In effect, this means that the 28 pupil rule may not be binding for a (unknown) subset of municipalities. To accommodate this potential pitfall, I will identify the empirical maximum class size rules of the specific municipalities and present results based on this along side with the primary specification results in Section 6 .

There are 98 municipalities in Denmark. Each municipality is divided into one or more school districts, where the affiliation of the pupil is determined by his/her official residential address.[8]

[6] Homeschooling satisfying official standards is a valid substitute but this is rarely practiced in Denmark.

[7] Admission of up to 30 pupils per class *during* the school year is accepted in order to counteract potential class divisions outside of the summer break.

[8] Note that the Danish school districts are more in line with a suggested school catchment

Approximately $14 - 15$ percent of Danish children are enrolled into private schools in the school years of $2009/2010 - 2011/2012$. Various types of independent private schools exist, such as religious or ideology based, while others simply constitute an alternative to the public school system. In Denmark, private schooling is heavily subsidized with almost 85 pct. of the expenditures covered by the local municipalities. Still, Bingley et al. (2007) note that Danish private schools are disproportionally attended by the children of highly educated parents. Private schools are mostly found in urban areas and in general they have a lower mean class size.

Generally, pupils are divided into classes when they enroll in preschool and follow the same class throughout the years of compulsory schooling with few exceptions, for example elective third language (most commonly German or French). These are usually not introduced until 7^{th} grade and onwards.

The Danish public school system is built on the principle that pupils should not be divided according to ability or social background (Wandall 2010). In fact, this is prohibited by law. This means that differential treatment is only justifiable when it offers each pupil the opportunity to achieve as much as possible regardless of his/her background. Consequently, there are no elite schools or classes in the public system. There is, however, considerable variation in ability within classes.

Since the school year of 2009/2010 pupils enrolled into Danish public schools have been subject to ten mandatory tests within different subjects across compulsory schooling. Acknowledging that learning processes may differ across linguistic and logical subjects this paper primarily studies test results in reading (2^{nd}, 6^{th}, and 8^{th} grade), math (3^{rd} and 6^{th} grade), and physics/chemistry (8^{th} grade). In order to be able to analyze the effects of class size across lower and upper primary as well as lower secondary school this last substitution is necessary as pupils are not tested in math after the 6^{th} grade. Furthermore, physics and math are often considered to be based on a somewhat similar mindset.

The tests are conducted in January through April in each year, they are compulsory[9], IT based, adaptive, and self-scoring, thus, the teacher is only presented with the test scores and is accordingly not able to bias the results.

The adaptiveness of the tests should ensure that ability of the pupils are very precisely evaluated including the top and bottom ability pupils, see Beuchert-Pedersen and Nandrup (2014). Also, the nature of the tests make them qualified for comparison both across and within individuals.

Even though there is no formal division between the different grades of the

area. The decision-making authority is the municipality and some school districts have more than one school. Also, since 2006 pupils are not required to attend the district school but they are entitled to.

[9]Dispensation of pupils can be granted if the school, in agreement with the parents, estimates that a pupil is unable to finish the test while obtaining a test score that is useful in the evaluation of the child's teaching plan. Thus, one has reason to suspect than mainly inferior pupils are exempt from taking the tests.

Danish compulsory schools, I choose, in accordance with the institutional settings of other countries, to divide the grades in three groups: lower primary school, consisting of preschool and 1^{st} through 3^{rd} grade; upper primary school, containing 4^{th}, 5^{th} and 6^{th} grade; and lower secondary school with grades 7 through 9. Thus, I am able to examine and compare the effects of class size in each of the three school levels. Additionally, it is municipalities rather than schools independently that finance the incurred expenditures associated with the maximum class size rule. This means that class size effects in the Danish setting are found without holding school budgets fixed; the costs of assignment according to the rule are met across all schools in the municipality (Bingley et. al. 2007).

4 Data and Identification

4.1 Data sources and sample selection

School enrollment into the Danish schools is registered in the beginning of each school year (early September). As such, this registry contains yearly class and school information of all individuals admitted in the Danish school system and from this it is possible to construct beginning of the school-year class sizes and enrollment counts of each grade in each school.

To this data, I match register-based information linking pupils to parents, along with a rich set of pupil and parent characteristics. The added data include information of parents' educational level, age, and yearly earnings in the year the child turns six years old, civil status of the mother, ethnicity, date of birth, birth weight and gestation length, as well as degree of urbanization of the school municipality, and the number of siblings etc. Pupil academic performance is measured by the national test scheme. This outcome is standardized to mean 0 and standard deviation 1 in accordance with Beuchert-Pedersen and Nandrup (2014). School information and test data are reported to UNI-C, an agency under The Danish Ministry of Children and Education, while data on annual earnings come from the Income Tax Register. Both registers are maintained by Statistics Denmark.

Using beginning of the school-year enrollment is of course not accurate seeing as pupil performance is measured in January through April. However, beginning of the school-year enrollment is less likely to be affected by the behavior of parents or school administrators (Angrist and Lavy 1999). Also, it is likely that class size during the school year is what matters for education attainment rather than just the class size at the time of the tests. This naturally implies that there is a slight chance that some pupils have changed school in the intervening time and thus, do not take the test in the class they are registered to. However, unless extraordinary circumstances are present, it is not unreasonable to assume that parents generally choose to time the school changes of their children during the summer period, where there is a natural break in the lessons.

The population sample consists of pupils enrolled into Danish public schools

	Total number of observations	Percentage of total sample
Pupils enrolled into public ordinary classes in a relevant grade (2, 3, 6, and 8) in the school years of 2009/2010 − 2011/2012$^{a)}$	1,288,741	100.00%
Test-score observations available	923,188	71.63%
Valid class reporting available	898,271	69,70%
Hereof:		
Pupils tested in reading, 2^{nd} grade	152,095	11.80%
Pupils tested in math, 3^{rd} grade	153,368	11.90%
Pupils tested in reading, 6^{th} grade	154,464	11.99%
Pupils tested in math, 6^{th} grade	154,510	11.99%
Pupils tested in reading, 8^{th} grade	142,387	11.04%
Pupils in tested physics/chemistry, 8^{th} grade	141,447	10.97%

Notes. $^{a)}$ indicates that pupils in the 6^{th} and 8^{th} grade are are included twice, as they are tested twice at this level.

Table 1: Sample selection procedure of the estimation samples

in all relevant grades subject to reading, math, and physics/chemistry tests $(2^{nd}, 3^{rd}, 4^{th}, 6^{th}$, and $8^{th})$ and who have completed these tests in the school years of 2009/2010 through 2011/2012[10]. Nearly all pupils appear multiple times as they have completed multiple tests in this three year period. Only very few pupils have erroneously completed the same test twice or completed a test of a different grade level. Hence, I follow five cohorts of pupils for a maximum of three subsequent school years. Table 1 summarizes the sampling procedure.[11]

4.2 Descriptives and Identifying Variation

Table 2 summarizes the descriptives of the six relevant subsamples by test information along side with the within pupil subsamples. Average enrollment count in the relevant grades is $50 - 61$ with an average class size of approximately 21 pupils, however, on average 0.5 pupils in each class do not take the test. Across the subsamples, enrollment counts are approximately 10 individuals larger in

[10]This scope is chosen as test results are not available before 2009/2010. Moreover, in October 2012 the Danish Ministry of Children and Education launched a large experiment including two teachers in the daily teaching covering public school in 18 Danish municipalities.

[11]A relatively large fraction of the missing test scores is due to technical problems in the implementation process (Wandall, 2010). In particular in 2010 the test system was shut down for two weeks and all test bookings in this period was lost. Unfortunately, not all schools managed to rebook the lost test slots, thus, not all pupils managed to complete the compulsory test. The 2013 evaluation report (Rambøll (2013)) establishes that the missing test results in connection to the break down are randomly distributed across pupils.

	Estimation sample							
Variable	Mean	S.d.	Min	Max	Mean	S.d.	Min	Max
	Reading, 2^{nd} ***grade***				***Math,*** 3^{rd} ***grade***			
	152,095 *obs.,* 8,028 *classes*				153,368 *obs.,* 8,279 *classes*			
Enrollment	50.57	22.38	1	145	49.98	21.77	1	135
Class size	21.19	4.08	1	61	21.17	4.05	1	54
Test class size	20.54	4.29	1	59	20.48	4.27	1	54
	Reading, 6^{th} ***grade***				***Math,*** 6^{th} ***grade***			
	154,464 *obs.,* 8,259 *classes*				154,510 *obs.,* 8,282 *classes*			
Enrollment	50.14	21.29	1	159	50.23	21.33	1	159
Class size	21.36	3.93	1	68	21.34	3.89	1	68
Test class size	20.90	4.12	1	51	20.88	4.09	1	47
	Reading, 8^{th} ***grade***				***Physics/chemistry,*** 8^{th} ***grade***			
	142,387 *obs.,* 7,757 *classes*				141,447 obs., 7,737 classes			
Enrollment	60.79	22.56	1	182	60.73	22.62	1	183
Class size	21.85	3.61	1	60	21.84	3.60	1	60
Test class size	21.12	3.90	1	54	21.13	3.89	1	54

Notes. Data is based on pupils enrolled in normal classes of the Danish public school system in the school years of 2009/2010 to 2011/2012 and tested through the national test scheme.

TABLE 2: INTRODUCTORY DESCRIPTIVE STATISTICS OF SUPSAMPLES OF THE FULL ESTIMATION SAMPLE

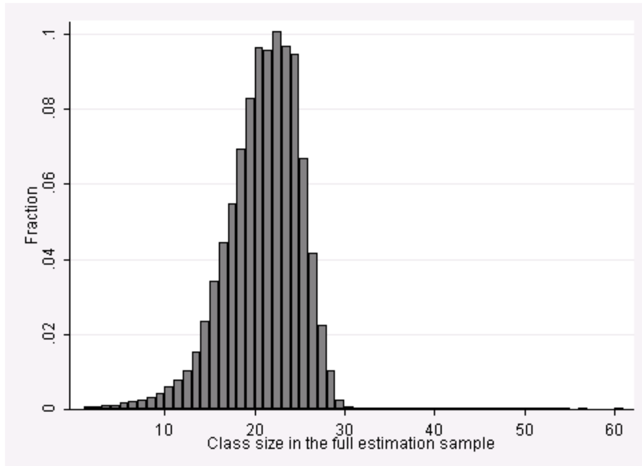
the 8^{th} grade while mean class size is increased with around 0.5 pupils. This is most likely caused by schools admitting pupils from other schools that only teach lower and upper primary school. Finally, note that descriptives of subsamples of different tests performed in the same grades are very similar.

Figure 1 illustrates the distributions of class sizes in the full estimation sample. A few (3.93 pct.) of the classes in the sample are very small (below 14 pupils) while 0.4 pct. are larger than 28 pupils, and of these only 95 exceed the official maximum of 30. These are mostly results of experimental classes with for example both a teacher and a pedagogue/teacher aid assigned to the class, or they could potentially be results of erroneous reportings of class names. I have chosen not to exclude these, because I am hesitant to condition on the potentially endogenous variable. Furthermore, all result in Section 6 are robust to the exclusion of 'too' large classes. The modal value of class size is 23.

As originally carried out by Angrist and Lavy (1999), I use the exogenous variation in class sizes created by the maximum class size rule of 28 pupils as instrument for the potential endogeneity bias of class size. Following the authors, the predicted class size function, assuming cohorts are divided into classes of equal size of grade g in school s in year t[12], is given by (1):

$$f_{gst} = e_{gst} / \left(floor\left(\left(e_{gst} - 1 \right) / 28 \right) + 1 \right) \quad (1)$$

[12]For simplicity, the t subscript is omitted in the remainder of this paper, but all instruments are based on the enrollment count of grades in the relevant years.



The figure shows the distribution of the pooled class sizes in grades 2, 3, 6, and 8 of the Danish public school system in the school years of 2009/2010 – 2011/2012.

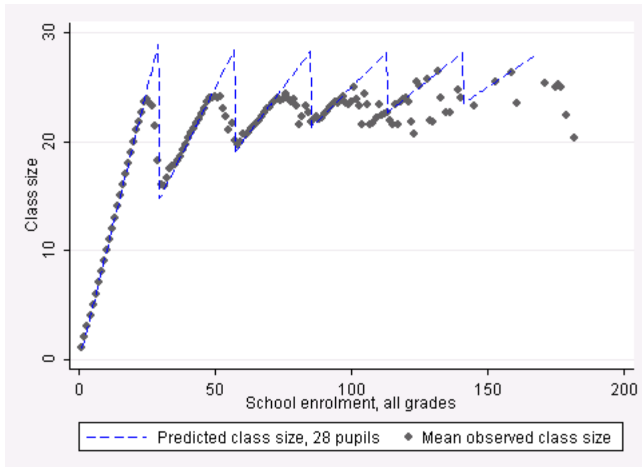
Figure 1: Distribution of class sizes, the full estimation sample

where e_{gst} denotes the enrollment count of the grade, and 28 is the maximum number of pupils in one class. The function $floor\,(n)$ returns the largest integer less than or equal to n, for any number n. (1) reflects that according to the maximum class size-rule enrollments of up to 28 pupils are assigned to one class while enrollments between 29 and 56 are divided into two classes of 14.5–28 pupils each, etc.

Figure 2 illustrates the relationship between school enrollment on the grade level on the horizontal axis and predicted as well as mean observed class sizes on the vertical axis in the full estimation sample. The line captures predicted class size while the dots mark observed mean class size of each enrollment count. Even though predicted class size based on at most 28 pupils per class is presumably not the only factor explaining class size, it clearly has large explanatory power over observed class size - at least below enrollment counts of 100.[13] Figure 2 shows how the probability of treatment (being assigned to a small class) in a fuzzy regression discontinuity context should be higher to the right of the cutoff rather than to the left. But treatment is not guaranteed (thus, 'fuzzy').

As identification only arises when the rule binds, I am disregarding useful variation by including all municipalities. However, obtaining information of suggested class size rules on a municipality level would require substantial resources. Furthermore, these suggested guidelines are, as previously mentioned, not mandatory and the school is still obliged to enroll pupils moving into the school district.

[13] See Figure A.1 of the Appendix for a more detailed grade-level overview of the relation between predicted and mean observed class size. The strong pattern is largely consistent across grades with a somewhat poorer fit for the eighth grade.



The figure shows predicted and mean observed class size for pupils in grades 2, 3, 6, and 8 in Danish public schools for the school years of 2009/2010-2011/2012. The predicted class size function indicated by the blue line is based on administrative rules ensuring a maximum of 28 pupil per class.

FIGURE 2: PREDICTED AND MEAN OBSERVED CLASS SIZE BY ENROLLMENT, THE FULL ESTIMATION SAMPLE

4.2.1 Inclusion of Controls

The detailed register-based information of the Danish population allows me to include rich set of controls in the regressions. Specifically, I incorporate controls regarding birth information, the parents' socioeconomic status, school characteristics and class characteristics of the child.

Birth information of the child includes indicator variables for quarter of birth, birth order of the child by the mother, multiple born (e.g. twin), gender, and age as well as birth weight and length of gestation. Previous studies have found severe long-term effects of being born prematurely or with low birth weight. For example, Black et al. (2007) show that higher birth weight significantly increases IQ scores of $18-20$-year-old Norwegian males as well as education attainment and adult earnings, while Clark et al. (2008) find evidence of lower cognitive abilities as well as poorer self-regulation for prematurely born children.

Socioeconomic status of the family includes indicator variables of the highest attained educational level of each parent in the child's sixth year, the marital status of the mother and the ethnicity of the child (both western and non-western immigrant and descendant hereof), as well as the logarithm of each parent's yearly earnings, the age of the child's parents and the number of siblings based on the births of the mother at the year the child turns six years old.

School characteristics include indicator variables for degree of urbanization of the school municipality and school year as well as functions of enrollment into the different grades of the school.

Furthermore, I add *class characteristics,* such as mean academic performance, mean education and log-earnings of the parents, share of girls and non-western immigrants among the other pupils in the class (excluding pupil i), to certain model specifications. A number of studies have suggested both negative disruption effects and positive spill-over effects on readings and math scores of (especially classroom) peers in both primary and secondary school (e.g. Burke and Sass 2013).

A complete list of controls including descriptive statistics hereof is found in the Appendix (Table A.1).

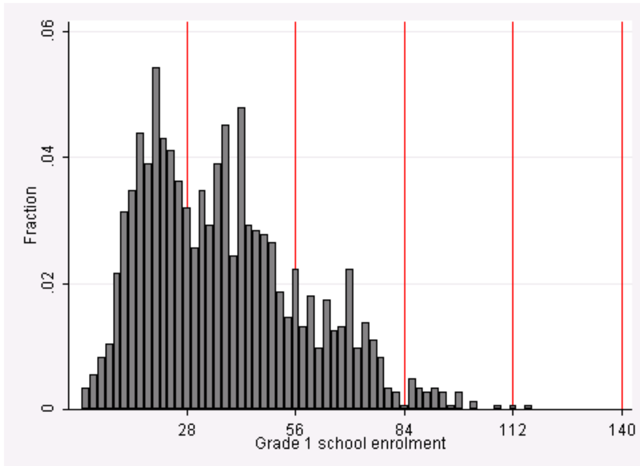
4.2.2 Is the Regression Discontinuity Design Valid at School Level?

In a regression discontinuity context, random assignment of treatment intensities may be undone by parent/administrator sorting when the treatment is public knowledge (McCrary 2008). Thus, validity of the instrument relies heavily on the absence of discontinuities in the distribution of grade enrollments at the thresholds created by the 28-pupil rule as this indicates sorting by municipality administrators.

Because of discontinuities in the enrollment count of Swedish schools Fredriksson et al. (2013) are compelled to disregard enrollment on a school level and focus on school-district enrollment instead. The problem arises as Swedish legislation encourages adjustment of school catchment areas within school districts such that the “needs” of the pupils as well as resources at the public schools are utilized optimally. Urquiola and Verhoogen (2009) documents an extreme case of bunching on the basis of the maximum class size rule in Chilean subsidized private schools. The number of schools to the right of the cutoff points are approximately five times larger than to the left.

In Denmark, the decision-making authority regarding school districts and school catchment areas lies with the municipality. Thus, municipalities are entitled to change the school catchment areas and school districts if deemed necessary. In practice this entitlement is implemented very differently across Danish municipalities, and I have only been able to find very few examples of a municipal council revising school districts yearly in order to utilize school capacities. Furthermore, a liberal choice of schools should at least partly offset this.

Figure 3 illustrates the distribution of enrollments on school level for 1^{st} graders in the school year of 2009/2010. As municipalities can only adjust school catchment areas before cohorts enroll into schools and kindergarten was not made mandatory until 2009, 1^{st} grade enrollment is the most relevant distribution to examine. By visual inspection one cannot reject the presence of a slight drop in the fraction of schools with enrollments into grade 1 around the first cutoff, while around the subsequent thresholds there is no clear evidence of discontinuities. In short, I do not find clear indications of bunching below the cutoff points in the distribution of grade enrollment, but I cannot reject the existence of a small (albeit important) potential problem of manipulation of enrollment counts around the first cutoff. Schools are not able to manipulate



The figure shows the distribution of school enrollment into the 1^{st} grade of Danish public schools in the school year of 2009/2010. Vertical lines indicate the cutoff points created by multiples of the 28-pupil rule.

FIGURE 3: DISTRIBUTION OF ENROLLMENT IN GRADE 1, 2009/2010

enrollment, however. This can only happen at the municipality level.

It is highly probable that better schools face increased demand, because parents caring about schooling inputs would selectively choose schools based on school quality. This is, however, likely to be a smooth function process and not reflecting the up-and-down pattern created by the 28 pupils rule (Angrist and Lavy 1999). Nonetheless, enrollment is likely to be related to pupil performance for reasons other than changing class sizes. A necessary assumption for identification, the independence assumption, requires that there is no direct effect of the instrument (or its underlying assigning variable, enrollment) on pupil performance, except through its impact on class size, thus, the need for including sufficient controls for enrollment effects is apparent. A natural implication of this assumption is that parents and administrators do not selectively exploit the administrative rules. However, the liberal Danish legislation regarding choice of schools reduces the costs of the parents of manipulating enrollment and class size for at least two reasons:

Since 2006 pupils have been entitled but not obliged to enroll in the district school. This means that parents can apply for enrollment in the other schools of the residential municipality - or even of other municipalities. These schools are then required to accept the pupil provided that they have available places (i.e. they do not have to introduce a new class based on the maximum class size rule) (Undervisningsministeriet, 2010). If the popularity of the school is larger than the number of free seats, the pupils are selected according to a list of criteria including traveling distance and the presence of siblings at the school. Also, parents can choose to enroll their children in heavily subsidized private schools (see Section 3).

The low costs of changing schools is potentially problematic as parents may be more likely to exploit the maximum class size rule and undo the random assignment of class sizes. But to the extend that parents are not able to *precisely* manipulate the assignment variable the variation in treatment near the cutoff points should be randomized (Lee and Lemieux 2010). Intuitively, parents may be able to predict class size crudely based on maximum 28 pupils in each class, but as treatment depends on the enrollment of all other children in the class, it would be very risky to actively choose schools based on enrollments just above the cutoffs. This becomes increasingly riskier in larger schools, but is seems unlikely that parents would select their children into very small public schools, usually located in the country side, to be certain of a small class size as it would commonly imply large transportation costs and potentially poorer family characteristics of the classmates. Additionally, there may be other costs related to small schools such as less specialization and diversity of the teacher staff and less flexibility of the school day. To some extend, this applies to the municipality administrators as well. They are able to estimate the number of children in each school district based on the previous year, but they cannot completely predict the number of children moving in to the municipality nor where they will live or their school preferences. Hence, there is reason to believe that the instrument has greater validity in small intervals around the enrollment thresholds[14].

Unfortunately, data on school resources are not available, and consequently I am unable to examine whether schools compensate pupils in larger classes e.g. by use of remedial training or extra supervision.

In order to assess the validity of the instrument based on potential parent manipulation, Table 3 shows the results of regressing the pooled indicator for being above an administrative threshold on selected baseline variables (the corresponding estimates of the remaining predetermined characteristics are found in Table A.2 of the Appendix). For simplicity I have chosen to pool the instruments such that the *above* cutoff equals unity when grade enrollment in a school supersedes one of the cutoffs. Column (1) shows regression results of the full estimation sample for grades 2, 3, 6 and 8. Here, only a few baseline characteristics are unrelated with the pooled instrument. However, based on the corresponding $p-$values of the ± 4 discontinuity sample in column (2), none of the predetermined characteristics, except the age of the mother, the indicator of father's education being vocational and length of gestation, are related to the instrument.[15]

[14] In the following, the ± 4 discontinuity sample refers to a sample including pupils enrolled in schools with a grade enrollment of ± 4 pupils around the three lower thresholds (28, 56, and 84). The sample is restricted to these threshold based on the aforementioned greater predictive power of the instruments here (see Figure 2). Descriptives of the ± 4 discontinuity sample are found in Table A.3 of the Appendix.

[15] Performing an OLS regression of the *above* dummies on all baseline covariates simultaneously in the ± 4 discontinuity sample yield similar results. Only the coefficients of birth weight and length of gestation are statistically significant, however, because of the small number of clusters in the discontinuity sample an F-statistic cannot be calculated. Adjusting the

Baseline covariate	(1) p-value, *above* cutoff **Full estimation sample**	(2) p-value, *above* cutoff ± 4 **discontinuity sample**
Female	.039	.542
Non-western immigrant	.001	.245
Mother's education:		
− Vocational	.256	.191
− Higher	.000	.205
Father's education:		
− Vocational	.900	.085
− Higher	.000	.687
Mother's log-earnings	.001	.399
Father's log-earnings	.003	.940
Mother's age	.000	.070
Father's age	.000	.204
Number of siblings	.051	.899
Separated parents	.001	.238
No. of observations	898,271	178,082

Notes. The above cutoff indicator equals 1 if the school enrollment at grade level exceeds a thresholds created by the 28 pupil rule up to +14 pupils (+4 pupils in the discontinuity sample). Columns report the p-values for t-tests of the significance of the pooled class size instrument by separate OLS regressions on the variables listed in each row. The following controls are also included in the regressions: Year and enrollment segment fixed effects, indicator variables of degree of urbanization of the school municipality and linear and square controls for grade enrollment interacted with separate thresholds (only for the full sample). Standard errors adjusted for clustering by enrollment count are in parenthesis.

TABLE 3: BALANCING OF COVARIATES

5 Estimation Strategy

Because of nonrandom selection of pupils into classes, class sizes are potentially endogenous and hence simple OLS estimates of effects on test scores may be biased and without causal interpretation. However, by exploiting the exogenous variation in class size induced by administrative rules, it is possible to interpret the effects of class size on pupil achievement causally (Angrist and Lavy 1999), given validity of the instrument.

The class size effects of pupil test scores are assumed to be explained by a standard reduced-form education production function:

$$\theta_{icgs\tau} = \boldsymbol{X}_{icgs}\boldsymbol{\alpha}_2 + \alpha_1 CS_{cgs\tau} + \varphi_\tau + g_\tau^k(e_{gs}) + \varepsilon_{icgs\tau} \qquad (2)$$

where θ_{icgst} denotes the standardized test score of individual i in class c of grade g at school s at enrollment segment τ. $\boldsymbol{X}_{icgs}$ is a vector of controls

standard errors for clusters by schools instead yields a F-statistic of approximately 67, thus rejecting that the covariates are jointly unrelated to the instrument. This is not surprising seeing as the number observations in the regression is quite large.

including characteristics of pupil i (family information, gender, birth weight etc.), degree of urbanization of the school municipality and characteristics of class c[16] etc. See Table A.1 for a complete list of controls. CS_{cgs} denotes the observed class size and the residual $\varepsilon_{icgs\tau}$ is pupil specific. I have included segment fixed effects, φ_τ, to accommodate different patterns around the separate enrollment thresholds[17]. Also, I have allowed the coefficients of the enrollment functions, $g_\tau^k(e_{gs})$, where k is the order of the polynomial, to vary by segment.

The preferred specification of this paper does not include the expected class size as a function of enrollment from (1). Rather I exploit a dummy variables approach taking on the value 1 if grade enrollment is above the threshold and zero otherwise. This highlights the quasi-experimental identification strategy of the RD design and excludes the smooth variation in the predicted class size between thresholds. More specifically: indicator variables, $above28 = \mathbf{1}(28 < e \leq 42)$, $above56 = \mathbf{1}(56 < e \leq 60)$ etc.[18], act as instruments, thus when e is to the right of the cutoff point the probability of treatment (small expected class size) is high, otherwise there is low probability of treatment. Heinesen and Browning (2003) argue that this is the most appropriate specification because only variation in the instrument around the cutoff points is used, while Fredriksson et al. (2013) is concerned that this specification results in a potential efficiency loss as the variation of treatment intensities, as a result of varying jumps in predicted class size, are not utilized. The authors conclude that the efficiency loss appears limited. Given the binary instruments class size is assumed to be explained by

$$CS_{cgs\tau} = \boldsymbol{X}_{icgs}\boldsymbol{\gamma}_2 + \boldsymbol{\gamma}_1\boldsymbol{above}_{gs} + \phi_\tau + q_\tau^k(e_{gs}) + \upsilon_{icgs\tau} \qquad (3)$$

where $\boldsymbol{above}_{gs}$ denotes the vector of dummy instruments, ϕ_τ are the segment fixed effects, $q_\tau^k(e_{gs})$ the enrollment polynomial of the k^{th} order, and $\upsilon_{icgs\tau}$ is the residual from the regression of CS on $\boldsymbol{X}$, the instrument indicators, along with segment fixed effects, ϕ_τ, and functions of enrollment, q_τ^k. It captures the remaining factors that are correlated with enrollment that are very likely also correlated with the test achievements of pupils. By allowing the enrollment polynomials to vary by segment, I follow Fredriksson et al. (2013) and effectively consider each threshold as a different experiment.

The model (2)-(3) is estimated using two stage least squares (2SLS), where the estimates of α_1 is of primary interest.

In subsection 4.2.2 I have already touched upon the identifying assumptions for causality of the IV estimand. Additional to the independence assumption (also known as the non-parametric version of the exclusion restriction), necessary assumptions include the stable unit treatment value assumption (SUTVA)

[16] Class characteristics are not included in all specifications.

[17] Each segment consists of grades with enrollments in an interval ± 14 pupils around threshold τ: $\varphi_\tau = \mathbf{1}(e_{gs} \in \bar{e}_\tau \pm 14)$, where $\bar{e}_\tau = \{56, 84, 112, 140, 168\}$. The first segment also include enrollments below 15 pupils: $\varphi_{28} = \mathbf{1}(e_{gs} \leq 42)$.

[18] Note that based on the pattern of Figure 2, it is questionable whether *above* indicators of the fourth, fifth and sixth cutoff points are valid instruments. This applies for f_{gs} as well and motivates a thorough analysis of subsamples.

and monotonicity. SUTVA is a common although, not a trivial assumption in the literature on causality. Monotonicity in this setting requires that for each child attending a school with a grade enrollment count above the administrative thresholds, the class size is *at most as large* as it would have been if the grade enrollment of the school was below the cutoff. As the independence assumption the monotonicity assumption is non-verifiable because of the counterfactual nature of the observations[19]. In both cases it applies that the stronger the instrument the less sensitive the IV estimand is to violations of the assumption (Angrist and Imbens 1995).

I refer to the parameter α_1 as the average causal response (also known as the local average treatment effect (LATE) for binary treatments). It captures a weighted average treatment effect to a unit change in class size for the unidentified subpopulation of pupils whose treatment status is affected by the instrument (see Angrist and Imbens (1995) for theoretical derivation). The weight attached to the average effect of a unit change in treatment is proportional to the number of pupils who, because of the administrative rules, are induced to attend a class with n or fewer pupils instead of one with more than n. These pupils are commonly known as the group of compliers. The group of compliers need not be representative of the entire population of pupils nor is it possible to identify because one naturally only observes a single counterfactual treatment status. Furthermore, membership of this group varies with the choice of instrument (Wooldridge 2002).

6 Empirical Results

This section quantifies the effect of class size on math and reading scores using the empirical approach outlined in Section 5. All reported standard errors are clustered to account for group structures within grade enrollment of the residuals.[20] Because of the fuzzy RD design and the discreteness of the assignment variable, the conventional IV sampling errors ignore this structure and may overstate the precision of the IV estimators.

6.1 Specification Analysis

Table 4 shows the 2SLS estimates of the effect of class size on reading and math scores using a selection of enrollment controls. For the sake of clarity, the specification analysis is presented only with the pooled test score information for reading (the 2^{th}, 6^{th}, and 8^{th} grade) and math/physics (grades 3, 6, and 8), respectively. In addition, I provide information of the first-stage F-test of

[19]It does, however, have testable implications for multivalued treatment intensities. See Appendix B for an analysis.

[20]Clustering on the grade enrollment level in IV estimations is suggested by Lee and Card (2008) and performed in Fredriksson et al. (2013). This yields 146 clusters in the full estimation sample, a considerably higher level compared to clustering on school grade level where my instrument varies. Thus, standard errors are slightly larger but the difference is modest.

Model	(1)	(2)	(3)	(4)	(5)	(6)
	Reading test scores, grades 2, 6, and 8 ($N = 448,946$)					
Partial R^2	.1409	.1935	.0618	.0742	.2017	.1879
F-test for instruments	10.01	21.82	16.45	15.84	20.29	31.72
Class size	$-.0045^{*}$	$-.0055^{***}$	$-.0070^{*}$	$-.0043$	$-.0052^{***}$	$-.0052^{***}$
	(.0024)	(.0020)	(.0039)	(.0031)	(.0018)	(.0019)
	Math/physics test scores, grades 3, 6, and 8 ($N = 449,325$)					
Partial R^2	.1438	.1955	.0640	.0725	.2031	.1910
F-test for instruments	10.14	19.00	13.89	12.47	17.74	26.75
Class size	$-.0072^{***}$	$-.0063^{***}$	$-.0060$	$-.0002$	$-.0059^{***}$	$-.0055^{***}$
	(.0027)	(.0021)	(.0040)	(.0034)	(.0019)	(.0020)
Enrollment controls:						
1^{st}-order polynomials	Yes		Yes		Yes	
2^{nd}-order polynomials		Yes		Yes		Yes
Interaction w/ threshold (combined)			Yes	Yes		
Interaction w/ thresholds (separately)					Yes	Yes
Class characteristics	No	No	No	No	No	No

Notes. The estimates are based on pupils enrolled in normal classes in the Danish public school system in the school years of 2009/2010 to 2011/2012 and tested through the national test system. All test score measures are standardized. Class size controls are contemporaneous. In addition to the control variables listed in the table, all specifications include fixed effects for enrollment segment and the remaining controls from Table A.1. Standard errors adjusted for clustering by enrollment count are in parentheses. Asterisks indicate statistical significance at the ***1%, **5%, and *10% level, respectively.

Table 4: INSTRUMENTAL VARIABLES ESTIMATES OF THE EFFECT OF CONTEMPORANEOUS CLASS SIZE, DIFFERENT ENROLLMENT CONTROLS

(3). Table 4 includes estimates of six enrollment specifications. Columns (1) and (2) restrict the enrollment polynomials to be the same across segments, while columns (3) and (4) allow the polynomials to differ above and below all thresholds combined. The remaining columns (5) and (6) are the fully flexible specifications. Intuitively, I favor these specifications because they account for the difference in the slopes of the predicted class size function across each threshold, separately.

Overall the results of Table 4 suggests quite stable effects, only the specifications with enrollment polynomials interacted with combined thresholds seems out of line. Also, the partial R^2 drops considerably in these specifications. The remaining IV estimates on math scores suggest a class size effect ranging between -0.006 and -0.007 of a standard deviation and between -0.005 and -0.007 of a standard deviation for reading scores. All effects are significant on the 10 percent level.

Following Staiger and Stock (1997), I can clearly reject that the instruments do not enter the first stage regression, however only marginally for the specification of column (1). Moreover, the fraction of the variation in the endogenous class size that is explained only by the instruments, represented by the partial

Model	(1)	(2)	(3)	(4)	(5)	(6)
	2^{th} grade reading scores			3^{rd} grade math scores		
Class size	$-.0041$	$-.0094^{***}$	$-.0033^{***}$	$-.0032$	$-.0047$	$-.0014$
	(.0036)	(.0029)	(.0008)	(.0037)	(.0031)	(.0009)
No. of observations		152,095			153,368	
	6^{th} grade reading scores			6^{th} grade math scores		
Class size	$-.0010$	$-.0046^{*}$	$-.0024^{***}$	$-.0054$	$-.0090^{***}$	$-.0090^{***}$
	(.0035)	(.0026)	(.0009)	(.0039)	(.0031)	(.0027)
No. of observations		154,464			154,510	
	8^{th} grade reading scores			8^{th} grade physics scores		
Class size	.0062	.0028	0.0015	.0047	.0009	$-.0011$
	(.0050)	(.0041)	(.0026)	(.0042)	(.0039)	(.0036)
No. of observations		142,387			141,447	
Baseline covariates	No	Yes	Yes	No	Yes	Yes
Class characteristics	No	No	Yes	No	No	Yes

Notes. The estimates are based on pupils enrolled in regular classes in the Danish public school system in the school years of 2009/2010 to 2011/2012 and tested via the national test system. All test score measures are standardized. Class size controls are contemporaneous. In addition to the control variables listed in the table, all specifications include fixed effects for enrollment segments and linear and squared controls for grade enrollment into schools interacted with separate thresholds. Baseline covariates include the remaining controls from Table A.1. Standard errors adjusted for clustering by enrollment count are in parentheses. Asterisks indicate statistical significance at the ***1%, **5%, and *10% level, respectively.

Table 5: Estimates of class size effects in the second−eighth grade.

R^2, is reasonable in the preferred specifications ranging from 18% to 20%. Thus, it seems that the enrollment specifications displayed in column (2), (5) and (6) are all quite appropriate. I favor the fully flexible specification in column (6) for later analyses, which also seem to yield slightly more conservative estimates.[21]

6.2 Main Results

Table 5 presents IV estimates of the impact of class size on pupil academic achievement in grades 2 through 8 using the fully flexible enrollment specification from column (6) of Table 4. Compared to this, the 2SLS regressions in column (3) and (6) include class controls, while specifications in column (1) and (4) only include segment fixed effects and linear and square enrollment controls interacted with segments and thresholds.[22]

The point estimates of Table 5 are remarkably similar across columns. If the maximum class size rule in fact do produce experimental variation in class size, the 2SLS estimates of class size effects should be robust to the inclusion

[21] On the grade level, the specifications in columns (2),(5), and (6) of Table 4 practically only differ by a significant negative class size effect on 3^{rd} grade math scores in specification (2) and (5).

[22] F-statistics for instruments are omitted from the table. They vary in between 24 and 35, thus, the presence of weak instruments are clearly rejected.

of control variables, they should only improve the precision of the estimates. Only the class size effects on physics scores in the 8^{th} grade change signs while remaining insignificant. The maximum class size rule is likely to have less bite in the 8^{th} grade because an accumulating fraction of pupils have transferred to private school while the mobility of pupils drops as graduation approaches. All standard errors are reduced as the level of controls increases confirming that the precision of estimates are enhanced. Also, controlling for the characteristics of classmates seems to slightly decrease the class size effect for reading.

Moreover, Table 5 reveals a noteworthy consistency in the main estimates across the various outcomes. Although modest, all coefficients of column (2) and (5) are negative apart from the 2SLS estimates on 8^{th} grade physics and reading scores. Placement in a small class seem to improve test scores in primary school. Compared to this, results of the OLS specification (Table A.4 of the Appendix) suggest an upward bias of the OLS estimates. This indicates a compensatory allocation of class size. I.e. children with poorer skills are typically placed in smaller classes and vice versa. The OLS estimates of class size on reading scores vary between significant $-.0012$ and significant .0075 of a standard deviation (on 2^{nd} and 8^{th} grade reading scores, respectively). Only the 3^{rd} grade point estimate is significantly negative where the corresponding 2SLS estimate remains insignificant.

Column (2) and (5) of Table 5 indicate a more adverse effect of contemporaneous class size on reading scores in the lower primary school of 0.9 percent of a standard deviation compared to half of that in the upper primary. On the other hand, the class size effects on math scores appear larger in the upper primary school. The 2SLS estimates of class size effects on 3^{rd} grade math scores are insignificantly negative. The corresponding effects on 6^{th} grade math scores are twice as large and suggest that a one-pupil increase in class size would reduce math scores with 0.9 percent of a standard deviation. To examine the statistical significance of these differentials, Table 6 presents results of the preferred specification where class size is interacted with grade levels. The main effects thus pertain to pupils in the upper primary school (6^{th} grade). Specifically, grade indicators are interacted with class size and instrument as well as enrollment control functions and segment in order to allow enrollment polynomials to differ across grade levels. Thus, adequately controlling for enrollment effects on the academic achievement of pupils.

In all specifications the main effect is significantly negative on at least the 10% significance level. While the differences between lower and upper primary school are all statistically insignificant, in all specifications except column (4) results suggest that 8^{th} grade test performance is significantly less adverse affected by a class size increment, more or less offsetting the negative main effect.

While I find significant, albeit modest, evidence that reducing class size improves the academic achievements in subjects such as reading and math/physics of pupils in the Danish public school system, I cannot reject that these effects are constant across primary school. Furthermore, the estimates suggest that class size effects of 8^{th} graders are negligible. Results are robust to both the exclusion of class sizes above 28 and schools that primarily act as lower secondary

Independent variable	(1) 2SLS, reading scores	(2)	(3) 2SLS, math scores	(4)
Interaction	−.0048	−.0019	.0038	.0045
(lower primary school)	(.0038)	(.0013)	(.0045)	(.0030)
Main effect	−.0044*	−.0020**	−.0086***	−.0075***
	(.0026)	(.0008)	(.0032)	(.0025)
Interaction	.0078*	.0040*	.0099**	.0064
(lower secondary school)	(.0044)	(.0021)	(.0045)	(.0040)
No. of observations	448,946		449,325	
Class characteristics	No	Yes	No	Yes

Notes. The estimates are based on pupils enrolled in regular classes in the Danish public school system in the school years of 2009/2010 to 2011/2012 and tested via the national test system. All test score measures are standardized. Class size controls are contemporaneous. The lower primary school interaction term pertains to the 2^{nd} grade in column (1)-(2) and the 3^{rd} grade in column (3)-(4). The main effect pertains to the 6^{th} grade while the lower secondary school interaction term denotes the 8^{th} grade. In addition to the control variables listed in the table, all specifications include fixed effects for enrollment segments interacted with school level and linear and squared controls for grade enrollment into schools interacted with both separate thresholds and school level. Baseline covariates include the remaining controls from Table A.1. The instruments of column $(1)-(4)$ are interacted with school levels as well. Standard errors adjusted for clustering by enrollment count are in parentheses. Asterisks indicate statistical significance at the ***1%, **5%, and *10% level, respectively.

TABLE 6: ESTIMATES OF CLASS SIZE EFFECTS IN THE SECOND−EIGHTH GRADE, INTERACTION SPECIFICATIONS

schools.

6.2.1 Reducing the Bandwidth

What happens if the choice of bandwidth is narrowed and the model specification focuses explicitly on the variation around the cutoffs? As the instruments are valid only when the maximum class size rule triggers a change in class size, I would only expect cross sectional IV estimates to be unbiased in a small interval around the cutoffs.[23] Table 7 presents 2SLS estimation results of the ± 4 discontinuity the three lower segments (see footnote 14). Intuitively, when the sample is adequately narrow around the cutoffs, segment fixed effects should be sufficient controls of enrollment.

The interpretation of the estimates in general and the discontinuity estimates in particular should be carefully conducted. Restricting the discontinuity sample to the three lower segments of enrollment obviously causes the corresponding estimates to be driven by the smaller schools. But even if higher enrollment segments were included in the analysis the weights of the 2SLS estimate on the weighted average causal effect is still larger for observations where the instrument affects the endogenous class size more (Angrist et al. 1996).

The evidence of Table 7 still suggests adverse effects of increasing class size.

[23] Furthermore, the likelihood of parents being able to undo the random assignment of treatment intensity are smaller just around the thresholds (Lee and Lemieux 2010).

Model	(1)	(2)	(4)	(5)
2SLS regression	**±4 discontinuity sample**			
	2^{th} grade reading scores		3^{rd} grade math scores	
Class size	−.0190**	−.0060***	.0038	.0040
	(.0081)	(.0019)	(.0196)	(.0058)
No. of observations	31,495		33,554	
	6^{th} grade reading scores		6^{th} grade math scores	
Class size	−.0216***	−.0090***	−.0258***	−.0167**
	(.0082)	(.0031)	(.0093)	(.0082)
No. of observations	31,977		32,006	
	8^{th} grade reading scores		8^{th} grade physics scores	
Class size	.0065	.0050	−.0100	−.0126
	(.0121)	(.0058)	(.0107)	(.0109)
No. of observations	30,186		29,864	
Class characteristics	No	Yes	No	Yes

Notes. The estimates are based on pupils enrolled in regular classes in the Danish public school system in the school years of 2009/2010 to 2011/2012, that have been tested via the national test system, and, for the discontinuity samples, are enrolled in schools with a grade count of ±4 around the three lower cutoffs: 28, 56, and 84. Test scores are standardized. Class size controls are contemporaneous. In addition to the control variables listed in the table, all specifications include fixed effects for enrollment segments and the remaining controls from Table A.1. Standard errors adjusted for clustering by enrollment count are in parentheses. Asterisks indicate statistical significance at the ***1%, **5%, and *10% level, respectively.

TABLE 7: ESTIMATES OF CLASS SIZE EFFECTS IN THE SECOND−EIGHTH GRADE, DISCONTINUITY SAMPLE

In particular, discontinuity estimates imply that the 'true' effect of class size on the group of compliers may be considerably larger compared to the 2SLS estimates of the full sample. Coefficients of class size are as large as .02 of a standard deviation on primary school reading scores, as well as .03 of a standard deviation on 6^{th} grade math scores. Once again the inclusion of class characteristics considerably reduces the effect of a class size increment on primary level reading scores. Consequently, full-sample estimates are likely conservative estimates of the class size effect.

An interaction analysis analogous to the one from Table 6 reveals significant 6^{th} grade class size effects once again. However, due to imprecisely estimated coefficients, none of the class size effects in the lower primary and secondary school of the ±4 discontinuity sample are significantly different from this main effect.

6.2.2 Robustness Tests

Is it probable that only contemporaneous class size inputs affect the test achievements of pupils? In general the answer is no. On the other hand the class sizes of pupils are highly correlated with previous class size observations, thus including for example class size lagged one year is likely to cause problems with

	(1)	(2)	(3)	(4)
Independent variable	**2SLS regression**			
	2^{th} grade reading scores		3^{rd} grade math scores	
Class size (2-year avg.)	$-.0123^{***}$	$-.0041^{***}$	$-.0076^{**}$	$-.0022^{**}$
	(.0035)	(.0010)	(.0036)	(.0009)
No. of observations	143,889		147,515	
	6^{th} grade reading scores		6^{th} grade math scores	
Class size (2-year avg.)	$-.0060^{*}$	$-.0028^{**}$	$-.0121^{***}$	$-.0117^{***}$
	(.0031)	(.0011)	(.0036)	(.0030)
No. of observations	150,232		150,248	
	8^{th} grade reading scores		8^{th} grade physics scores	
Class size (2-year avg.)	.0007	$-.0015$	$-.0006$	$-.0033$
	(.0072)	(.0037)	(.0069)	(.0062)
No. of observations	138,699		137,781	
Class characteristics	No	Yes	No	Yes

Notes. The estimates are based on pupils enrolled in regular classes in the Danish public school system in the school years of 2009/2010 to 2011/2012 and tested via the national test system. All test score measures are standardized. Class size controls are three-year averages. In addition to the control variables listed in the table, all specifications include fixed effects for twice lagged enrollment segments and linear and squared controls for twice lagged grade enrollment into schools interacted with separate thresholds as well as the remaining controls from Table A.1. Standard errors adjusted for clustering by twice lagged enrollment count are in parentheses. Asterisks indicate statistical significance at the ***1%, **5%, and *10% level, respectively.

Table 8: Estimates of class size effects, three-year average class size measure

multicollinearity in the data[24]. Rather, an average class size measure, corresponding to the one of Fredriksson et al. (2013), is sensible to employ. Class information are available from the school year of 2007/2008, thus, providing me with the opportunity of incorporating class sizes two years previous to the test observation for almost every pupil in the sample.

Results of Table 8 are obtained by using the average of the pupils' class sizes over the course of three years as the endogenous class size measure. For example, for pupils tested in literacy in the 2^{nd} grade the average class size that the pupil experienced during kindergarten, 1^{st} and 2^{nd} grade is used. The interpretation of the 2SLS estimates is then the effect of a one-pupil increase during these past three years. The instrument is based on enrollment two the school years prior to that of the test for the simple reason that enrollments in the test year are potentially endogenous to class sizes in the previous year etc. Hence, I cannot validly treat enrollment in the test year as exogenous (Fredriksson 2013). Also, the included enrollment controls are also based on two school years prior to the test observation.

Compared to Table 5 the inclusion of earlier class sizes does not alter the results on reading scores as well as 6^{th} grade math scores, while class size effects on 3^{rd} grade math scores becomes significantly negatively affected by a change

[24]The correlation between the two class size measures are as high as .73 for pupils in the 2^{nd} grade.

in average class size. In general, the results of Table 8 are of slightly greater magnitude compared to those of the contemporaneous class size measure. Overall, historical class size inputs are undoubtedly of importance when evaluating the effect of such inputs and contributing all these effects to contemporaneous class size is error-prone. Empirically, however, the evidence of the Danish school system suggests only modest changes in the results indicating that the coefficients to contemporaneous class size may be a fair proxy of the effect of (at least three-year) average class size. Similar results are found when treating the class size two school years prior to that of the test as the endogenous variable to be studied.

As a final robustness test, I use a bandwidth of ± 14 pupils around each threshold to estimate the discontinuities in enrollment into 1^{st} grade *at* the two lower cutoffs on the basis of three different maximum class size rules (24, 26, and 28). I employ this strategy to each municipality separately. This exercise should provide evidence of municipalities that are certainly abiding by the 28 pupil rule and not a lower self-imposed maximum, thus, strengthening the validity of the instruments. That is, municipalities for which the discontinuity in the 1^{st} grade enrollment around the first or second threshold is significant on the 5 percent level *and* is numerically larger when thresholds are based on the 28 pupil rule compared to 24 or 26. Table 9 presents the results of employing the preferred enrollment specification on children attending schools in the 31 municipalities that in all likelihood abide only by the government imposed 28 pupil rule. Although reducing the sample sizes to around a third of the full samples, these results are in line with earlier findings presented in this paper; 8^{th} grade ability measures are not significantly affected by a class size reduction, whereas the literacy of primary schoolers and 6^{th} grade math abilities are significantly improved. Again, the inclusion of classmates' characteristics reduces the class size effect on primary level reading outcomes.

6.2.3 Heterogeneity

To examine whether class size effects are heterogeneous across identifiable subpopulations, I present results where class size is interacted with gender, parental income/education and immigrant status. Here, I interact, for example gender, with the treatment intensity and the instrument as well as the enrollment control functions and enrollment segment. Once again the latter ensures that as many effects of enrollment on test achievements as possible are excluded from the first stage estimations.

The results of Table 10 reveal little evidence of systematic effects of class size across pupil characteristics. In the first column there is a significant negative effect of class size in IV regression for boys only on 6^{th} grade math scores while coefficients of the female interaction term are almost all negative. Particularly, girls' 3^{rd} grade math scores and 8^{th} grade literacy levels may benefit more from a small class size compared to boys. The negative interaction effect could be an implication of the somewhat ancient 'nice girl syndrome'; quiet girls may be

	(1)	(2)	(3)	(4)
Sample	**Municipalities abiding by the 28 pupil rule**			
	2^{th} grade reading scores		3^{rd} grade math scores	
Class size	$-.0071^{**}$	$-.0022^{**}$	$-.0019$	$-.0008$
	(.0035)	(.0011)	(.0046)	(.0013)
No. of observations	50,932		52,281	
	6^{th} grade reading scores		6^{th} grade math scores	
Class size	$-.0073^{*}$	$-.0037^{**}$	$-.0088^{**}$	$-.0075^{**}$
	(.0037)	(.0017)	(.0038)	(.0033)
No. of observations	53,091		52,806	
	8^{th} grade reading scores		8^{th} grade physics scores	
Class size	$-.0005$	.0002	.0015	.0001
	(.0048)	(.0025)	(.0055)	(.0054)
No. of observations	48,495		48,186	
Class characteristics	No	Yes	No	Yes

Notes. The estimates are based on pupils enrolled in regular classes in Danish public schools located in municipalities abiding by a maximum class size rule of 28 pupils in the school years of 2009/2010 to 2011/2012 and tested via the national test system. All test score measures are standardized. Class size controls are contemporaneous. In addition to the control variables listed in the table, all specifications include fixed effects for enrollment segments and linear and squared controls for grade enrollment into schools interacted with separate thresholds as well as the remaining controls from Table A.1. Standard errors adjusted for clustering by enrollment count are in parentheses. Asterisks indicate statistical significance at the ***1%, **5%, and *10% level, respectively.

TABLE 9: ESTIMATES OF CLASS SIZE EFFECTS, MUNICIPALITIES ABIDING BY THE 28 PUPIL RULE

considered more well behaved, but in a large class they may just disappear in the crowd and correspondingly their needs in terms of learning abilities may be neglected.

The main effect in the third column pertains to ethnic Danes and immigrants from other western countries. The class size effect of the nonwestern-immigrants-or-descendants-hereof-interaction term in the fourth column indicates that pupils from other cultural backgrounds in the Danish public school system do not seem to suffer particularly under large class sizes. None of the interaction coefficients is significant and moreover it is of varying sign. It is a common concern that children from disadvantaged backgrounds are more adversely affected by decreasing school resources. However, it seems that schools (and teachers in particular) are observant of these children when the class size is large, preventing them from falling further behind. Correspondingly, I find on empirical evidence that children from low earnings families[25] are more adversely affected by a larger class size in Denmark. In fact, for 8^{th} grade reading ability the opposite holds. A similar pattern emerges when interacting class size with the education level of the parents instead (omitted from Table 10). Interestingly, for children from the high end of the parents' earnings distribution a class size reduction appears to be more beneficial in terms of 6^{th} grade math

[25]The 'highest earnings' variable is defined as the highest earnings of the pupil's mother and the father. If the parents are divorced the income of the mother is used.

achievements compared to pupils from the two middle quartiles. This suggests that high-earning parents do not compensate their children in 6^{th} grade math corresponding to class size. Rather the opposite is true. Perhaps socioeconomically advantaged parents are more inclined to accept that a larger class size means less time for meeting their pupils' needs (without providing extra tutoring at home). Likewise, they may be more likely to ensure that their child gets individual attention by teachers when class sizes are small.

The last three columns of Table 10 present the results for urbanization of the school municipality interactions where the main effect pertains to country and small city (below $10,000$ citizens) municipalities, 8^{th} graders' reading skills in the capital city of Copenhagen is significantly (marginally) negatively affected by a larger class size compared to8^{th} graders from country or small city municipalities. 8^{th} graders from large city municipalities, on the other hand, are more positively affected by a larger class. This suggests that lower secondary teachers working the capital city may be better at utilizing small classes when it comes to reading (or worse at teaching large classes). Also, lower secondary teachers in the other large cities may be additionally motivated by the 'teaching hurdle' when class sizes are larger.

6.3 Implications and Comparisons

I realize that the estimated class size effects presented here cannot be readily compared across grades except under additional assumptions. By only including contemporaneous class size in the specifications, this term is in effect capturing all contributions of previous class sizes to pupil performance (unless one is willing to assume that only contemporaneous class size matters to the production of current achievement or that the class size input is constant over time). For example, 8^{th} grade class size potentially captures the effect of 8 (9) years of class sizes while 3^{rd} grade class size incorporate much less. Furthermore, the material taught differs across grade levels, and many acquired abilities in the early grade may very well be complements for accumulating later skills. However, given the class sizes already encountered by pupils, the results of this paper provides insight into the immediate class size effects in the Danish public school system. Furthermore, results are fairly robust to the inclusion of previous class sizes in terms of both an average class size measure and previous class sizes alone.

Overall, the results presented here reveal significantly positive effects of class size reductions in the Danish public school system, specifically in the lower and upper primary school. Reading and physics abilities of pupils in the lower secondary school, however, seem largely unaffected by a change in class size. Results suggest that inclusion of class characteristics are relevant to obtain precise unbiased estimates of the class size effects using a quasi-experimental IV setting. Controlling for the characteristics of classmates reduces the magnitude of the primary level reading estimates, but unfortunately not many previous studies have been capable of including these and those that have been able consider long-term outcomes such as length of education etc.

The preliminary findings of this paper suggest that placement in a small

Table 10: Heterogeneous effects of class size

	Gender		Immigrant status		Highest earnings quartile			Degree of urbanization		
Outcome variable	Main	Interact. (*female*)	Main	Interact. (*nonwestern*)	Interact. $(1^{st}\,Q)$	Main	Interact. $(4^{th}\,Q)$	Main	Interact. (*large city*)	Interact (*capital*)
2^{nd} grade reading scores	−.0021	−.0024	−.0035***	.0044	.0005	−.0039***	.0020	−.0033***	.0009	.0018
$(N = 152,095)$	(.0018)	(.0031)	(.0009)	(.0064)	(.0031)	(.0014)	(.0030)	(.0012)	(.0017)	(.0029)
3^{rd} grade math scores	.0004	−.0037*	−.0013	.0004	.0039	−.0028	.0027	−.0021**	.0030*	−.0001
$(N = 153,368)$	(.0015)	(.0022)	(.0009)	(.0045)	(.0039)	(.0020)	(.0044)	(.0009)	(.0017)	(.0034)
6^{th} grade reading scores	−.0005	−.0038	−.0026***	.0020	.0036	−.0025	−.0043	−.0021*	.0005	−.0015
$(N = 154,464)$	(.0017)	(.0025)	(.0010)	(.0063)	(.0042)	(.0018)	(.0038)	(.0013)	(.0024)	(.0032)
6^{th} grade math scores	−.0095***	.0008	−.0090***	−.0008	.0035	−.0073**	−.0144***	−.0095**	.0025	−.0015
$(N = 154,510)$	(.0032)	(.0038)	(.0027)	(.0049)	(.0030)	(.0030)	(.0049)	(.0039)	(.0066)	(.0059)
8^{th} grade reading scores	.0060	−.0092**	.0010	.0030	.0102*	−.0024	−.0036	−.0005	.0104**	−.0072*
$(N = 142,387)$	(.0038)	(.0046)	(.0024)	(.0080)	(.0060)	(.0033)	(.0064)	(.0020)	(.0045)	(.0043)
8^{th} grade physics scores	.0020	−.0062	.0002	−.0097	−.00097	−.0003	−.0056	.0001	.0030	−.0042
$(N = 141,447)$	(.0053)	(.0051)	(.0039)	(.0077)	(.0060)	(.0042)	(.0056)	(.0045)	(.0059)	(.0107)

Notes. The estimates are based on pupils enrolled in regular classes in the Danish public school system in the school years of 2009/2010 to 2011/2012 and tested via the national test system. All test score measures are standardized. Class size controls are contemporaneous. All regression include fixed effects for enrollment segments, segment fixed effects interacted with interaction term and linear and squared controls for grade enrollment into schools interacted with both separate thresholds and interaction terms and the remaining controls from Table A.1 including class characteristics. The dummy instruments of the 2SLS regressions are interacted with interaction terms as well. Standard errors adjusted for clustering by enrollment count are in parentheses. Asterisks indicate statistical significance at the ***1%, **5%, and *10% level, respectively. See the text for further explanation.

class during upper primary school increases both math and reading scores. A one-pupil reduction in 6^{th} grade class size is associated with an increase in literacy of 0.2 percent of a standard deviation and an increase in math ability of 0.9 percent of a standard deviation. These effects do not seem to differ significantly across primary school. A class size reduction equivalent to that of the Project STAR (seven pupils on average) suggests that lower primary reading skills would improve by up to 0.02 of a standard deviation once class characteristics are controlled for. This short-run effect on test scores is notably smaller compared to the results of the STAR experiment. Krueger (1999) finds achievement gains in terms of SAT scores of 0.20 of a standard deviation from attending a small class in kindergarten and 0.19 from attending a small class in all four years (kindergarten through grade 3). Finn and Achilles (1999) find that small class assignment improves reading skills by 0.19 of a standard deviation for 2^{nd} graders. Neither of these studies include controls for the characteristics of classmates. Comparing class size estimates without class characteristics presented here, a seven pupil class size reduction would improve 2^{nd} graders reading skills by 0.07 of a standard deviation. Even applying the estimation results from the ± 4 discontinuity sample still reveals class size effects in the lower end of the scale for 2^{nd} graders (0.13 of a standard deviation).

Although not pertaining to the same grade levels, Fredriksson et al. (2013) recalculate their estimated class size effects on cognitive skills in the upper primary level of Swedish school in terms of a seven pupil reduction. This yields a class size effect of 0.23 of a standard deviation. Also, Angrist and Lavy (1999) compare their estimates to Project STAR finding the class size effects of 5^{th} graders on reading scores to be approximately 0.18 of a standard deviation[26] while the corresponding result for 4^{th} graders is considerably smaller and presented only in terms of the distribution of class means ($.13\sigma$). Based on the authors' recalculation for 5^{th} graders, I would estimate the 4^{th} grade effect for pupils to approximately 0.08 of a standard deviation. By comparison, the class size effect on 6^{th} graders math scores presented here is 0.06 of a standard deviation for a seven-pupil reduction in class size increasing to 12 percent of a standard deviation if results of the ± 4 discontinuity sample is used and class characteristics are included (18 percent when class controls are omitted). Acknowledging that the IV estimates of class size effects in the full sample may be driven by variation outside of the administrative thresholds, the ± 4 discontinuity coefficients are potentially closer to the 'true' class size effects. However, these are still in the lower end compared to the similar studies summarized above.

The Danish national test system is an adaptive test system on the contrary to test measures used in the cited studies. The advantage of adaptive tests are that they measure the true ability of the pupil within specific fields of the subjects regardless of the pupil's place in the skill distribution. This property may distort the test performance of particularly high and low ability pupils

[26]Note, that the authors use a class size reduction of eight pupils when comparing their results.

compared to common tests where results are measured as the simple fraction of correct answers. Furthermore the test system was introduced for evaluation purposes as a tool to help teachers identify the shortcomings of each pupil. As such the test environment is rather informal compared to many other tests, and the only preparation the pupils receive for the tests is a short introduction to the computer system carrying out the tests[27]. While the pupil's parents are informed of the test results, this is presented on a crude scale from 1 to 5. In short, the nature of the test system may cause some of the modest magnitudes of the results. Also, let us not forget that there is still significant evidence that class size has an adverse effect on test scores albeit smaller than comparable studies.

The heterogeneity analysis is a common extension of many studies of class size effects, but the results are vague. Fredriksson et al. (2013) find no significant gender differences in the effects of class size on cognitive and non-cognitive ability measures, while effects on non-cognitive abilities and the probability of earning a bachelor's degree are significantly larger for pupils from the upper quartile of parent income. Also, Hoxby (2000) finds significantly larger effects on pupils in high income schools. Heinesen and Browning (2007) note that the point estimates of the ± 4 discontinuity sample of Danish 8^{th} graders vary considerably across subgroups of pupils, although the only significant effect is on the probability of completing a secondary education on pupils where neither parent has a higher education. Nye et al. (2000) find that smaller class sizes benefit girls significantly more than boys only in the kindergarten, and they find no statistical difference for children of minority backgrounds. The results presented here indicate that girls, in some cases may benefit slightly more from a class size reduction compared to boys. Furthermore, children from backgrounds that are usually considered disadvantaged are in general not more adversely affected by a large class size. And finally certain pupils of resourceful parents or who are attending a school in the capital city may benefit more from being in a small class.

7 Conclusion

The main contribution of this paper is to extensively analyze the effects of class size *across* all school levels during compulsory school. Earlier studies have primarily been concerned with class size effects of close grade levels, and as such there is little evidence of how the effects of class sizes behaves across school grades.

The estimated effects reveal significantly negative (albeit modest) impacts of larger class sizes in the Danish public school system in the school years of 2009/2010−2011/2012, particularly in the lower and upper primary school. Furthermore, the results suggest that inclusion of class characteristics are relevant to obtain precise estimates of the class size effects, though it implies a reduction

[27] How to answer questions, how to get to the next question etc.

in the magnitude of the estimated class size effects on literacy. Unfortunately, not many previous studies have been capable of including these.

To gain insight into the effect of class size on pupil academic achievement I employ a commonly known instrumental variables approach, that exploits a quasi-experimental nature to estimate causal effects under weak non-parametric assumptions. While the nonparametric assumptions required for causal interpretation of the IV estimands are rather weak, the testing of auxiliary hypothesis cannot completely reject that these are not met - at least for the full estimation sample. Samples only including pupils attending a school of which the grade enrollment is close to exogenous class size thresholds imposed by administrative rules seem more appropriate given the assumptions. Also, these estimates are generally of greater magnitude, thus, findings based on samples, in which one is unable to obtain an adequately small sample around the enrollment thresholds, may be error-prone.

For primary school, all effects of a class size increment on literacy are significantly negative and in the range of $0.2 - 0.9$ percent of a standard deviation, furthermore, I am able to reject that the results do not differ for 8^{th} graders. The corresponding estimated effect when narrowing the bandwidths is between 0.6 and 2 percent of a standard deviation when class characteristics are controlled for. The effect of a class size increment on 6^{th} grade math abilities are in the upper end of these scales. Moreover, results suggest that primary school class size effects on math scores are of equal size, while an increment in 8^{th} grade class size has a significantly less adverse impact on physics scores.

The findings of this paper suggest a beneficial impact from reducing class size on pupils in primary school, particularly for 6^{th} graders, whereas no significant effect are found for the older, lower secondary pupils. However, the results on younger pupils are modest when compared to similar studies from other countries as well as in absolute values. As such, other initiatives, for example introducing a second teacher or pedagogue to certain lessons or increasing the number of teaching hours in key subjects, may be more cost-effective compared to simple class size reductions.

References

Angrist, J. and G. Imbens (1995), "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity", *Journal of the American Statistical Association*, Vol. 90, No. 430 (June), pp. 431 − 442

Angrist, J., G. Imbens, and D. Rubin (1996), "Identification of Causal Effects Using Instrumental Variables", *Journal of the American Statistical Association,* Vol. 91, No. 434 (June), pp. 444 − 455

Angrist, J. and V. Lavy (1999), "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", *The Quarterly Journal of Economics,* Vol. 114, No. 2, pp. 533 − 575

Beuchert-Pedersen and Nandrup (2014), "The National Tests - A Practical Guide", unpublished manuscript.

Bingley, P, V. Jensen and I. Walker (2007), "The Effect of School Class Size on Length of Post-Compulsory Education: Some Cost-Benefit Analyses", *IZA working paper series*, No. 1603

Black, S., P. Devereaux and K. Salvanes (2007), "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes", *The Quarterly Journal of Economics*, Vol. 122, No. 1 (February), pp. 409 − 439

Blatchford, P., H. Goldstein, C. Martin, and W. Browne (2002), "A Study of Class-Size Effects in the English School Reception Year Classes", *British Educational Research Journal*, Vol. 28, No. 2 (April), pp. 169 − 185

Burke, M. and T. Sass, "Classroom Peer Effects and Student Achievement", *Journal of Labour Economics*, Vol. 31, No. 1, pp. 51 − 82

Carneiro, P. and J. Heckman (2003), " Human Capital Policy", *Inequality in America*, The MIT Press, 1^{st} edition

Clark, C., L. Woodward, J. Horwood, and S. Moor (2008), "Development of Emotional and Behavioral Regulation in Children Born Extremely Preterm and Very Preterm: Biological and Social Influences", *Child Development*, Vol. 79, No. 5, pp. 1444 − 1462

Cunha, F. and J. Heckman (2007), "The Technology of Skill Formation", *The American Economic Review*, Vol. 97, No. 2 (May), pp. 31 − 47

Datar, A. and B. Mason (2008), "Do Reductions in Class Size 'Crowd Out' Parental Investment in Education?", *Economics of Education Review*, 27, pp. 712 − 723

Finn, J. and C. Achilles (1999), "Tennessee's Class Size Study: Findings, Implications, Misconceptions", *Educational Evaluation and Policy Analysis*, Vol. 21, No. 2 (Summer), Special Issue: "Class Size: Issues and New Findings", pp. 97 − 109

Fredriksson, P., B. Öckert and H. Oosterbeek (2013), " Long-term Effects of Class Size", *The Quarterly Journal of Economics*, Vol. 128, No. 1, pp. 249 − 250

Hanushek, E. (1996), "Measuring Investment in Education", *The Journal of Economic Perspectives,* Vol. 10, No. 4, pp. 9 − 30

Heinesen, E. (2010), "Estimating Class-Size Effects Using Within-School Variation in Subject-Specific Classes", *The Economic Journal*, No. 120 (June), pp. 737 − 760

Heinesen, E. and M. Browning (2007), "Class Size, Teacher Hours and Educational Attainment", *Scandinavian Journal of Economics*, Vol. 109, No. 2, pp. 415 − 438

Hoxby, C. (2000), "The Effects of Class Size on Student achievement: New Evidence from Population Variation", *The Quarterly Journal of Economics*, Vol. 115, No. 4, pp. 1239−1285

Krueger, A. (1999), "Experimental Estimates of Education Production Functions", *Quarterly Journal of Economics,* Vol. 114, No. 2 (May), pp. 497 – 532

Lazear, E. (2001), "Educational Production", *The Quarterly Journal of Economics*, Vol. 116, No. 3, pp. 557 – 586

Lee, D. and D. Card (2008), "Regression Discontinuity Inference with Specification Error", *Journal of Econometrics*, Vol. 142, Issue 2, pp. 655 – 674

Lee, D. and T. Lemieux (2010), "Regression Discontinuity Designs in Economics", *Journal of Economic Literature*, Vol. 48, No. 2 (June), pp. 281 – 355

McCrary, J. (2008), "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test", *Journal of Econometrics*, Vol. 142, pp. 698 – 714

Nikolajsen, L. and M. Molsgaard (2012), "Notat vedr. elevtal", UNI-C. Available at http://cms.uvm.dk/sitecore/shell/Controls/Rich%20Text%20Editor/~/media/UVM/Filer/Stat/PDF12/121122%20Elevtalsnotat%202011_12.ash (June 23, 2013)

Nye, B., L. Hedges and S. Konstantopoulos (2000), "Do Disadvantaged Children Benefit More from Small Classes? Evidence from the Tennessee Class Size Experiment", *American Journal of Education*, Vol. 109, No. 1, pp. 1 – 26

Staiger, D. and J. Stock (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica,* Vol. 65, No. 3, pp.. 557 – 586

Todd, P. and K. Wolpin (2003), "On the Specification and Estimation of the Production Function for Cognitive Achievement", *The Economic Journal*, Vol. 113, No. 485, pp. $F3 - F33$

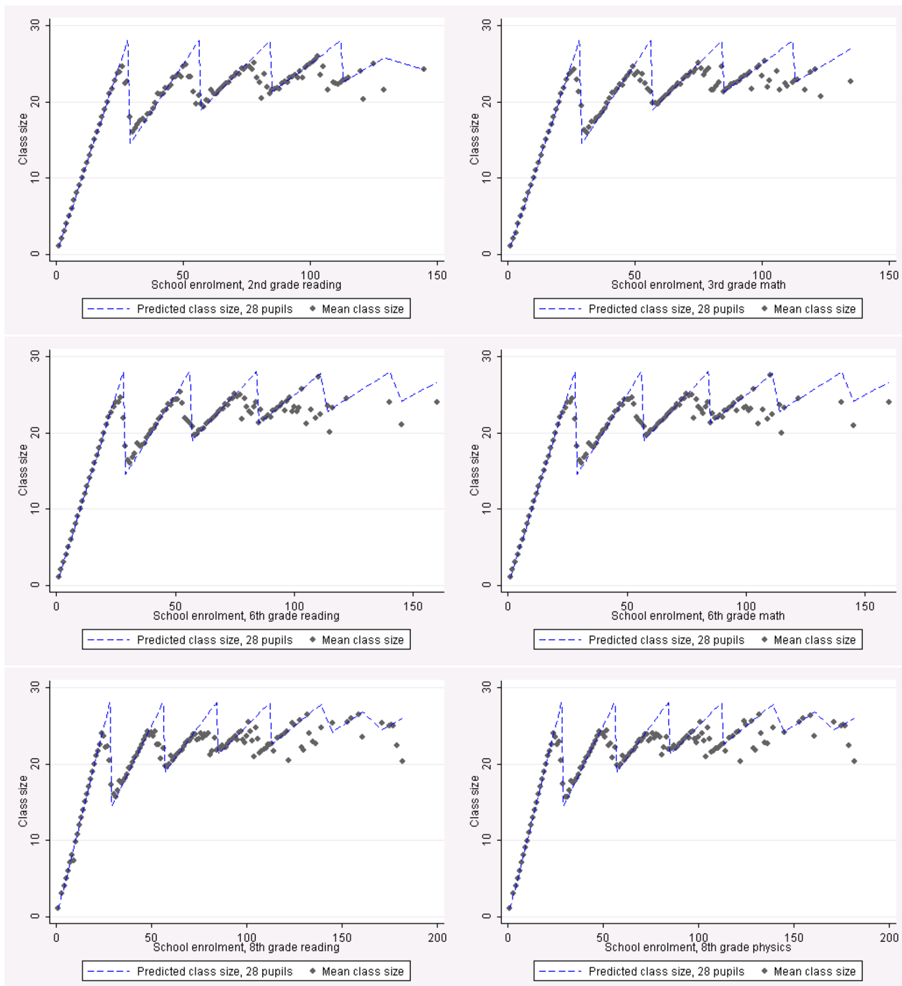
Undervisningsministeriet (2010), "Fælles mål 2009 - Elevernes Alsidige Udvikling", *Undervisningsministeriets håndbogsserie*, nr. 4 2010

Urquiola, M. and E. Verhoogen (2009), "Class-Size Caps, Sorting, and the Regression-Discontinuity Design", *The American Economic Review*, Vol. 99, No. 1 (March), pp. 179 – 215

Wandall, J. (2010), "National Test in Denmark, CAT as a Pedagogic Tool", *Journal of Applied Testing Technology*, Vol. 12, Article 3

Wooldridge, J. (2002), "Econometric Analysis of Cross Section and Panel Data", The MIT Press, 1^{st} edition

A Appendix: Full Descriptives and Results



The figure shows predicted and mean observed class size for pupils in grades 2, 3, 6, and 8, separately, in Danish public schools for the school years of 2009/2010-2011/2012. The predicted class size function indicated by the blue line is based on administrative rules ensuring a maximum of 28 pupil per class. The top right and left graphs illustrates predicted and mean observed class size by enrollment for pupils tested in 2^{nd} grade reading and 3^{rd} grade math, respectively. The middle left and right panel depict the corresponding figures for 6^{th} grade math and reading samples while the bottom panels illustrates mean and predicted class size by enrollment for 8^{th} graders' reading (left) and physics (right) test samples.

Figure A.1: Predicted and mean observed class size by enrollment, separate grade levels

Sample Variable		Full sample of pupils Mean	 Std. dev.
Outcome	Average standardized θ	0.017	0.981
Instruments	Above 28	0.199	
	Above 56	0.204	
	Above 84	0.059	
	$Above 112$	0.006	
	$Above 140$	0.001	
Controls	Class size	21.450	3.880
	Enrollment	53.557	22.518
Family information at age 6	*Mother's education*: None/missing	0.045	
	− Basic	0.258	
	− Vocational	0.367	
	− Higher	0.330	
	Father's Education: None/missing	0.076	
	− Basic	0.237	
	− Vocational	0.410	
	− Higher	0.277	
	Other		
	·Mother's log-earnings	9.844	4.737
	Mother's age	34.788	7.391
	Father's log-earnings	10.190	4.900
	Father's age	36.184	10.436
	Family with a single mother	0.155	
	Number of siblings	1.261	0.867
Pupil controls	Girl	0.492	
	Western immigrant (or descendant hereof)	0.019	
	Nonwestern immigrant (or descendant hereof)	0.100	
	Birth weight (g)	3298.145	1007.525
	Length of gestation (days)	199.722	125.989
	Born in the first quarter of the year	0.242	
	Born in the second quarter of the year	0.252	
	Born in the third quarter of the year	0.262	
	Born in the last quarter of the year	0.235	
	Firstborn	0.420	
	Second born	0.371	
	Third born or later	0.197	
	Multiple born	0.038	
	Age dummies (omitted here)	−	
School controls	Municipality in the capital area	0.307	
	Municipality with a large city	0.352	
	Municipality with smaller cities	0.264	
	Municipality on the countryside	0.052	
Class controls	Class average θ	0.019	0.410
	Class average mother's education: None/missing	0.044	0.055
	− Basic	0.250	0.130
	− Vocational	0.355	0.143
	− Higher	0.320	0.170
	Class average father's education: None/missing	0.074	0.068
	− Basic	0.229	0.115
	− Vocational	0.397	0.115
	− Higher	0.269	0.165
	Class average mother's log-earnings	9.541	1.810
	Class average father's log-earnings	9.877	1.706
	Class average fraction of girls	0.476	0.110
	Class average fraction of nonwestern immigrants	0.101	0.151
No. of observations		898, 271	

TABLE A.1: SAMPLE MEANS OF THE FULL ESTIMATION SAMPLE AND THE WITHIN SAMPLE

Baseline covariate	(1) *p*-value, *above* cutoff **Full estimation sample**	(2) *p*-value, *above* cutoff ± 4 **discontinuity sample**
First-born	.222	.174
Second-born	.258	.778
Born third or later	.079	.535
Multiple born	.704	.519
Born in the first quarter	.413	.677
Born in the second quarter	.531	.340
Born in the third quarter	.996	.414
Born in the fourth quarter	.793	.522
Birth weight	.097	.359
Length of gestation	.071	.037
Mother has no education	.008	.427
Basic education of the mother	.002	.818
Father has no education	.006	.261
Basic education of the father	.002	.213
Western immigrant	.069	.283
No. of observations	$898,271$	$178,082$

Notes. The above cutoff indicator equals 1 if the school enrollment at grade level exceeds a thresholds created by the 28 pupil rule up to $+14$ pupils ($+4$ pupils in the discontinuity sample). Columns report the p-values for t-tests of the significance of the pooled class-size instrument by separate OLS regressions on the variables listed in each row. The following controls are also included in the regressions: Year and enrollment segment fixed effects, indicator variables of degree of urbanization of the school municipality and linear and square controls for grade enrollment interacted with separate thresholds (only for the full estimation sample).

Table A.2: Balancing of remaining covariates

Variable	Mean	S.d.	Min	Max	Mean	S.d.	Min	Max
	±4 pupil discontinuity sample							
	Reading,* 2nd *grade				***Math,* 3rd *grade***			
	31,495 *obs.,* 1,682 *classes*				33,554 *obs.,* 1,794 *classes*			
Enrollment	49.45	19.37	25	88	49.55	19.30	25	88
Class size	20.60	4.55	6	32	20.85	5.02	1	54
Test class size	20.00	4.61	1	32	20.20	5.09	1	54
	Reading,* 6th *grade				***Math,* 6th *grade***			
	31,977 *obs.,* 1,687 *classes*				32,006 *obs.,* 1,696 *classes*			
Enrollment	51.09	19.64	25	88	51.18	19.70	25	88
Class size	21.05	4.70	1	32	21.04	4.70	1	32
Test class size	20.63	4.72	1	32	20.61	4.73	1	32
	Reading,* 8th *grade				***Physics/chemistry,* 8th *grade***			
	30,186 *obs.,* 1,725 *classes*				29,864 *obs.,* 1,715 *classes*			
Enrollment	57.52	19.34	25	88	57.62	19.31	25	88
Class size	20.74	4.62	5	60	20.75	4.61	5	60
Test class size	20.07	4.65	1	54	20.09	4.65	1	54

Notes. The table summarizes descriptive statistics of selected variables of subsamples from the subsamples of ±4 pupils around the thresholds. Only observations around the three lower cutoffs (28, 56, and 84) are included.

TABLE A.3: INTRODUCTORY DESCRIPTIVE STATISTICS, ±4 DISCONTINUITY SAMPLE

Model	(1)	(2)	(4)	(5)
OLS regression				
	2^{th} grade reading scores		3^{rd} grade math scores	
Class size	$-.0002$	$-.0012^{**}$	$-.0000$	$-.0010^{**}$
	(.0019)	(.0006)	(.0016)	(.0004)
No. of observations	150,642		151,736	
	6^{th} grade reading scores		6^{th} grade math scores	
Class size	$.0039^{***}$	.0004	.0022	$-.0013$
	(.0014)	(.0005)	(.0017)	(.0015)
No. of observations	152,731		152,856	
	8^{th} grade reading scores		8^{th} grade physics scores	
Class size	$.0075^{***}$	.0016	$.0049^{***}$	.0017
	(.0016)	(.0008)	(.0016)	(.0015)
No. of observations	141,175		140,242	
Class characteristics	No	Yes	No	Yes

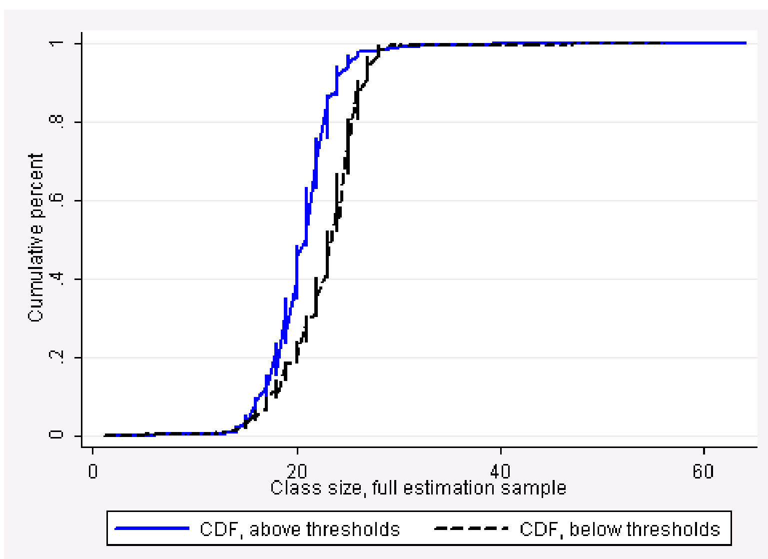
Notes. The estimates are based on pupils enrolled in regular classes in the Danish public school system in the school years of 2009/2010 to 2011/2012, that have been tested via the national test system. Test scores are standardized. Class size controls are contemporaneous. In addition to the control variables listed in the table, all specifications include fixed effects for enrollment segments and linear and square controls for grade enrollment into schools interacted with separate thresholds. Baseline covariates include the remaining controls from Table A.1. Standard errors adjusted for clustering by enrollment count are in parentheses. Asterisks indicate statistical significance at the ***1%, **5%, and *10% level, respectively.

TABLE A.4: OLS ESTIMATES, PREFERRED ENROLLMENT SPECIFICATION

B The Monotonicity Assumption

Recall that the monotonicity assumption implies that class sizes in grades below thresholds are never smaller than they would otherwise have been. Although this is a non-verifiable restriction it does have a testable implication in the case of multivalued treatments (Angrist and Imbens 1995). Namely that the empirical cumulative distribution functions of class size given enrollments above and below the cutoffs, respectively, should not cross. Thus, class sizes below thresholds are generally not smaller than those above. Figure B.1 illustrates the CDF of class size above and below the administrative thresholds. Once again, I have chosen to pool the thresholds for simplicity of the illustration. Also, schools with grade enrollments below 15 are excluded from the sample as class sizes above thresholds can never be smaller than class sizes in these. Based on figure B.1 it would appear that class sizes in grades below enrollment thresholds generally are larger compared to those above.

Violations of the monotonicity assumption causes the IV estimand to be biased because the group of people who are inversely affected by the instrument



The figure shows the cumulative distribution functions of class size given grade enrollments above and below enrollment threshold, respectively. The class sizes pertain to classes in grades 2, 3, 6 and 8 of the Danish public school system that are tested in the national test system in the school years of 2009/2010 − 2011/2012. Schools with grade enrollments less than or equal to 14 are excluded from the sample. Below (above) thresholds implies that enrollments are within −14 (+14) pupils of an administrative thresholds.

Figure B.1: Empirical CDF of class size given enrollments above and below thresholds is nonempty. The size of the bias involves two factors (Angrist et al. 1996). The larger the proportion of pupils who are inversely affected by the instrument the larger the bias, and the larger the difference between the average causal effect of the instrument on the test outcome between these and the group of compliers, the larger the bias. Thus, if the average causal effect is identical for the two groups, violations of the monotonicity assumption do not result in bias to the class size effect.