Inside the black box of class size effects: Behavioral responses to class size variation*

Peter Fredriksson Björn Öckert Hessel Oosterbeek

Preliminary, please do not quote

Abstract

Experimental and quasi-experimental evidence on class size provide estimates of the total policy effect of class size variation. Since a class size intervention may trigger changes in the other inputs of the human capital production function, this total policy effect can be smaller or larger than the direct effect of a class size change. This paper examines the behavioral responses of pupils and parents to class size variation in Swedish primary schools. Parents respond to larger classes by helping their children more with their homework and by moving more frequently. Students respond to larger classes by spending less time on their homework and by reading less. We construct a simple model to derive the conditions under which our findings are consistent with the negative effect of class size on achievement reported in Fredriksson et al. (2012b).

JEL-codes: I21, I28, J24, C31
Keywords: Class size, pupils, parents, teachers, regression discontinuity

1 Introduction

There is mounting experimental and quasi-experimental evidence that larger classes is detrimental for short run educational outcomes.[1] This evidence should be interpreted as "policy effects", i.e., the total effect of an exogenous change in class size that includes the direct (or

*This version: May 2012. Fredriksson is affiliated with Stockholm University, IZA, IFAU, and Uppsala Center for Labor Studies (UCLS); Öckert with the Institute for Labour Market Policy Evaluation (IFAU) and UCLS; Oosterbeek with the University of Amsterdam.

[1]See, for instance, Angrist and Lavy (1999), Krueger (1999), Urquiola (2006), and Fredriksson et al. (2012b). Of course, there are also studies finding no effect (e.g., Hoxby 2000) but they are fewer in number.

ceteris paribus) effect of class size as well as indirect effects coming from responses of other inputs to the human capital production function; see Todd and Wolpin (2003). Of course, for evaluation purposes we are often most interested in the total policy effect. But experimental evidence of class size on achievement provides limited information on the nature of the human capital production function.

We study the behavioral responses to quasi-experimental variation in class size. We focus on pupils and parents, but also provide some descriptive evidence on teachers. Behavioral responses among parents, pupils, and teachers may magnify or moderate the direct impact of class size. Take parents as an example. If parents, in general, respond to larger classes by helping out with home work more, such behavior moderates the impact of class size. One might also expect parents to be differentially effective in providing homework assistance. If high-educated parents are more effective (or compensate more), their behavior contributes to the observed distributional impact of school resources.

To estimate the effects of class size we exploit variation in class size attributable to a maximum class size rule in Swedish primary schools. This maximum class size rule gives rise to a (fuzzy) regression discontinuity design. We apply this identification strategy to data covering the cohorts born in 1967, 1972, 1977, and 1982. In Fredriksson et al. (2012b), we showed that larger classes in the last three years of primary school (age 10 to 13) are detrimental for cognitive test scores at ages 13 and 16, and for wages and earnings at age 27 to 42. The focus here is on understanding the short run effect on test scores at age 13. This short run effect on test scores does not differ by family background.

Previous research on these issues has not convincingly dealt with selection problems caused by the non-random allocation of pupils and teachers to classes. Nevertheless, there are three papers that are closely related to what we do and make an attempt to deal with the selection issues. Bonesrønning (2004) uses a maximum class size rule in Norwegian lower secondary schools to investigate how parental effort responds to variation in class size. He concludes that parental effort decreases when class size increases, thereby reinforcing the negative effect of class size on achievement. This would thus suggest that reductions in class size and parental effort are complements in the education production function. The nice feature of this paper is that it is the first to examine the relation between class size and parental effort. The empirical analysis itself is, however, questionable. None of the specifications control for school enrollment at the grade level, first stage estimates of class size on the instrument are not reported, and reported effects are often insignificant, while the text suggests they are significant.

Datar and Mason (2008) use panel data from kindergarten and first grade students in the U.S. to examine whether class size has an impact on parent-child interactions, parent-school interactions and parent-financed activities. Larger classes appear to be associated with less parent–child interaction, the same level of parent–school interaction, and more parent-financed activities. The magnitudes of these effects are between 3% and 7% of a standard deviation. To account for endogeneity, the authors use child fixed effects combined with instrumental variables, where actual class size is instrumented by average class size at the grade level. It is not clear that the instrument used is valid.[2]

Houtenville and Conway (2008) use a value-added approach to investigate whether school spending has an impact on parental effort. They use data on eighth grade U.S. students and a variety of measures of parental effort. They consistently find that parental effort is reduced when per-pupil expenditure increases, suggesting that school and parental resources are substitutes in educational production. But the results are hard to interpret for a variety of reasons. First, they include per-pupil spending, class size, and teacher wages at the same time, suggesting that the coefficient on per-pupil spending (which is the only one they report) captures the relationship between other resources than teachers and parental effort. Second, they only report ordered probit coefficients so it is impossible to say anything about the magnitudes. Third, and most importantly, the value-added specification relies on very restrictive assumptions that are hard to justify (Todd and Wolpin 2003).

Relative to previous literature we make three contributions. First, and foremost, we use credible quasi-experimental variations to estimate the effect of class size on behavioral responses.[3] Second, we examine responses along more dimensions than in previous work. Third, building on Albornoz et al. (2011) we develop a theoretical model which is consistent with our results; we use this model to infer how teachers behave in larger classes.

We find that high-educated parents respond to larger classes by helping their children more with homework, and that all parents respond by being more likely to move to another school district. All students, independently of their family background, respond to larger classes by spending less time on homework and by reading less.

[2]The authors note that this instrument has been used in previous studies (Wößmann and West, 2006). This does not make it more convincing, though. The first stage coefficient equals 0.984 indicating that actual class size and average class size at the grade level are almost identical. This makes it difficult to argue that average class size at the grade level is not endogenous while actual class size is. We treat average class size as the endogenous variable to be instrumented. In Fredriksson et al. (2012b) we show that if average class size at the school level is assumed exogenous, class size is unrelated to the school achievement. This suggests that class sizes are not randomly allocated across schools.

[3]Even though Bonesrønning (2004) uses the same type of variation as we are using, it is far from clear that his instrument is valid. In contrast, we make sure that the design satisfies all validity and robustness checks that are nowadays standard for the application of regression discontinuity designs (e.g. Imbens and Lemieux, 2008; Lee and Lemieux, 2010; McCrary, 2008)

The model identifies the behavior of teachers as being crucial for the responses we observe among parents and students. In larger classes, teachers devote less attention to any given child. Parents compensate for this reduction using the means where they are most apt. The compensatory adjustment by parents can never counterbalance the reduction in teacher attention. Pupils, therefore, respond by providing less effort. We also present some descriptive evidence on the behavior of teachers in small and large classes, to see if teacher behavior is largely consistent with the model. We find this to be the case.

The paper proceeds as follows. In Section 2 we describe the relevant institutions of the Swedish schooling system. Section 3 describes the data and Section 5 examines the validity and strength of our instrumental variable approach. Section 6 reports the empirical findings. Section 7 presents the model we use to interpret the results. In Section 8 we present descriptive evidence on teacher behavior. Section 9 summarizes and concludes.

2 Institutional background[4]

In this section we describe the institutional setting pertaining to the cohorts we are studying (the cohorts born 1967-1982). During the relevant time period, earmarked central government grants determined the amount of resources invested in Swedish compulsory schools and allocation of pupils to schools was basically determined by residence.[5] Compulsory schooling was (and still is) 9 years. The compulsory school period was divided into three stages: lower primary school, upper primary school and lower secondary school. Children were enrolled in lower primary school from age 7 to 10 where they completed grades 1 to 3; after that they transferred to upper primary school where they completed grades 4 to 6. At age 13 students transferred to lower secondary school.

The compulsory school system had several organizational layers. The primary unit in the system was the school. Schools were aggregated to school districts.[6] School districts typically had one lower secondary school and at least one primary school. The catchment

[4]This section follows Fredriksson et al. (2012b) closely.

[5]This changed in the 1990s with the introduction of decentralization and school choice. From 1993 onwards compulsory schools are funded by the municipalities; see Björklund et al. (2005) for a description of the Swedish schooling system after decentralization. Du Rietz et al. (1987) contains an excellent description of the school system prior to decentralization, on which we base this section.

[6]We use the term "school district" for want of a better word. The literal translation from Swedish would be "principal's district" (*Rektorsområde*). Note that these school districts are very different from U.S. school districts. The prime responsibility of the school district was to allocate teachers over classes within district. Unlike U.S. school districts, they cannot raise funding on their own and there is no school board. In the Swedish context, the municipality is the closest analogy to U.S. school districts.

area of a school district was determined by a maximum traveling distance to the lower secondary school. The recommendations concerning maximum traveling distances were stricter for younger pupils, and therefore there were typically more primary schools than lower secondary schools in the school district. There was at least one school district in a municipality.

The municipalities formally ran the compulsory schools. But central government funding and regulations constrained the municipalities substantially. The municipalities could top-up on resources given by the central government; but they could not employ additional teachers. The central government introduced county school boards in 1958 to allocate central funding to the municipalities. In addition, the county school boards inspected local schools.[7]

Maximum class size rules have existed in Sweden in various forms since 1920. Maximum class sizes were lowered in 1962, when the compulsory school law stipulated that the maximum class size was 25 at the lower primary level and 30 at the upper primary and lower secondary levels.[8]

We focus on class size in upper primary school, i.e., grades 4 to 6. More precisely, the main independent variable in our analyses is the average of the class sizes students experience in grades 4, 5 and 6.[9] There main reason for this focus is data availability. We do not have precise information on schools (and hence school districts) attended for lower primary school.

The maximum class size rule at the upper primary level stipulated that classes were formed in multiples of 30; 30 students in a grade level in a school yielded one class, while 31 students in a grade level in a school yielded two classes, and so on.[10] We will use this rule for identification in a (fuzzy) regression discontinuity (RD) design. This method has been applied in several previous studies to estimate the causal effect of class size.[11]

Implementing the RD design must be done with care, however. The compulsory school law from 1962 opened up for adjustment of school catchment areas within school district such that empty class rooms would be filled. In that process, the county school boards were instructed to take the "needs" of the pupil population into account. Thus, it is likely that the school catchment areas are adjusted within school districts to favor disadvantaged pupils. In

[7]In the late 1970s, Sweden was divided into 24 counties and around 280 municipalities.

[8]The fine details of the rule were changed in 1978. Prior to 1978, the rule was formulated in terms of maximum class size. From 1978 onwards, a resource grant (the so called base resource) governed the number of teachers per grade level in a school. The discontinuity points were not changed.

[9]Hence, if a student is in a class of 25 pupils in grade 4, in a class of 24 students in grade 5 and in a class of 23 students in grade 6, the average class size to which this student was exposed in second stage primary school equals 24 ($= (25+24+23)/3$).

[10]There have always been special rules in small schools. In such areas, the rules pertained to total enrollment in 2 or 3 grade levels.

[11]The seminal paper is Angrist and Lavy (1999). See also Gary-Bobo and Mahjoub (2006); Hoxby (2000); Leuven et al. (2008); Urquiola and Verhoogen (2009).

a companion paper we show that such sorting takes place, rendering the RD design at the school level invalid.[12] Because of these problems, we implement the RD design at the school district level rather than at the school level. The virtue of the school district level is that pupils were assigned to a school district given their residential address, and that district boundaries were fixed due to regulations concerning maximum traveling distances. A problem with the school district analysis is that the maximum class size rule has less bite in multiple school districts. For that reason we focus on districts containing one upper primary school, which we refer to as one-school districts. We provide evidence that the RD design at the school district level is valid in Section 5.

The RD design requires, inter alia, that other school resources do not exhibit the same discontinuous pattern. There is no such pattern. In the mid 1980s, for instance, central government money for teachers amounted to 62 percent of the overall grant. The only other major grant component (27 percent of the grants) was aimed at supporting disadvantaged students. This grant was tied to the overall number of compulsory school students in a municipality and there were no discontinuities in the allocation of the grant.

3 Data

The key data source is the so-called ETF-project which is run by the Department of Education at Göteborg University; see Härnquist (2000) for a description of the data. Among other things, the data contain cognitive test scores at age 13 for roughly a 10 percent sample of the cohorts born 1967, 1972, and 1982. In addition, there is information on a 5 percent sample for the cohort born in 1977. For all cohorts, a two-stage sampling procedure was used. In the first stage, 30 out of the 280 municipalities were systematically selected; the selection criteria were based on population size and political majority. In the second stage, classes were randomly sampled within municipality. This sampling procedure implies that comparisons across municipalities for a given cohort are not valid, but comparisons within municipalities are valid. For this reason all analyses condition on municipality by cohort fixed effects.

To these data we have matched register information maintained by Statistics Sweden. The added data include information on class size (from the Class register) and parental informa-

[12] In Fredriksson et al. (2012a) we show that there is bunching around the cut-offs when school enrollment is the forcing variable. In particular it is more likely that schools are found just below than just above the cut-offs. Moreover, expected class size according to the rule predicts parental education; more children with well-educated parents are found just below the kink when school enrollment is the forcing variable.

tion (which is made possible by the multi-generational register containing links between all parents and their biological or adopted children). Class size is measured at the school level.

The cognitive tests at age 13 are traditional "IQ-type" tests. We constructed a measure based on scores for verbal skills and logical skills. The verbal test involves finding a word having the opposite meaning as a given word. The logical test requires the respondent to fill in the next number in a sequence of numbers. We refer to this measure as "cognitive skills" for short, it is standardized such that the mean is zero and the standard deviation equals one.

Data from the UGU-project also contain student and parental responses regarding issues related to the school. Students are, for instance, asked about their educational expectations, occupational aspirations, and time spent on homework and reading outside school. Parents are, e.g., asked about whether they help their children in doing homework. These questionnaires were distributed at age 13 when the pupils were in 6th grade. Unfortunately, indicators of teacher behavior is more sparse; we return to this issue on Section 8.

Data on parental characteristics come mainly from the Educational Register and the Income Tax Register, both maintained by Statistics Sweden. The Educational Register records the highest attained education level for the resident population.[13] The Income Tax Register contain, inter alia, data on annual earnings, which are based on income statements made by employers. We measure parental characteristics before age 10, i.e., prior to the class size variation that we are studying.

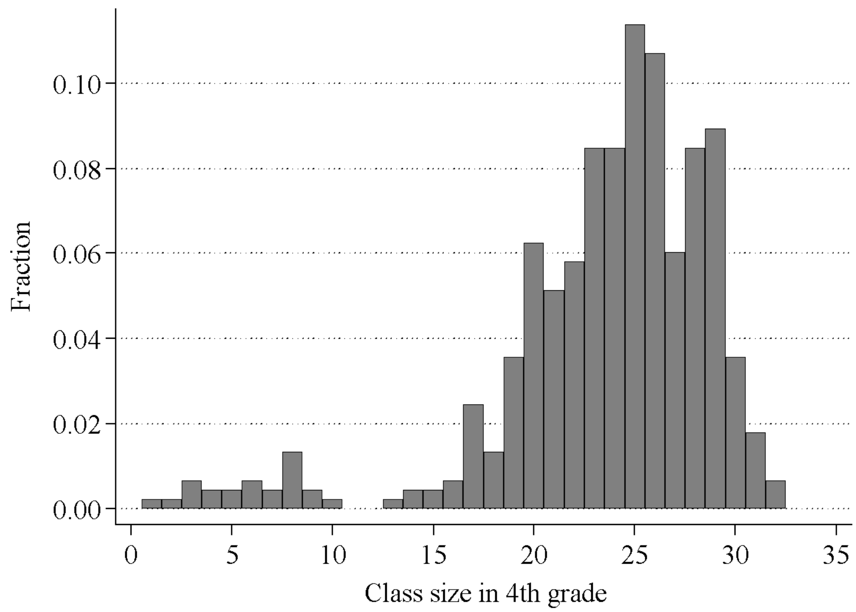
Table A1 in the Appendix reports descriptive statistics for all individuals together and broken down by parental education and income. The second part of the table shows that average class size in grades 4-6 is almost 24 pupils and that this is somewhat below the predicted average class size of 26 in these grades. Figure 1 shows the distribution of actual class size in grade 4. There are few very small classes (below 15) and few classes (2%) exceed the official maximum class size of 30.

4 Estimation strategy

To gain precision we pool the data from the different enrollment thresholds in the following way. Define the thresholds, $\bar{E}_\tau$, as $\bar{E}_\tau = \{30, 60, 90, 120\}$ and the indicator variable $I_{d\tau} = I(E_d \in \bar{E}_\tau \pm W)$. Thus $I_{d\tau} = 1$ if district enrollment (d indexes school districts) belongs to

[13] The register is complete for individuals with an education from Sweden. Information for immigrants stems from separate questionnaires to new arrival cohorts. The underlying data include information on the courses taken at the university level, which implies that this is a relatively accurate measure of years of schooling even for those who do not have a complete university degree.

Figure 1. Distribution of class size in grade 4



segment τ, where each segment is defined as enrollment counts within $\pm W$ of $\bar{E}_\tau$. Our default specification has $W = 15$, but conceptually $W = 1, ..., 30$.[14] Define normalized enrollment as $e_{d\tau} = (E_d - \bar{E}_\tau) I_{d\tau}$ and the treatment indicator

$$Above_{d\tau} = I(e_{d\tau} > 0) \tag{1}$$

For an individual i, the outcome equation of interest is

$$y_{id\tau} = \beta CS_{d\tau} + \alpha_\tau + f_\tau(e_{d\tau}) + \varepsilon_{id\tau} \tag{2}$$

where we use $Above_{d\tau}$ as the instrument for class size ($CS_{d\tau}$):

$$CS_{d\tau} = \gamma Above_{d\tau} + \delta_\tau + g_\tau(e_{d\tau}) + \nu_{d\tau} \tag{3}$$

To accommodate different patterns around different thresholds, we include segment fixed effects (α_τ and δ_τ) and allow the coefficients on the enrollment polynomials ($f_\tau(e_{d\tau})$ and $g_\tau(e_{d\tau})$) to vary by segment. This approach parallels analyses of randomized experiments

[14] With $W > 15$, the same observation is used as treated for one threshold and control for the next. E.g.: for $W \geqslant 17$, a district with enrollment equal to 47 belongs to the treated group at the threshold of 30 and to the control group at the threshold of 60.

with conditional random assignment (e.g. Krueger 1999 and Black et al. 2003), where each threshold is regarded as a different experiment.[15]

Notice that the endogenous variable in our analysis is the average of the class sizes student experience in grades 4, 5 and 6, while the instrument is derived from enrollment in grade 4. There are two reasons for this. The first reason is that enrollment in 5th and 6th grade are potentially endogenous to class size in 4th grade. Therefore, we cannot validly treat enrollment in 5th and 6th grade as exogenous. Enrollment in 4th grade can arguably be treated as exogenous since 3rd (lower primary school) and 4th grade (upper primary school) belong to different stages of compulsory school. The transition between lower primary and upper primary school often implies a change of school, and class size rules are different in lower primary and upper primary school. Given that enrollment in 5th and 6th grade are potentially endogenous we have no instruments for class size in grades 5 and 6. The second reason is that class sizes in grades 4, 5, and 6 are highly correlated. The correlation between class size in grades 4 and 5 is 0.79 and the correlation between class size in grades 4 and 6 is 0.57. Attributing all effects only to class size in grade 4 would not be correct. By focusing on the average of the class sizes in grades 4, 5 and 6, the instrumental variables (IV) estimates reflect the effects of an increase of class size by one pupil during three years.

5 Validity of the instrument

A threat to the validity of the RD design is bunching on one side of the cut-offs, since that indicates that the forcing variable is manipulated. Urquiola and Verhoogen (2009) document an extreme example of bunching in the context of a maximum class size rule in Chile. In their data there are at least five times as many schools just below than just above the cut-offs.

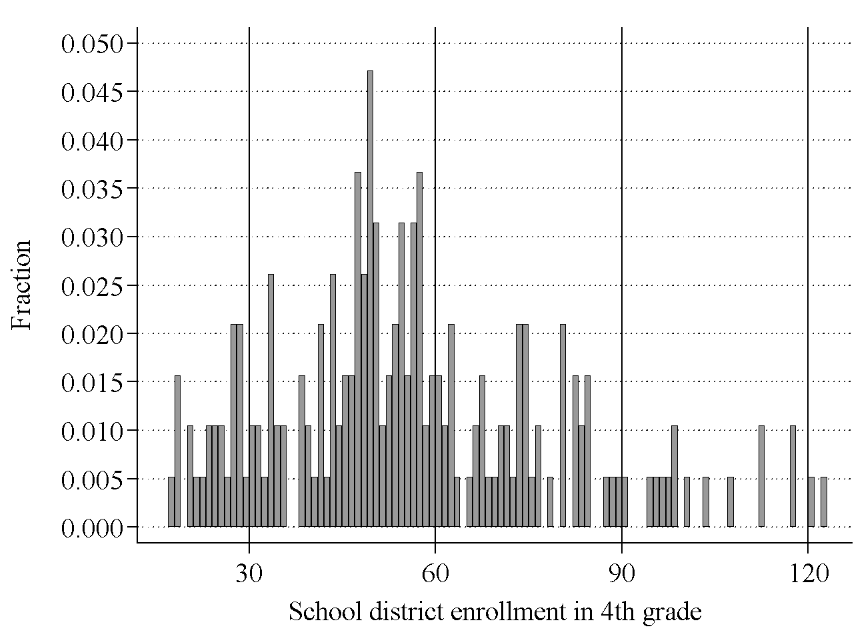
Figure 2 shows the distribution of enrollment in grade 4 in one-school districts. Visual inspection reveals no suspect discontinuities in the distribution of the forcing variable. The McCrary (2008) density test confirms this: we cannot reject the hypothesis that there is no shift in the discontinuity.[16]

A more direct way to assess whether the instrument is valid is to examine if pre-determined characteristics are balanced across observations above and below the thresholds. Figure 3

[15]Potentially, there would be an efficiency gain of utilizing information on treatment intensity, which varies since the sizes of the jumps in expected class size vary across segments. This is not true in our setting, however. The reason is that there is more noise in small school districts (where there is more variation in expected treatment intensity).

[16]To implement the test we used a bin size of 1 student and a bandwidth of 5 students. The estimated log difference in the height of the density is 0.19 with a standard error of 0.57.

Figure 2. Distribution of enrollment in grade 4 in one-school districts



shows that this is the case for parental education: the estimated discontinuity is -0.076 with a standard error of 0.369. Analogous plots for other covariates show very similar pictures.

Table 1 addresses the question of the balancing of pre-determined covariates more formally. The first two columns show that the baseline covariates we consider are highly relevant predictors of cognitive ability at age 13 and adult wages (observed at age 27-42). For instance, children who have more educated mothers score higher on the cognitive test (a year of education is associated with an increase in test scores of 0.069 standard deviations) and go on to have higher wages (a year of education is associated with a 0.6 percent increase in wages).

Column (3) of Table 1 shows the result of regressing the instrument on all baseline covariates.[17] The next to last row contains the result of an F-test of the hypothesis that all the coefficients on baseline covariates are jointly zero. The message of this F-test is that pre-determined characteristics are unrelated to the instrument (the p-value is 0.70). In column (4) we test whether the coefficient on the instrument is zero in a regression of each individual characteristic on the instrument. Again, pre-determined characteristics are unrelated to the instrument.

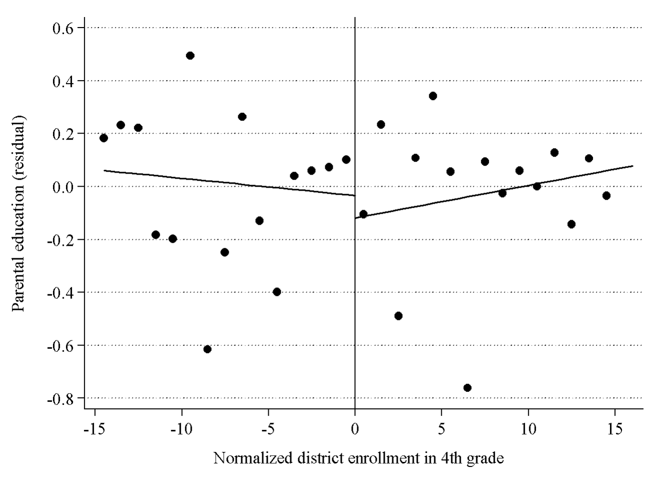
[17]These results come from regressions where we control for: segment fixed effects; linear controls for normalized school district enrollment, where the slopes are allowed to differ above and below the thresholds as well as across segments; and cohort by municipality fixed effects. We justify this specification in detail below.

Table 1. Balancing of covariates

	Cognitive ability age 13	ln(wage) age 27-42	Above threshold	p-value
	(1)	(2)	(3)	(4)
Female	-0.0020	-0.1422***	0.0027	0.433
	(0.0253)	(0.0107)	(0.0035)	
Month of birth	-0.0227***	-0.0017	0.0005	0.453
	(0.0035)	(0.0014)	(0.0007)	
Immigrant	-0.4616***	0.0222	0.0113	0.566
	(0.0585)	(0.0226)	(0.0168)	
Mother's years of education	0.0687***	0.0064**	0.0004	0.981
	(0.0059)	(0.0023)	(0.0010)	
Father's years of education	0.0597***	0.0135***	-0.0009	0.665
	(0.0051)	(0.0018)	(0.0010)	
Parental income (SEK 100,000s)	0.0384***	0.0112***	0.0002	0.947
	(0.0074)	(0.0026)	(0.0018)	
Mother's age at birth	0.0189***	0.0027***	-0.0004	0.471
	(0.0023)	(0.0009)	(0.0007)	
Number of siblings	-0.0728***	-0.0057*	-0.0011	0.709
	(0.0116)	(0.0045)	(0.0022)	
Parents separated	-0.1066***	-0.0305**	-0.0053	0.580
	(0.0299)	(0.0118)	(0.0057)	
p-value of F-test	0.000	0.000	0.698	
N	5,116	3,185	5,920	

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. Columns (1)-(3) report results of OLS regressions on the variables listed in the rows. These regressions also include the following control variables: fixed effects for segment, linear controls for school district enrollment interacted with threshold and segment, and cohort by municipality fixed effects. Cognitive ability at age 13 is standardized. Above threshold (the instrument for class size) is an indicator equalling unity if school district enrollment in 4th grade exceeds the class size rule threshold in the enrollment segment. Independent variables are pre-determined parent and student characteristics. The p-value reported at the bottom of columns (1)-(3) is for an F-test of the joint significance of the variables listed in the table. Each row of column (4) reports a p-value from separate OLS regressions of the pre-determined variable (listed in the corresponding row) on the instrument, and the same set of control variables as in columns (1)-(3). The p-value is for a t-test of the significance of the class size instrument. Standard errors adjusted for clustering at the enrollment counts (77 clusters) are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 percent level.

Figure 3. Parental education by enrollment in grade 4



Notes: The figure shows residual parental education, after controlling for municipality by cohort fixed effects, by 1-student bins. The regression lines were fitted to individual data. Discontinuity at threshold: -0.076 (standard error: 0.369).

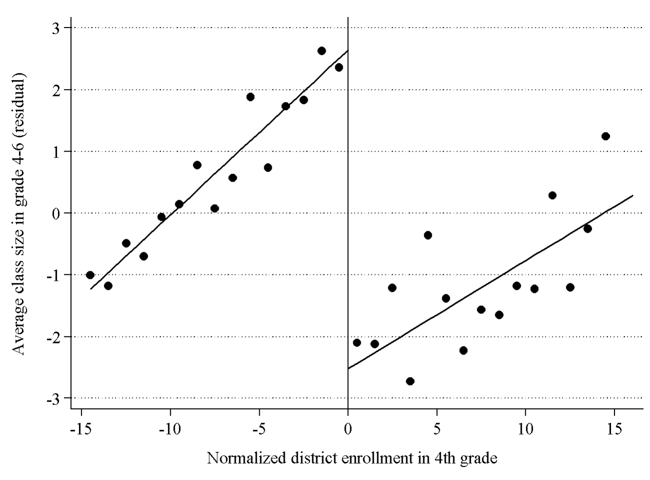
Figure 4 shows a graphical representation of the first stage. There is a clear, and statistically significant, jump at the threshold. Districts that have just surpassed one of the thresholds have classes that are systematically smaller than classes just below the threshold. The discontinuity at the threshold is -5.21 with a standard error of 0.85. When we control for predetermined characteristics, which is valid given the results in Table 1, the estimate of the discontinuity at the threshold does not change at all (the estimate is -5.22, with a standard error 0.85).

6 The effects of class size

6.1 Short-term cognitive outcomes

Here we reproduce a sub-set of the estimates in Fredriksson et al. (2012b). Table 2 shows OLS and IV estimates of the effect of class size on cognitive skills at age 13. The OLS estimate in the first column is a very precisely estimated zero. IV estimates are presented for six different specifications of the function $f(E_{jd})$. Columns (2) to (4) include a linear, quadratic and cubic controls for enrollment, respectively. The fifth column allows for linear splines in enrollment, and the sixth column allows for quadratic splines. In the final column, the sample has been restricted to districts in which enrollment is at most 5 pupils away from a cut-off. The estimates in columns (2) to (7) are all very similar, implying that it does not

Figure 4. Class size by enrollment in grade 4



Notes: The figure shows residual class size, after controlling for municipality by cohort fixed effects, by 1-student bins. The regression lines were fitted to individual data. Discontinuity at threshold: -5.207 (standard error: 0.848).

matter much how we exactly control for the forcing variable. From now on we will always include a third order polynomial in school district enrollment. The results for other outcome variables are also insensitive to this choice.

Table 3 looks at differences in the impact of class size on short-term cognitive outcomes by parental education and income. The first column repeats the result from column (4) in Table 2 for the entire sample . Columns (2) and (3) present results separately for children with low- and high-educated parents, and columns (4) and (5) present results separately for pupils with low (below median) and high (above median) income parents.

The estimate for children with high-educated parents is slightly higher in absolute size than for children with less-educated parents, but the estimates are not statistically different from one another. The effect on cognitive ability is very similar for pupils with low income parents and pupils with high income parents.[18] We therefore conclude that the estimates do not differ by parental background.

What lies behind these estimates? The remainder of the paper investigates how students, parents, and teachers respond to variation in class size. After presenting the evidence we piece together a story which is consistent with what we observe. The key ingredient in this story is the behavior of teachers in small and large classes.

[18]Consistent with this, unconditional quantile regression estimates shows that the entire distribution is uniformly shifted to the left when class size increases; see Fredriksson et al. (2012b).

Table 2. IV and first stage estimates, different enrollment controls

Model	(1)	(2)	(3)	(4)	(5)	(6)
	Cognitive ability, age 13 ($N = 5,116$)					
RF: Above threshold	0.2546***	0.2405***	0.2493***	0.2089**	0.2144**	0.3858***
	(0.0732)	(0.0771)	(0.0766)	(0.0908)	(0.0876)	(0.1223)
IV: Class size grades 4-6	-0.0471***	-0.0457***	-0.0454***	-0.0317*	-0.0330**	-0.0628**
	(0.0163)	(0.0165)	(0.0161)	(0.0147)	(0.0146)	(0.0292)
	Average class size grades 4-6 (first stage) ($N = 5,920$)					
Above threshold	-5.4215***	-5.2766***	-5.5303***	-6.7143***	-6.6254***	-6.3740***
	(0.8899)	(0.8715)	(0.8729)	(0.8064)	(0.7523)	(1.4522)
F-test for instrument	37.12	36.65	40.14	69.32	77.56	19.26
Enrollment controls						
- 1st order polynomial	√		√		√	
- 2nd order polynomial		√		√		√
Interacted with segments			√	√	√	√
Interacted with thresholds					√	√
No. of districts	191	191	191	191	191	191

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. Above threshold equals unity if school district enrollment in 4th grade exceeds the class size rule threshold in the enrollment segment. Cognitive ability at age 13 is standardized. In addition to the control variables listed in the table, all models include: fixed effects for enrollment segments, cohort by municipality fixed effects, gender, dummy variables for month of birth, dummy variables for mother's and father's educational attainment, parental income, mother's age at child's birth, indicators for being a first or second generation Nordic immigrant, indicators for being a first or second generation non-Nordic immigrant, an indicator for having separated parents and the number of siblings. Standard errors adjusted for clustering at the enrollment counts (77 clusters) are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 percent level, respectively.

Table 3. IV estimates of class size in 4th-6th grade on cognitive outcomes

	Parents' income		
	Interact. (1st Q)	Main effect	Interact. (4th Q)
Cognitive ability, age 13	0.0057	-0.0335**	0.0061
	(0.0220)	(0.0163)	(0.0197)

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. Class size in grades 4-6 is instrumented with Above threshold (=1 if school district enrollment in 4th grade is above the class size rule threshold in the enrollment segment). The model includes the following controls for school district enrollment in grade 4: fixed effects for segment; linear controls for enrollment which are interacted with threshold and segment. In addition the model includes the following baseline characteristics: cohort×municipality fixed effects, gender, dummy variables for month of birth, dummy variables for mother's and father's educational attainment, a second order polynomial in parental income, mother's age at child's birth, indicators for being a first or second generation Nordic immigrant, indicators for being a first or second generation non-Nordic immigrants, an indicator for having separated parents, and the number of siblings. Standard errors adjusted for clustering by enrollment count (77 clusters). ***/**/*=the estimates are significantly different from zero at the 1/5/10 percent level.

6.2 The exclusion restriction

An important question is whether the instrument is excludable from the outcome equation. If it is not, we can only causally interpret the reduced form estimates shown in Table 2. To provide evidence on the validity of the exclusion restriction, we examine if districts respond in other ways to the rule. Results are presented in columns (1)-(6) of Table 4. In column (1), we examine whether the probability of being assigned to remedial training is affected by the instrument. If schools respond to the instrument, we would expect it to be lower in districts that have surpassed one of the thresholds. We find no such evidence, however. Column (2) examines if the probability of being assigned to an age integrated class is affected by the instrument. Again, we find no evidence that this is an issue.

In columns (3) and (4) we examine the possibility that there may be greater scope for tracking when a threshold is surpassed, since surpassing a threshold implies the addition of another class. To address this issue we construct two dissimilarity indices (Duncan and Duncan, 1955) which relate class composition to school composition. Column (3) considers segregation in terms of parental education while column (4) considers parental income. In both cases segregation is unrelated to the instrument.[19]

[19]Notice that the standard errors are biased downwards in columns (3) and (4). The bias comes from the fact that the indices has complete evenness as the baseline. Since classes are small units the appropriate baseline is random unevenness. To generate the appropriate baseline one should simulate the baseline by randomly allocating individuals to units; see Carrington and Troske (1997) on these points. Since our estimates are not

In columns (5) and (6) we relate teacher characteristics to the instrument. The rule is unrelated to teacher experience; see column (5). But there is some evidence that the share of teachers with a college degree is lower in districts having surpassed one of the thresholds; see column (6). This may be a source of concern. Note, however, that the reduction in teacher credentials is arguably driven by the decrease in class size; moreover, there is little credible evidence suggesting that teacher credentials affect student performance. Nevertheless, smaller classes come with less educated teachers. If anything, this would tend to reduce our estimate of class size relative to an ideal experiment conducted in our context.

An issue that affects the interpretation of the IV estimates is whether class size in grades 4-6 is correlated with class sizes in the other stages of compulsory school. Columns (7) and (9) in Table 4 address this issue by showing results from regressions of class size in lower primary school (grades 1-3) and class size in lower secondary school (grades 7-9) on the instrument. The estimates show that class sizes in the other stages of compulsory school are unrelated to the instrument. Dividing the estimates in column (9) with the first-stage estimate in column (8), we find that a pupil increase in class size in upper-primary school leads to an (insignificant) 0.10 increase of class size in lower-secondary school. The correlation with class size in lower primary school (obtained analogously) is 0.11, which is also insignificant.

Given the evidence in Table 4 we focus on IV estimates from here on. And we interpret these IV estimates as the effects of one pupil change throughout upper primary school (grades 4-6).

6.3 Main results

For the main results we focus on measures of pupil and parental effort as well as parental resources. Pupil and parental effort is either measured by the time they devote to certain activities (time spent on homework and time spent on help with homework) or the frequency with which they engage in certain activities (how many times they read outside school within a certain time frame). Parental resources is measured by residential mobility.

We pool information from four cohorts. The questions regarding effort are framed somewhat differently across the cohorts. To make the responses comparable across cohorts we percentile rank the outcomes.

In terms of heterogeneity, we focus on parental education from now on (heterogeneous effects with respect to parental income have an analogous flavor, but differ less across groups).

significant even with complete evenness as the baseline we have refrained from simulating the data.

Table 4. Other responses to the instrument

	Remedial	Age inte-	Class composition		Teacher characteristics		Class size		
	training	grated class	education	income	experience	education			
Grade	4	4	4	4	4	4	1-3	4-6	7-9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Above threshold	-0.062	0.054	0.033	0.009	-0.312	-0.036*	-0.745	-6.625***	-0.693
	(0.054)	(0.047)	(0.026)	(0.022)	(0.622)	(0.019)	(1.034)	(0.752)	(0.867)
No. of districts	191	191	191	191	191	191	191	191	191
No. of individuals	4,346	5,920	5,920	5,920	5,834	5,834	5,896	5,920	5,920

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. Remedial training equals one if the pupil attends remedial training, age integrated class is the share of pupils in the school who are placed in an age-integrated class, class composition with respect to education is the dissimilarity index for parental education, class composition with respect to income is the dissimilarity index for parental income, teacher experience is the average years of experience for teachers in grades 4-6 in the school, teacher education is the share of teachers with a college degree in grades 4-6 in the school and class size is the average class size in the grades. Above threshold equals unity if school district enrollment in 4th grade exceeds the class size rule threshold in the enrollment segment. All models include the following controls for school district enrollment in grade 4: fixed effects for segment; linear controls for enrollment which are interacted with threshold and segment. In addition all models include the following baseline characteristics: cohort×municipality fixed effects, gender, dummy variables for month of birth, dummy variables for mother's and father's educational attainment, parental income, mother's age at child's birth, indicators for being a first or second generation Nordic immigrant, indicators for being a first or second generation non-Nordic immigrants, an indicator for having separated parents, and the number of siblings. Standard errors adjusted for clustering at the enrollment counts (77 clusters) are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 percent level, respectively.

Table 5. IV estimates of class size in 4th-6th grade on pupil effort

		Parents' education	
Dependent variable	All	Low	High
Effort			
Time spent on homework	-0.611*	-0.542	-0.625*
(percentile ranked, 0-100)	(0.347)	(0.452)	(0.351)
Reading outside school	-0.483**	-0.573*	-0.466
(percentile ranked, 0-100)	(0.239)	(0.330)	(0.301)
N	5,xxx	2,xxx	2,xxx

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. The outcomes are derived from pupil questionnaires distributed at age 13. Actual class size in grades 4-6 is instrumented with the expected class size in grade 4 as predicted by the class size rule at the school district level. All models control for cohort×municipality fixed effects, gender, dummy variables for month of birth, dummy variables for mother's and father's educational attainment, a third order polynomial in parental income, mother's age at child's birth, indicators for being a first or second generation immigrant, having separated parents and the number of siblings, and a third order polynomial of school district enrollment in grade 4. High educated parents means that one of the parents has a least 3 years of university-preparatory upper-secondary education. Standard errors adjusted for clustering at the school district×cohort level are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 per cent level of confidence, respectively.

Pupil responses Table 5 shows the effects of class size on pupil effort.[20] The first column presents estimates for all observations together, while the other two columns present estimates by parental education.

The results in the first row shows that an increase in class size reduces time spent on homework; the effects do not vary significantly across the distribution of parental education. There are two interpretations of this result. Both of them rely on the fact that monitoring homework is costly for the teacher. In the first interpretation, teachers give the same amount of homework (or more) in a larger class; since pupils know that it requires more effort on the part of the teacher to monitor whether they have done their homework or not, students respond by doing less homework. Another interpretation is that teachers respond to a class

[20]We also have other information from pupil questionnaires that could be interpreted as relating to effort (e.g. whether the pupil sometimes gives up in school). But we focus on the cardinal measures since they: (i) contain more variation; and (ii) have a straightforward interpretation. When we relate class size to an indicator for whether the pupil sometimes gives up in school, we find no effects (the estimates have the expected sign, however). But it is somewhat hard to interpret this result. The answer to a question on whether pupils' give up in school depends on the reference point. A likely reference point is the behavior of other pupils in the class. If a class size increase shifts the entire effort distribution uniformly within class, these answers will convey no information.

size increase by giving less homework since teachers know that it requires more effort to monitor students in a larger class.

The second row shows that reading outside school is reduced because of an increase in class size. The response is somewhat larger at the lower end of the family background distribution (although not significantly so). One possible interpretation is that teachers also have a key role in encouraging reading outside school.

What magnitudes are involved? Consider an increase of class size by five students (which corresponds roughly to the standard deviation in class size in the 4th grade). Such a change would reduce homework by 6 percent and reading outside school by 5 percent relative to the medians in the two distributions.

Parental responses We look at two dimensions of parental response: parental help with homework and the probability of moving the child to another school district (within municipality) or another municipality. Regarding the mobility decision, it is likely that both of these moves require moving house. But the move to another school district is likely to reflect a response to class size to a greater extent than a move to another municipality.

The point estimate suggests that an increase in class size increases parental help with homework. But this estimates is not statistically significant. The reason it is not statistically significant is that the effects vary so much with respect to education of the parents. For parents below the median there is no effect whatsoever. For parents above the median, the effect of class size is to increase parental help with homework. This is a striking result given that pupil effort on homework is reduced, which implies that there is less opportunity for parents to help. The estimate for high-educated parents implies that help with homework increases by 6 percent relative to the median in the distribution when class size increases by five pupils.

The second row shows that children change school district as a result of an increase in class size. This is a big impact, considering that the mobility rate across school districts within municipality is 6 percent. Note in particular that less educated parents move their children to other school districts in response to an increase in class size. All parents thus respond to class size. Less-educated parents are not equally apt to help out with homework; hence they refrain from doing that. But they do place their children in other school districts. High-educated are apt to help out with homework; moreover they move their children to other school districts as a response to a class size increase.

The last row shows that parents do not move to another municipality in response to a change in class size. We think this result lends credence to the result on school district

Table 6. IV estimates of class size in 4th-6th grade on parental responses

		Parents' education	
Dependent variable	All	Low	High
Help with homework	0.330	-0.017	0.630**
(percentile ranked, 0-100)	(0.256)	(0.300)	(0.318)
Move			
To another school district	0.013**	0.014*	0.011*
(=1 if "yes")	(0.005)	(0.007)	(0.006)
To another municipality	0.002	-0.002	0.003
(=1 if "yes")	(0.003)	(0.004)	(0.003)
N	5,xxx	2,xxx	2,xxx

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. The outcomes are derived from parent questionnaires distributed when the pupils were 13 years-of-age. Actual class size in grades 4-6 is instrumented with the expected class size in grade 4 as predicted by the class size rule at the school district level. All models control for cohort×municipality fixed effects, gender, dummy variables for month of birth, dummy variables for mother's and father's educational attainment, a third order polynomial in parental income, mother's age at child's birth, indicators for being a first or second generation immigrant, having separated parents and the number of siblings, and a third order polynomial of school district enrollment in grade 4. High educated parents means that one of the parents has a least 3 years of university-preparatory upper-secondary education. Standard errors adjusted for clustering at the school district×cohort level are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 per cent level of confidence, respectively.

mobility. Mobility across municipality borders are likely tied to job opportunities of the parents, while mobility within a municipality is likely tied to the school quality for the child.

6.4 Other results

Expectations about performance is most likely a mediating factor in educational production. Another potential mediating factor is the class room environment (e.g. Lazear 2001).

We have some information related to pupil expectations and educational aspirations. First, children are asked whether they think they do well in school. Second, they are asked about their future educational attainment. Again, the precise formulation of the second question varies across cohorts and therefore we percentile rank this outcome.

The first two rows in Table 7 shows the results for these two outcomes. We find no effects on average. However, the expectations among children with high-educated parents are revised downwards when class size increases. In interpreting these results, it is important to point out that pupil expectations are likely heavily influenced by the expectations that parents and teachers have about their performance. That children of high-educated parents respond that they do not do well in school is probably related to failure relative to the expectations that teachers and parents have on them. Similarly, the fact that we do not find anything for children of less-educated parents may be related to low expectations on them anyhow.

The last two rows in Table 7 shows the results for measures that we think are related to the class room environment. The outcome in the next to last row is meant to capture distractions. Somewhat surprisingly, the evidence seems to suggest less distractions in a larger class. However, it is not entirely clear that this inference is valid. Another interpretation is that larger classes require greater focus on the part of students. And, therefore, the effect of an increase in class size is that they think about other things to a lesser extent.

The final row shows that pupils are less comfortable to answer questions in larger classes; this effect is concentrated among children with less educated parents. A possible interpretation of this group has an aversion to speaking in front of larger audiences (which by definition increases in larger classes).

6.5 Summary and discussion

Pulling the main results together we find that student effort is reduced in larger classes, despite parental attempts to compensate for an increase in class size. The reduction is student effort is most likely an important reason for the reduction in student performance caused by larger classes.

Table 7. IV estimates of class size in 4th-6th grade on pupil expectations and the class room environment

Dependent variable	All	Parents' education: Low	Parents' education: High
Expectations			
Think they do well in school	-0.008	-0.001	-0.015**
(=1 if "yes")	(0.005)	(0.007)	(0.006)
Educational aspirations	-0.296	-0.079	-0.625*
(percentile ranked, 0-100)	(0.281)	(0.345)	(0.351)
Classroom environment			
Do not think about other things	0.011*	0.011	0.009
(=1 if "yes")	(0.006)	(0.008)	(0.008)
Not scary to answer questions	-0.006**	-0.011**	-0.002
(=1 if "yes")	(0.003)	(0.005)	(0.003)
N	5,xxx	2,xxx	2,xxx

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. The outcomes are derived from pupil questionnaires distributed at 13 years-of-age. Actual class size in grades 4-6 is instrumented with the expected class size in grade 4 as predicted by the class size rule at the school district level. All models control for cohort×municipality fixed effects, gender, dummy variables for month of birth, dummy variables for mother's and father's educational attainment, a third order polynomial in parental income, mother's age at child's birth, indicators for being a first or second generation immigrant, having separated parents and the number of siblings, and a third order polynomial of school district enrollment in grade 4. High (low) educated parents ... High (low) income parents means that the parents' total earnings is above (below) the median. Standard errors adjusted for clustering at the school district×cohort level are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 per cent level of confidence, respectively.

Only high-skilled parents help out with homework. Yet we observe the same reduction in student achievement (and effort). Assuming that parental help is useful, this may suggest that the direct effect of class size is larger for children of high-skilled parents than for children of less-educated parents.

An interesting (and difficult) question is whether parental help is actually useful. We have not seen any credible evidence on this. To shed some light on the issue, we follow Datar and Mason 2008 and estimate a child fixed effects model for the cohort born 1972 which is the only cohort where we observe performance in two grades – grades 3 and 6. Thus we relate child achievement to parental help with homework and average class size at the school by grade level. These estimates are plagued by an obvious simultaneity concern: negative innovations in achievement likely cause more help, yielding a downward bias in the estimate. Despite this concern, we find that parental help with homework improves achievement. The estimate is statistically significant (at the 5 percent level) but relatively small: A standard deviation increase in parental help improves a achievement by 0.02 standard deviations. The magnitude of this estimate should be interpreted carefully since it is arguably biased downwards.

Our purpose next is to piece together a model which is able to generate the gist of the empirical results we observe. The key ingredient of the model is that parental and teacher effort provides incentives for students to work hard. And since working hard improves achievement, the model thus takes as given that parental effort is good for achievement.

7 A simple model of educational performance

7.1 Basic set-up

In this section we set up a model to guide the interpretation of the empirical findings reported above. The model is a simplified version of the model in Albornoz et al. (2011). The main ingredients of the model are the following. Children care about their short-run utility rather than the future value of their human capital. Parents and teachers try to provide the right incentives for student learning. Providing incentives for learning is a time-consuming activity, which is determined by parents and teachers simultaneously. In equilibrium, the actions of parents and teachers are substitutes. The key assumption generating this result is that greater efforts on the part of teachers reduce the marginal benefit of parental effort. The actions of parents and teachers (and their effectiveness) determine the skill impact of class size variations.

The behavior of children The skills of the child are given by

$$\theta^c = \alpha e \tag{4}$$

where $\alpha = \alpha(\theta^p)$, $\alpha'(\cdot) > 0$, denotes the ability of the child from a family with parental skill θ^p and e effort. Children care about short-run utility rather than being forward-looking permanent income maximizers.[21] They choose effort to maximize

$$U^c = ce - \frac{1}{2}e^2 \tag{5}$$

The quadratic term captures the disutility of providing effort. The marginal return to increasing effort (c) is determined by the reward structure offered by parents and teachers. There are several ways to think of c. In general, it is determined by all parental and teacher actions that induce effort on the part of students. Thus it can represent encouragement from parents/teachers. It can also represent monitoring by, e.g., teachers (to see this, just reformulate the utility function as $U^c = -c(1-e) - (1/2)e^2$ which generates the same behavior on the part of students). Moreover, it can represent parental help with homework, in which case we can think of ce as effort in efficiency units. We assume that

$$c = (1-\gamma)c^p + \gamma c^t \tag{6}$$

where c^p denotes the incentives generated by the parent and c^t the incentives generated by the teacher. The reward structure c is thus a weighted average of the actions by parents and teachers. The parameter γ serves as an indicator of the efficiency of teachers relative to parents in generating the incentives that children respond to. Notice that the functional form in (6) is inconsequential for the results. The key assumption is that the marginal return to increasing parental effort is decreasing in teacher effort (see Albornoz et al. 2011).

The behavior of parents Each parent has one child. Parents care about the (future) wage of their children (which depends on θ^c) and their own welfare (W^p): $U^p = \beta w(\theta^c) + W^p$ where $\beta \leq 1$ is a discount factor. Parents' own welfare is given by $W^p = (1 - c^p e/\theta^p)w(\theta^p)$, where $w(\theta^p)$ is the opportunity cost of time for the parent. So inducing effort from the child is a time-consuming activity. Parents differ in their ability to generate a given reward; more talented parents (as indexed by θ^p) use less effort to generate a given reward. We also

[21] We think this assumption is realistic rather than restrictive. Primary school children need parents and teachers to provide the link between what they learn in school and future earnings.

assume that θ^p influences the opportunity cost of time: talented parents are likely to have higher opportunity costs of time and so $w'(\theta^p) > 0$. We thus specify the utility of the parent as

$$U^p = \beta w(\theta^c) + (1 - \frac{c^p e}{\theta^p}) w(\theta^p) \tag{7}$$

The parent chooses the reward, c^p, to maximize (7).

The behavior of teachers Teachers care about the average human capital in the class and their own welfare (W^t): $U^t = \bar{\theta}^c + W^t$. There are several possible motivations for including average human capital of students into teachers utility function. One straightforward reason is that teacher wages depend on the performance the students in the class; another is that they take pride in what they do. We assume that teachers' own welfare is given by $W^t = (1 - c^t \sum e)$, where the sum runs over all children in the class. Again, providing incentives for children is a time-consuming activity; this cost is larger in a larger class (this is the congestion externality in this model; cf. Lazear 2001).[22] In sum, teacher utility is given by

$$U^t = \frac{1}{n} \sum \theta^c + (1 - c^t \sum e) \tag{8}$$

where n denotes class size. The teacher chooses a single reward level c^t that maximizes (8). We assume that teachers do not differentiate the reward across pupils within the class, based for example on parental education.

Structure of the game The virtue of our empirical strategy is that class size is as good as randomly assigned. Thus, in the first stage, "nature" assigns class size. In the second stage, parents and teachers simultaneously decide on the reward structure, i.e., they determine c^p and c^t respectively. After observing c^p and c^t, children determine their effort. Given the structure, we solve the game by backward induction, and then examine what happens to equilibrium responses when class size is changed.

7.2 Equilibrium

Children maximize (5) taking (6) as given. Optimal effort is given by

$$e^* = (1 - \gamma) c^p + \gamma c^t \tag{9}$$

[22] In principle we could have made W^t dependent on teacher skills as in W^p. But since we do not have data on teacher skills we ignore this possibility.

Parents take (9) into account when determining c^p. An optimal choice of c^p satisfies

$$c^{p*} = \max\left\{\frac{1}{2}\left[\beta\frac{w'(\theta^c)\alpha(\theta^p)\theta^p}{w(\theta^p)} - \frac{\gamma}{1-\gamma}c^t\right], 0\right\}$$

This condition implies that if teachers provide a lot of incentives, or if the incentives generated by the teacher are more efficient relative to the parent (γ is high), then it is more likely that parents are at the corner solution $c^p = 0$. Skilled parents are more likely to offer rewards if the first term in square brackets is increasing in θ^p. This will be the case under general conditions.[23] To simplify the exposition we impose the assumption that wages are linear in skills $w^j = q\theta^j$, $j = c, p$. The optimal choice for parents is then given by

$$(\theta^p > \underline{\theta}):\ c^{p*} = \frac{1}{2}\left[\beta\alpha(\theta^p) - \frac{\gamma}{1-\gamma}c^t\right]; \quad (\theta^p \leq \underline{\theta}):\ c^{p*} = 0 \tag{10}$$

Thus, all parents with skills above $\underline{\theta}$ will supply effort to improve the learning outcomes of their children.

Teachers also take (9) into account when determining c^t. The optimal choice of c^t is given by

$$c^{t*} = \frac{1}{2}\left[\frac{\bar{\alpha}}{n} - \frac{1-\gamma}{\gamma}\bar{c^p}\right] \tag{11}$$

where $\bar{\alpha} = (1/n)\sum\alpha$ denotes the average ability of the students taught by the teacher, and $\bar{c^p} = (1/n)\sum c^p$ the average rewards provided by the parents of the students that are taught by the teacher.[24]

Equilibrium responses to changes in class size It follows from (10) and (11) that c^t is decreasing in class size, i.e.,

$$\frac{\partial c^{t*}}{\partial n} = -\frac{2}{3}\frac{\bar{\alpha}}{n^2} < 0 \tag{12}$$

The reason c^{t*} is decreasing in class size is that it requires more teacher effort to provide a given c^t when class size increases. Whereas the teacher does not differentiate his or her response across students, parental and student responses differ by skill. From now on we will separate the analysis with respect to parental skill, holding the class composition fixed

[23] The condition is that $(1 + \alpha'(\theta^p)\theta^p/\alpha(\theta^p)) > \partial \ln w/\theta^p$. This restriction is very weak. To see this note that Fredriksson et al. (2012b) estimate the wage return to skill ($\partial \ln w/\theta^p$) to 0.084 and Grönqvist et al. (2010) estimate the intergenerational correlation in skills ($\alpha'(\theta^p)\theta^p/\alpha(\theta^p)$) to 0.5.

[24] In deriving (11) we ignore the possibility of a corner solution. With a corner solution, the model would be equivalent to home schooling, which is hardly realistic.

(which implies that $\partial c^t/\partial n$ does not vary across skill groups). High-skilled parents are those with $\theta^p > \underline{\theta}$; low-skilled parents are those with $\theta^p \leq \underline{\theta}$. For the high-skilled, we obtain

$$(\theta^p > \underline{\theta}): \; \frac{\partial c^{p*}}{\partial n} = -\frac{\gamma}{2(1-\gamma)}\frac{\partial c^{t*}}{\partial n} > 0; \; \frac{\partial e^*}{\partial n} = \frac{\gamma}{2}\frac{\partial c^{t*}}{\partial n} < 0; \; \frac{\partial \theta^c}{\partial n} = \frac{\alpha(\theta^p)\gamma}{2}\frac{\partial c^{t*}}{\partial n} < 0 \quad (13)$$

For the low-skilled, we get

$$(\theta^p \leq \underline{\theta}): \; \frac{\partial c^{p*}}{\partial n} = 0; \; \frac{\partial e^*}{\partial n} = \gamma\frac{\partial c^{t*}}{\partial n} < 0; \; \frac{\partial \theta^c}{\partial n} = \alpha(\theta^p)\gamma\frac{\partial c^{t*}}{\partial n} < 0 \quad (14)$$

In sum, the model predicts that the parents of the high-skilled will try to compensate for the reduction of class size to a greater extent that the parents of the low-skilled. Student effort should decrease more among the children of the low-skilled. But since there are complementarities between student ability and effort in the production of skills, it is ambiguous whether child skills are reduced more among the low-skilled than among the high-skilled.[25]

The results we presented in the previous section are well in line with the model we present here. In particular, the model is consistent with the average responses we observe in the data. Almost all distributional impacts are also correctly predicted. What we get wrong relative to the evidence is that the model predicts a smaller (absolute) impact on pupil effort among the high-skilled than among the low skilled. However, if incentives for teachers are geared towards the lower end, the model would also be consistent the evidence on this last point.

In the model, the behavior of teachers is key. In the next section, we present descriptive evidence on the behavior of teachers in large and small classes.

8 Evidence on teacher behavior

Here we present some descriptive evidence on teacher behavior in small and large classes. For the 1982 cohort there is some limited information on teacher behavior and attitudes. We have information on leave of absence and attitudes vis-a-vis particular teaching methods. Teachers are asked about their attitudes on homework and testing, as well as on whether they think that pupils are responsible for learning.

There are two problems with these data. First, we cannot estimate the effects with the same rigor as for the other outcomes. Second, we cannot get at the question of whether teachers target attention to particular groups in larger classes.

[25] Comparing the last expressions of equations (13) and (14), this depends on $\alpha(\theta^p|\theta^p > \underline{\theta})/2 \gtreqless \alpha(\theta^p|\theta^p < \underline{\theta})$.

The model predicts that total effort increases in large classes (but effort directed towards any given pupil of course decreases). If total effort increases, we expect more stress and hence more leave of absence. Estimates to be added...

9 Conclusion

This is the first paper that looks at behavioral responses of parents and students following an increase in class size using a credible identification strategy. When class size increases, pupil effort is reduced and parental efforts to help their children increase.

We think the behavior of teachers is key to understand the simultaneous occurrence of both the above results. Therefore, we present a stylized model that has this flavor. The model builds on the assumption that providing incentives for learning (for any given student) is costlier for teachers in larger classes. Most parents anticipate this, and try to compensate by providing greater incentives (and help) for learning. The increase in parental efforts, notwithstanding, student effort is reduced in larger classes.

To provide some descriptive evidence on the role of teachers in small and large classes. This evidence is broadly consistent with the gist of the model (to be confirmed).

According to our evidence, the total policy effect of larger classes is negative. This total policy effect may be smaller or larger than the direct (ceteris paribus) impact of class size. The effect is magnified via pupil effort, but the effect is moderated via parental help with homework. Our evidence thus implies that parental inputs and school resources are substitutes in educational production.

The total policy effect does not vary by parental education. At the same time, we observe that only high-educated parents respond by exerting more effort on help with homework. The detrimental impact of class size would thus have been greater for this relatively advantaged group had we not seen this parental response.

References

Albornoz, F., Berlinski, S., and Cabrales, A. (2011). Motivation, resources and the organization of the school system. Unpublished working paper.

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

Björklund, A., Clark, M. A., Edin, P.-A., Fredriksson, P., and Krueger, A. (2005). *The Market Comes to Education in Sweden - An Evaluation of Sweden's Surprising School Reforms*. New York: Russell Sage Foundation.

Black, D. A., Smith, J. A., Berger, M. C., and Noel, B. J. (2003). Is the threat of reemployment services more effective than the services themselves? *American Economic Review*, 93(4):1313–1327.

Bonesrønning, H. (2004). The determinants of parental effort in education production: Do parents respond to changes in class size? *Economics of Education Review*, 23:1–9.

Carrington, W. J. and Troske, K. R. (1997). On measuring segregation in samples with small units. *Journal of Business & Economic Statistics*, 15(4):402–409.

Datar, A. and Mason, B. (2008). Do reductions in class size "crowd out" parental investment in education? *Economics of Education Review*, 27:712–723.

Du Rietz, L., Lundgren, U. P., and Wennås, O. (1987). Ansvarsfördelning och styrning på skolområdet. DsU 1987:1, Stockholm: Ministry of Education.

Duncan, O. D. and Duncan, B. (1955). A methodological analysis of segregation indices. *American Sociological Review*, 20:210–217.

Fredriksson, P., Öckert, B., and Oosterbeek, H. (2012a). The devil is in the (institutional) detail: Sorting and the RD design in a public school system. Unpublished manuscript, Stockholm University.

Fredriksson, P., Öckert, B., and Oosterbeek, H. (2012b). Long-term effects of class size. *Quarterly Journal of Economics*, 128(1):forthcoming.

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

Grönqvist, E., Öckert, B., and Vlachos, J. (2010). The intergenerational transmission of cognitive and non-cognitive abilities. Discussion Paper 7908, CEPR.

Houtenville, A. J. and Conway, K. S. (2008). Parental effort, school resources, and student achievement. *Journal of Human Resources*, XLIII:437–453.

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

Härnquist, K. (2000). Evaluation through follow-up. In Jansson, C., editor, *Seven Swedish Longitudinal Studies in the Behavioral Sciences*. Forskningsrådsnämnden, Stockholm.

Imbens, G. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142:615–635.

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

Lazear, E. (2001). Educational production. *Quarterly Journal of Economics*, 116(3):777–803.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48:281–355.

Leuven, E., Oosterbeek, H., and Rønning, M. (2008). Quasi-experimental estimates of the effect of class size on achievement in Norway. *Scandinavian Journal of Economics*, 110:663–693.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142:698–714.

Todd, P. and Wolpin, K. I. (2003). Towards a unified approach for modeling the production function for cognitive achievement. Features:F3–F33.

Urquiola, M. (2006). Identifying class size effects in developing countries: Evidence from rural Bolivia. *Review of Economics and Statistics*, 88(1):171–176.

Urquiola, M. and Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review*, 99:179–215.

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

Appendix

Table A1: Descriptive statistics, 1967-1982 birth cohorts

Variable [# individuals]	One-school districts	Full sample
Female	0.495	0.488
$[N = 5,920; N = 29,371]$	(0.500)	(0.500)
Mother's years of education	11.226	11.000
$[N = 5,920; N = 29,371]$	(2.7614)	(2.708)
Father's years of education	11.096	10.743
$[N = 5,920; N = 29,371]$	(3.078)	(2.982)
Parental income	476,268	456,418
$[N = 5,920; N = 29,371]$	(232,763)	(204,826)
Cognitive ability, age 13	0.009	0.002
$[N = 5,116; N = 25,856]$	(1.022)	(1.001)
Non-cognitive ability, age 13	0.028	0.011
$[N = 4,681; N = 23,864]$	(1.006)	(0.998)
Academic achievement, age 16	0.021	0.003
$[N = 5,755; N = 28,610]$	(1.015)	(1.001)
Cognitive ability, age 18 (men only)	0.048	0.004
$[N = 2,455; N = 12,949]$	(1.010)	(0.999)
Non-cognitive ability, age 18 (men only)	-0.042	0.004
$[N = 2,313; N = 12,184]$	(0.986)	(1.001)
Years of schooling, age 27-42	13.519	13.491
$[N = 5,588; N = 27,771]$	(2.614)	(2.608)
Bachelor's degree, age 27-42	0.272	0.269
$[N = 5,920; N = 29,371]$	(0.4453)	(0.4433)
Earnings, age 27-42	232,248	242,372
$[N = 5,920; N = 29,371]$	(176,675)	(179,520)
P(Earnings>0), age 27-42	0.906	0.911
$[N = 5,920; N = 29,371]$	(0.292)	(0.285)
ln(Wage), age 27-42	10.148	10.156
$[N = 3,185; N = 16,283]$	(0.279)	(0.274)

Table A1 - continued: Descriptive statistics, 1967-1982 birth cohorts

	One-school districts	Full sample
Class variables		
Class size in grade 4	24.337	23.329
	(3.843)	(3.437)
Class size in grade 4>30	0.024	0.026
	(0.112)	(0.115)
Average class size grades 4-6	24.357	24.066
	(3.489)	(3.990)
School district variables		
Enrollment 4th grade	63.457	105.985
	(23.436)	(38.807)
Above class size rule threshold	0.408	0.470
	(0.492)	(0.499)
N individuals	5,920	29,371
N schools	191	1,129
N school districts	191	697
N clusters (enrollment counts)	77	165

Note: The data are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. All measures of cognitive ability, non-cognitive ability and academic achievement have been standardized in the full sample. The educational outcomes are measured 2009, while the labor market outcomes have been averaged over the 2007-2009 period. Wages are restricted to wage-earners. Standard deviations are in parentheses.