Parental responses to public investments in children: Evidence from a maximum class size rule[*]

Peter Fredriksson   Björn Öckert   Hessel Oosterbeek

Abstract

We study differential parental responses to variation in class size induced by a maximum class size rule in Swedish schools. In response to an increase in class size: (i) only high-income parents help their children more with homework; (ii) all parents are more likely to move their child to another school; and (iii) only low-income children find their teachers harder to follow when taught in a larger class. These findings indicate that public and private investments in children are substitutes, and help explain why the negative effect of class size on achievement in our data is concentrated among low-income children.

JEL-codes: I21, I28, J24, C31

Keywords: Class size, parental responses, social background, regression discontinuity

1   Introduction

Parents and schools interact to shape the educational outcomes of children. In this context, parents are important for two reasons. First, parental characteristics and the home environment influence cognitive skill development, implying that children may be differentially affected by variation in school inputs for given parental behavior. Second, parents may be differentially able to adjust to variation in school inputs by altering their investment behavior.

This paper examines how parental investment behavior responds to variation in class size. By doing so we shed light on the extent of substitutability of household and school inputs in the human capital production function. To date, there is rather limited knowledge about how this question.

The major contribution of the paper, however, is that we pay particular attention to differential responses across the parental income distribution. This is an issue that has not been addressed completely credibly before; it is also a question of substantive policy interest. If high-income (high-skill) parents compensate for variation in school inputs to a greater extent, their children may be less susceptible to what is going on in school. For this reason, there may be a case for targeting public interventions to children with disadvantaged backgrounds. This reasoning also helps us understand the literature evaluating the effects of education interventions, where the typical result is that the effects are larger for pupils from low-income families than for pupils from high-income families.[1]

Two previous studies have used credible identification methods to examine parental responses to education interventions. Das et al. (2013) use experimental variation in primary school grants in India and find that household education spending is unaffected when the grant is unanticipated and goes down when when the grant is anticipated. Pop-Eleches and Urquiola (2013) use a regression discontinuity design to assess the effects and behavioral responses of being admitted to a better high school in Romania. They find that parents reduce effort when their children attend a better school.

Relative to these two papers we make two main contributions. First, and as noted above, we focus explicitly on whether the responses of parents to an educational intervention vary with parental income. Second, we use a credible identification method to examine responses of parents to class size variation, a policy tool that policy makers can control directly.[2] [3]

---

[1]In the context of the class size literature, see for example: Summers and Wolfe (1977), Angrist and Lavy (1999), Krueger (1999), and Heinesen (2010).

[2]See, amongst others: Angrist and Lavy (1999); Krueger (1999); Hoxby (2000); Krueger and Whitmore (2001); Finn et al. (2003); Urquiola (2006); Fredriksson et al. (2013).

[3]Two previous studies have looked at responses of parents to class size variation. Bonesrønning (2004)

To structure our analysis we develop a model in the spirit of Becker and Tomes (1986) and Todd and Wolpin (2003). Parents derive utility from own consumption and the future income of their children. They choose the levels of private investments in their children and of assets transferred to their children. Assuming that parents cannot leave debt to their children, the responses among unconstrained parents (i.e. high-income parents) identifies whether public and private investments are complements or substitutes. If they are substitutes, high-income parents reduce their private investments more in response to an increase of public investments than low-income parents under plausible assumptions. This may, in turn, translate into a smaller total impact of an increase of public investments on the human capital of high-income children than on the human capital of low-income children.

In the empirical analysis we study how parents' help with homework and their decision to move their child to another school respond to variation in class size. We present estimates for the entire sample and separate estimates for children from families with below and above median income. To get a sense of what is going on when class size is changed, we also look at the impact on the classroom environment and on teacher perceptions.

The source of variation we exploit stems from a maximum class size rule in Swedish upper 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. For these cohorts we have access to rich data from questionnaires among 5% and 10% samples providing information about the behavior of teachers, students and parents. We supplement this information with register data. In Fredriksson et al. (2013), we have used data from these cohorts to show that larger classes in upper 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.

We find that only high-income parents respond to larger classes by helping their children more with homework.[4] This is clear evidence that parental (time) investments and public investments are substitutes in the human capital production function. We also find that parents

---

uses a maximum class size rule in Norwegian lower secondary schools for identification. He finds that parental effort decreases when class size increases. However, since he fails to control for the forcing variable (school enrollment at the grade level), the results may be contaminated by parental sorting into schools of different sizes. Datar and Mason (2008) use panel data from kindergarten and 1st grade students in the U.S. to examine whether class size has an impact on parental involvement. Larger classes are associated with less parent–child interaction, the same level of parent–school interaction, and more parent-financed activities. They instrument actual class size with average class size at the grade level. The results are potentially contaminated by sorting across schools. Overall, we think the identification in these two studies is not entirely credible.

[4]This effect is not driven by changes in the frequency of homework. Pupils do not do more homework, and teachers do not change their views vis-a-vis homework, when exposed to a larger class.

respond to larger classes by moving their child to another school. This is evidence that public and private resources are substitutes in a broader sense. The magnitude of the mobility response is greater among low-income households (although not significantly so), perhaps because they are not able to provide parental help as effectively as high-income households. Furthermore, we document that teachers seem to assign more responsibility to students for their own learning in larger classes. While this response does not vary by social background, we believe that the effect of less teacher-student interaction is more detrimental for low-income students than for high-income students. Consistent with this, we show that students from low-income families find their teachers hard to follow when they explain in front of a large class.

The above results potentially provide an explanation for why the total effect of a class size intervention on achievement is greater among low-income children. In line with previous results, we find that a reduction of class size in upper primary school by one student during three years, increases academic achievement of students from below median income families by 6.6 percent of a standard deviation while this effect is only 2.1 percent of a standard deviation for students from high-income families.

The paper proceeds as follows. The next section presents the conceptual framework. Section 3 describes the relevant institutions of the Swedish schooling system. Sections 4 and 5 describe the data and the estimation strategy. Sections 6-8 present and discuss the results. Section 9 summarizes and concludes.

2 Conceptual framework

We set up a stylized model in the spirit of Becker and Tomes (1986) to analyze parental responses to changes in public investments in human capital. We focus on differential responses of different types of families. Let the human capital of the child be given by:

$$h = h(s, p, \theta) \tag{1}$$

where s denotes public investments, p private investments and θ the initial level of human capital. Human capital is increasing in each of the three arguments (i.e., $h_j > 0$, $j = s, p, \theta$). There are decreasing returns in public and private investments, $h_{ss} < 0$ and $h_{pp} < 0$, but (initially) we make no assumptions about the signs of the cross partials ($h_{sp}, h_{s\theta}, h_{p\theta}$).

We think of private investments in human capital as being investments in time for the parent. The "price" for investing in the child is then $\hat{p}(\theta_f)w_f\theta_f$, where $w_f\theta_f$ denotes the

4

earnings of a parent with human capital θ_f. More skilled parents have higher opportunity costs of time, but need less time to produce a given amount of the private investment, i.e., $\hat{p}'(\theta_f) < 0$. In practice, we define p in efficiency units and, hence, $p = \hat{p}\theta_f$.

Each parent is assumed to have one child. The preferences of the parent are given by

$$U = \ln c_f + \ln y$$

The parent thus cares equally about own consumption (c_f) and the (future) income of the child (y). The budget constraint for the parent is given by:

$$c_f = (1 - \hat{p})w_f\theta_f - a$$

where a denotes assets transferred to the child. The income of the child is given by:

$$y = wh + (1 + r)a$$

Parents can thus transfer resources in two ways: either by directly investing in child human capital or by financial investments. There are two additional non-negativity constraints, i.e., $a \geq 0$ and $p \geq 0$. We thus assume that parents cannot leave debt to their children. Also, $p = 0$ is a real possibility in our empirical setting, since we look at children who are 10-13 years old.

If we let $MRS = y/c_f$ denote the marginal rate of substitution between the parent's consumption and the child's consumption opportunities, the first-order conditions to this problem are given by:

$$U_p = \frac{w}{w_f}\frac{\partial h}{\partial p} - MRS \leq 0, \; p \geq 0, \; p \cdot U_p = 0 \tag{2}$$

$$U_a = (1 + r) - MRS \leq 0, \; a \geq 0, \; a \cdot U_a = 0 \tag{3}$$

It is instructive to define the marginal return to private investments as $1 + \rho(s, p, \theta) = \frac{w}{w_f}\frac{\partial h}{\partial p}$. Equations (2)-(3) then say that unconstrained parents invest in child human capital to the point where the marginal return to the investment $\rho(s, p, \theta)$ equals the financial return r. In what follows we compare the investment responses to changes in public resources among three types of families: (i) unconstrained parents ($a > 0$, $p > 0$); (ii) partially constrained parents ($a = 0$, $p > 0$); and (iii) constrained parents ($a = 0$, $p = 0$).

Unconstrained parents. Wealthier parents (who have high θ_f) are less likely to be constrained if the marginal rate of substitution is falling in parental skill. Let $\bar{\theta}_f$ denote the marginal parent, i.e., a parent that optimally sets $a^* = 0$. We have

$$\left.\frac{dMRS}{d\theta_f}\right|_{\theta_f=\bar{\theta}_f} = -\frac{MRS}{\bar{\theta}_f}\left[1 - \eta_\theta \beta + \frac{\hat{p}}{1-\hat{p}}\right] \tag{4}$$

where $\eta_\theta \equiv (\partial h/\partial \theta)(\theta/h) > 0$ is the elasticity of the human capital production function with respect to child skills and $\beta \equiv (\partial \theta/\partial \theta^f)(\theta^f/\theta)$ is the intergenerational elasticity of skills (we thus allow the child's initial human capital to depend on parental skill). A sufficient condition for this derivative to be negative is that $\eta_\theta \beta < 1$; $\eta_\theta \beta < 1$ means that parental earnings are more directly related to parental skill than child earnings. Since this seems like a very reasonable condition, we impose it. Thus, unconstrained parents are more skilled, i.e., $\theta_f \geq \bar{\theta}_f$. Let p^* denote the solution to the investment problem for unconstrained parents.

From (2)-(3) it follows that the investment response is given by

$$\frac{\partial p^*}{\partial s} = -\frac{\rho_s}{\rho_p}$$

where $\rho_p < 0$ because of decreasing returns and $\text{sign}(\rho_s) = \text{sign}(h_{ps})$. Thus if private and public investments are substitutes, private investments fall (rise) when there is an unexpected increase (decrease) in public investments.

Partially constrained parents. Now, consider partially constrained parents. Let the solution to their problem be denoted $p^c > 0$ and remember that $\theta_f < \bar{\theta}_f$. A partially constrained parent equates the gross rate of return to private investments with the marginal rate of substitution and, therefore, $p^c < p^*$. The response among partially constrained parents thus also reflects how the *MRS* is affected by changes in s and p. The investment response among partially constrained families is given by:

$$\frac{\partial p^c}{\partial s} = -\frac{\rho_s - \partial MRS/\partial s}{\rho_p - \partial MRS/\partial p}$$

Since $\partial MRS/\partial p > 0$, the denominator is negative $((\rho_p - \partial MRS/\partial p) < 0)$. If $\rho_s < 0$, private investments fall in response to an increase in public investments (since $\partial MRS/\partial s > 0$). If $\rho_s > 0$, $(\partial p^c/\partial s)$ is ambiguous in sign.

Constrained parents. There is also the possibility that parents are at a corner regarding their private investments. This is the case if

$$1 + \rho(s, p, \theta) < MRS(s, p, \theta, \theta_f)$$

The right-hand-side is decreasing in θ_f (see equation (4)). The derivative of the left-hand-side with respect to θ_f is given by

$$\frac{d(1+\rho)}{d\theta_f} = \frac{1+\rho}{\theta_f} \varepsilon_\theta \beta$$

where $\varepsilon_\theta \equiv (\partial\rho/\partial\theta)(\theta/(1+\rho))$ is the elasticity of the gross return with respect to child skills. If skill-begets-skill, $\varepsilon_\theta > 0$, the gross return unambiguously increases in parental skill.

Let $\underline{\theta}_f$ denote the parent who optimally chooses $p^c = 0$. The sign of

$$\Omega(\underline{\theta}_f) = \left.\frac{d(1+\rho)}{d\theta_f}\right|_{\theta_f=\underline{\theta}_f} - \left.\frac{dMRS}{d\theta_f}\right|_{\theta_f=\underline{\theta}_f}$$

determines which parents are more likely to be at a corner. If $\Omega > 0$ more skilled parents are more likely to set $p^c > 0$. We obtain

$$\Omega(\underline{\theta}_f) = \frac{1+\rho}{\underline{\theta}_f} \{\beta(\varepsilon_\theta - \eta_\theta) + 1\} \tag{5}$$

Now, we argued above that $\eta_\theta \beta < 1$. Therefore, $\varepsilon_\theta > 0$ implies that less-skilled parents are more likely to be at a corner.

A CES-example. To make some headway on (5), let us assume a CES production structure

$$h = \theta^{1-\mu} R^\mu$$

where the child's initial level of human capital (θ) and a CES-index in resources (R)

$$R = [\lambda p^\phi + (1-\lambda)s^\phi]^{1/\phi}$$

are inputs in the production of human capital. As usual $\phi \leq 1$ governs the elasticity of substitution ($1/(1-\phi)$). If $\phi = 1$, public and private investments are perfect substitutes. This production structure is sufficiently general for most of our purposes. It does invoke one

important additional assumption, however: the initial level of human capital and resources are complements. This implies that the return to private investment is increasing in θ (i.e. $\varepsilon_\theta \equiv (\partial\rho/\partial\theta)(\theta/(1+\rho)) > 0$). In particular $\varepsilon_\theta = \eta_\theta$ and, therefore, it is unambiguously the case that low-skilled parents are more likely to opt for the corner solution $p^0 = 0$.

Under the CES production structure we have that the investment response among unconstrained parents is given by

$$\frac{\partial p^*}{\partial s} = -\left(\frac{1-\lambda}{\lambda}\right)\left(\frac{(\phi-\mu)\kappa}{1-\phi+\kappa(\phi-\mu)}\right)\left(\frac{p}{s}\right)^{1-\phi}$$

where $0 \le \kappa = \frac{\lambda p^\phi}{\lambda p^\phi + (1-\lambda)s^\phi} \le 1$ is a measure of the relative effectiveness of private investments in producing skills. If $\phi > \mu$ ($\phi < \mu$) private and public investments are substitutes (complements). The investment response among partially constrained parents is given by:

$$\frac{\partial p^c}{\partial s} = -\left(\frac{1-\lambda}{\lambda}\right)\left(\frac{\phi\kappa}{1-\phi+\kappa(\phi+\mu)}\right)\left(\frac{p}{s}\right)^{1-\phi}$$

If $\phi > 0$, parents respond by reducing their investment. However, it is generally not possible to determine whether unconstrained parents respond more or less than partially constrained ones.[5]

The main message of this simple framework is that the responses among unconstrained families provide information on whether public and private resources are complements or substitutes in the production sense. In our empirical work we provide evidence consistent with the view that they are substitutes. If they are substitutes, parents (with positive investment) compensate for an increase (decrease) in public resources by reducing (increasing) their own investment.

If we take the CES production structure seriously and assume that private and public investments are substitutes, we can categorize parents in a way that is in broad accordance with our empirical findigs:

$$\begin{array}{lll} \theta_f < \underline{\theta}_f: & p^0 = 0 & \frac{\partial p^0}{\partial s} = 0 \\ \underline{\theta}_f \le \theta_f < \bar{\theta}_f & p^c > 0 & \frac{\partial p^c}{\partial s} < 0 \\ \bar{\theta}_f \le \theta_f & p^* > p^c & \frac{\partial p^*}{\partial s} < 0 \end{array}$$

In other words, more skilled parents have a higher probability of responding to a change in public investments.

---

[5]If $\phi \to 1$, however, $\frac{\partial p^*}{\partial s} < \frac{\partial p^c}{\partial s}$, unambiguously.

Direct and total effects on achievement. If $\frac{\partial p}{\partial s} \neq 0$ there is a wedge between the direct effect ($\frac{\partial h}{\partial s}$) and the total effect ($\frac{dh}{ds}$) of public investments on student outcomes (Todd and Wolpin, 2003; Pop-Eleches and Urquiola, 2013). The relation between these two effects is given by:

$$\frac{dh}{ds} = \frac{\partial h}{\partial s} + \frac{\partial h}{\partial p}\frac{\partial p}{\partial s} \tag{6}$$

Since $\frac{\partial h}{\partial p} > 0$, the total effect will exceed the direct effect in case private and public investments are complements and fall short of the direct effect in the case private and public investments are substitutes.

If the direct effect of public investments is the same for children from high-skilled and low-skilled parents ($h_{s\theta}\frac{\partial \theta}{\partial \theta_f} = 0$), the total effect of public investments will be smaller for children from high-skilled parents than for the children from low-skilled parents in case private and public investments are substitutes. In the CES-example, $h_{s\theta}\frac{\partial \theta}{\partial \theta_f} > 0$; therefore, we cannot unambiguously determine the relative sizes of the total effects across the distribution (see Appendix).

3 Institutional background[6]

Here 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. Compulsory schooling was (and still is) 9 years. The compulsory school period was divided into three stages: lower primary school (age 7-10, grades 1-3), upper primary school (age 10-13, grades 4-6) and lower secondary school (age 13-16, grades 7 to 9).

The compulsory school system had several organizational layers. The primary unit was the school. Schools were aggregated to school districts (note that these school districts are very different from U.S. school districts).[7] School districts typically had one lower secondary school and at least one primary school. The catchment 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

---

[6]This section follows Fredriksson et al. (2013) closely; see also Rietz et al. (1987).

[7]We use the term "school district" for want of a better word. The literal translation from Swedish would be "principal's district" (*Rektorsområde*). 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.

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 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.[8]

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.[9]

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. The reason for this focus is that our primary source of data are surveys that were administered when the students were in 6th grade.

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 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

---

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

[9]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.

[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).

[12]In Fredriksson et al. (2012) 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.

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 not adjusted in response to enrollment levels. Here we focus on districts containing one upper primary school, which we refer to as one-school districts.[13] We provide evidence that the RD design at the school district level is valid in Subsection 6.1.

4 Data

The primary data source for our study is the so-called ETF-project which is run by the Department of Education at Göteborg University; see Härnquist (2000) for a detailed description. This data source contains measures of school performance of students in the final year of upper primary school for roughly a 10 percent sample of the cohorts born in 1967, 1972, and 1982, and a 5 percent sample for the cohort born in 1977. The ETF-project also contains answers from parents, pupils and teachers to questions about issues related to the school. These questionnaires were distributed when the pupils were in 6th grade (age 13). Based on this information we constructed variables that measure: parental responses (e.g. help with homework), elements of the learning environment (e.g. whether it is easy to understand the teacher), and amount of homework.[14]

The data contain several measures of pupil performance at age 13. We collapse these into academic achievement and cognitive ability. Academic achievement is based on achievement tests in Swedish and mathematics. Cognitive ability is based on traditional "IQ-type" tests with logical and verbal elements.

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.[15] In the second stage, classes were randomly sampled within municipality. This sampling procedure implies that it is unlikely that comparisons across municipalities for a given cohort are valid, but comparisons within municipalities are likely

---

[13]In Fredriksson et al. (2013) we compared the results for one-school districts with the results for all school districts (which includes districts with more than one school). The magnitudes of the coefficient estimates are basically the same, but the estimates for all school districts are less precise. The reason for this is that the maximum class size rule has less bite in districts with more than one school. Table B1 in the Online Appendix compares one school districts to all districts in the sample we use. One-school districts tend to be located in cities to a greater extent. But in all other respects they are very similar to other school districts.

[14]See Table A1 for a detailed description of the creation of all variables. This table shows which questions were included in the questionnaires for the respective cohorts.

[15]Table B1 in the Online Appendix compares the districts in the ETF-data to all districts in the country. This shows that the districts in the ETF-data are very similar to all districts in the country.

valid. For this reason all analyses condition on municipality-by-cohort fixed effects.

To the ETF-data we have matched information from the Class Register, the Educational Register, and the Income Tax Register, all maintained by Statistics Sweden. From the Class register we get information on actual class size and school enrollment by grade (the assignment variable). Using links between parents and children we get information on parental income and education from the Income Tax Register and the Educational Register.

Figure 1 shows the distribution of enrollment in grade 4 across one-school districts. In total there are 205 districts with one school in our data.[16] Therefore, the distribution is "thin" at certain points. It is especially thin around the 4th threshold (enrollment of 120), which implies that this threshold does not contribute to the identification of the class size effects. We therefore restrict the sample to districts with 4th grade enrollment of at most 120 students (which excludes 5 schools), and are thus exploiting the variation in class size induced by the thresholds at 30, 60 and 90. The fact that there are few school districts in our data restricts what we can do with the data. For example it is not meaningful to estimate separate effects of class size at the different thresholds. Given the amount of information we have in the data, we must pool information across the different thresholds. Section 5 outlines our approach.

Table A2 in the Appendix reports descriptive statistics of all variables, for all individuals together and broken down by parental income. In square brackets below each variable the table also reports: the number of of students; the number of districts×cohorts; and the number of enrollment counts. Following Lee and Card (2008), we cluster the standard errors by enrollment in the regression discontinuity analyses.

Figure 2 shows the distribution of actual class size in grade 4. Very few classes are below the minimum of 15 and none is above the official maximum of 30.

5   Estimation strategy

The maximum class size rule generates thresholds at each multiple of 30 and thereby multiple discontinuities. In Fredriksson et al. (2013) we pooled the different thresholds using the analogy of aggregating several small scale experiments. Another, and more efficient, way of pooling information from different thresholds is to define a separate instrument at each threshold (see Angrist and Lavy (1999) and the web-appendix of Fredriksson et al. (2013)). With enrollment segment fixed effects (but no other covariates) this produces a minimum variance linear combination of Wald estimates from each threshold. This is the approach

---

[16]Since some of the outcome variables are not available for all cohorts, the number of districts×cohorts is sometimes smaller than 205.

Figure 1. Distribution of enrollment in grade 4 in one-school districts
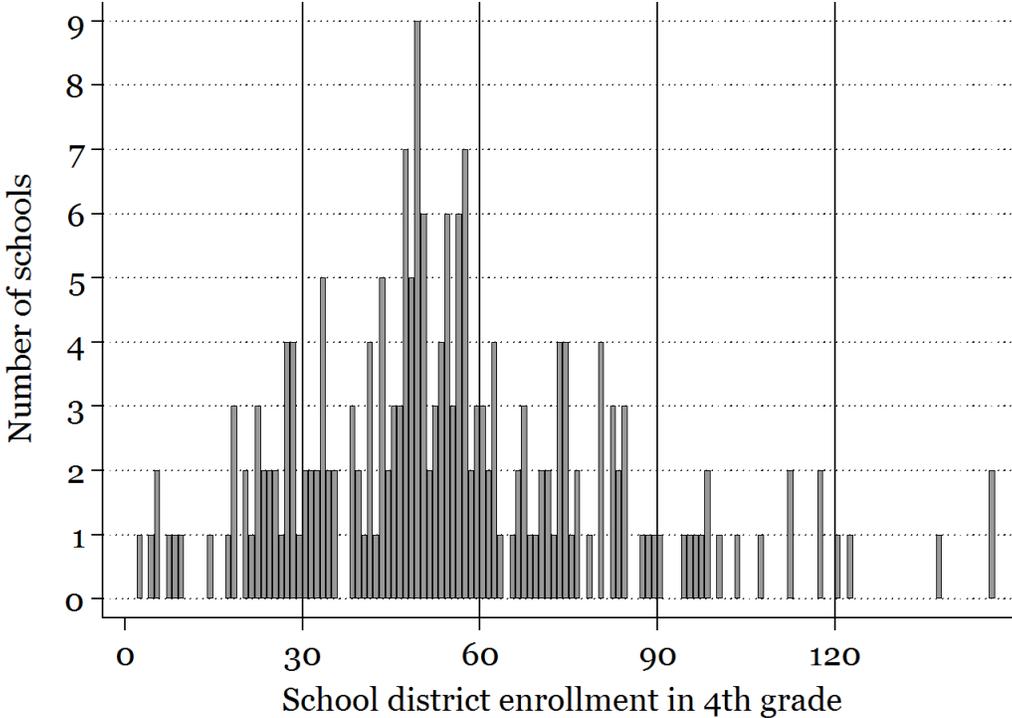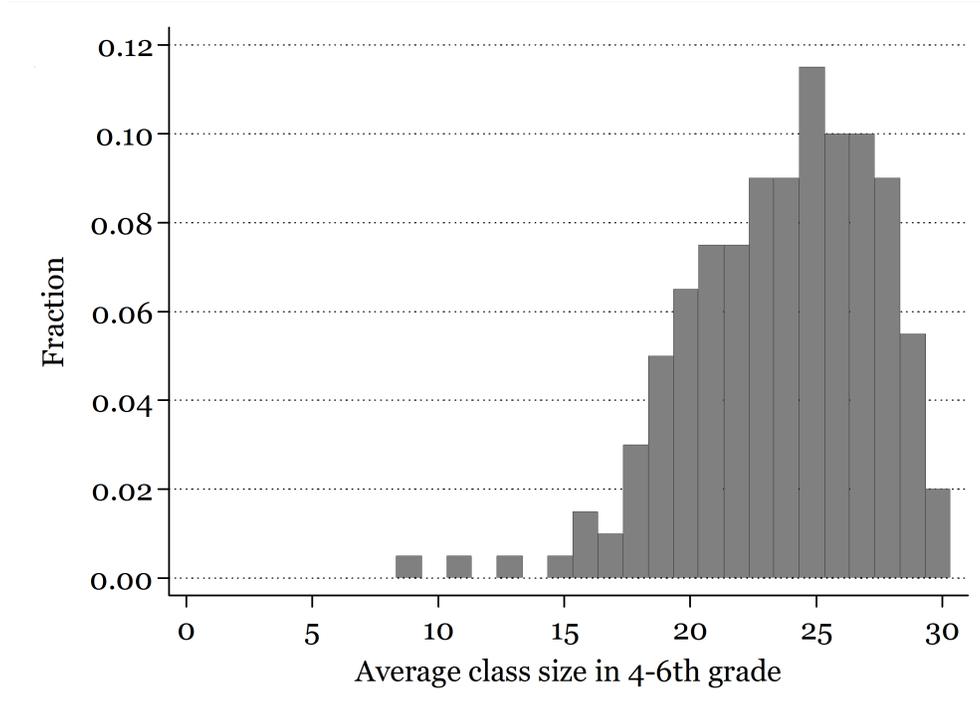
Figure 2. Distribution of class size in grades 4-6

we take here. In the appendix we demonstrate that the two approaches deliver very similar results; see Table A3. With separate instruments at each threshold, the estimates are more precise, however.

The RD-design also involves choices of bandwidths and of the control function for the assignment variable. To formalize our estimation approach we introduce the following notation: y_{id} is the outcome for student i in district d, e_d is district enrollment in the fourth grade, CS_d is average class size in grades 4 to 6 in the district, and Z_{td} is a dummy variable equal to 1 if 4th grade enrollment in the district is above the t-th threshold, with $t = \{1,2,3\}$. The equations that we estimate are (omitting further control variables):

$$y_{id} = \beta CS_d + f(e_d) + \alpha_\tau + \varepsilon_{id} \tag{7}$$

and

$$CS_d = \gamma_1 Z_{1d} + \gamma_2 Z_{2d} + \gamma_3 Z_{3d} + g(e_d) + \delta_\tau + v_d \tag{8}$$

where β is the effect of interest, γ_1, γ_2 and γ_3 are the first stage effects, $f(e_d)$ and $g(e_d)$ are functions of enrollment, α_τ and δ_τ are fixed effects for five different enrollment segments (1-15; 16-45; 46-75; 76-105; 106-120), and ε_{id} and v_d are the error terms. Due to the inclusion

14

of the segment fixed effects, the instrumental variables (Z_{td}) effectively contrast observations at most 15 students above a threshold with observations at most 15 students below a threshold. Observations on the first (1-15) and last (106-120) segments only contribute to the estimation of the enrollment controls.

The preferred specification of the control function should capture any direct effect of school enrollment on the outcome. Because of the relatively small numbers of observations available for most of the outcome measures, however, we are restricted in the flexibility of the control functions. In our main specification we control for enrollment linearly, and let the slopes vary across the three thresholds. Table A4 shows what happens to the results with alternative ways of controlling for enrollment. Overall, the results are fairly robust to alternative ways of controlling for the assignment variable.[17] Tables B8-B10 in the Online Appendix show that the results are robust to reducing the bandwidth to +/-10 pupils.

In all our analyses the endogenous variable is the average of the class sizes a student experiences in grades 4, 5 and 6, while the instruments are derived from enrollment in grade 4. There are two reasons for this. The first reason is that enrollment levels 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.[18] 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 estimates reflect

---

[17]Table A4 column (3) presents the results with linear control for school district enrollment interacted with threshold. The specifications in columns (1) and (2) of Table A4 are too restrictive in our view since the enrollment control function is not interacted with threshold. In general, it seems that the 2nd order polynomial control function provides too much flexibility relative to the amount of variation that we have in the data. When we move to a 2nd order polynomial, the F-statistic for instrument relevance drops substantially (although the instruments are never weak in the conventional sense); see columns (4) and (6) of Table A5. The relevant contender to the specification in column (3) is the specification in column (5). In column (5), the enrollment controls are also allowed to vary by segment. Notice that results are similar with this specification. Tables B4 and B5 in the Online Appendix show that this is also true within the subsamples of low-income and high-income students.

[18]In Section 7.2 we show that parents respond to increases class size by moving to another school, implying that enrollment in grades 5 and 6 is endogenous.

the effects of an increase of class size by one student during three years. This interpretation is clear-cut when the outcome variable is school performance measured in grade 6, which measures the accumulated skills and knowledge of the student. This is potentially not the case for other outcomes we consider such as "help with homework" or finding the teacher "easy to understand", which are also measured in grade 6. Here it may be more natural to relate them to class size in grade 6. However, this only amounts to rescaling the estimates; the precision of the estimates is driven by the precision of the reduced form. We prefer to rescale the estimates using the same factor to ease interpretation.

6 Balancing, first stage, and achievement

This section starts by presenting the results from balancing tests and of the first stage (Section 6.1). It then continues with the estimates of the impact of class size on cognitive outcomes at age 13 (Section 6.2). Section 7 examines the responses of parents while Section 8 looks at the impact on the class room environment.

6.1 Balancing and first stage

A threat to the validity of the RD design is bunching on one side of the cut-offs, since that indicates that the assignment variable is manipulated; see McCrary (2008).[19] Figure 1 shows the density of school district enrollment. Visual inspection reveals no suspect discontinuities in the distribution of the assignment variable. The McCrary (2008) density test confirms this: we cannot reject the hypothesis that the density is smooth at the discontinuity.[20]

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. Table 1 addresses this issue. Column (1) shows that the baseline covariates are highly relevant predictors of academic achievement at age 13. 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 0.058 standard deviation units). We strongly reject the hypothesis that the pre-determined covariates do not predict this outcome; the p-value of the F-test equals 0.0000.

The next three columns in Table 1 show the results of regressing each of the three instruments on all baseline covariates. Only one out of 27 coefficients is significantly different

---

[19]Urquiola and Verhoogen (2009) document an extreme example of bunching in the context of a maximum class size rule in Chile.

[20]To implement the test we pooled thresholds and 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.

from zero.[21] The fourth to last row reports the result of F-tests of the hypothesis that all the coefficients on baseline covariates are jointly zero. The p-values range from 0.32 to 0.76 and the message thus is that pre-determined characteristics are unrelated to the instruments.[22]

To illustrate balancing further we conducted the following exercise. On the basis of the estimates in column (1) of Table 1 we calculated predicted achievement scores. Predicted scores is thus an overall measure that takes the importance of each of the baseline characteristics into account. Figure 3 illustrates that predicted scores are balanced across all three thresholds. Moreover, and importantly, Figure 4 shows that predicted achievement scores are balanced across thresholds, separately for the low-income and the high-income group.[23]

Figure 5 shows the first stage graphically. There are clear, and statistically significant, jumps at all three thresholds. Districts that have just surpassed one of the thresholds have classes that are systematically smaller than districts just below the thresholds. When enrollment surpasses the first threshold, class size drops by 10 students, At the second threshold, as well as the third, class size falls by 5 students.[24]

6.2 Pupil achievement

We start by presenting estimates of the effect of class size on student performance measured at age 13 at the end of upper primary school. This is the short-term policy effect of class size on student performance, captured by the term dh/ds in equation (6).[25] Figure 6 gives a graphical illustration, and shows that at each threshold academic achievement jumps up when a threshold is surpassed.

---

[21]We cluster the standard errors by enrollment counts (81 clusters). However, when studying separate thresholds as in Table 1, the number of clusters contributing to identification is much fewer. In particular, there are only 18 clusters around the third threshold, which may produce too small standard errors.

[22]We have also performed another balancing test. We have asked whether surpassing one of the thresholds in upper primary school is related to class size in lower primary school, which is predetermined relative to class size in upper primary school. We cannot reject the hypothesis that class size in lower primary school is unrelated to the instruments (p-value = 0.646).

[23]Tables B2 and B3 in the Online Appendix show balancing results, akin to Table 1, separately for low-income and high-income students, respectively.

[24]Table A5 shows estimates of the first stage for various specifications of the control function. Tables B6 and B7 report first stage results for the subsamples of low-income and high-income students. In all cases the indicator for being above the third threshold becomes statistically insignificant in the most flexible specifications. The indicators for being above the first and second threshold are statistically significant throughout.

[25]These estimates are comparable but not the same as results reported in Fredriksson et al. (2013). Differences are due to inclusion of small school districts, a slightly different empirical strategy and a split of the sample into below and above median income parents. Table A3 in the Appendix shows estimates of the effects of class size on various short-term and long-term outcomes for the specification used in Fredriksson et al. (2013) and for the specification used in this paper. Table B11 in the Online Appendix reports the same results broken down by students from low and high-income families.

Figure 3. Predicted achievement scores by enrollment in grade 4
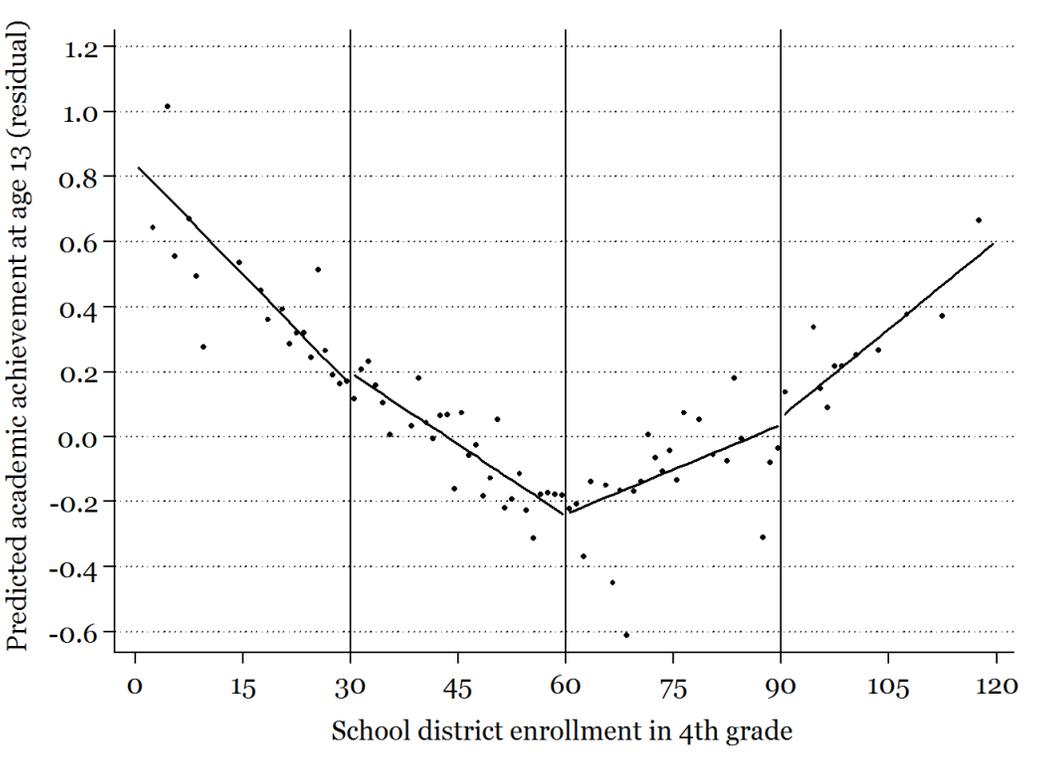
Notes: The figure shows residuals scores, after controlling for fixed effects for enrollment segments and municipality-by-cohort fixed effects, by enrollment in grade 4. The data pertain to one-school districts for cohorts born 1967, 1972, 1977 and 1982. The regression lines were fitted to individual data. Discontinuities at thresholds: 0.041 (0.058), 0.018 (0.096), 0.029 (0.134). P-value of the F-test for excluding the three instruments 0.905.

Figure 4. Predicted achievement scores for children with low and high-income parents by enrollment in grade 4
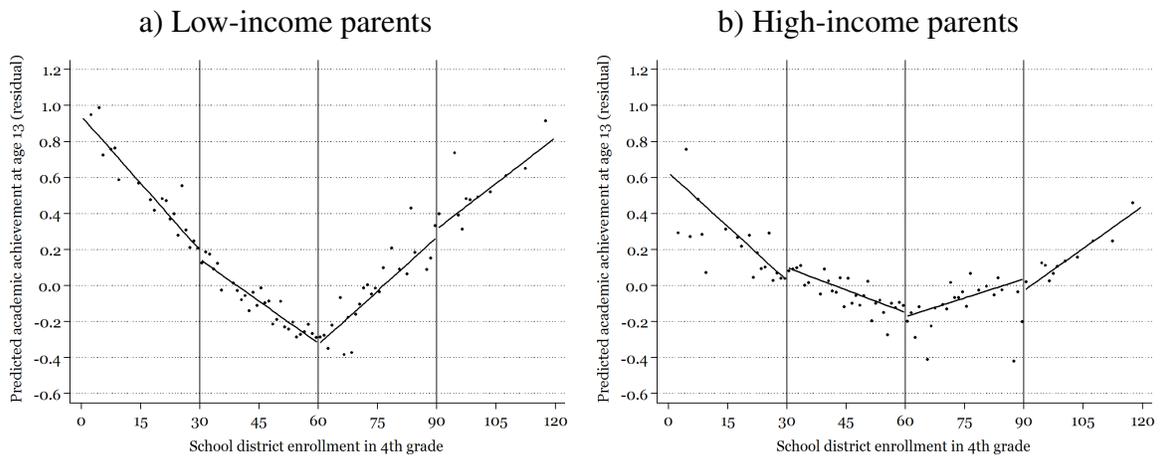
Notes: The figures show residuals scores, after controlling for fixed effects for enrollment segments and municipality-by-cohort fixed effects, by enrollment in grade 4. The data pertain to one-school districts for cohorts born 1967, 1972, 1977 and 1982. The regression lines were fitted to individual data. Panel a) shows achievement scores by enrollment for low-income parents (income below the median in the population); discontinuities at thresholds: -0.046 (0.043), 0.010 (0.063), 0.045 (0.101); p-value of the F-test for excluding the three instruments 0.731. Panel b) shows achievement scores by enrollment for high-income parents (income above the median in the population); discontinuities at thresholds: 0.078 (0.057), -0.016 (0.075), -0.060 (0.097); p-value of the F-test for excluding the three instruments 0.554.

Table 1. Balancing of covariates

	Academic achievement	Above 1st threshold	Above 2nd threshold	Above 3rd threshold	p-value
	(1)	(2)	(3)	(4)	(5)
Female	0.0474	0.0013	-0.0023	0.0038***	0.0814
	(0.0321)	(0.0017)	(0.0034)	(0.0013)	
Month of birth	-0.0236***	-0.0008	0.0005	-0.0002	0.2957
	(0.0041)	(0.0005)	(0.0005)	(0.0002)	
Immigrant	-0.2350***	0.0025	0.0112	-0.0032	0.9725
	(0.0631)	(0.0056)	(0.0178)	(0.0050)	
Mother's years of education	0.0583***	-0.0002	0.0009	0.0002	0.9558
	(0.0065)	(0.0006)	(0.0010)	(0.0005)	
Father's years of education	0.0462***	0.0004	-0.0015	0.0003	0.9276
	(0.0068)	(0.0006)	(0.0013)	(0.0006)	
Family income (SEK 100,000s)	0.0383***	0.0006	0.0020	-0.0012	0.8394
	(0.0098)	(0.0013)	(0.0018)	(0.0011)	
Mother's age at birth	0.0107***	-0.0001	0.0004	-0.0005	0.6553
	(0.0027)	(0.0003)	(0.0007)	(0.0003)	
Number of siblings	-0.0478***	-0.0005	-0.0003	-0.0013	0.6755
	(0.0120)	(0.0010)	(0.0022)	(0.0012)	
Parents separated, age 9	-0.1350***	0.0007	-0.0019	-0.0055	0.7142
	(0.0331)	(0.0039)	(0.0058)	(0.0045)	
p-value of F-test	0.0000	0.7575	0.3710	0.3170	
Number of students	4,707	6,009	6,009	6,009	
Number of districts×cohorts	187	200	200	200	
Number of clusters	79	81	81	81	

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982 in one-school districts. Academic achievement is standardized. Columns (1) – (4) report results of OLS regressions on the variables listed in the rows. These regressions also include the following control variables: fixed effects for enrollment segment, linear controls for school district enrollment interacted with threshold, and municipality-by-cohort fixed effects. Above 1st/2nd/3rd threshold are indicators equaling unity if school district enrollment in 4th grade exceeds the 1st/2nd/3rd threshold of the class size rule. Independent variables are pre-determined parent and student characteristics. The p-value reported at the bottom of the columns is for an F-test of the joint significance of the variables listed in the table. Each row of column (5) 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)-(4). The p-value is for a t-test of the significance of the class size instruments. Standard errors adjusted for clustering by enrollment count are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 percent level, respectively.
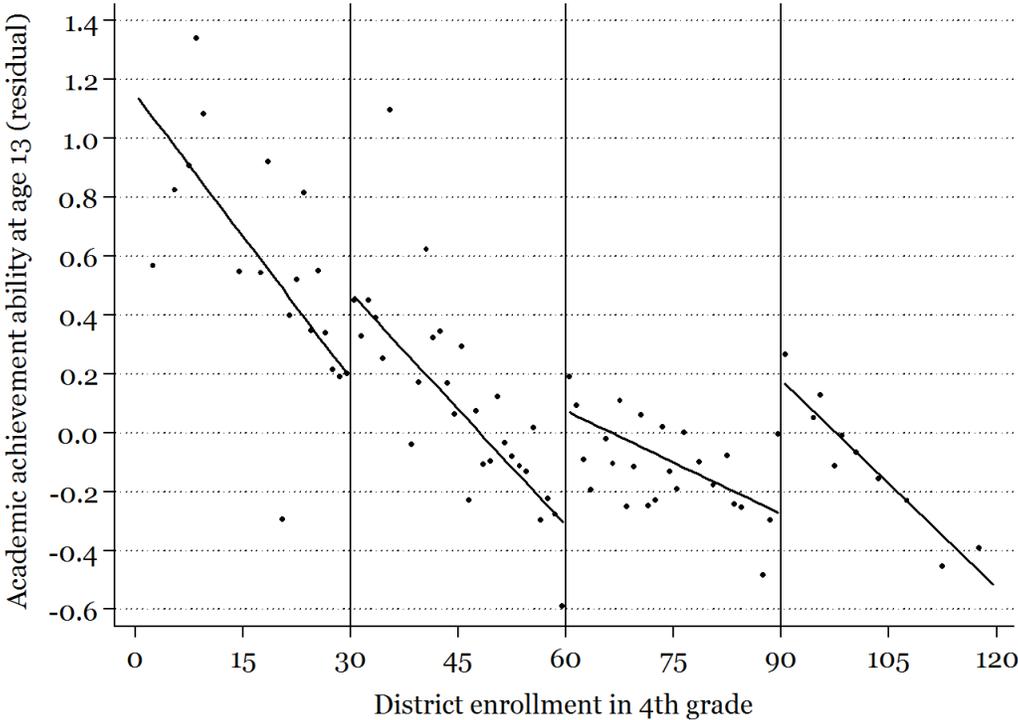
Figure 5. Class size by enrollment in grade 4
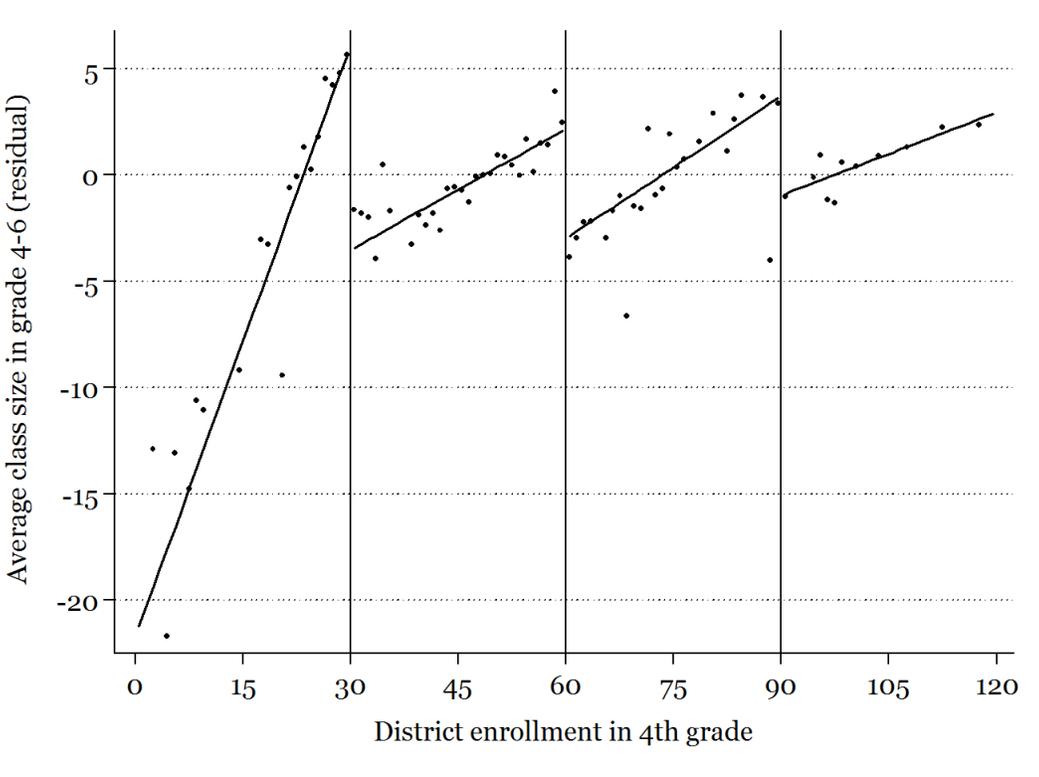
Notes: The figure shows residual class size, after controlling for municipality by cohort fixed effects, by 1-student bins. The regression lines were fitted to individual data. Discontinuities at thresholds: -9.984 (0.925), -5.137 (1.055) and -4.712 (1.414).

Figure 6. Academic achievement by enrollment in grade 4

Notes: The figure shows residual academic achievement at age 13, after controlling for fixed effects for enrollment segments, municipality-by-cohort fixed effects and baseline covariates, by enrollment in grade 4. The data pertain to one-school districts for cohorts born 1967, 1972, 1977 and 1982. The regression lines were fitted to individual data. Discontinuities at thresholds: 0.295 (0.149), 0.397 (0.121) and 0.448 (0.156).

Table 2. IV-estimates of class on school performance

Outcome [# pupils; #districts×cohorts; clusters]	All (1)	Income parents		Difference (2) and (3) (4)
		low (2)	high (3)	
Academic achievement, age 13	-0.0441***	-0.0661***	-0.0203	-0.0458*
[4,707; 187; 80]	(0.0122)	(0.0177)	(0.0149)	(0.0232)
Cognitive ability, age 13	-0.0268**	-0.0560***	-0.0025	-0.0534**
[5,197; 197; 82]	(0.0116)	(0.0151)	(0.0141)	(0.0206)

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982 in one-school districts. All ability measures are standardized. Average class size in grades 4-6 is instrumented by indicators for being above 1st, 2nd, or 3rd threshold of the class size rule. All models include the following controls for school district enrollment in grade 4: fixed effects for enrollment segment; linear controls for enrollment which are interacted with threshold. In addition all models include the following baseline controls: municipality-by-cohort 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 by enrollment count are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 percent level, respectively.

Table 2 presents the estimation results. For both outcome measures the effects of class size is significantly negative. A one student increase in class size during three years lowers academic achievement by 4.4 percent of a standard deviation and cognitive ability by 2.7 percent of a standard deviation. Splitting the sample by parental income, reveals that the negative effects are concentrated among students from low-income families.[26] For students from high-income families, the estimates have the expected negative sign but are not statistically significant. The larger negative effect of class size on the academic achievement of students from low-income families is consistent with results from other studies, including Angrist and Lavy (1999) and Krueger (1999). In the next section we ask whether the responses among parents contribute to this pattern.

7   Parental responses to class size

Are public and private investments in children complements or substitutes? Section 2 shows that the responses among high-income parents are proportional to the cross-derivative of the

---

[26]Splitting the sample by above and below median parental education (instead of income) gives very similar results, throughout.

Figure 7. Help with homework by enrollment in grade 4

a) Low-income parents	b) High-income parents
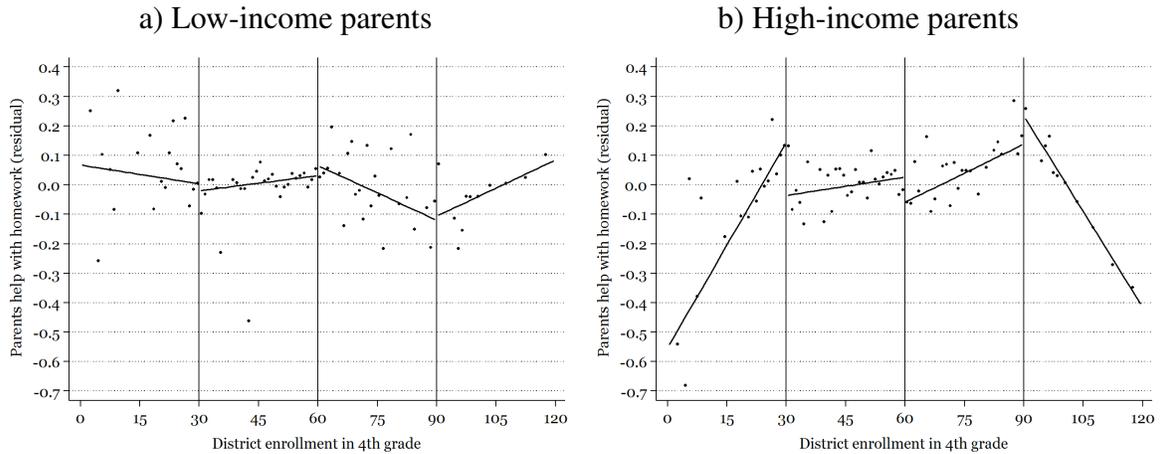
Notes: The figure shows residual "help with homework", after controlling for fixed effects for enrollment segments, municipality-by-cohort fixed effects and baseline covariates, by enrollment in grade 4. The data pertain to one-school districts for cohorts born 1967, 1972, 1977 and 1982. The regression lines were fitted to individual data. Panel a) shows parental help by enrollment for low-income parents (income below the median in the population); discontinuities at thresholds: -0.022 (0.058), 0.030 (0.048), 0.020 (0.122). Panel b) shows parental help by enrollment for high-income parents (income above the median in the population); discontinuities at thresholds: -0.192 (0.057), -0.083 (0.058), 0.081 (0.054).

production function. Thus, the response among high-income parents potentially identifies the answer to the above question in the most straightforward fashion.

We have access to two measures of parental responses. The first measure is whether parents help their children with homework (see Section 7.1). This is the measure that most closely resembles private investments in the model since it is a time investment on the part of parents. The second measure is the probability of moving the child to another school (see section 7.2)

7.1 Help with homework

Figure 7 plots parents' help with homework against enrollment in grade 4 for the two income groups. The left-hand panel clearly illustrates that low-income parents do not respond at all. Matters are different among high-income parents (see right-hand panel). In this group, there are clear drops in parental help at the first two thresholds. Since public resources increase when a threshold is surpassed (i.e. surpassing a threshold implies a class size reduction), this illustrates that parents adjust by reducing their (time) investment. At the third threshold this pattern is different, but the upward jump is not statistically significant .

24

Panel a) of Table 3 reports the estimated effects of class size on parental help with homework. The overall effect is significantly positive and indicates that a one student increase in class size increases the probability that a child receives help with homework by almost one percentage point. As the share of parents helping their children with homework in the sample equals 0.8, another way of phrasing the size of the effect is that a five student reduction in class size, increases the share of parents **not** helping their children with homework with 22 percent $((5 \times 0.0086)/(1 - 0.8))$. The results in columns 2 and 3 show that this effect is entirely caused by the response of high-income parents. In response to a class size reduction of five students, they are 43 percent $((5 \times 0.0154)/(1 - 0.82))$ more likely to not help their children with homework. Low-income parents do not respond at all. The difference across the two groups is statistically significant (see column 4).

Panel b) of Table 3 shows the effects on the amount of homework that students do. The amount of homework is clearly unaffected by an increase in class size. This is true on average as well as across the distribution. Consistently with this, teachers' views on homework are unaffected by an increase in class size. A regression with an indicator for whether teachers think homework is important for pupil learning as the dependent variable produces an estimate of -0.008 on class size, with a standard error of 0.010. Analogously, teachers do not change their views on exams when class size increases, which suggests that students do not spend more time at home preparing for exams in larger-sized classes.[27]

Table 3 thus implies that the response among high-income parents should be interpreted as an increase in the propensity to help with homework for a given amount of homework. The wider implication of this result is that public and private investments are substitutes in the human capital production function. According to the model in Section 2: it is for high-income (high-skilled) parents that we can identify the substitutability of these two input factors; moreover, the results are consistent with the view that low-income (low-skilled) parents are constrained by their (lack of) skills, implying that they are less efficient providers of parental help.

7.2 Moving to another school

Now let us turn to the mobility response to class size increases. Mobility is of course a form of parental investment. Nevertheless, it does not fit conveniently within our conceptual framework, since mobility is surrounded by other types of restrictions that we have not explicitly modeled. Remember that the school districts are defined on the basis of commuting distance.

---

[27]These effects do not vary across the parental income distribution.

25

Table 3. IV-estimates of class size on homework

Outcome		Income parents		Difference
[#pupils; #districts×cohorts; #clusters]	All	low	high	(2) and (3)
	(1)	(2)	(3)	(4)
a) Parental help				
Parents help child with homework	0.0086**	-0.0001	0.0154***	-0.0154**
[5,107; 197; 82]	(0.0043)	(0.0053)	(0.0046)	(0.0070)
b) Amount of homework				
Student does homework (hrs)	0.0071	0.0071	-0.0008	0.0079
[5,149; 197; 82]	(0.0183)	(0.0232)	(0.0232)	(0.0328)

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982 in one-school districts. Average class size in grades 4-6 is instrumented by indicators for being above 1st, 2nd, or 3rd threshold of the class size rule. All models include the following controls for school district enrollment in grade 4: fixed effects for enrollment segment; linear controls for enrollment which are interacted with threshold. In addition all models include the following baseline controls: municipality-by-cohort 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 by enrollment count are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 percent level, respectively.

Therefore, a change of school most likely implies moving house, and there are substantial transaction costs associated with a moving house. These transaction costs depend on housing tenure, where those living in rental housing (pre-dominantly low-income households) face lower transaction costs than those who own their house (pre-dominantly high-income households).

We define mobility as moving to another school (district) during the time when the child is aged 10 to 13. Figure 8 shows the mobility response for the two income groups. The left-hand panel shows clear and statistically significant drops in mobility for the low-income group when any of the three thresholds is surpassed. The pattern is slightly more erratic for the high-income group, but at least consistent with the view that increases in class size causes mobility also among the high-income group.

Table 4 shows the mobility response to class size increases. We find that parents are more likely to move to another school when their child is placed in a large class. The response magnitude is larger among low-income parents, but the difference across the two income groups is not significant.[28] Moving to another school district requires moving home. Nine percent

---

[28]The differential mobility response across the family income distribution is also evident in STAR. We have

Figure 8. Mobility by enrollment in grade 4

a) Low-income parents

b) High-income parents
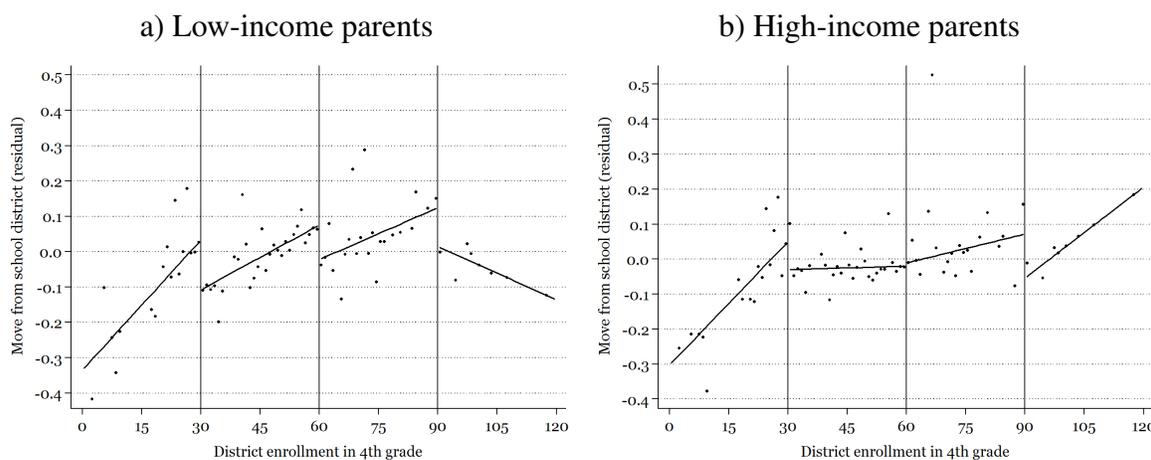
Notes: The figures shows the residual probability of moving school, after controlling for fixed effects for enrollment segments, municipality-by-cohort fixed effects and baseline covariates, by enrollment in grade 4. The data pertain to one-school districts for cohorts born 1972, 1977 and 1982. The regression lines were fitted to individual data. Panel a) shows mobility by enrollment for low-income parents (income below the median in the population); discontinuities at thresholds: -0.144 (0.034), -0.099 (0.049), -0.115 (0.054). Panel b) shows mobility by enrollment for high-income parents (income above the median in the population); discontinuities at thresholds: -0.085 (0.071), 0.010 (0.042), -0.123 (0.067).

of the parents move to another school district in the three-year period we examine. House mobility in general is of course much higher and equals 27 percent over a 3-year horizon. The estimation results indicate that a five-student increase in class size (during 3 years) boosts the mobility rate by 5.8 percentage points. Relative to a base of 9 percent this is substantial, and indicates that school quality is an important determinant of cross district mobility for families with children in the compulsory school age; however, relative to overall mobility, a 5-student increase in class size has a mobility impact that corresponds to 21 percent of overall mobility.

At first glance it maybe somewhat surprising that low-income parents respond to class size increases by moving to a greater extent than high-income parents. While low-income parents are likely to be financially constrained to a greater extent, it is also true that the transactions costs are lower for this group since they tend to live in rental housing to a greater extent than high-income households. Moreover, since high-income households (can and do) respond along the parental assistance margin, it is not necessary for them to move to the same

---

looked at the responses among the population randomized to small and regular sized classes in Kindergarten. As a proxy for mobility we used attrition, i.e. whether a student's test score is recorded in grade 3. The class size reduction in STAR reduces attrition among students on free lunch by 5 percentage points (t-ratio 2.43). Among pupils not eligible for free lunch there was a reduction by 0.8 percentage points (with a t-ratio 0.47).

Table 4. IV-estimates of class size on mobility

Outcome [#pupils; #districts×cohorts; #clusters]	All (1)	Income parents		Difference (2) and (3) (4)
		low (2)	high (3)	
Move to another school district [4,370; 162; 75]	0.0117*** (0.0034)	0.0159*** (0.0042)	0.0070 (0.0050)	0.0090 (0.0065)
Change in class size for movers [4,370; 162; 75]	-7.3569** (2.9874)	-7.7013** (3.0104)	-4.5330* (2.8063)	-3.1683 (4.1155)

Note: The estimates are based on representative samples of individuals born in 1972, 1977 or 1982 in one-school districts. Average class size in grades 4-6 is instrumented by indicators for being above 1st, 2nd, or 3rd threshold of the class size rule. All models include the following controls for school district enrollment in grade 4: fixed effects for enrollment segment; linear controls for enrollment which are interacted with threshold. In addition all models include the following baseline controls: municipality-by-cohort 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 by enrollment count are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 percent level, respectively.

extent as for the low-income households.

An interesting question in this context is whether families move to smaller-sized classes, conditional on moving. The final row of Table 4 shows results from an IV-regression where the dependent variable is the difference between actual class size and expected class size (the class size for non-movers) and where mobility is instrumented by the class size rule. The estimate answers the question: By how much does mobility (caused by surpassing a threshold) contribute to reducing class size? The identifying assumption (apart from random assignment of the instrument) is that the class size rule affects the difference between actual and expected class size only through its effect on mobility, which is (almost) true by construction. The results show that the movers are indeed moving to a school with smaller class size. Again there is no statistical difference across the parental income distribution.

It is also interesting to note that class size is reduced in excess of what would be expected from a random choice of destination school. The nature of the RD design implies that the origin school has maximum class size, i.e. 30. This implies that we would also expect a decline in class size with a random choice of destination school. With a random choice of destination school (within municipality) class size would be reduced by 4.4 students which is less than the 7.4 students documented in Table 3.

To summarize Sections 7.1 and 7.2, we have found that parents compensate for increases in class size. This suggests that public and private resources are substitutes in the production of human capital. High-income parents and low-income households use different adjustment margins: high-income (high-skilled) parents respond to an increase in class size by helping out more with homework; low-income parents do not respond at all along this dimension. For high-income children, this change in behavior implies that the effect of class size on achievement is lower than it would otherwise be. All parents move school in response to a class size increase. This response is not statistically different across the two groups, but it is twice as large in the low-income group.

8   The learning environment and the total effect of class size

This section begins by examining how variation in class size affects the learning environment as perceived by students and teachers. This informs us about what actually changes in the classroom when the number of students changes. In terms of the conceptual framework of Section 2, these outcomes can be thought as relating to the factors that determine the sign and magnitude of the direct effect $\partial h/\partial s$ in equation (6). We close this section by discussing what implications the evidence on parental responses and the evidence on the learning environment have for the total effect of a class size intervention.

8.1   The learning environment

Panel a) of Table 5 reports IV estimates of the effect of class size on teachers' views on pupil responsibility (unfortunately we do not have information on what teachers actually do). The results indicate that teachers in larger classes assign more responsibility for learning to the pupils. Columns (2) and (3) show that teacher perceptions about pupil responsibility do not vary by socio-economic characteristics of the school. Independently of the characteristics of the school, teachers assign more responsibility for learning to the pupils when they teach in a larger class.

Panel b) of Table 5 pursues the same theme by examining how pupil responses change as a consequence of an increase in class size. Column (1) shows that pupils, on average, find it harder to understand full-class teaching in larger classes. This effect is entirely due to students from low-income families; a 5-student increase in class size reduces the probability that a low-income students finds the teacher easy to understand by 7 percentage points. This should be compared to the sample mean of around 85 percent of the students finding it easy

to understand when the teacher explains in front of class. Along this dimension, pupils from high-income families are unaffected by class size. Low-income students also appear to ask teachers for help less frequently in larger classes; this effect is, however, not significant at the 10% level.[29]

Our interpretation of these findings is that teachers in larger classes use less intensive teaching methods (assign more responsibility for learning to the students), probably implying more frontal teaching and less personalized instruction. As a result of an increase in class size, students from low-income families find it harder to understand when taught in full class. Students from high-income families, apparently do not to experience a change in the learning environment when the class size changes. One way to interpret this is that the direct effect of class size on achievement is larger (in absolute size) for students from low-income families than for students from high-income families. We cannot exclude, however, that high-income students are unaffected because they are helped more with homework by their parents.

8.2 The total effect of class size variation – discussion

An increase in class size, by construction, implies less direct pupil/teacher interaction. Consistent with this, teachers think that pupils have to take greater responsibility for their learning when class size increases. Although this teacher response does not vary across contexts, we believe that this result is important for understanding the differential net impact of class size. The fact that students appear to be left on their own to a greater extent in a larger class may be particularly detrimental for low-income students. In other words, it seems likely that the effect of less student-teacher interaction is larger (in absolute value) for low-income students. In line with this interpretation, we find that only low-income students find their teachers harder to understand when the teacher explains in front of a larger class.

We have also demonstrated that public and private investments in child human capital are substitutes rather than complements. This implies that the total effect of class size variation is smaller (in absolute) size than the direct effect of class size variation. Differentially skilled parents adjust along different margins. High-income parents help their children with homework to a greater extent in response to an increase in class size, whereas low-income do not adjust at all along this margin. All parents move to another school in response to larger-sized classes; this is particularly true among low-income parents, although the estimates across groups are not statistically different from one another.

---

[29]Notice, that we do not have a lot of variation to exploit for this outcome. 96% of students answer that they ask for help when they do not understand. This share does not vary by family income.

Table 5. IV-estimates of class size on learning environment: teacher and pupil perceptions

Outcome [# pupils; #districts×cohorts; clusters]	All (1)	Family income below median (2)	Family income above median (3)	Difference (2) and (3) (4)
a) Teacher perceptions				
Teacher thinks it is important that pupils take responsibility for their learning [1,848; 93; 56]	0.0155** (0.0072)	0.0170** (0.0074)	0.0143** (0.0066)	0.0027 (0.0099)
b) Pupil perceptions				
Pupil thinks it is easy to understand when teacher explains in front of class [2719; 83; 52]	-0.0055*** (0.0019)	-0.0138*** (0.0042)	0.0025 (0.0040)	-0.0163*** (0.0058)
Pupil asks the teacher for help if (s)he do not understand [2802; 83; 52]	-0.0015 (0.0015)	-0.0043 (0.0029)	-0.0003 (0.0030)	-0.0040 (0.0041)

Notes: Panel a) is based on (the teachers of) a representative samples of individuals born in 1982 in one-school districts. Panel b) is based on representative samples of individuals born in 1967 and 1972. In panel a) columns (2) and (3) present separate estimates for schools where family income is below/above median. In panel b) columns (2) and (3) present separate estimates for pupils whose parents earn below/above median. Average class size in grades 4-6 is instrumented by indicators for being above 1st, 2nd, or 3rd threshold of the class size rule. All models include the following controls for school district enrollment in grade 4: fixed effects for enrollment segment; linear controls for enrollment which are interacted with threshold. In addition all models include the following baseline controls: municipality-by-cohort 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 by enrollment count are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 percent level, respectively.

Which of these adjustment margins is most efficient? This question is of course impossible to answer definitively. Let us just make two remarks in this regard. First, the differential responses across the income distribution could be interpreted as high-income parents choosing the most effective response. They could have moved, but since they had access to a more effective response, they chose not to do so to the same extent as low-income parents. Low-income parents are constrained by their skills; therefore they could not provide effective parental help and had to resort to moving instead. Second, and along the same lines, it is not entirely clear that moving to another school is to the benefit of the student. The move comes with a reduction in class size, but it also comes with some disruption (changes of school, teachers, peers and neighborhood). The net effect of these changes is unclear.[30]

From an efficiency point of view, we are mostly interested in how the direct effect of class size varies in the distribution (since the envelope theorem implies that the indirect responses of parents are not relevant). If parental help is the more effective response, then the total effects are more biased at the higher end of the distribution. In the CES-example of Section 2, the direct effect class size is given by

$$\frac{\partial h}{\partial s} = \frac{\mu(1 - \kappa(p,s))h(p,s,\theta)}{s}$$

For given s, high-income children have more human capital than low-income children. Since p is higher for high-income children, its relative importance in the production of human capital is also larger, i.e., $\kappa = \frac{\lambda p^\phi}{\lambda p^\phi + (1-\lambda)s^\phi}$ is higher for high-income children than for low-income children.[31] Since these two factors move in opposite direction we cannot sign the relative size of the direct effects across the distribution. Nevertheless, the fact that κ is higher for high-income children, suggests that what is going on in school should be less (more) important for high-income (low-income) children. The result that low-income children seem to be more affected by the changes in the learning environment could thus be the result of different private investments across the distribution.

From a distributional point of view, we care about the total effects of class size variation. Based on our analysis, we would attribute the variation in these total effects across the distribution to two factors: first, low-income children seem to be more affected by the class-room environment; second, high-income parents are better able to compensate for any changes in the learning environment. This interpretation rests on the assumption that other factors of

---

[30]Of course, if parents are informed, the net benefit should be positive. Nevertheless, using data from Texas, Hanushek et al. (2004) find that the disruption associated with moving is negative for student achievement.

[31]Remember that our estimates suggest that $\phi > \mu > 0$, which implies that κ is increasing in p.

production do not change when class size increases. In Fredriksson et al. (2014), we show that other school resources, such as remedial education and teacher quality, are unaffected by class size changes.

9 Conclusion

A recently emerging literature uses credible identification to estimate how private investments in children respond to variation school resources in general (e.g., Das et al. (2013) and Pop-Eleches and Urquiola (2013)).This paper contributes to this literature by: i) examining whether parental responses differ by social background, and ii) investigating parental responses to variation in class size.

Our key finding is that an increase in class size causes high-income families to help their children more with homework. For low-income families we do not find such an effect. We also demonstrate that all parents are more likely to move their child to another school when exposed to a larger-sized class, and that only low-income children find their teachers harder to follow when taught in a larger class. These findings indicate that public and private investments in children's human capital are substitutes, and help explain why the negative effect of class size on achievement in our data is concentrated among low-income children.

Our paper thus shows that parents compensate for what is going on in schools, and that parents choose different adjustment margins depending on their own skills and income. Parental choices of response margins realistically depend on comparative advantage and the restrictions parents face. Therefore, it is likely that the magnitudes of the responses, and whether one particular adjustment margin is used, vary somewhat over contexts. As an example, one may note that a market for private tutoring in Sweden was in practice non-existent during the relevant time period. Private tutoring may clearly be an important adjustment margin in other settings. On the basis of our analysis, we expect that parents would use private tutoring to compensate for an increase in class size (potentially at the expense of own help with homework), and that this response would primarily be found among high-income households.

References

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.

Becker, G. S. and Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labor Economics*, 4(3):S1–S39.

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.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., and Sundararaman, V. (2013). School inputs, household substitution, and test scores. *American Economic Journal: Applied Economics*, 5:29–57.

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

Finn, J. D., Pannozzo, G. M., and Achilles, C. M. (2003). The "why's" of class size: Student behavior in small classes. *Review of Educational Research*, 73(3):321–368.

Fredriksson, P., Öckert, B., and Oosterbeek, H. (2012). 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. (2013). Long-term effects of class size. *Quarterly Journal of Economics*, 128(1):249–285.

Fredriksson, P., Öckert, B., and Oosterbeek, H. (2014). Inside the black box of class size: Mechanisms, behavioral responses, and social background. Discussion paper 8019, IZA.

Hanushek, E. A., Kain, J. F., and G, S. (2004). Disruption versus tiebout improvement: the costs and benefits of switching schools. *Journal of Public Economics*, 88:1721–1746.

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

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.

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

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

Lee, D. S. and Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 142:655–674.

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

Pop-Eleches, C. and Urquiola, M. (2013). Going to a better school: Effects and behavioral responses. *American Economic Review*, 103:1289–1324.

Rietz, L. D., Lundgren, U. P., and Wennas, O. (1987). Ansvarsfordelning och styrning på skolomradet. DsU 1987:1, Stockholm: Ministry of Education.

Summers, A. A. and Wolfe, B. L. (1977). Do schools make a difference? *American Economic Review*, 67:639–652.

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.

Appendix

A.1 Total effects of a change in s under the CES production structure

In general, the direct effect of a change in *s* is given by

$$\frac{\partial h}{\partial s} = \frac{\mu(1-\kappa)h}{s}$$

where $\kappa = \frac{\lambda p^\phi}{\lambda p^\phi + (1-\lambda)s^\phi} \in [0,1)$. For constrained parents ($\theta_f < \underline{\theta}_f$), we thus have

$$\frac{dh^0}{ds} = \frac{\partial h^0}{\partial s} = \frac{\mu h^0}{s}$$

For partially constrained parents ($\underline{\theta}_f \leq \theta_f < \bar{\theta}_f$)

$$\frac{dh^c}{ds} = \frac{\partial h^c}{\partial s} + \frac{\partial h^c}{\partial p^c}\frac{\partial p^c}{\partial s} = \frac{\mu(1-\kappa^c)h^c}{s}\left[\frac{1-\phi+\mu\kappa^c}{1-\phi+\kappa^c(\phi+\mu)}\right]$$

And for unconstrained parents ($\bar{\theta}_f \leq \theta_f$)

$$\frac{dh^*}{ds} = \frac{\partial h^*}{\partial s} + \frac{\partial h^*}{\partial p^*}\frac{\partial p^*}{\partial s} = \frac{\mu(1-\kappa^*)h^*}{s}\left[\frac{1-\phi}{1-\phi+\kappa^*(\phi-\mu)}\right]$$

If $\phi > \mu$, the total effects are all positive and smaller than the direct effects. If $\kappa(\phi-\mu) > 1-\phi$, then $\left.\frac{\partial p^*}{\partial s}\right|_{\theta_f=\bar{\theta}_f} < \left.\frac{\partial p^c}{\partial s}\right|_{\theta_f=\bar{\theta}_f}$ and, consequently, $\left.\frac{dh^*}{ds}\right|_{\theta_f=\bar{\theta}_f} < \left.\frac{dh^c}{ds}\right|_{\theta_f=\bar{\theta}_f}$.

A.2 Further tables

Table A1. Construction of outcome variables

Table	Outcome	Cohorts	Data source/Survey question	Measure
2	Academic achievement	1967, 1972, 1977, 1982	Achievement tests in Swedish and mathematics	Standardized average test score by cohort
	Cognitive ability	1967, 1972, 1977, 1982	Tests for verbal and logical abilities	Standardized average test score by cohort
3	Parents help child with homework	1967, 1972, 1977, 1982	Pupil question: "Does someone at home help you with homework or school work?" (1967 and 1972) "How many times/week…") (1977 and 1982)	Dummy variable for receiving any help with homework
	Pupil do homework (hours)	1967, 1972, 1977, 1982	Pupil question: "How often do you do homework or other school work at home?" (times/week for 1967 cohort) (hours/week for 1972, 1977 and 1982 cohorts)	Number of hours/week spent on homework. The values for the 1967 cohorts have been predicted from response on frequency of homework.
4	Move from school district	1972, 1977, 1982	Administrative records on pupil mobility	Dummy variable for moving both school district and housing
5	Teacher thinks that it is important that pupils take responsibility for their own learning	1982	Teacher question: "How important is it that pupils take responsibility for their own learning?" (1-5)	Percentile ranked response divided by 100
	Pupil thinks it is easy to understand when the teacher explains in front of the class	1967, 1972	Pupil question: "Do you think it is hard to understand when the teacher explains in front of the class?" (Y/N)	Dummy variable for negative response
	Pupil asks the teacher for help if (s)he do not understand	1967, 1972	Pupil question: "Do you ask the teacher for help when you do not understand?" (Y/N)	Dummy variable for positive response

Table A2. Descriptive statistics

| Variable | All | Income parents | |
[#pupils; #districts×cohorts; #clusters]		Low	High
Class size grades 4-6	24.121	23.830	24.382
	(3.642)	(3.717)	(3.554)
Academic achievement, age 13 (SD)	0.032	-0.170	0.209
[4712; 187; 80]	(1.002)	(1.027)	(0.945)
Cognitive ability, age 13 (SD)	0.005	-0.223	0.202
[5201; 197; 82]	(1.023)	(1.039)	(0.966)
Parents help child with homework	0.798	0.772	0.821
[5112; 197; 82]	(0.401)	(0.419)	(0.384)
Pupil does homework (hours)	2.735	2.721	2.747
[5154; 197; 82]	(1.669)	(1.715)	(1.628)
Parents move school districts	0.090	0.110	0.071
[4389; 162; 75]	(0.286)	(0.313)	(0.257)
Teacher thinks that it is important that pupils take responsibility for their own learning (rank 0/1) [1853; 93; 56]	0.516 (0.183)	0.515 (0.183)	0.516 (0.183)
Pupil thinks it is easy to understand when the teacher explains in front of the class [2719; 83; 52]	0.873 (0.333)	0.855 (0.352)	0.888 (0.315)
Pupil asks the teacher for help if (s)he do not understand [2802; 83; 52]	0.956 (0.204)	0.956 (0.206)	0.957 (0.202)

Note: The table shows means (standard deviations) of outcome variables and are based on representative samples of individuals born in 1967, 1972, 1977 or 1982 in one-school districts. Standardized variables are normalized to have mean zero in the population. Median income, which is used to divide households into low and high-income households, is calculated from the population.

Table A3. IV-estimates of class size on different outcomes

Outcome [# individuals]	Current spec (1)	QJE-spec (2)
Cognitive ability, age 13	-0.0268**	-0.0330**
[N=5,197; N=5,116]	(0.0116)	(0.0146)
Non-cognitive ability, age 13	-0.0177	-0.0265**
[N=4,740; N=4,681]	(0.0113)	(0.0118)
Academic achievement, age 16	-0.0265***	-0.0233***
[N=5,377; N=5,318]	(0.0085)	(0.0101)
Years of schooling, ages 27-42	-0.0318	-0.0545*
[N=5,669; N=5,588]	(0.0234)	(0.0256)
P(Bachelor's degree), ages 27-42	-0.0059	-0.0076*
[N=5,669; N=5,588]	(0.0042)	(0.0043)
Earnings, ages 27-42	-0.0149**	-0.0117*
[N=6,009; 5,920]	(0.0057)	(0.0061)
P(earnings>0), ages 27-42	-0.0035	-0.0016
[N=6,009; N=5,920]	(0.0022)	(0.0024)
ln(Wage), ages 27-42	-0.0065***	-0.0063*
[N=3,227; N=3,185]	(0.0022)	(0.0033)
Number of districts×cohorts	200	191

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982 in one-school districts. All ability measures are standardized. The educational outcomes are measured in 2009, while the labor market outcomes have been averaged over the 2007-2009 period. Earnings effects (and their standard errors) are divided by average earnings level to facilitate interpretation. The ln(wage) estimates are restricted to wage-earners. Average class size in grades 4-6 is instrumented by indicators for being above 1st, 2nd, or 3rd threshold of the class size rule in column (1). In column (2), average class size in grades 4-6 is instrumented by an indicator for being above any threshold of the class size rule. All models include the following controls for school district enrollment in grade 4: fixed effects for enrollment segment; linear controls for enrollment which are interacted with threshold. In column (2), the controls for enrollment are also interacted with segment. In addition all models include the following baseline controls: municipality-by-cohort 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 by enrollment count (84 clusters) are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 percent level, respectively.

Table A4. IV estimates of class size, different enrollment controls

Model [#pupils; #districts×cohorts; #clusters]	(1)	(2)	(3)	(4)	(5)	(6)
			Academic achievement			
Class size grades 4-6 [4707; 187; 80]	-0.0592***	-0.0516***	-0.0441***	-0.0419***	-0.0415***	-0.0606**
	(0.0154)	(0.0134)	(0.0122)	(0.0119)	(0.0114)	(0.0270)
			Parents help child with homework			
Class size grades 4-6 [5107; 197; 82]	0.0063	0.0097**	0.0086**	0.0029	0.0057	0.0037
	(0.0046)	(0.0043)	(0.0043)	(0.0041)	(0.0039)	(0.0052)
		Easy to understand when the teacher explains before the class				
Class size grades 4-6 [2719; 83; 52]	-0.0077***	-0.0046**	-0.0055***	-0.0094***	-0.0107***	-0.0044
	(0.0026)	(0.0020)	(0.0019)	(0.0030)	(0.0026)	(0.0068)
Enrollment controls						
Polynomials:						
- 1st order	✓				✓	
- 2nd order		✓				
Interacted with thresholds			✓	✓	✓	✓
Interacted with segments					✓	✓

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982 in one-school districts. Average class size in grades 4-6 is instrumented by indicators for being above 1st, 2nd, or 3rd threshold of the class size rule. All models include fixed effects for school district enrollment in grade 4, municipality-by-cohort 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 by enrollment count are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 percent level, respectively.

Table A5. First-stage estimates, different enrollment controls

	(1)	(2)	(3)	(4)	(5)	(6)
Above 1st threshold	-6.535***	-8.575***	-10.040***	-9.202***	-9.379***	-9.569***
	(1.359)	(1.508)	(0.900)	(1.492)	(1.260)	(1.962)
Above 2nd threshold	-6.215***	-5.587***	-5.201***	-8.106***	-7.005***	-7.842***
	(0.942)	(0.897)	(1.069)	(1.179)	(0.960)	(1.446)
Above 3rd threshold	-6.139***	-3.074*	-4.684***	-2.737	-3.612	-0.367
	(1.056)	(1.808)	(1.305)	(2.626)	(2.240)	(3.316)
F-test for instruments	18.30	20.48	47.89	39.49	40.82	18.16
Enrollment controls						
Polynomials:						
- 1st order	✓		✓		✓	
- 2nd order		✓		✓		✓
Interacted with thresholds			✓	✓	✓	✓
Interacted with segments					✓	✓
Number of pupils	6,009	6,009	6,009	6,009	6,009	6,009
Number of districtsxcohorts	200	200	200	200	200	200
Number of clusters	82	82	82	82	82	82

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982 in one-school districts. All models include fixed effects for school district enrollment in grade 4, municipality-by-cohort 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 by enrollment count are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 percent level, respectively.