Economics Working Papers

2015-07

Do class size effects differ across grades?

Anne Brink Nandrup

AARHUS UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

Do class size effects differ across grades?

This paper contributes to the class size literature by analyzing whether short-run class size effects are constant across grade levels in compulsory school. Results are based on administrative data on all pupils enroled in Danish public schools. Identification is based on a government-imposed class size cap that creates exogenous variation in class sizes. Significant (albeit modest) negative effects of class size increases are found for children on primary school levels. The effects on math abilities are statistically different across primary and secondary school. Larger classes do not affect girls, non-Western immigrants and socioeconomically disadvantaged pupils more adversely than other pupils.

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

1. Introduction

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. This paper evaluates the short-run effects of class size on pupil abilities within both mathematics and reading across different grade levels in compulsory school. While there exists a vast literature identifying negative short- and medium-run effects of class size increases in primary and secondary school, previous studies are often concerned with only one or a few close grades in the same setup (e.g. Finn and Achilles (1999) on preschool to grade 3, Angrist and Lavy (1999) on 4^{th} and 5^{th} graders, Heinesen and Browning (2007) and Dee and West (2011) on 8^{th} graders, Krassel and Heinesen (2014) on 10^{th} graders, and Fredriksson, Öckert and Oosterbeek (2013) on an average of 4^{th}-6^{th} graders). Unfortunately, comparisons of these effects are complicated by varying institutional settings as well as incomparable outcome measures. Thus, the literature provides little empirical insight into the mechanisms of class size effects across the years of compulsory schooling. This paper attempts to remedy this by employing the same identification method across grade levels in compulsory school using directly comparable measures of pupil ability as the outcomes of interest.

The effect of class size is unlikely to be constant across grade levels for several reasons. For example, smaller classes may in particular benefit pupils in the lower grades: Younger pupils may depend more on adult supervision and help, therefore, peer-tutoring or group work may be more effective in older grades

(Blatchford and Mortimore 1994). The self-control of pupils may increase with age and also a number of other psychological and hormonal factors change as pupils mature. Mischel and Mischel (1983) show that older children can create a more favourable environment for effective self-control. As such, older pupils may be less inclined to participate in interrupting behaviour. Conversely, parents may be more qualified to assist their children with homework, supplementary reading etc. in the early school years, thus effectively reducing the need for teacher one-on-one time for younger pupils.

The quality of public compulsory schooling has increasingly been at the center of attention in many countries. This follows partly from the recognition that the formation of human capital has important implications for both the individual and society. As such, early test score measures of pupils' academic achievement are possibly strongly related to measures of sustained success in adulthood such as wage and length of education (Todd and Wolpin 2003).

Additionally, budget limitations in many countries leave school administrators and politicians preoccupied with creating better schools within tighter budgets. Class and school sizes are a recurrent issue when considering means to cut compulsory schooling expenditures. These school inputs are readily measured and are in general considered easier to manipulate. Furthermore, increasing class sizes comprise large budget savings; In OECD, teachers' salaries alone constitute 62% of compulsory schooling expenditure (OECD 2012). Such cost reductions may come at a price, however. Recent work by Fredriksson, Öckert and Oosterbeek (2013) suggests significant adverse long-term effects from class size increases in upper primary school.

Exploiting test results from the unique Danish national test system in combination with detailed register-based data, this paper identifies the effects of changes in class size on test results for three different levels of compulsory schooling: lower and upper primary and lower secondary school. Following in the footsteps of Angrist and Lavy (1999), I employ a fuzzy regression discontinuity design arising from a government-imposed maximum class size rule of 28 pupils. I apply this identification strategy to data covering pupils in Danish public schools between 2009/2010 and 2011/2012. As learning processes likely differ across linguistic and logical subjects, the effects of increased class sizes on reading and math abilities are studied separately.

Results show significant (albeit modest) negative effects of increasing class sizes in the Danish public schools where the average class size is 21 with a modal value of 23. Most effects of a class size increase in primary school are significantly negative, whereas none of the lower secondary level estimates are significant. More importantly, under certain circumstances I am able to reject that the results do not differ across grades levels. I employ a wide range of robustness tests to underpin the validity of the results presented here.

The remainder of this paper is organized as follows. Section 2 summarizes the institutional setting of the paper while Section 3 presents the available data as well as the identifying variation of the IV estimates. The identification strategy is described in Section 4 while Section 5 presents and discusses the results. Section 6 summarises and concludes.

2. Institutional setting

There are 98 municipalities in Denmark, each of which is divided into one or more school districts. Pupil residential address determines school district affiliation. However, since 2006 pupils are only entitled, but not required, to attend the district school. Public schools are free and financed by local municipalities through a combination of municipality income tax and a between-municipality redistribution scheme subsidizing expenditures in low-income municipalities. Furthermore, public schools are subject to a government imposed maximum class size rule of 28 pupils per classroom.[1] But class sizes vary considerably across schools and cohorts. The school structure implies that municipalities, rather than schools independently, finance the expenditures associated with the maximum class size rule. Approximately 86% of Danish children were enroled into public schools in 2009/2010-2011/2012.

Children are taught from the calendar year they turn six years old, beginning with preschool (preschool was optional before 2009). Public schools typically contain grades 0-9 (smaller schools may only contain grades 0-7). Pupils are generally divided into classes when they enrol in preschool (grade 0) and follow the same class throughout compulsory school with few exceptions, for example elective third language. These subjects are usually not introduced until grade 7. Teachers are subject-specific and typically follow classes through (parts of) compulsory school. The public school system builds on the principle that pupils cannot be tracked according to ability or social background. Consequently, there are no elite schools or classes in the public system.

There is no formal division of the levels in Danish compulsory schools, but following the literature this paper denotes three overall grade levels: lower primary school (grades 1-3), upper primary school (grades 4-6) and lower secondary school (grades 7-9).

In 2010, ten mandatory tests were introduced to public compulsory schools, including reading tests (grades 2, 6 and 8), math tests (grades 3 and 6), and physics/chemistry (grade 8).[2] There is no math test in lower secondary school, but as physics and math are often considered to be based on somewhat similar mindsets, test results for physics/chemistry act as substitutes here.

The mandatory tests are electronic, adaptive and self-scoring, thus, teachers are not able to bias the results. Test adaptivity ensures that pupil ability within subjects is very precisely determined compared to regular linear tests, see Beuchert-Pedersen and Nandrup (2014). Also, the nature of the tests makes them qualified for comparison both across and within individuals.

3. Data and identification

3.1 *Data*

The data set contains registry data on all pupils in the Danish school system and their test results, maintained by The Danish Ministry of Children and Education, combined with registry data on pupils and their parents, maintained by Statistics Denmark. School enrolment is registered annually in September and

Table 1. Summary statistics of key variables

Variable	Reading sample Mean	S.d.	Math (physics/chemistry) sample Mean	S.d.
	Grade 2 ($N = 150,065$)		*Grade 3* ($N = 152,800$)	
Class size	21.20	4.06	21.19	4.03
Enrolment	50.65	22.21	50.06	21.70
Std. test result	0.01	0.99	0.01	0.99
	Grade 6 ($N = 153,810$)		*Grade 6* ($N = 153,846$)	
Class size	21.37	3.91	21.36	3.86
Enrolment	50.25	21.23	50.35	21.27
Std. test result	0.04	0.96	0.03	0.97
	Grade 8 ($N = 141,938$)		*Grade 8* ($N = 140,975$)	
Class size	21.87	3.58	21.86	3.57
Enrolment	60.97	22.38	60.92	22.45
Std. test result	0.05	0.94	0.02	0.98

allows one to construct beginning-of-the-school year class sizes and enrolment counts of all grade levels in all schools. Test results are obtained in January through April and are available from the school year of 2009/2010. Thus, results are based on all mainstream classroom pupils (i.e. excluding special needs pupils and alternative class divisions) in tested grade levels of public schools in the school years of 2009/2010-2011/2012 ($4,259$ schools$\times$years). Because of unobserved test results $70,753$ observations (7.44%) are dropped from the sample. Thus, in total the sample consists of $893,434$ observations in either grades $2,3,6$ or 8.

The explanatory variable of interest is class size. However, test results may be observed up to eight months after class registration, which means that class size may have changed meanwhile. Conversely, beginning-of-the-school year class size and enrolment may be more 'exogenous' because they are less likely to be affected by parents. Also, one may argue that class size during the school year is just as important for ability accumulation. Table 1 shows summary statistics of the key explanatory variables in the six subpopulations. Test results are standardized to mean 0 and standard deviation 1 across subjects, cohorts and years. Across subsamples 8^{th} grade enrolment is around 10 pupils higher, correspondingly class sizes increase with approximately 0.5 pupil. Excluding schools arbsorbing lower secondary pupils from small schools that do not offer grades 8 and 9, class sizes and enrolment counts are similar across subsamples.

A few classes (5.61%) contain below 14 pupils while only 0.36% are larger than the 28-pupil cap size. The latter are not excluded from the sample, because I am hesitant to condition on the endogenous variable. All results in Section 6 are robust to the exclusion of 'too' large classes.

Other explanatory variables include: school characteristics and detailed pupil-specific information such as birth information, family information and socioeconomic status (from the years of the pupil's sixth birthday). A complete list of controls including descriptive statistics hereof is found in the Appendix (Table A.1), where regression results demonstrate that the included controls indeed are relevant predictors of pupils' test results.

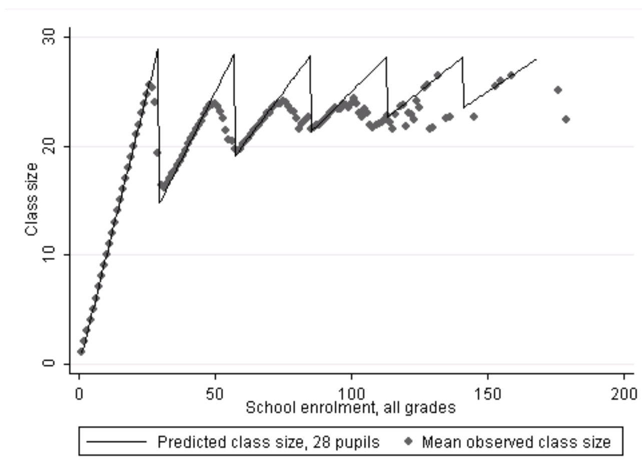
3.2 *Identifying variation*

The causal effect of class size is rather difficult to study because the majority of class size variations is the result of choices made by parents, school administrators, teachers and politicians on a local or national level. Thus, class size is potentially correlated with other determinants of pupil achievement. As originally suggested by Angrist and Lavy (1999), this paper uses exogenous variation in class sizes created by the 28-pupil rule as an instrument for the endogenous class size variable. Following the authors, the expected class size, assuming cohorts are divided into classes of equal size of grade g in school s in year t[3], is given by (1):

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

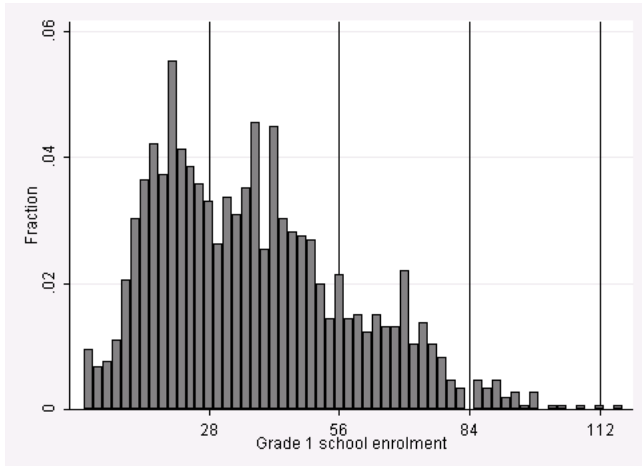
where e_{gst} denotes the grade level enrolment and $floor\left(n\right)$ the largest integer less than or equal to n. (1) reflects that subject to the 28-pupil rule enrolments of up to 28 pupils are assigned to one class while enrolments between 29 and 56 are divided into two classes of 14.5−28 pupils each, etc.

Figure 1 illustrates the relationship between grade enrolment and expected (solid line) and mean observed (dots) class sizes in the full estimation sample. Even though f_{gst} is presumably not the only factor contributing to actual class size, it is a strong predictor - at least below enrolment counts of 100.[4] Figure 1 shows how the probability of treatment (assignment to a small class) in a fuzzy regression discontinuity context should be higher to the right of the cutoff than to the left. The fuzzy design indicates that treatment is not guaranteed.



The sample includes mainstream class room pupils in grades 2, 3, 6, and 8 in Danish public schools (2009/2010 − 2011/2012). The expected class size indicated by the line is based on the 28-pupil rule.

Figure 1. Expected and mean observed class size by enrolment, the full estimation sample



The sample includes mainstream classroom 1^{st} graders in Danish public schools. Vertical lines indicate thresholds created by multiples of the 28-pupil rule.

Figure 2. Distribution of grade 1 enrolment, 2009/2010

3.3 *Is the regression discontinuity design valid at school level?*

In a regression discontinuity context, random assignment of treatment intensity may be undone by administrator sorting when the assigment rule is public knowledge. 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 differently across Danish municipalities, and there is only very few examples of yearly school district revisions. Because of discontinuities in the enrolment count of Swedish schools, Fredriksson, Öckert and Oosterbeek (2013) are compelled to focus on school district enrolment rather than on the school level. The discontinuities arise as Swedish legislation encourages adjustment of school catchment areas within school districts to utilise demography and school resources optimally. Also, Urquiola and Verhoogen (2009) document an extreme case of bunching based on class size caps in Chilean subsidized private schools. To examine the Danish setting, Figure 2 illustrates the distribution of grade 1 enrolments in the school year of 2009/2010. Grade 1 is chosen because municipalities can only adjust school catchment areas before cohorts enrol in schools. Further, preschool was not mandatory until 2009. By visual inspection, there is no clear indication of bunching below the thresholds caused by administrator sorting. Furthermore, the free school choice should at least partly offset this.

4. Estimation strategy

Exploiting the exogenous variation in class sizes induced by the 28-pupil rule, it is possible to interpret the effects of class size on pupil achievement causally

(Angrist and Lavy 1999). The model is estimated by two-stage least squares (2SLS):

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

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

where $\theta_{icgs\tau}$ denotes the standardized test result of individual i in class c of grade g at school s at enrolment segment τ. $\boldsymbol{X}_{icgs}$ is a vector of controls (including pupil, parental and school characteristics and cohort dummies). CS_{cgs} denotes observed class size and the residuals, $\varepsilon_{icgs\tau}$ and $\upsilon_{icgs\tau}$, are idiosyncratic. (2) and (3) include segment fixed effects, φ_τ and ϕ_τ, to accommodate different patterns around the separate enrolment thresholds.[5] Also, the coefficients of the second order enrolment polynomials, $g_\tau\left(e_{gs}\right)$ and $q_\tau\left(e_{gs}\right)$, are allowed to vary by segment.

Class size is instrumented by the binary indicators $\boldsymbol{above}_{gs}$ equalling one for grade enrolments above thresholds and zero otherwise: $above28 = \mathbf{1}\left(28 < e \leq 42\right)$, $above56 = \mathbf{1}\left(56 < e \leq 60\right)$ etc. This setup highlights the quasi-experimental identification strategy of the RD design and ignores the smooth variation in the expected class size between thresholds. Heinesen and Browning (2007) argue that this is the most appropriate specification because only variation in the instrument around thresholds is used. By allowing the enrolment polynomials to vary by segment, I follow Fredriksson, Öckert and Oosterbeek (2013) and effectively consider each threshold as a different experiment.

The coefficient α_1 is of primary interest. It captures a weighted average treatment effect to a unit change in class size for the unknown subpopulation of pupils whose treatment status is affected by the instrument in a setting where class size effects are heterogeneous and non-linear (Angrist and Imbens 1995). The weights are proportional to the number of pupils who, because of the 28-pupil rule, are induced to attend a smaller class. Identification arises given independence and monotonicity assumptions: Independence requires that treatment is as good as randomly assigned, which is closely related to the exclusion restriction that requires instruments not to affect test results other than through their effect on class size. The monotonicity assumption requires that class size given enrolment above a threshold is never larger than it would otherwise have been. In all cases these assumptions are non-verifiable but the stronger the instruments the less sensitive the IV estimand is to violations of the assumptions (Angrist and Imbens 1995).

Better schools likely face increased demand, thus, enrolment and instruments are potentially related to pupil ability for reasons other than class size. This relation is, however, expected to be a smooth function of enrolment and highlights the need for including sufficient controls for enrolment effects (Angrist and Lavy 1999). In Denmark, a free school choice effectively reduces school transfer costs. This is potentially problematic as parents may be more inclined to exploit the 28-pupil rule and undo the random assignment of class sizes. However, if parents are not able to *precisely* manipulate the assignment variable, the variation in treatment near the thresholds should be randomized (Lee and Lemieux 2010).

Intuitively, parents can roughly predict class size based on district size, but as treatment depends on the enrolment of all other children in the school cohort, it would be risky to actively choose schools based on enrolments in small intervals around thresholds - particularly as school transfers always involve costs (disruptions, loss of peers etc.).

Generally speaking, parents can evade the 28-pupil rule in two ways: when enroling their child into compulsory school and by school transfers during school years. In 2008-2011 more than 26% of the Danish public schools had a different number of classes on the first grade level compared to the year before. Thus, before entering compulsory school parents may have difficulties anticipating the class size based on previous years. However when pupils transfer schools during the school years, class sizes in the receiving schools are already observed. There are three reasons why this is less problematic. Firstly, identification is based on beginning-of-the-school year class sizes, thus sorting during the school year does not affect the results. Secondly, for schools with grade enrolments in small intervals around the thresholds it would still be risky to predict class size. Finally, choice of school is presumably based on many other factors than just class size. In the end, the school headmaster decides which class to enrol a new pupil in, given that there are several classes on the grade level. It 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. Besides larger transportation costs and potentially poorer family characteristics of classmates, countryside schools are generally associated with less flexibility and less specialisation and diversity of teachers.

To examine the pupil observable characteristics above and below thresholds, Table 2 tests the significance of the coefficient to the instrument when regressing selected baseline variables separately on a pooled version of the instrument.[6] Column (1) shows the results for all pupils. Only some covariates are unrelated to the pooled instrument, thus, one has reason to suspect that treatment is completely not randomly assigned across thresholds. However, in a ± 4 pupil interval around the three lower cutoffs, few of the predetermined characteristics are related to the instrument (column 2). Also, when regressing all baseline covariates on being above a cutoff for ± 4 pupils around the three lower cutoffs, all coefficients are jointly insingificant (p-value .195, omitted here).

Because of limited data on other school inputs, estimated class size effects should be interpreted as 'total policy effects' (Todd and Wolpin 2003). I.e. the *ceteris paribus* effect of a class size increase plus an indirect effect through the responses of other inputs. Although one is usually interested in total policy effects, these estimates provide little insight into the education production function. Among others, Pop-Eleches and Urquiola (2013) finds that parents respond to higher school quality by reducing effort.

5. Results

This section quantifies the effect of class size on math and reading abilities using the empirical approach outlined in Section 4. All reported standard er-

Table 2. Balancing of covariates

Selected baseline covariates	(1) p-value, *above* cutoff All pupils	(2) p-value, *above* cutoff ± 4 pupils around lower cutoffs
Female	.899	.603
Non-Western immigrant	.001	.029
Birthweight	.133	.189
Mother's education:		
− Basic	.028	.457
− Vocational	.341	.912
− Higher	.003	.302
Mother's log-earnings	.000	.160
Father's log-earnings	.005	.542
Mother's age	.001	.119
Single mother	.011	.101
Number of siblings	.023	.859
No. of observations	893,434	188,061

Notes. The *above* cutoff indicator equals 1 if the grade enrolment exceeds a threshold created by the 28-pupil rule up to +14 pupils (+4 pupils in (2)). Columns report the p-values for t-tests of the significance of the pooled instrument by separate OLS regressions of each variable listed on the instrument. The regressions also include: Year and enrolment segment fixed effects, and linear and square controls for grade enrolment interacted with separate thresholds (only column (1)). Standard errors are adjusted for clustering by enrolment count.

rors are clustered to account for group structures of the residuals within grade enrolment.[7]

5.1 *Main results*

Table 3 presents 2SLS estimates of the class size effect on pupil ability across compulsory school. Results in the full sample are obtained using a flexible enrolment specification where second order polynomials of enrolment are allowed to differ above and below each threshold to fully account for enrolment effects.[8] Whereas segment fixed effects should control sufficiently for enrolment effects when the sample is limited to close intervals around thresholds. Columns (1) and (2) present results for the full sample, while (3) and (4) include ± 4 pupil intervals around the three lower cutoffs. Specifications (1) and (3) do not include other baseline covariates.

If the 28-pupil rule produces experimental variation in class size, the 2SLS estimates should be robust to the inclusion of controls; they should only improve the precision of the estimates. Particularly coefficients for the ± 4 pupil intervals around the three lower thresholds are robust to the inclusion of controls. Also, F-statistics of excluded instruments clearly reject the null of weak instruments.

Although modest in magnitude, all coefficients in columns (2) and (4) of Table 3 are negative. Thus, increasing class size seems to harm pupil ability in both reading and math, though only significantly in primary school. Compared to this, results of an OLS specification suggest a compensatory allocation of class size. The OLS estimates of class size effects vary between significant $-.0010$ and significant .0056 standard deviations (on 2^{nd} and 8^{th} grade reading results, respectively) (Table A.2).

Table 4 presents the results in a setting where class size is interacted with grade levels, thus allowing one to study the significance of the differences in class size effects directly. Here, the main effect pertain to pupils in the upper

Table 3. 2SLS estimates, class size effects in grades 2-8

Outcome variable	(1)	(2)	(3)	(4)
	All pupils		**±4 pupils around lower cutoffs**	
2^{nd} grade reading score	$-.0015$	$-.0071^{***}$	$-.0047$	$-.0100^{***}$
	(.0035)	(.0030)	(.0045)	(.0036)
F-test (excl. instruments)	41.50	43.14	49.02	53.77
No. of observations	150,065		31,061	
3^{rd} grade math score	$-.0026$	$-.0044$	$-.0073$	$-.0058$
	(.0039)	(.0029)	(.0056)	(.0040)
F-test (excl. instruments)	27.80	28.08	28.97	28.66
No. of observations	152,800		33,404	
6^{th} grade reading score	.0001	$-.0044$	$-.0123^{***}$	$-.0125^{***}$
	(.0039)	(.0029)	(.0045)	(.0037)
F-test (excl. instruments)	33.30	33.87	36.59	35.80
No. of observations	153,810		31,543	
6^{th} grade math score	$-.0033$	$-.0078^{**}$	$-.0144^{***}$	$-.0149^{***}$
	(.0039)	(.0034)	(.0050)	(.0041)
F-test (excl. instruments)	33.39	33.92	36.89	35.73
No. of observations	153,846		31,543	
8^{th} grade reading score	.0027	$-.0014$	$-.0060$	$-.0077$
	(.0059)	(.0044)	(.0080)	(.0067)
F-test (excl. instruments)	17.87	18.42	26.75	29.36
No. of observations	141,938		30,424	
8^{th} grade physics-/chemistry score	.0039	$-.0004$	.0008	$-.0014$
	(.0044)	(.0039)	(.0055)	(.0048)
F-test (excl. instruments)	16.39	16.44	29.21	31.96
No. of observations	140,975		30,086	
2^{nd} order polynomial of enrolment	√	√		
Enrolment interacted w/ segment and threshold	√	√		
All controls		√		√

Notes. The estimates are based on pupils in mainstream classrooms in the Danish public schools in 2009/2010 − 2011/2012. Specifications (3) and (4) only include pupils enroled in schools with a grade enrolment of ±4 around the three lower cutoffs: 28, 56, and 84. In addition to the control variables listed in the table, specifications (1) and (2) include fixed effects for enrolment segments and linear and squared controls for grade enrolment into schools interacted with separate thresholds. Specifications (3) and (4) only include segment fixed effects. Controls include the remaining covariates from Table A.1. Standard errors adjusted for clustering by enrolment count are in parentheses, $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 4. 2SLS estimates, class size effects in grades 2-8, interaction specification

Independent variable	(1)	(2)	(3)	(4)
	2SLS, reading ability		2SLS, math ability	
Interaction	−.0039	.0022	.0032	.0091
(lower primary school)	(.0045)	(.0058)	(.0044)	(.0056)
Main effect	−.0034	−.0119***	−.0075**	−.0148***
	(.0031)	(.0036)	(.0035)	(.0041)
Interaction	.0023	.0044	.0078	.0133**
(lower secondary school)	(.0045)	(.0064)	(.0049)	(.0061)
No. of observations	446,113	93,028	447,621	95,033
All pupils	√		√	
±4 pupils around cutoffs		√		√
All controls	√	√	√	√

Notes. Table note (3) applies. The lower primary school interaction term pertains to grade 2 in columns (1) − (2) and grade 3 in columns (3) − (4). The main effect pertains to grade 6 while the lower secondary school interaction term denotes the 8^{th} grade. In addition to the control variables listed in the table, specifications (1) and (2) include segment fixed effects interacted with school level and linear and squared controls for grade enrolment into schools interacted with both separate thresholds and school level. Specifications (3) and (4) only include segment fixed effects interacted with school level. Instruments are interacted with school levels as well. Standard errors adjusted for clustering by enrolment count are in parentheses, $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

primary school (grade 6). Specifically, grade indicators are interacted with class size and instrument as well as enrolment control functions and segment fixed effects.

Table 4 shows that increased class sizes in upper primary school generally decrease test results. However, the difference across school levels is only significant for lower secondary math abilities.

Note that the interpretation and comparison of these class size effects are not entirely straightforward except under additional assumptions. By only including contemporaneous class size in the specifications, this term is in effect capturing contributions of previous class sizes as well. Still, the results provided here offer insight into the class size effects of compulsory school. Also, the results of Tables 3 and 4 are robust to alternative specifications of class size. Replacing beginning-of-school year class size with a two-year average, to accomodate the hypothesis that not only contemporaneous class size may affect ability accumulation, does not change the conclusions (see Table A.2). Here, enrolment and instrument specifications are based on the year prior to the outcome measure. Standard errors are slightly increased, which causes more imprecise estimates. Likewise, replacing class size, enrolment and instruments with the corresponding grade 1 information (only feasible for grades 2 and 3 results) does not change the magnitude of the estimates (results are available from the author).

Some municipalities may choose to operate under a lower class size cap than 28. To accomodate this potential pitfall, I use a bandwidth of ±14 pupils around the two lower thresholds to estimate the discontinuities in grade 1 class sizes on the basis of different maximum class size rules. This strategy is applied separately to each municipality and provides evidence of municipalities abiding by the 28-pupil rule. The exercise leaves 48 municipalities, but results are unchanged although of greater magnitude (see Table A.2).

5.2 *Heterogeneity*

To examine whether class size effects are heterogeneous, Table (5) presents results of the ± 4 pupil sample around the three lower cutoffs where class size is interacted with gender, parental income and immigrant status. Here, for example, gender is interacted with the class size and the instruments as well as enrolment segment.

Table 5 reveals little evidence of systematic effects of class size across pupil characteristics. Neither girls nor immigrants of non-Western countries (or descendents hereof) are more adversely affected by increased class sizes. If anything, 2^{nd} grade reading skills of non-Western immigrants seem to be slightly improved by a larger class. A general concern is that children from disadvantaged backgrounds are more adversely affected by a decrease of 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, children from low-earnings families[9] are not more adversely affected by larger class sizes in Denmark. A similar pattern emerges when interacting class size with the education level of the parents (omitted here). Interestingly, class size effects on upper primary math abilities of children from high-earning families appear to be more negative.

6. Conclusion

This paper extensively analyses the effects of class size across grade levels in compulsory school. Previous studies are primarily concerned with class size effects of close grade levels, therefore little evidence of how the class size effects behave across grades exists. To gain insight, I employ a well-known fuzzy RD design approach exploiting exogenous variation in class sizes based on a maximum class size rule.

The results are based on administrative data of all Danish pupils in public compulsory schools and reveal significantly negative (albeit modest) impacts of class size increases at the primary school level but not at the lower secondary level. Thus, the findings suggest that marginal class size increases in grade 8 may not be harmful to the learning environment, whereas pupils in grade 6 may in particular benefit from a class size decrease. However, the beneficial impact is modest when compared to the literature as well as in absolute values. As such, other initiatives, for example introducing a second teacher in the classroom or increasing instruction time of key subjects, may be more cost-effective compared to mere class size reductions. Furthermore, larger class sizes do not seem to increase inequality; girls, non-Western immigrants and socioeconomically disadvantaged pupils are not more adversely affected than other pupils.

Table 5. Heterogeneous effects of class size

	(1) Gender		(2) Immigrant status		(3) Highest earnings quartile		
Outcome variable	Main	Interact. (*female*)	Main	Interact. (*nonwestern*)	Interact. (1^{st} Q)	Main	Interact. (4^{th} Q)
2^{nd} grade reading scores	−.0061	−.0000	−.0085	.0348***	.0053	−.0099	.0001
($N = 31,061$)	(.0086)	(.0043)	(.0079)	(.0127)	(.0083)	(.0084)	(.0043)
3^{rd} grade math scores	−.0244	.0003	−.0236	.0111	.0139*	−.0278	.0107
($N = 33,404$)	(.0203)	(.0039)	(.0185)	(.0220)	(.0072)	(.0201)	(.0080)
6^{th} grade reading scores	−.0254*	−.0032	−.0268**	.0035	−.0001	−.0254**	−.0019
($N = 31,542$)	(.0141)	(.0043)	(.0130)	(.0190)	(.0092)	(.0131)	(.0050)
6^{th} grade math scores	−.0262*	.0044	−.0231*	.0042	.0040	−.0203	−.0156**
($N = 31,543$)	(.0144)	(.0049)	(.0140)	(.0101)	(.0049)	(.0142)	(.0077)
8^{th} grade reading scores	.0200	−.0074	.0205	.0163	.0157*	.0095	.0154
($N = 30,424$)	(.0189)	(.0058)	(.0152)	(.0167)	(.0089)	(.0183)	(.0106)
8^{th} grade physics scores	−.0043	−.0101	−.0078	.0021	−.0014	−.0079	−.0021
($N = 30,086$)	(.0139)	(.0076)	(.0127)	(.0172)	(.0105)	(.0129)	(.0075)

Notes. Estimates are based on schools with a grade enrolment of ±4 pupils within the three lower thresholds. Regressions include controls from Table A.1 and segment fixed effects interacted with interaction term. The instruments are interacted with interaction terms as well. Standard errors adjusted for clustering by enrolment count are in parentheses, $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Notes

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

[2]Pupils are also tested in reading (grade 4), English (grade 7), and biology and geography (grade 8).

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

[4]The strong pattern is largely consistent across grades with a somewhat poorer fit for the eighth grade.

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

[6]For simplicity, p-values are from regressions on a pooled binary indicator for being above *any* threshold. Results carry through for regressions on each *above*-indicator separately (available on request from the author).

[7]Clustering by the assigning variable is suggested by Lee and Card (2008) and performed in Fredriksson, Öckert and Oosterbeek (2013). This yields 136 clusters in the full estimation sample, a considerably higher level compared to clustering on school grade by year level where the instrument varies. However, the ensuing difference is modest.

[8]A short specification analysis is presented in the Appendix (Table A.2).

[9]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.

References

Angrist, Joshua, and Guido Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity." *Journal of the American Statistical Association* 90 (430) : 431 – 442.

Angrist, Joshua, and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *The Quarterly Journal of Economics* 114 (2) : 533 – 575.

Beuchert-Pedersen, Louise, and Anne Nandrup. 2014. "The National Tests - A Practical Guide." Unpublished manuscript, Aarhus University.

Blatchford, Peter, and Peter Mortimore. 1994. "The Issue of Class Size for Young Children in Schools: What Can We Learn from Research?" *Oxford Review of Education* 20 (4) : 411 – 428.

Dee, Thomas, and Martin West. 2011. "The Non-Cognitive Returns to Class Size." *Educational Evaluation and Policy Analysis* 33 (1) : 23 – 46.

Finn, Jeremy, and Charles Achilles. 1999. "Tennessee's Class Size Study: Findings, Implications, Misconceptions." *Educational Evaluation and Policy Analysis* 21 (2, Summer Special Issue: "Class Size: Issues and New Findings"): 97 – 109.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek. 2013. "Long-term Effects of Class Size." *The Quarterly Journal of Economics* 128 (1) : 249 – 250.

Heinesen, Eskil, and Martin Browning. 2007. "Class Size, Teacher Hours and Educational Attainment." *Scandinavian Journal of Economics* 109 (2) : 415 – 438.

Krassel, Karl, and Eskil Heinesen. 2014. "Class-size Effects in Secondary School." *Education Economics* 22 (4) : 412 – 426.

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

Lee, David, and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2 Jun): 281 – 355.

Mischel, Harriet, and Walter Mischel. 1983. “The Development of Children’s Knowledge of Self-Control Strategies.” *Child Development*, no. 54 : 613 – 619.

OECD. 2012. *Education at a Glance 2012: Highlights*. OECD Publishing. doi: 10.1787/eag_highlights-2012-en.

Pop-Eleches, Christian, and Miguel Urquiola. 2013. “Going to a Better School: Effects and Behavioral Responses.” American Economic Review 103 (4) : 1289 – 1324.

Todd, Petra, and Kenneth Wolpin. 2003. “On the Specification and Estimation of the Production Function for Cognitive Achievement.” *The Economic Journal* 113 (485) : *F*3 – *F*33.

Urquiola, Miguel, and Eric Verhoogen. 2009. “Class-Size Caps, Sorting, and the Regression-Discontinuity Design.” *The American Economic Review* 99 (1 June): 179 – 215.

Appendix

Table A.1. Sample means and OLS regression of outcome on controls

	Mean	SD	OLS Coeff.	SE
Outcome				
Average standardized test result	.027	.971		
Instruments				
Above 28-threshold	.198			
Above 56-threshold	.207			
Above 84-threshold	.061			
Above 112-threshold	.006			
Above 140-threshold	.001			
Controls				
Class size	21.468	3.854		
Enrolment	53.689	22.430		
Mother's education				
- None/missing	.045			
- Basic	.257		0.002	(0.010)
- Vocational	.367		0.055***	(0.010)
- Higher	.331		0.315***	(0.010)
Father's education				
- None/missing	.075			
- Basic	.236		0.023**	(0.009)
- Vocational	.410		0.076***	(0.010)
- Higher	.278		0.342***	(0.009)
Mother's logearnings	9.843	4.738	0.006***	(0.000)
Mother's age	34.763	7.445	0.014***	(0.000)
Father's logearnings	10.189	4.901	0.005***	(0.000)
Father's age	36.160	10.469	0.004***	(0.000)
Single mother	.155		-0.084***	(0.003)
Number of siblings	1.262	.866	0.011***	(0.002)
Girl	.492		0.032***	(0.003)
Western immigrant (or descendant hereof)	.020		0.016*	(0.009)
Non-Western immigrant (or descendant hereof)	.099		-0.324***	(0.006)
Birthweight (g)	3298.906	1006.609	0.000***	(0.000)
Length of gestation (days)	199.637	126.030	-0.000***	(0.000)
Born in the first quarter of the year	.242			
- second quarter	.252		-0.040***	(0.003)
- third quarter	.262		-0.090***	(0.003)
- fourth quarter	.235		-0.070***	(0.003)
First-born	.420		0.288***	(0.004)
Second-born	.371		0.125***	(0.003)
Born third or later	.196			
Multiple-born	.038		0.074***	(0.007)
Age indicators (omitted here)	-			
d_2010	.316		0.006	(0.006)
d_2011	.342		0.005	(0.006)
d_2012	.342			
Municipality with smaller cities (below 10,000)	.169			
- with a large city	.518		0.027***	(0.008)
- in the capital area	.313		0.027***	(0.008)
Observations	893,434		893,434	
R^2			0.12	

Notes. Information related to parents and family structure are registered in the year of the pupil's sixth birthday. In addition to the control variables listed in the table, the regression includes segment fixed effects and linear and squared controls for grade enrolment into schools interacted with separate thresholds. Standard errors adjusted for clustering by enrolment count are in parentheses, $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A.2. Auxiliary results

Outcome variable	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS, 2yr class size	(5) 2SLS, rule abiding
2^{nd} grade reading score	−.0010	−.0094***	−.0088***	−.0065	−.0105**
	(.0014)	(.0030)	(.0029)	(.0067)	(.0049)
F-test (excl. instruments)		94.60	51.86	36.67	19.53
No. of observations		150, 065		30, 570	15, 736
3^{rd} grade math score	−.0010	−.0065**	−.0065**	−.0049	−.0046
	(.0015)	(.0029)	(.0029)	(.0061)	(.0058)
F-test (excl. instruments)		92.77	25.43	50.84	20.97
No. of observations		152, 800		32, 881	16, 753
6^{th} grade reading score	.0033***	−.0043	−.0040	−.0077	−.0189***
	(.0012)	(.0031)	(.0030)	(.0047)	(.0045)
F-test (excl. instruments)		82.74	26.23	89.84	20.99
No. of observations		153, 810		31, 396	16, 793
6^{th} grade math score	.0013	−.0097***	.0087***	−.0110*	−.0242***
	(.0017)	(.0031)	(.0031)	(.0064)	(.0046)
F-test (excl. instruments)		82.54	25.12	90.98	21.45
No. of observations		153, 846		31, 390	16, 710
8^{th} grade reading score	.0056***	−.0011	−.0019	−.0068	−.0068
	(.0014)	(.0058)	(.0053)	(.0097)	(.0050)
F-test (excl. instruments)		30.86	10.15	18.48	35.70
No. of observations		141, 938		30, 287	15, 254
8^{th} grade physic/chemistry score	.0024	−.0008	.0008	.0003	−.0061
	(.0015)	(.0054)	(.0049)	(.0084)	(.0080)
F-test (excl. instruments)		33.05	10.35	20.81	35.97
No. of observations		140, 975		29, 949	15, 031
Instrument					
−Expected class size (f_{gs})		✓			
−Binary *above* indicators			✓	✓	✓
Enrolment specifications:					
−2^{nd} order polynomial	✓	✓	✓	✓	✓
−Interacted w/ segment and threshold	✓			✓	✓
Full set of controls	✓	✓	✓	✓	✓

Notes. The estimates are based on pupils in mainstream classrooms in the Danish public schools in 2009/2010−2011/2012. Specifications (4) and (5) only include pupils enroled in schools with a grade enrolment of ±4 around the three lower cutoffs: 28, 56, and 84. In (4) enrolment and instruments are based on enrolment two years before the outcome is observed. Class size is a three-year average. (5) includes municipalities that likely abide by the 28-pupil rule. In addition to the enrolment specifications listed in the table, all regressions include segment fixed effects. Controls include the remaining covariates from Table A.1. Standard errors adjusted for clustering by enrolment count are in parentheses, $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Economics Working Papers

2014-18: Ritwik Banerjee: On the Interpretation of Bribery in a Laboratory Corruption Game: Moral Frames and Social Norms

2014-19: Ritwik Banerjee and Nabanita Datta Gupta: Awareness programs and change in taste-based caste prejudice

2014-20: Jos Jansen and Andreas Pollak: Strategic Disclosure of Demand Information by Duopolists: Theory and Experiment

2014-21: Wenjing Wang: Do specialists exit the firm outsourcing its R&D?

2014-22: Jannie H. G. Kristoffersen, Morten Visby Krægpøth, Helena Skyt Nielsen and Marianne Simonsen: Disruptive School Peers and Student Outcomes

2014-23: Erik Strøjer Madsen and Yanqing Wu: Globalization of Brewing and Economies of Scale

2014-24: Niels-Hugo Blunch and Nabanita Datta Gupta: Social Networks and Health Knowledge in India: Who You Know or Who You Are?

2014-25: Louise Voldby Beuchert and Anne Brink Nandrup: The Danish National Tests – A Practical Guide

2015-01: Ritwik Banerjee, Tushi Baul, and Tanya Rosenblat: On Self Selection of the Corrupt into the Public Sector

2015-02: Torben M. Andersen: The Nordic welfare model and welfare services - Can we maintain acceptable standards?

2015-03: Nina Neubecker, Marcel Smolka and Anne Steinbacher: Networks and Selection in International Migration to Spain

2015-04: Sylvanus Kwaku Afesorgbor and Kaleb Girma Abreha: Preferential Market Access, Foreign Aid and Economic Development

2015-05: Ritwik Banerjee: Corruption, Norm Violation and Decay in Social Capital

2015-06: Christian Giødesen Lund and Rune Vejlin: Documenting and Improving the Hourly Wage Measure in the Danish IDA Database

2015-07: Anne Brink Nandrup: Do class size effects differ across grades?