

Fredriksson, Peter; Öckert, Björn; Oosterbeek, Hessel

Working Paper

Long-term effects of class size

IZA Discussion Papers, No. 5879

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Fredriksson, Peter; Öckert, Björn; Oosterbeek, Hessel (2011) : Long-term effects of class size, IZA Discussion Papers, No. 5879, Institute for the Study of Labor (IZA), Bonn, <https://nbn-resolving.de/urn:nbn:de:101:1-201108093062>

This Version is available at:

<https://hdl.handle.net/10419/51922>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

IZA DP No. 5879

Long-Term Effects of Class Size

Peter Fredriksson
Björn Öckert
Hessel Oosterbeek

July 2011

Long-Term Effects of Class Size

Peter Fredriksson

*Stockholm University, IFAU, UCLS
and IZA*

Björn Öckert

IFAU and UCLS

Hessel Oosterbeek

University of Amsterdam

Discussion Paper No. 5879
July 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Long-Term Effects of Class Size^{*}

This paper evaluates the long-term effects of class size in primary school. We use rich administrative data from Sweden and exploit variation in class size created by a maximum class size rule. Smaller classes in the last three years of primary school (age 10 to 13) are not only beneficial for cognitive test scores at age 13 but also for non-cognitive scores at that age, for cognitive test scores at ages 16 and 18, and for completed education and wages at age 27 to 42. The estimated effect on wages is much larger than any indirect (imputed) estimate of the wage effect, and is large enough to pass a cost-benefit test.

JEL Classification: I21, I28, J24, C31

Keywords: class size, regression discontinuity, cognitive skills, non-cognitive skills, educational attainment, earnings

Corresponding author:

Peter Fredriksson
Stockholm University
Department of Economics
SE-106 91 Stockholm
Sweden
E-mail: peter.fredriksson@ne.su.se

^{*} We gratefully acknowledge valuable comments from Helena Svaleryd and seminar participants in London, Mannheim, Paris, Stockholm and Uppsala.

1 Introduction

This paper evaluates the effects of class size in primary school on long-term outcomes, including completed education, earnings and wages at age 27-42. While there is a large literature estimating the short-term effects of class size, estimates of long-term effects of class size are sparse.¹ To judge the effectiveness of class size reductions, it is vital to know whether short-term effects on cognitive skills (if any) persist or fade-out, and whether these effects translate into economically meaningful improvements in labor market outcomes.

Three previous studies examine long-term effects of class size. Krueger and Whitmore (2001) analyze the long-term effects of small classes using information from students who participated in the Tennessee STAR experiment. In this experiment, students and their teachers were randomly assigned (within school) to different classrooms in grades K-3. Some students were randomly assigned to a class of around 15 students while others were assigned to a class of around 22 students. Attendance of a small class in grades K-3 increases the likelihood of taking college-entrance exam, especially among minorities. Test scores are also slightly higher.

Chetty et al. (2011) also use the STAR experiment and link the original data to administrative data from tax returns. Among their main results is that students in small classes are significantly more likely to attend college and exhibit improvements on other outcomes. However, smaller classes do not have a significant effect on earnings at age 27. The point estimate is even negative, but rather imprecise. The upper bound of the 95% confidence interval is an earnings gain of 3.4 percent. The authors compare this with a prediction of the expected earnings gain based on the estimated impact of small classes on test scores and the cross-sectional correlation between test scores and earnings (see also Schanzenbach, 2007). This implies a positive effect of 2.7 percent, which – as the authors stress – lies within the 95% confidence interval of the directly estimated impact of small classes on earnings.²

Bingley et al. (2010) apply a similar “two-stage” method using Danish data. They first estimate the impact of class size on the amount of schooling combining a (not so strict) maximum class size rule and family fixed effects, and find that a 5 percent reduction in

¹Findings of short-term effects vary across countries, by age of the pupils and by empirical approach. Most studies that focus on class size in primary school and use a credible empirical strategy find that class size has a negative effect on cognitive achievement measured shortly after exposure. Well-known studies showing such effects are Angrist and Lavy (1999) for Israel, Krueger (1999) for the United States and Urquiola (2006) for Bolivia. An equally well-known study finding no impact on US data is Hoxby (2000).

²Chetty et al. (2011) do not only use the STAR experiment to examine the long-term effects of class size, but also investigate the long-term impact of other characteristics of the class in which people were placed in grades K-3.

class size (one student) in grade 8 increases completed schooling by half a week. They then estimate the effect of the amount of schooling on earnings using data from twins, and find a return to schooling of 8%. Together these two pieces of evidence suggest that class size has a negative effect on earnings.

While the studies of Chetty et al. (2011), Schanzenbach (2007) and Bingley et al. (2010) are suggestive of a negative long-term effect of class size on adult earnings, the evidence reported therein is by no means conclusive. First, there is no guarantee that the higher test scores (in Chetty et al. and Schanzenbach) induced by smaller classes *cause* an increase in earnings. Indeed, the negative direct estimate reported by Chetty et al. (2011) indicates that this need not be the case.³ Second, the two-stage strategy assumes that the effect of class size on earnings only works through test scores or educational attainment. This need not be the case. For instance, if a reduction in class size has a positive effect on non-cognitive skills, and these non-cognitive skills have a positive effect on earnings conditional on educational attainment, the estimates reported in Bingley et al. (2010) are biased downward.⁴

Using unique Swedish data, we trace the effects of changes in class size in primary school on cognitive and non-cognitive achievement at ages 13, 16, and 18, as well as on long-term educational attainment and wages observed when individuals are aged 27–42. We exploit variation in class size attributable to a maximum class size rule in Swedish primary schools. This maximum class size rule gives rise to a regression discontinuity design. We apply this identification strategy to data covering the cohorts born in 1967, 1972, 1977, and 1982. The focus on these cohorts is motivated by the fact that we have information on cognitive achievement at the end of primary school for a 5–10 percent sample of these cohorts. To these data we match individual information on educational attainment and earnings. Educational attainment and earnings are observed in 2007–2009.

We find that smaller classes in the last three years of primary school (age 10 to 13) are beneficial for cognitive and non-cognitive test scores at age 13 and for cognitive test scores at ages 16 and 18. Moreover, the effects on cognitive (and non-cognitive) scores do not fade out substantively over time.⁵ We also find that smaller classes increase completed education and wages at age 27 to 42. The wage effect is stronger for individuals with parents who have

³To avoid confusion, a negative effect of *small classes* on earnings implies a positive effect of *class size* on earnings.

⁴Bingley et al. (2010) also assume that cognitive skills only affects earnings via their effect on educational attainment.

⁵There are several papers documenting that the effects of early interventions on cognitive test scores fade out fairly rapidly over time. This pattern appears in STAR, the Perry and Abcederian pre-school demonstrations, and Head-start (see Almond and Currie, 2010, for a survey of the pre-school interventions).

income above the median. We compare the direct estimate of the wage effect to estimates obtained using the indirect methods of previous studies, and find that the direct estimate is much larger than the indirect “imputed” estimates. We conduct a cost-benefit analysis and find that a reduction in class size from 25 to 20 pupils has an internal rate of return of almost 20%.

The paper proceeds as follows. In Section 2 we describe the relevant institutions of the Swedish schooling system. Section 3 describes the data and Section 4 presents results concerning the validity and strength of our instrumental variable approach. Section 5 presents and discusses the empirical findings. Section 6 summarizes and concludes.

2 Institutional background

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

The compulsory school system had several organizational layers. The primary unit in the system was the school. Schools were aggregated to school districts.⁷ School districts 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 there were typically more primary schools than lower secondary schools in the school district. There was at least one school district in a municipality.

The municipalities formally ran the compulsory schools. But central government funding and regulations constrained the municipalities substantially. The municipalities could top-up

⁶This changed in the 1990s with the introduction of decentralization and school choice. From 1993 onwards compulsory schools are funded by the municipalities; see Björklund et al. (2005) for a description of the Swedish schooling system after decentralization. Du Rietz et al. (1987) contains an excellent description of the school system prior to decentralization, on which we base this section.

⁷We use the term “school district” for want of a better word. The literal translation from Swedish would be “principal’s district” (*Rektorsområde*).

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 6.¹⁰ The reason to focus on these grades is data availability. We can only assign a student to a school (and a school district) in grades 4-6; this information is not available for lower grades.

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.¹¹ We will use this rule for identification in a regression discontinuity (RD) design. This method has been applied in several previous studies to estimate the causal effect of class size.¹² Since the law allowed the county school boards to adjust the borders of its school catchment areas in a way that may systematically favor disadvantaged pupils, we will implement the RD design at the school district level rather than at the school level. The boundaries of the school district are given by the rules on maximum travelling distances to schools; they are thus fixed to a much greater extent than the boundaries of the school catchment area. We provide evidence that the RD design at the school district level is valid in Section 4.

The treatment we have in mind is an increase in class size by one pupil throughout upper primary school (grades 4-6). The instrument used to identify this effect is predicted class size according to school district enrollment in grade 4 and the maximum class size rule. To be able to attribute our findings to average class size in grades 4 to 6, class size in these grades

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

⁹The fine details of the rule were changed in 1978. In 1978 a resource grant – the so called *base resource* – was introduced. The base resource governed the number of teachers per grade level in a school. This discontinuous funding rule had the same discontinuity points as the 1962 compulsory school law.

¹⁰Hence, if a student is in a class of 25 pupils in grade 4, in a class of 24 students in grade 5 and in a class of 23 students in grade 6, the average class size to which this student was exposed in second stage primary school equals 24 ($= (25 + 24 + 23)/3$).

¹¹There have always been special rules in small schools and school districts. In such areas, the rules pertained to total enrollment in 2 or 3 grade levels.

¹²The seminal paper is Angrist and Lavy (1999). See also Gary-Bobo and Mahjoub (2006); Hoxby (2000); Leuven et al. (2008); Urquiola and Verhoogen (2009).

should not be correlated with class sizes in other stages of compulsory school. If class sizes in other stages would be positively correlated with class size in upper primary school, we would overestimate the effect of class size at that level. We think that a correlation with class size in lower primary school is unlikely since the rule was given in multiples of 25. Unfortunately, we do not have the data to verify this statement. We are able to test whether class size in grades 4 to 6 has an “effect” on class size in grades 7 to 9, however. It turns out that the instrument has no significant effect on class size in lower secondary school (estimate 0.14; standard error 0.18).

For the RD design to be credible, other school resources should not exhibit the same discontinuous pattern. This is indeed not the case. The base resource – the discontinuous funding rule which governed the number of teachers – was the largest component of central government grants for running expenses. In the mid 1980s for instance, the base resource amounted to 62 percent of these grants. The only other major grant component (27 percent of the grants) was aimed at supporting disadvantaged students. This grant was tied to the overall number of students in compulsory school in a municipality and there were no discontinuities in the allocation of the grant.

3 Data

The key data used in this paper come from the so-called UGU-project which is run by the Department of Education at Göteborg University; see Härnquist (2000) for a description of the data. The data contain cognitive test scores at age 13 for roughly a 10 percent sample of the cohorts born 1967, 1972, and 1982. In addition, there is information on a 5 percent sample for the cohort born in 1977.

To these data we have matched register information maintained by Statistics Sweden. The added data include information on class size (from the Class register), parental information (which is made possible by the multi-generational register containing links between all parents and their biological or adopted children), and medium-term and long-term outcomes. The medium-term outcomes are individuals’ test scores (at age 16) and scores on cognitive and non-cognitive tests (at age 18). Long-term outcomes are completed education, earnings and wages measured in 2007-2009. The cognitive and non-cognitive test scores at age 18 are only available for men since they are derived from the military enlistment.

The cognitive tests at age 13 are traditional “IQ-type” tests. We constructed a measure based on scores for verbal skills and logical skills. The verbal test involves finding a word having the opposite meaning as a given word. The logical test requires the respondent to fill

in the next number in a sequence of numbers. We refer to this measure as “cognitive skills” for short, it is standardized such that the mean is zero and the standard deviation equals one. The measure of non-cognitive skills at age 13 is based on a questionnaire about the pupils’ situation in school. We form an index based on four questions reflecting the pupils’ perceived self-confidence, persistence, self-security and expectations.¹³ The index is standardized to mean zero and standard deviation one.

Academic achievement at age 16 is measured as test scores at the end of lower secondary school. The achievement tests involve Maths, Swedish, and English. These achievement tests were used to anchor subject grades at the school level: the school average test result thus determined the average subject grade at the school level. Also this outcome is standardized to mean zero and standard deviation one.

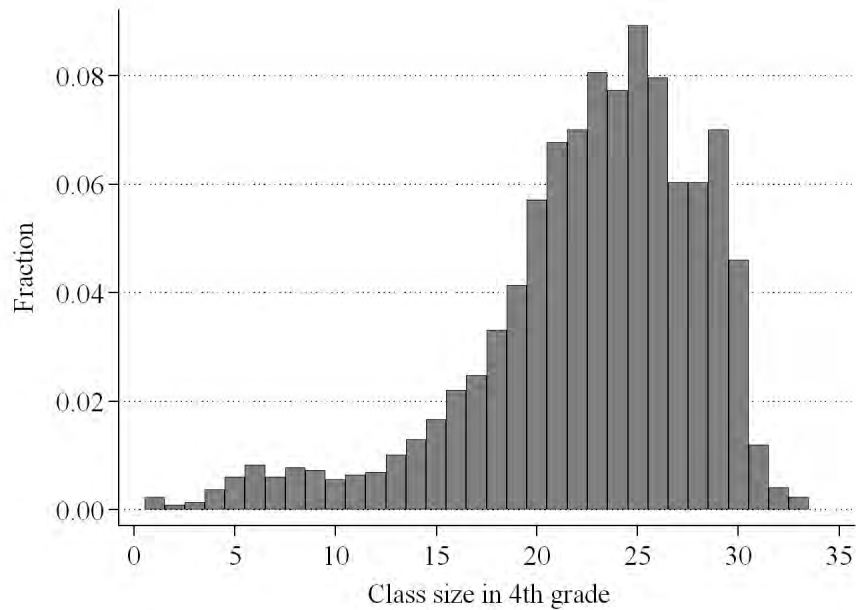
The military enlistment cognitive test is very similar in nature to the test administered at age 13; see Carlstedt and Mårdberg (1993) for a description of the Swedish military enlistment battery. It is designed to measure general ability and it is similar to the AFQT (Armed Forces Qualifications Test) used in the US. We again constructed a standardized measure based on the verbal and logical parts of this test. Upon enlistment, army recruits also have a 20 minutes interview with a psychologist who assesses their non-cognitive functioning. Details of the psychologists’ assessments are classified and we have only access to a single score for non-cognitive ability. This overall score is based on four underlying items and a conscript is given a high score if considered to be emotionally stable, persistent, socially outgoing, willing to assume responsibility, and able to take initiatives. Motivation for doing the military service is, however, explicitly not a factor to be evaluated.

Data on educational attainment come from the Educational Register maintained by Statistics Sweden. This register records the highest attained education level for the resident population.¹⁴ We construct two measures based on this. The first is years of completed schooling, the second a binary indicator for having at least a Bachelor’s degree. This measure is analogous to the college indicator used in studies based on the STAR experiment. Data on annual

¹³The questions are “Do you think that you do well in school?” (self-confidence), “Do you give up if you get a difficult task to do in school?” (persistence), “Do you think that it is unpleasant to have to answer questions in school?” (self-security) and “Do you get disappointed if you get bad results in a test?” (expectations). To ensure that this index is comparable with the psychological evaluation at age 18 (see below), we formed the index by weighting the indicators by the estimated parameters from a regression of non-cognitive skills at age 18 and our four indicators of non-cognitive skills at age 13.

¹⁴The register is complete for individuals with an education from Sweden. Information for immigrants stems from separate questionnaires to new arrival cohorts. The underlying data include information on the courses taken at the university level, which implies that this is a relatively accurate measure of years of schooling even for those who do not have a complete university degree.

Figure 1. Distribution of class size in grade 4



earnings come from the Income Tax Register, while data on wages stem from the Wage Register; both of these registers are maintained by Statistics Sweden. Earnings are based on income statements made by employers. The wage data relate to those who are employed in October/November in a particular year and are measured in full-time equivalent wages. We use earnings and wage data from 2007-2009; individuals of the oldest (1967) cohort are then 42 years old and individuals of the youngest (1982) cohort are 27 years old. Earnings and wages are therefore measured at an age when they correlate highly with lifetime income (Böhlmark and Lindquist, 2006).

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

Table 1 reports results from regressions of long-term outcomes ($\ln(\text{wage})$ and years of schooling) on cognitive and non-cognitive test scores measured at age 13, separately for men and women. This shows high correlations between short-term and long-term outcomes. A one standard deviation increase in the cognitive test score of men, for instance, is associ-

Table 1. Cross-sectional correlations between cognitive and non-cognitive scores at age 13 and long-term outcomes, by gender

Model	ln(Wage)			Years of schooling		
	(1)	(2)	(3)	(4)	(5)	(6)
<u>Men</u>						
- cognitive	0.090*** (0.003)		0.082*** (0.003)	1.181*** (0.020)		1.124*** (0.021)
- non-cognitive		0.051*** (0.003)	0.029*** (0.003)		0.498*** (0.022)	0.221*** (0.020)
<i>N</i>	6738	6738	6738	12775	12775	12775
<u>Women</u>						
- cognitive	0.078*** (0.002)		0.073*** (0.002)	1.158*** (0.022)		1.098*** (0.022)
- non-cognitive		0.037*** (0.002)	0.019*** (0.002)		0.500*** (0.023)	0.246*** (0.022)
<i>N</i>	8032	8032	8032	12177	12177	12177

Note: Per gender, each column reports estimates from OLS regressions based on representative samples of individuals born in 1967, 1972, 1977 or 1982. All models control for cohort×municipality fixed effects. ***=the estimates are significantly different from zero at the 1 percent level of confidence.

ated with a wage increase of 9 percent. Interestingly, and importantly, if cognitive and non-cognitive test scores are included jointly, both are highly significant (see also Lindqvist and Vestman, 2011). These estimates are useful later on when we compare our direct estimates of the effect of class size on long-term outcomes to the estimates that are obtained when we apply the two-step approach of Chetty et al. (2011).¹⁵

4 Validity and strength of the instrument

Validity of the instrument A threat to the validity of the RD design is bunching on one side of the cut-offs, since that indicates that the forcing variable is manipulated. Urquiola and Verhoogen (2009) document an example of this in the context of a maximum class size rule in Chile. In their data there are at least five times as many schools just below than just above the cut-offs. Schools apparently want to avoid the fixed cost of starting a new classroom. They also show that at the cut-off point there are jumps in household income and in mothers' schooling; schools that just passed the cut-off points serve children from better-off families.

¹⁵Table A2 in the appendix reports correlations for each pair of outcome variables. Almost all outcomes are positively correlated and correlations are often substantial. Cognitive ability at age 13 is highly correlated with academic achievement at age 16 and with cognitive ability at age 18 (both above 0.7), but also the correlations with completed years of schooling and log wages at age 27-42 are above 0.3.

Figure 2. Distribution of grade 4 school district enrollment

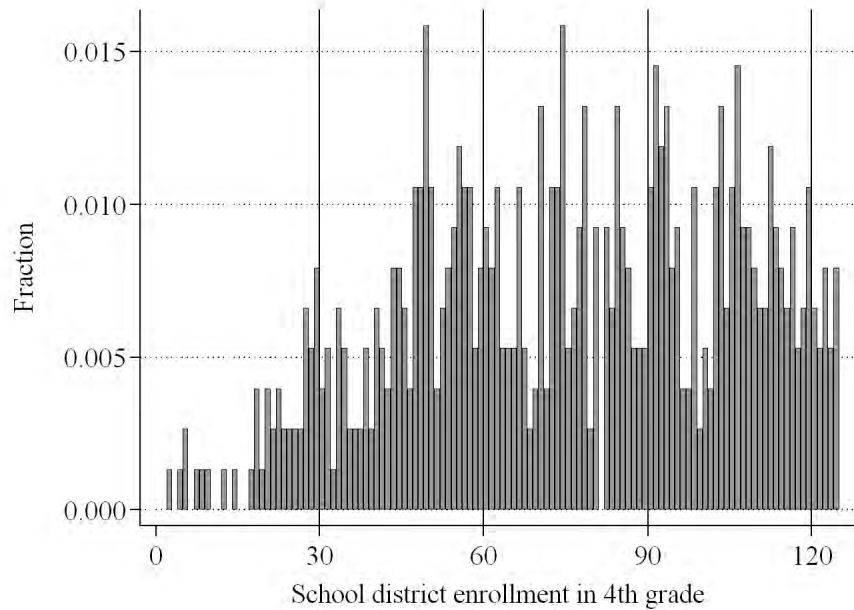


Figure 2 shows the distribution of school district enrollment in grade 4. Visual inspection reveals no suspect discontinuities in the distribution of the forcing variable. A formal test confirms this. We examined if the instrument (with a cubic control for enrollment) can predict the number of observations at different enrollment counts, and found that it cannot.¹⁶

The validity of the RD design can also be examined in other ways. If the instrument – expected class size as predicted by the class size rule – is valid, background variables should be unrelated to it. To test this, we first constructed a composite measure of background variables. We regressed cognitive skills at age 13 on an intercept, gender, dummy variables for month of birth, dummy variables for mother’s and father’s educational attainment, a third order polynomial in parental income, mother’s age at child’s birth, indicators for being a first or second generation immigrant, having separated parents and the number of siblings. We standardized the predicted value of this regression and use that as the composite measure of pupils’ backgrounds. Table 2 reports IV estimates of the “effect” of average class size in 4th to 6th grade on the composite measure of a pupil’s background for several specifications of the enrollment controls. Average class size in grades 4 to 6 is instrumented by predicted class size in grade 4. None of the effects is significantly different from zero, and with the exception of the specification without any control for enrollment, all point estimates are close to zero.

¹⁶We also plotted the distribution of enrollment in grade 4 at the school level. Consistent with the ability to redraw the borders of school catchment areas, we find that schools bunch just before the kink.

Table 2. Specification test: IV estimates of class size on predicted cognitive skills at age 13

Model	(1)	(2)	(3)	(4)	(5)	(6)
Average class size 4th-6th grade	0.023 (0.017)	-0.006 (0.020)	-0.001 (0.020)	0.000 (0.020)	0.005 (0.020)	0.007 (0.021)
Enrollment controls						
Polynomial:						
- 1st order		✓			✓	
- 2nd order			✓			✓
- 3rd order				✓		
Interacted with break-points					✓	✓
F-test (p-value)	0.097	0.187	0.224	0.229	0.197	0.342
<i>N</i>	31,590	31,590	31,590	31,590	31,590	31,590

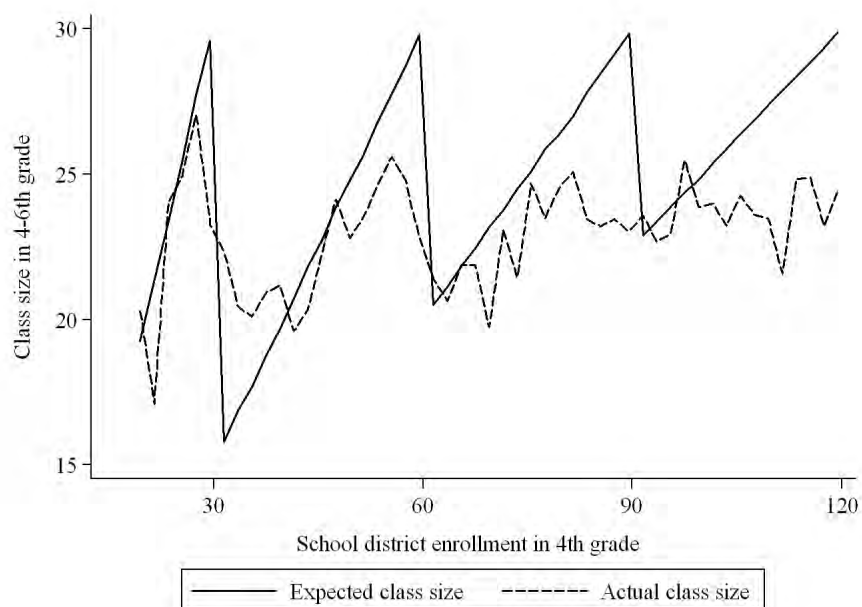
Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. All models controls for cohort×municipality fixed effects. Actual class size in grades 4-6 is instrumented with the expected class size in grade 4 as predicted by the class size rule at the school district level. Predicted cognitive skills at age 13 comes from a regression of cognitive skills on an intercept, gender, dummy variables for month of birth, dummy variables for mother’s and father’s educational attainment, a third order polynomial in parental income, mother’s age at child’s birth, indicators for being a first or second generation immigrant, having separated parents and the number of siblings. The predicted cognitive skills have been standardized. The relation between the instrument and separate background variables are presented in Table A3 in the appendix. The F-test is a joint test that all background variables are unrelated to the instrument, and is based on a separate regression of the instrument on all the background variables. Standard errors adjusted for clustering at the cohort×school district level are in parentheses.

We also analyzed the relation between the instrument and each background variable. Results are presented in Table A3 in the appendix, and confirm the validity of the instrument. The same is true for the results from separate regressions of the instrument on all the background variables. Table 2 reports the p-values of the F-test for joint significance of the background variables. In short, our RD approach survives all common specification tests.

Strength of the instrument Figure 3 illustrates the relations between school district enrollment in 4th grade on the horizontal axis, and actual and expected class size on the vertical axis. The solid line shows expected class size in case class size would be entirely determined by the maximum class size rule, the dashed line pertains to actual class size. When school district enrollment reaches a multiple of 30, actual average class size falls. This is particularly the case when school district enrollment passes 30 and when it passes 60.

For the full sample the first stage estimate in a specification with a third degree polynomial of school district enrollment in grade 4 is 0.335 (with s.e. 0.051). The first stage estimate

Figure 3. Expected and actual class size in grades 4-6 by school district enrollment in grade 4

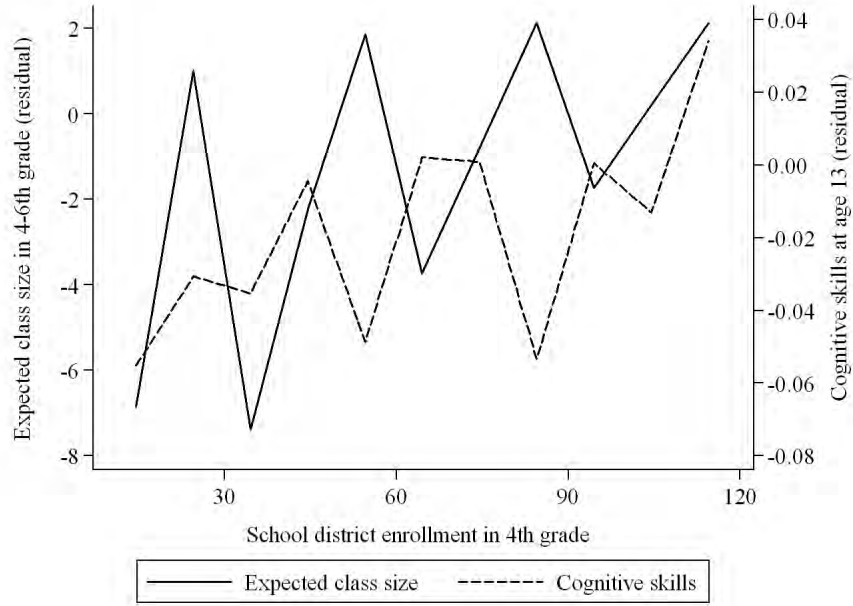


is very similar in the various sub-samples that we will consider: for women it is 0.337 (s.e. 0.052), for men 0.333 (s.e. 0.053), for individuals with low-income parents 0.321 (s.e. 0.052) and for individuals with high-income parents 0.347 (s.e. 0.057). F-values are all around 40. While this is lower than the F-values in some studies that use the maximum class size rule at the school level (Angrist and Lavy, 1999; Leuven et al., 2008), it is substantially above the “critical value” of 10 which is often seen as the critical value for the presence of a weak instrument problem.

5 The effects of class size

We start with a graphical analysis. Figure 4 presents the relation between the class size rule and cognitive skills at age 13. To remove some noise in the outcome measures, we use the residual from a regression where we control for pre-determined characteristics (see Table 3 for a list of these characteristics). The solid line represents the relation between 4th grade enrollment in the school district and expected class size in grades 4 to 6 based on the maximum class size rule. The dashed line represents the relation between 4th grade enrollment in a school district and cognitive skills at age 13. The values of expected class size and the skill test are computed at multiples of 10 of 4th grade enrollment. Hence, for school

Figure 4. Expected class size and cognitive skills by school district enrollment



district enrollment equal to 5, the graph indicates the expected class size and the residual skills averaged over enrollment from 1 to 10. Likewise, at school district enrollment equal to 15, the graph indicates expected class size and residual skills averaged over enrollment from 11 to 20, and so on.

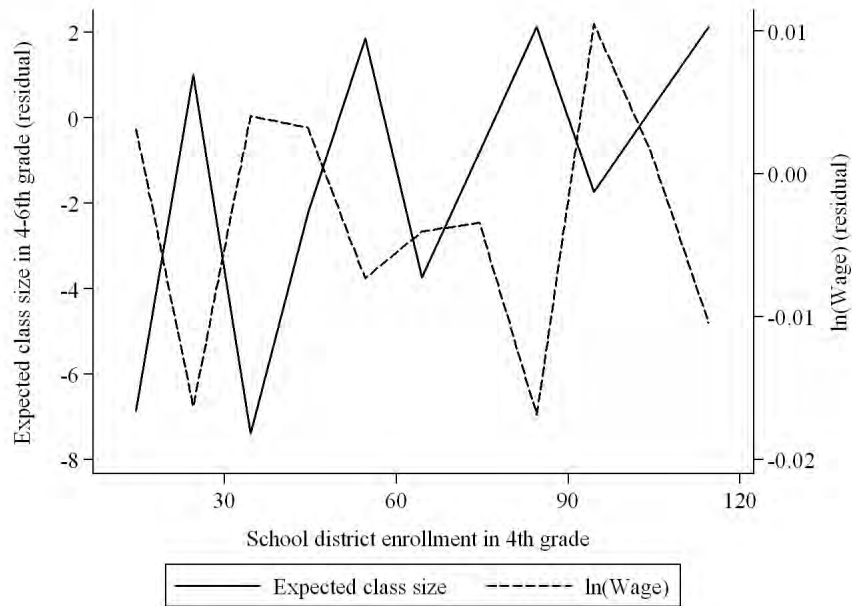
The dashed line mirrors the solid line. When expected class size increases, the score on the cognitive skill test decreases, and vice versa. This suggests a negative impact of class size on the cognitive score. Figure 5 shows the same relationships, but here wages at age 27-42 is the outcome variable. Also here the dashed line mirrors the solid line, suggesting that there is also a negative impact of class size on wage earnings.

To further investigate the impact of class size on various outcome variables, we will adopt the following specification of the outcome equation throughout.

$$y_{ijsdm} = \alpha_{jm} + \beta CS_{js} + f(E_{jd}) + \gamma X_i + \varepsilon_{ijsdm} \quad (1)$$

In equation (1), y denotes the outcome which varies by individual (i), cohort (j), school (s), school district (d) and municipality (m). By including cohort by municipality fixed effects (α_{jm}) we identify the effect of class size (β) using the variation across school districts within municipality for each cohort. There are two main reason for including the cohort by municipality fixed effects. First, and most importantly, we want to make sure that we compare the

Figure 5. Expected class size and wages by school district enrollment



comparable; notice that the specification analysis is conditional on the cohort by municipality fixed effects. Second, these fixed effects account for the sampling of the data.¹⁷ CS_{js} denotes the actual average class size to which pupils of cohort j in school s were exposed during their three grades (4, 5 and 6) in upper primary school, $f(E_{jd})$ is a function in enrollment in grade 4 for cohort j , in school district d , and X individual characteristics. ε is the error term where we allow for clustering at the school district by cohort level. In all regressions, actual average class size CS is instrumented by predicted class size in grade 4, where the prediction is based on total enrollment in grade 4 in the school district.

5.1 Short-term and medium-term outcomes

Table 3 shows OLS and IV estimates of the effect of class size on cognitive skills at age 13. The OLS estimate in the first column is a very precisely estimated zero. IV estimates are presented for six different specifications of the function $f(E_{jd})$. Columns (2) to (4) include a linear, quadratic and cubic controls for enrollment, respectively. The fifth column allows for linear splines in enrollment, and the sixth column allows for quadratic splines. In the final column the sample has been restricted to districts in which enrollment is at most 5 pupils

¹⁷For each cohort, the sampling procedure was to select 30 (out of some 300) municipalities. The cohort by municipality fixed effects are included to adjust for the fact that the selected municipalities may be non-representative of the wider population.

away from a cut-off. The estimates in columns (2) to (7) are all very similar, implying that it does not matter much how we exactly control for the forcing variable. From now on we will always include a third order polynomial in school district enrollment. The results for other outcome variables are also insensitive to this choice.

Table 4 presents estimates of the impact of class size on short-term and medium-term outcomes. Each row refers to a different dependent variable and each column to a different sample or subsample. The first column presents results for all 30,818 observations together. Columns (2) and (3) present results separately for women and men, and columns (4) and (5) present results separately for pupils with low (below median) and high (above median) income parents.

The first two rows relate to short-term outcomes, cognitive and non-cognitive ability measured at the end of primary school when students are 13 years old. The point estimates for the entire sample are negative and significantly different from zero. Placement in a small class during grades 4 to 6 increases cognitive as well as non-cognitive ability at age 13. Reducing class size by 1 pupil increases the score for cognitive ability by 0.020 SD units and the score for non-cognitive ability by 0.017 SD units.

The effect on cognitive ability is larger for boys than for girls, and is very similar for pupils with low income parents and pupils with high income parents. Also the effect on non-cognitive ability at age 13 is larger for boys than for girls. For non-cognitive ability, the effects are also different for pupils from different social background. Children from high income parents benefit from a class size reduction, while the impact is basically zero for children from low income parents.

The second time we have an outcome measure is at the end of lower secondary school when pupils are 16 years old. This is three years after pupils left primary school. Only academic achievement has been measured, and the results are reported in the third row. For all students together, we find a significantly negative impact of class size on academic achievement. The size of the effect is equal to the effect on cognitive ability measured immediately at the end of the exposure to class size in grades 4 to 6. There is thus no evidence of fade-out. The effects are very similar for boys and girls, and for children from high and low income families.

The next time outcomes are measured is at age 18. Since the results come from the military draft they are only available for boys. For all boys together the impact on cognitive ability is negative and significant at the 11%-level. The magnitude of the impact is very similar to the impact measured at age 16, suggesting again that the effect does not fade out. Splitting the sample by parents' income gives negative point estimates for both groups, but

Table 3. OLS and IV estimates of class size on cognitive skills at age 13, different enrollment controls

Model	OLS			IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Average class size 4th-6th grade	0.001 (0.002)	-0.019** (0.009)	-0.021** (0.009)	-0.020** (0.010)	-0.020* (0.010)	-0.021* (0.011)	-0.024** (0.012)
Enrollment controls							
Polynomial:							
- 1st order		√			√		
- 2nd order			√			√	
- 3rd order				√			
Interacted with break-points					√		
Districts with enrollment ± 5 pupils from cut-off						√	
N	30,818	30,818	30,818	30,818	30,818	30,818	15,798

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

Table 4. IV estimates of class size in 4th-6th grade on short-term and medium-term outcomes

Dependent variable	All	Women	Men	Parents' income	
				Low	High
Cognitive ability, age 13	-0.020** (0.010)	-0.012 (0.012)	-0.028** (0.012)	-0.021 (0.012)	-0.019* (0.011)
Non-cognitive ability, age 13	-0.017* (0.010)	-0.011 (0.012)	-0.023* (0.013)	-0.003 (0.015)	-0.024** (0.010)
Academic achievement, age 16	-0.020** (0.009)	-0.019 (0.012)	-0.022* (0.012)	-0.017 (0.013)	-0.021** (0.011)
Cognitive ability, age 18	.	.	-0.018 (0.011)	-0.029* (0.016)	-0.009 (0.015)
Non-cognitive ability, age 18	.	.	-0.017 (0.013)	-0.009 (0.016)	-0.024 (0.018)
<i>N</i>	30,818	15,076	15,742	15,271	15,547

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. All measures of cognitive ability, non-cognitive ability and academic achievement have been standardized. The abilities at age 18 pertain to men only. Actual class size in grades 4-6 is instrumented with the expected class size in grade 4 as predicted by the class size rule at the school district level. All models control for cohort×municipality fixed effects, gender, dummy variables for month of birth, dummy variables for mother's and father's educational attainment, a third order polynomial in parental income, mother's age at child's birth, indicators for being a first or second generation immigrant, having separated parents and the number of siblings, and a third order polynomial of school district enrollment in grade 4. High (low) income parents means that the parents' total earnings is above (below) the median. There is small internal attrition (less than 1 percent) for the separate ability tests. Standard errors adjusted for clustering at the school district×cohort level are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 per cent level of confidence, respectively.

– somewhat surprisingly – the estimate is only significant for sons of low income parents. The estimate of the effect on non-cognitive ability at age 18 is also negative but imprecisely estimated. This is true for all boys together as well as for both subgroups.

5.2 *Long-term outcomes*

We now turn to the effects of class size in upper primary school on long-term outcomes. Table 5 shows the results. Again, the first column presents estimates for all observations together, while the other four columns present estimates for subgroups.

The results in the first row suggest that a reduction of class size increases years of completed schooling. The estimate is, however, only statistically significant for women. A reduction of class size by one pupil during the last three years of primary school increases completed years of schooling of women by three weeks. The point estimates for individuals with low and high income parents are very similar, but lack precision.

In the next row, education is measured as a binary indicator of having obtained a Bachelor's degree or higher. All point estimates are negative but they are only statistically significant for women and for people with high income parents. For these groups, every one pupil reduction in class size in upper primary school increases the probability to have completed at least a Bachelor's degree by 1 percentage point.

In the next row we look at log wages in full-time equivalents as outcome measure. For all observations together, we find a 0.7 percent increase in wages for each one pupil reduction in class size. This estimate is significantly different from zero at the 5%-level. This is the most important finding of this paper. No previous study has been able to demonstrate significantly negative effects of class size in primary school on wage earnings of adults, using a credible identification approach.

Breaking the sample down by gender, gives negative point estimates for both genders. While both estimates are negative, it is only significantly different from zero for men. Each one pupil reduction in class size increases the wages for men by 1 percentage point. Breaking down the sample by parental income, reveals that the negative effect is entirely concentrated among individuals with high income parents. The estimate for this group indicates a 1.3 percent wage increase for each one pupil reduction in class size. This estimate is significant at the 1%-level.

The final two rows present results for annual earnings. The fourth row shows that class size variations have no effect on the probability of working (having positive annual earnings). This is true on average as well as for the various subgroups we consider. Since the probability

Table 5. IV estimates of class size in 4th-6th grade on long-term outcomes

Dependent variable	All	Women	Men	Parents' income	
				Low	High
Years of schooling, age 27-42	-0.031 (0.021)	-0.063** (0.027)	-0.007 (0.027)	-0.031 (0.029)	-0.034 (0.027)
P(Bachelor's degree), age 27-42	-0.005 (0.004)	-0.010* (0.005)	-0.002 (0.004)	-0.003 (0.005)	-0.009* (0.005)
ln(Wage), age 27-42	-0.007** (0.003)	-0.004 (0.003)	-0.010** (0.005)	0.000 (0.003)	-0.013*** (0.004)
P(Earnings>0), age 27-42	-0.000 (0.002)	-0.003 (0.003)	0.002 (0.003)	-0.004 (0.004)	0.004 (0.003)
Earnings, age 27-42	-0.004 (0.005)	-0.013* (0.007)	0.003 (0.007)	0.004 (0.007)	-0.008 (0.007)
<i>N</i>	30,818	15,076	15,742	15,271	15,547

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. The educational outcomes are measured 2009, while the labour market outcomes have been averaged over the 2007-2009 period. The earnings estimates are expressed as shares of average earnings for the group. The ln(wage) estimates are restricted to wage-earners. Actual class size in grades 4-6 is instrumented with the expected class size in grade 4 as predicted by the class size rule at the school district level. All models control for cohort×municipality fixed effects, gender, dummy variables for month of birth, dummy variables for mother's and father's educational attainment, a third order polynomial in parental income, mother's age at child's birth, indicators for being a first or second generation immigrant, having separated parents and the number of siblings, and a third order polynomial of school district enrollment in grade 4. High (low) income parents means that the parents' total earnings is above (below) the median. Standard errors adjusted for clustering at the school district×cohort level are in parentheses. ***/**/*=the estimates are significantly different from zero at the 1/5/10 per cent level of confidence, respectively.

of working is unaffected by variations in class size, the wage effects are not driven by the fact that wages are observed for the selected subsample of workers.

The fifth (and final) row examines annual earnings directly. We include those with zero earnings and, to facilitate interpretation, express earnings as a percentage of the average (in the sample, or sub-sample). The earnings effects are driven by the effect on wages, the probability of working, as well as any effect on annual hours. Given that we find no effect on the probability of working, it is not surprising to see that there is no significant effect on earnings on average. For women, however, earnings appear to increase by 1.3% relative to the average when class size is reduced by one.

5.3 *Effects over the distributions of early test scores and wages*

The results in the previous subsections show negative effects of class size on short/medium-term and long-term outcomes. The short/medium-term effects are somewhat larger for men than for women. The long-term effects are mainly concentrated among men and individuals with high income parents. In this subsection we inquire heterogeneity of the impact of class size further by examining how it varies over the distributions of the outcome variables (see Angrist and Imbens, 1995).

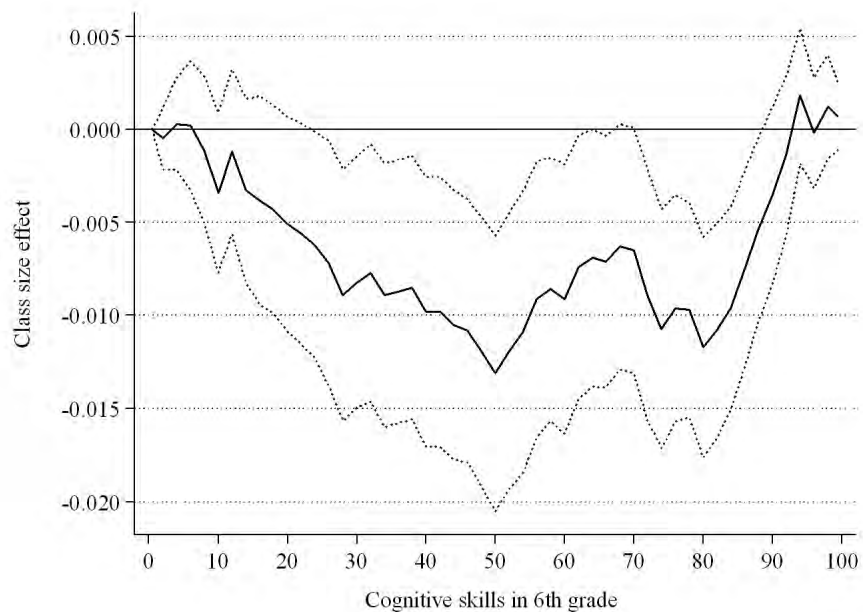
Figure 6 shows the effect of increasing class size by one student on the probability of having cognitive skills above the corresponding percentile rank on the horizontal axis. The effects are negative across almost the entire distribution, which is consistent with children of low income parents (who are more likely to end up below the median of the cognitive test score distribution) and children of high income parents (who are more likely to end up above the median of the cognitive test score distribution) experiencing very similar effects of class size on cognitive ability.

Figure 7 plots the effects of class size on the probability of having a wage above the corresponding percentile rank on the horizontal axis. Here we observe that the negative effect of class size is concentrated above the median of the wage distribution. This is consistent with the different estimates for children from low income parents and children from high income parents we reported above.

5.4 *Implications*

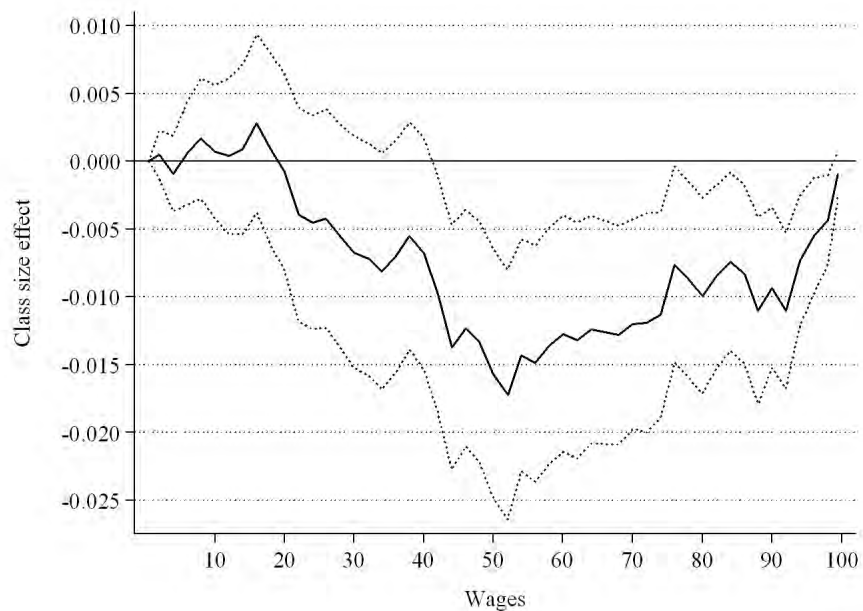
Comparison with indirect estimates Chetty et al. (2011) present an indirect estimate of the effect of class size on wage earnings by multiplying the effect of class size on cognitive ability with the cross-sectional correlation between cognitive ability and wage earnings. Following

Figure 6. The effect of class size in 4-6th grade over the distribution of cognitive skills at age 13 (percentile ranks)



Note: The solid line shows how one additional student in 4th-6th grade affects the probability of having cognitive skills above the percentile rank reported on the horizontal axis. The dotted lines show the 90% confidence interval.

Figure 7. The effect of class size in 4-6th grade over the distribution of wages at age 27-42



Note: The solid line shows how one additional student in 4th-6th grade affects the probability of having a wage above the percentile rank reported on the horizontal axis. The dotted lines show the 90% confidence interval.

this approach using our data, gives an estimate of $-0.02 \times 8.4\% = -0.17\%$.¹⁸ When we add the “imputed” impact of non-cognitive skills, the estimate increases to $(-0.02 \times 7.8\%) + (-0.017 \times 2.4\%) = -0.20\%$. If we would instead follow Bingley et al. (2010) and use the impact of class size on completed years of education in the two-stage procedure, the estimates are -0.12% for men and women together and -0.25% for women.¹⁹ All these indirect estimates are substantially below the estimate of -0.7% per pupil that we find when we estimate the wage effect directly.

Cost-benefit analysis The ultimate question is now whether the benefits of class size reduction outweigh the costs of such an intervention. Important here is that the costs are incurred when children are 10 to 13 years old, while the benefits in terms of wage earnings only start to accrue when these children are adults that entered the labour market. A cost-benefit analysis shows that for all reasonable discount rates the present value of the benefits exceeds the present value of the costs. In calculating the benefits we focus on the wage effect. The wage effect is arguably a better estimate of how individuals’ productivity is affected by a class size reduction than the earnings effect. The variation in annual earnings reflect preferences and labor supply choices to a greater extent than wages.

Assume average class size during upper primary school is reduced from 25 to 20. This increases the number of teachers from 4 per 100 pupils to 5 per 100 pupils, thereby increasing the per pupil wage costs by 1% of teachers’ average wage during three years. There are also costs involved with overhead and extra classrooms; say that this adds one third to the extra costs of teachers. The present value of the costs - starting when pupils are 10 years old - is then: $\sum_{t=0}^2 0.01w(1 + \frac{1}{3})/(1+r)^t$, where w is the annual wage of a teacher and r the discount rate. Assume further that average wages in the country are approximately equal to the average teacher wage, and that people work from age 21 until age 65. The present value of the benefits is then: $\sum_{t=10}^{54} 0.035w/(1+r)^t$, where 0.035 is five times our estimate of the effect of a one pupil reduction of class size on wage earnings. The internal rate of return (the discount rate that equalizes the present values of costs and benefits) is equal to 0.186. For discount rates below this value, the net present value of a 5 pupil reduction in class size is positive.

These calculations assume that the same quality teachers can be hired at a constant wage rate, and that the supply of more skilled labour does not affect the wage return to the class size reduction. The internal rate of return would be lower if one of these assumptions does

¹⁸The -0.02 comes from Table 4, the 8.4% is the average of 9.0% (for men) and 7.8% (for women) from Table 1.

¹⁹This combines the class size effects from the first row of Table 5 and a rate of return to education of 4%.

not hold. But even if we double the the costs and cut the benefits in half, the internal rate of return is still quite high: 0.094. This all implies that in the context of Sweden of the 1980s, a class size reduction in upper primary school would have been a beneficial intervention.

6 Conclusion

This is the first paper that documents significantly negative effects of class size in primary school on adult wage earnings. Previous attempts were plagued by lack of precision (Chetty et al., 2011) or unavailability of directly linked data on labor market outcomes (Schanzenbach, 2007; Bingley et al., 2010). The size of the effect of class size on wages is of the same order of magnitude as the effects of class size on short-term and medium-term cognitive and non-cognitive skills. We thus find no evidence of fading-out.

Our estimates of the wage effects of class size are much larger than estimates obtained using a two-stage procedure. The wage effects are substantive, and given that we measure wages at age 27-42 these effects can be considered as permanent effects. Using our estimates of the wage effects in a cost-benefit analysis reveals that the present value of the benefits outweigh the directly incurred costs. The internal rate of return is almost 20%.

Many previous studies have found negative effects of class size in primary school on short-term achievement. None of these studies has been able to demonstrate that these effects may be long-lasting. There is no reason to believe that the permanence of the impact of class size that we find, is attributable to specificities of the Swedish context. There is for instance no strong intertemporal correlation of class size in Sweden.²⁰ There is also no evidence that the return to skill is higher in Sweden than elsewhere.²¹

References

Almond, D. and Currie, J. (2010). Human capital development before age five. Working Paper No. 15827, NBER.

Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal

²⁰Recall that the “impact” of class size in upper primary school on class size in lower secondary school is only 0.14 (with s.e. 0.18).

²¹Using data from the International Adult Literacy Survey, Leuven et al. (2004) estimate wage regressions for 15 different countries. In a specification with only years of education and experience (squared) the return to education in Sweden is 0.034 which is lower than in any of the other 14 countries. Including a measure of cognitive skill lowers the return to education in Sweden to 0.028, again lower than in any other country. The return to cognitive skill from this regression is in Sweden very close to the average of the 15 countries.

- effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442.
- Angrist, J. D. and Pischke, J. K. (2009). *Mostly Harmful Econometrics: An Empiricist's Perspective*. Boston: Harvard University Press.
- Angrist, J. D. and Pischke, J. K. (2014). *Mastering Metrics: The Art and Science of Regression Analysis and Causal Inference*. Boston: Harvard University Press.
- Bingley, P., Jensen, V. M., and Walker, I. (2010). The effect of class size on education and earnings: Evidence from Denmark. Unpublished Working Paper.
- Björklund, A., Clark, M., Edin, P., Fredriksson, P., and Krueger, A. (2005). *The Market Comes to Education in Sweden - An Evaluation of Sweden's Surprising School Reforms*. New York: Russell Sage Foundation.
- Böhlmark, A. and Lindquist, M. J. (2006). Life-cycle variations in the association between current and lifetime income: Replication and extension for Sweden. *Journal of Labor Economics*, 24:879–896.
- Carlstedt, B. and Mårdberg, B. (1993). Construct validity of the Swedish enlistment battery. *Scandinavian Journal of Psychology*, 34:353–362.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from project STAR. *Quarterly Journal of Economics*, 126(4).
- Du Rietz, L., Lundgren, U., and Wennås, O. (1987). Ansvarsfördelning och styrning på skolområdet. Technical Report DsU 1987:1, Stockholm: Ministry of Education.
- Gary-Bobo, R. J. and Mahjoub, M. B. (2006). Estimation of class-size effects, using Maimonides' rule and other instruments: The case of French junior high schools. Discussion Paper No. 5754, CEPR.
- 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.
- Leuven, E., Oosterbeek, H., and Rønning, M. (2008). Quasi-experimental estimates of the effect of class size on achievement in Norway. *Scandinavian Journal of Economics*, 110:663–693.
- Leuven, E., Oosterbeek, H., and van Ophem, H. (2004). Explaining international differences in male wage inequality by differences in demand and supply of skill. *Economic Journal*, 144:478–498.
- Lindqvist, E. and Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics*, 3(1):101–128.
- Schanzenbach, D. W. (2007). What have researchers learned from Project STAR? *Brookings Papers on Education Policy*, 2006/2007:205–228.
- 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

Table A1: Descriptive statistics, 1967-1982 birth cohorts

	All	Girls	Boys	Income parents	
				Low	High
Girl	0.49 (0.50)	1.00 (0.00)	0.00 (0.00)	0.49 (0.50)	0.49 (0.50)
Mother's years of education	10.97 (2.70)	10.99 (2.71)	10.96 (2.69)	10.08 (2.36)	11.87 (2.71)
Father's years of education	10.70 (2.97)	10.70 (2.98)	10.70 (2.96)	9.67 (2.43)	11.73 (3.11)
Cognitive ability, age 13	0.00 (1.00)	0.03 (0.99)	-0.03 (1.01)	-0.19 (1.00)	0.18 (0.96)
Non-cognitive ability, age 13	0.00 (1.00)	-0.03 (1.00)	0.03 (1.00)	-0.09 (1.05)	0.09 (0.94)
Academic achievement, age 16	0.00 (1.00)	0.10 (0.97)	-0.10 (1.02)	-0.20 (1.00)	0.20 (0.96)
Cognitive ability, age 18	.	.	0.00 (1.00)	-0.21 (1.01)	0.20 (0.95)
Non-cognitive ability, age 18	.	.	0.00 (1.00)	-0.17 (0.96)	0.16 (1.01)
Years of schooling, age 27-42	13.48 (2.60)	13.87 (2.61)	13.11 (2.53)	12.98 (2.50)	13.97 (2.60)
Bachelor's degree, age 27-42	0.27 (0.44)	0.32 (0.47)	0.21 (0.41)	0.20 (0.40)	0.33 (0.47)
Earnings, age 27-42	241730 (177644)	202548 (135533)	279127 (203183)	220101 (163461)	263350 (188310)
Full-time employed, age 27-42	0.91 (0.28)	0.91 (0.28)	0.91 (0.29)	0.89 (0.31)	0.93 (0.25)
ln(Wage), age 27-42	10.15 (0.27)	10.09 (0.23)	10.23 (0.30)	10.11 (0.24)	10.19 (0.29)
<i>N</i>	30,818	15,076	15,742	15,271	15,547

Note: The data are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. All measures of cognitive ability, non-cognitive ability and academic achievement have been standardized. The abilities at age 18 pertain to men only. The educational outcomes are measured 2009, while the labour market outcomes have been averaged over the 2007-2009 period. Wages are restricted to wage-earners. High (low) income parents means that the parents' total earnings is above (below) the median. There is small internal attrition (less than 1 percent) for the separate ability tests. Standard deviations are in parentheses.

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

	All	Girls	Boys	Income parents	
				Low	High
<u>Class variables</u>					
Class size in grade 4	23.66 (4.61)	23.67 (4.57)	23.65 (4.64)	23.28 (4.66)	24.04 (4.52)
Class size in grade 4>30	0.02 (0.11)	0.02 (0.11)	0.03 (0.11)	0.02 (0.10)	0.03 (0.12)
Average class size grades 4-6	23.84 (4.13)	23.85 (4.11)	23.84 (4.15)	23.49 (4.19)	24.20 (4.04)
<u>School district variables</u>					
Enrollment 4th grade	106.58 (42.42)	106.32 (42.62)	106.84 (42.23)	105.09 (41.26)	108.08 (43.50)
Expected class size 4-6 grade	26.12 (2.84)	26.12 (2.84)	26.13 (2.84)	26.07 (2.89)	26.18 (2.79)
N individuals	30,818	15,076	15,742	15,271	15,547
N schools	1,291	1,291	1,291	1,291	1,291
N school districts	757	757	757	757	757

Note: The data are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. Standard deviations are in parentheses.

Table A2: Correlation matrix

	Cognitive ability, age 13	Non-cognitive ability, age 13	Academic achievement, age 16	Cognitive ability, age 18	Non-cognitive ability, age 18	Years of schooling, age 27-42	Bachelor's degree, age 27-42	Earnings, age 27-42	Full-time employed, age 27-42
Cognitive ability, age 13	1.00								
Non-cognitive ability, age 13	0.28	1.00							
Academic achievement, age 16	0.79	0.27	1.00						
Cognitive ability, age 18	0.74	0.26	0.78	1.00					
Non-cognitive ability, age 18	0.24	0.19	0.28	0.32	1.00				
Years of schooling, age 27-42	0.45	0.19	0.53	0.53	0.28	1.00			
Bachelor's degree, age 27-42	0.35	0.13	0.41	0.39	0.20	0.87	1.00		
Earnings, age 27-42	0.16	0.12	0.17	0.20	0.23	0.17	0.17	1.00	
Full-time employed, age 27-42	0.04	0.03	0.04	0.03	0.10	0.14	0.15	0.43	1.00
ln(Wage), age 27-42	0.30	0.17	0.33	0.33	0.26	0.28	0.24	0.80	-0.02

Note: The table show bivariate correlations between short-run, medium-run and long-run outcomes. The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982.

Table A3: Specification test: IV estimates of class size on different background variables

	(1)	(2)	(3)	(4)	(5)	(6)
Background variable:						
Woman	-0.000 (0.003)	0.001 (0.008)	0.003 (0.004)	0.002 (0.004)	0.003 (0.004)	0.002 (0.004)
Immigrant	-0.010** (0.005)	0.002 (0.006)	0.001 (0.006)	-0.000 (0.006)	-0.002 (0.006)	-0.001 (0.007)
Month of birth	-0.007 (0.022)	-0.001 (0.028)	-0.001 (0.028)	-0.010 (0.028)	-0.009 (0.030)	-0.035 (0.031)
Parents' years of education	0.032 (0.040)	-0.018 (0.048)	-0.026 (0.047)	-0.007 (0.046)	0.010 (0.046)	0.019 (0.049)
Parents' labour income	3240 (2407)	-822 (3017)	44 (3061)	-249 (3118)	-12 (3402)	-625 (3676)
Enrollment controls						
Polynomial:						
- 1st order		✓			✓	
- 2nd order			✓			✓
- 3rd order				✓		
Interacts with break-points					✓	✓
<i>N</i>	31,590	31,590	31,590	31,590	31,590	31,590

Note: The estimates are based on representative samples of individuals born in 1967, 1972, 1977 or 1982. All models controls for cohort×municipality fixed effects. Actual class size in grades 4-6 is instrumented with the expected class size in grade 4 as predicted by the class size rule at the school district level. Standard errors adjusted for clustering at the cohort×school district level are in parentheses.