Long-term effects of class size

Peter Fredriksson
Björn Öckert
Hessel Oosterbeek

WORKING PAPER 2012:5

Long-term effects of class size[*]

Peter Fredriksson   Björn Öckert   Hessel Oosterbeek

Abstract

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-codes: I21, I28, J24, C31

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

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-

---

[1]Findings 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).

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 Krueger 2003 and 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 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 for 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

---

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

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 observed 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 (fuzzy) 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 and non-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 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

---

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

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

[5]There 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 Abecederian pre-school demonstrations, and Head-start (see Almond and Currie, 2010, for a survey of the pre-school interventions).

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 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 there were typically more primary schools than lower secondary schools in the school district. There was at least one school district in a municipality.

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

---

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

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

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

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] There main reason for this focus is data availability. We do not have precise information on schools (and hence school districts) attended for lower primary school.

The maximum class size rule at the upper primary level stipulated that classes were formed in multiples of 30; 30 students in a grade level in a school yielded one class, while 31 students in a grade level in a school yielded two classes, and so on.[11] 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.[12]

Implementing the RD design requires some care, however. The compulsory school law from 1962 opened up for adjustment of school catchment areas such that empty class room would be filled. Moreover, the county school boards were instructed to take the "needs" of the pupil population into account when adjusting the catchment areas. Thus, it is likely that the catchment areas are adjusted to favor disadvantaged pupils. In a companion paper we show that such that sorting takes place, rendering the RD design at the school level invalid.[13] Because of these problems, we 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 traveling 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

---

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

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

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

[13]In Fredriksson et al. (2011) 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.

able to attribute our findings to average class size in grades 4 to 6, class size in these grades should not correlate 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. To shed light on this issue, we examined the class size data across the various stages of compulsory school. We found that the correlations are small and statistically insignificant.[14]

For the RD design to be credible, other school resources should not exhibit the same discontinuous pattern. There is no such pattern. 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 compulsory school students 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. Among other things, the data contain cognitive test scores at age 13 for roughly a 10 percent sample of the cohorts born 1967, 1972, and 1982. In addition, there is information on a 5 percent sample for the cohort born in 1977.[15]

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

---

[14]For instance, we found that the "effect" of increasing class size in upper-secondary school on class size in lower secondary school was 0.07 (s.e. 0.15). This correlation was estimated controling for a 2nd order polynomial in school district enrollment in the 4th grade interacted at the break-points.

[15]For the cohorts born 1967, 1972, and 1982, the sampling procedure was to sample roughly 30 (out of some 300) municipalities and then to randomly sample classes within municipality.

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.[16] 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.[17] We construct two measures based on this. The first is years of completed schooling,

---

[16]The 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.

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

Figure 1. Distribution of class size in grade 4
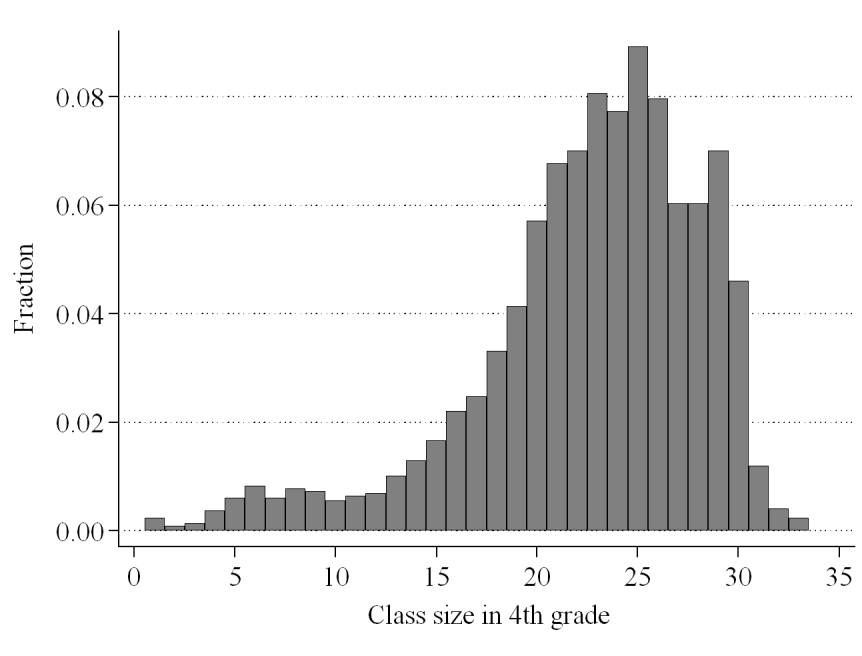
the second a binary indicator for having at least a Bachelor's degree. This measure is analogous to the college indicator used in studies based on the STAR experiment (see Krueger and Whitmore, 2001; Schanzenbach, 2007; Chetty et al., 2011). Data on annual earnings come from the Income Tax Register, 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.

Figure 2. Distribution of grade 4 school district enrollment
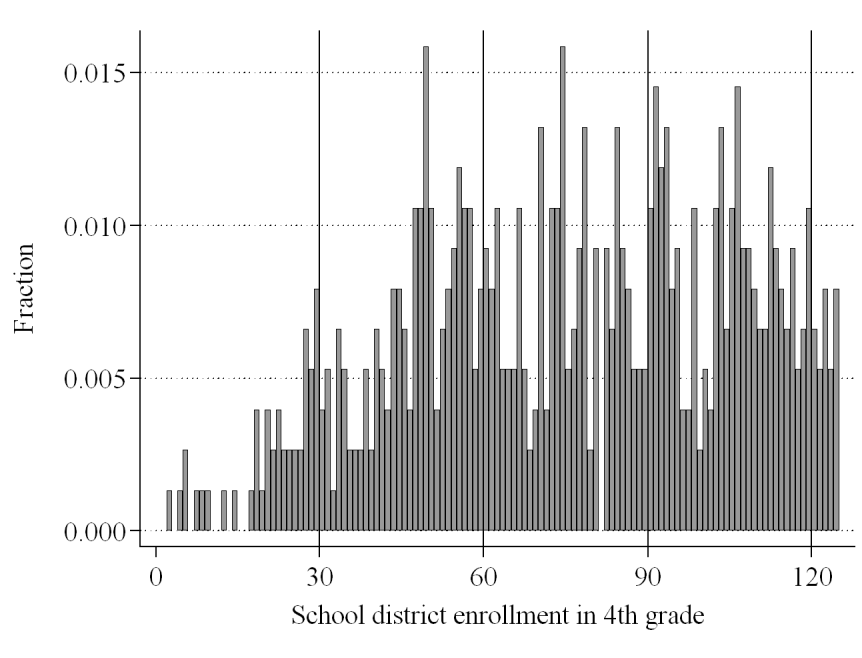
4 Validity and strength of the instrument

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.

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.

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

9

Table 1. 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	-0.006	-0.001	0.000	0.005	0.007
	(0.017)	(0.020)	(0.020)	(0.020)	(0.020)	(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
N	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.

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 1 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. We also analyzed the relation between the instrument and each background variable. Results are presented in Table A2 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 1 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.

Figure 3. Expected and actual class size in grades 4-6 by school district enrollment in grade 4
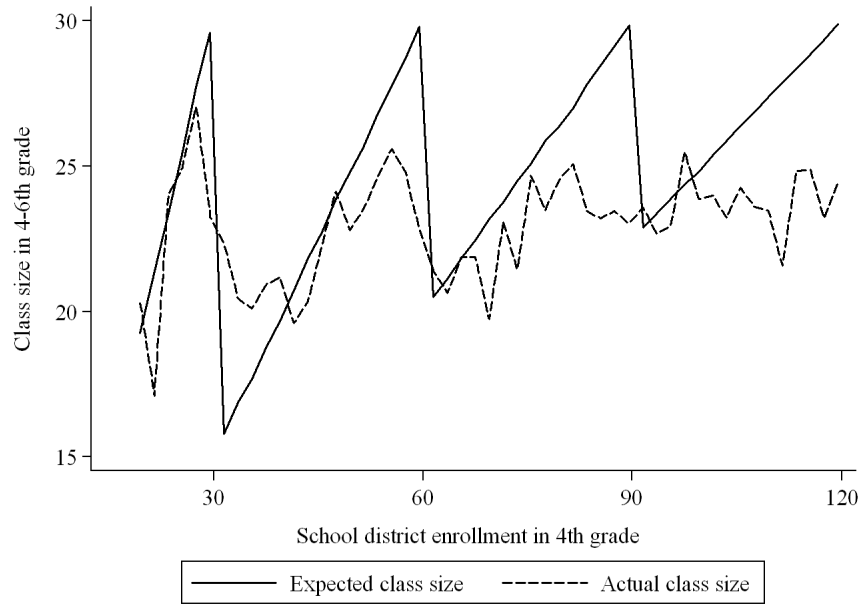
Strength of the instrument   Figure 3 illustrates the relations between school district enroll-ment in 4th grade on the horizontal axis, and actual and expected class size on the vertical axis. The solid line shows expected class size in case class size would be entirely determined by the maximum class 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 is very similar in the various sub-samples that we 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 (see Stock and Yogo 2005).

Figure 4. Expected class size and cognitive skills by school district enrollment
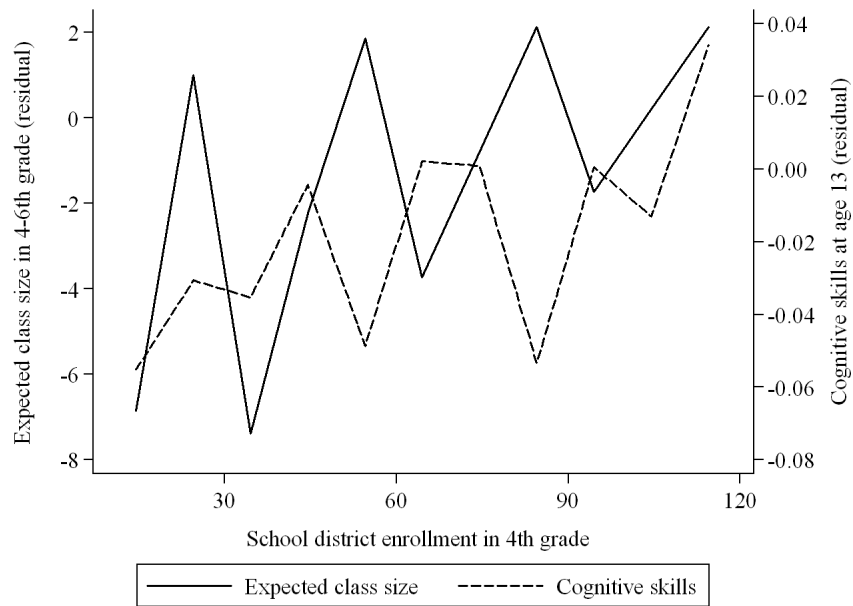
5   The effects of class size

We start with a graphical analysis for a subset of our outcome variables. To remove some noise we follow Angrist and Lavy (1999) by examining the residuals from regressions where we control for pre-determined characteristics (see Table 3 for a list of these characteristics) averaged over multiples of 10 of 4th grade enrollment.[18] Figure 4 presents the relation between the class size rule and cognitive skills at age 13. The solid line represents the relation between 4th grade enrollment in a school district and expected class size in grades 4 to 6 based on the maximum class size rule, while the dashed line shows the relation between between 4th grade enrollment in a school district and cognitive skills at age 13.

On the whole, 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, again that there is a negative impact of class size on wages.

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

---

[18]In Figures 4-5 the enrollment axes represent interval mid points.

12

Figure 5. Expected class size and wages by school district enrollment
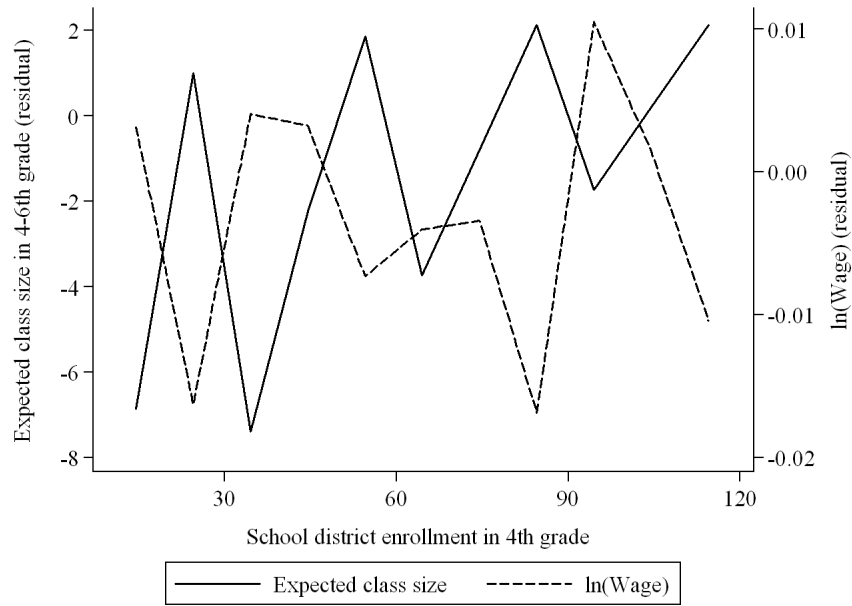
$$y_{ijsdm} = \alpha_{jm} + \beta CS_{js} + f(E_{jd}) + \gamma X_i + \varepsilon_{ijsdm} \qquad (1)$$

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. The main reason for including the cohort by municipality fixed effects is that we want to make sure that we compare the comparable.[19] 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.

The presentation of the results obtained from estimation of (1) is organized as follows. In Section 5.1 we present results for the short- and medium-term outcomes observed at ages 13, 16, and 18. Section 5.2 then turns to the long-term outcomes observed when the individuals

---

[19]It is not a priori obvious that a comparison of school districts across municipalities would be balanced in terms of observed and unobserved characteristics. It turns out, however, that the municipality by cohort fixed effects do not affect the results at all. The results do not change if we exclude these fixed effects.

are prime-aged. Section 5.3 examines whether the effects vary over the outcome distributions. Section 5.4 discusses our results in relation to previous studies and Section 5.5 discusses the quantitative implications of the results.

5.1 Short-term and medium-term outcomes

Table 2 shows OLS and IV estimates of the effect of class size on cognitive skills at age 13. The OLS estimate in the first column is a very precisely estimated zero. IV estimates are presented for six different specifications of the function $f(E_{jd})$. Columns (2) to (4) include a linear, quadratic and cubic controls for enrollment, respectively. The fifth column allows for linear splines in enrollment, and the sixth column allows for quadratic splines. In the final column the sample has been restricted to districts in which enrollment is at most 5 pupils away from a cut-off. The estimates in columns (2) to (7) are all very similar, implying that it does not matter much how we exactly control for the forcing variable. From now on we will always include a third order polynomial in school district enrollment. The results for other outcome variables are also insensitive to this choice.

Table 3 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 sub-sample. 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

14

Table 2. 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	-0.019**	-0.021**	-0.020**	-0.020*	-0.021*	-0.024**
	(0.002)	(0.009)	(0.009)	(0.010)	(0.010)	(0.011)	(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.

15

Table 3. IV estimates of class size in 4th-6th grade on short-term and medium-term outcomes

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

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 – 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 4 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 only statistically significant for women, however. 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

17

Table 4. 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	-0.063**	-0.007	-0.031	-0.034
	(0.021)	(0.027)	(0.027)	(0.029)	(0.027)
P(Bachelor's degree), age 27-42	-0.005	-0.010*	-0.002	-0.003	-0.009*
	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
ln(Wage), age 27-42	-0.007**	-0.004	-0.010**	0.000	-0.013***
	(0.003)	(0.003)	(0.005)	(0.003)	(0.004)
P(Earnings>0), age 27-42	-0.000	-0.003	0.002	-0.004	0.004
	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)
Earnings, age 27-42	-0.004	-0.013*	0.003	0.004	-0.008
	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
N	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 labor 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.

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 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 sub-sample of workers.

The fifth (and final) row examines annual earnings directly. We include those with zero earnings; to facilitate interpretation we express earnings as a percentage of the average (in the sample, or sub-sample). The earnings effects can be decomposed into the effects on wages, the probability of working, and annual hours. Given that we find no effect on the probability of working, the difference between the wage effect and the earnings effect comes from the hours response. Thus, the earnings effect for women comes entirely from variation in annual hours. When class size is reduced by one, earnings increase by 1.3% relative to the average; this effect is driven by an increase in female labor supply.

5.3   Effects on 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 affects the unconditional quantiles of the distributions of early test scores and wages (see Firpo et al. 2009).

Take the effects on the wage distribution as an example. The Firpo et al. (2009) approach then builds on two set of estimates: (i) estimate of the shift in the cumulative distribution function (CDF) of wages when class size is increased by one pupil; (ii) estimates of the unconditional density of the wage distribution. It is straightforward to estimate both of these

objects. To estimate the shift in the CDF we first percentile rank wages and then redefine the outcome variable such that it equals unity if the individual wage exceeds the wage at a given percentile; then this indicator variable is regressed on class size using the specification outlined in equation (1) (see Angrist and Imbens 1995 for an analogous exercise). The unconditional wage density is estimated using a kernel density estimator. The class size effect on the unconditional wage quantile is given by the shift in the CDF at that quantile multiplied by the inverse of the density at that quantile. The unconditional quantile regression estimates shows how the wage quantile is affected by a unit increase in class size.

Figure 6 shows the first set of unconditional quantile regression estimates.[20] It pertains to cognitive skills at age 13 and shows the class size effect on early test scores at different points of the percentile ranked test score distribution (the horizontal axis). Apart from at the end-points of the distribution, the effects are negative throughout the distribution. This 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. The lack of effect at the extremes of the distribution is presumably driven by floor and ceiling effects in the tests that generate our measure of cognitive skills.

Figure 7 plots the effects of class size on the wage quantiles (the horizontal axis). Here, the negative effects of class size are 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. We discuss this striking result further in the next subsection.

5.4 Comparison to previous studies

A brief summary of our major results is that a class size reduction equivalent to STAR (7 students) would improve cognitive skills at age 13 by 0.14 of a standard deviation (SD) and non-cognitive skills at age 13 by 0.12 SD on average. The effect on cognitive skills does not vary by parental income but the effect on non-cognitive skills is entirely concentrated among individuals whose parents earn above the median. In the long-term, a class size reduction equivalent to STAR improves wages by 4.9% on average; the effect is entirely driven by the impact for individuals whose parents earned above the median. How do these results compare to the previous literature?

---

[20]The confidence intervals are not corrected for estimation of the density.

Figure 6. Unconditional quantile regression estimates of the effect of class size on cognitive skills at age 13
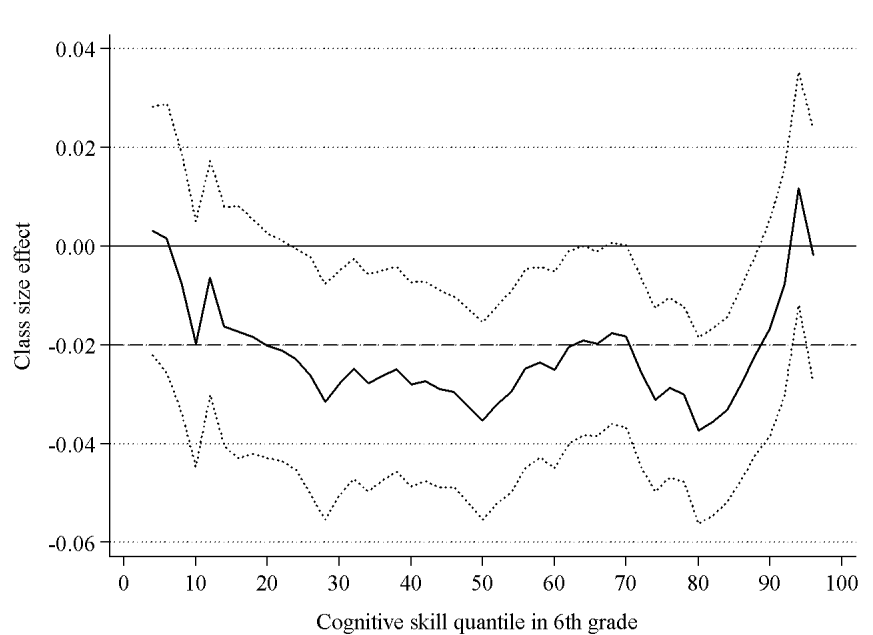
Note: The solid line shows how one additional student in 4th-6th grade affects the test score quantile. The dotted lines indicate the 90% confidence interval. The dashed line shows the relevant IV-estimate from Table 3.

Figure 7. Unconditional quantile regression estimates of the effect of class size on log wages at age 27-42
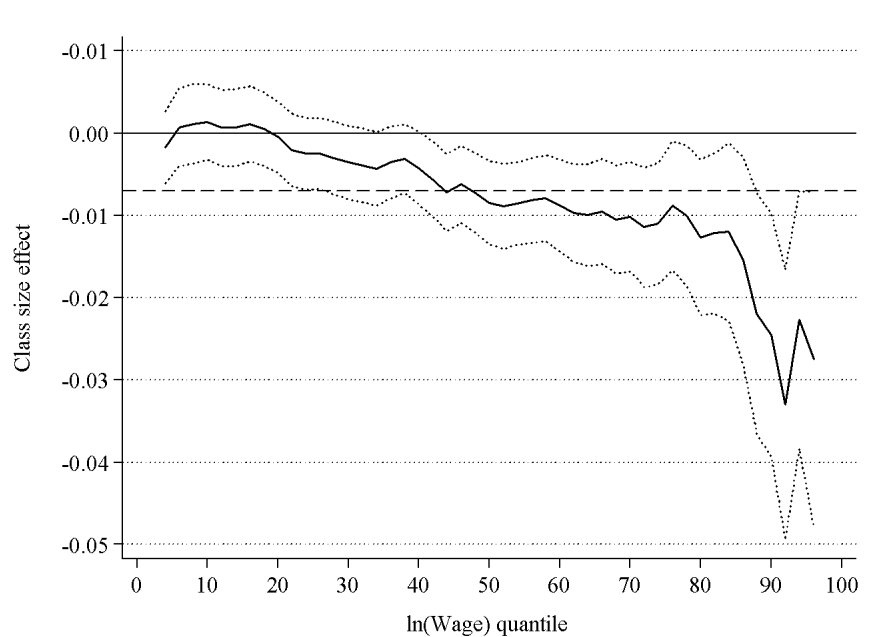
Note: The solid line shows how one additional student in 4th-6th grade affects the wage quantile. The dotted lines indicate the 90% confidence interval. The dashed line shows the relevant IV-estimate from Table 4.

21

The short-run effect on test scores is slightly smaller than the typical estimate from STAR; Krueger (1999) reports a short-run estimate in Kindergarten of 0.2 SD. Our estimate of 0.14 SD is in the mid range of the difference-in-differences estimates for Stockholm 6th graders reported in Lindahl (2005).

Analyses based on STAR data suggest a bigger short-run impact on test scores for students from low-income families (those eligible for free lunch), a finding that we do not reproduce. Notice that there is no a priori reason for believing that the effects should be greater for those having a weaker background. Meghir and Rivkin (2010) make this point using a slight extension of the Lazear (2001) model. The extension allows the value of instruction time to depend on initial achievement. Ultimately, the differential impact of class size reductions depends on aspects of the human capital production function, the targeting of the curriculum, how teacher effort is distributed in the class room etc.[21]

There is not much evidence on how class size reductions affect non-cognitive skills. The evidence based on STAR in Dee and West (2008) suggests that the effects on behavior in the 4th grade are about as large as the ones we report; there is no evidence of an impact on 8th grade behavior, however. Using other data, Dee and West examine whether the effects vary for by socioeconomic status; they find no such differences.

That we find a bigger impact among individuals who have high-income parents makes sense given that it is only in this group that we find an effect on non-cognitive skills. Since non-cognitive skills is valued on the labor market (see Lindqvist and Vestman, 2011), we should also expect bigger wage effects for individuals with high-income parents. Interestingly, the evidence on heterogeneous long-term effects reported in Chetty et al. (2011) point in the same direction as our wage estimates. Chetty et al. find that the small class effect on a summary index of various beneficial outcomes is more than twice as large for individuals who were not eligible for free lunch in comparison to individuals who were eligible for free lunch.

The fact that we find no wage effect for individuals from low-income families is more surprising, given that a number short- and medium term outcomes are affected in the expected direction. We think that Swedish wage-setting institutions contribute to this finding. At the lower end of the wage distribution (where individuals from low-income families are more likely to end up), wages are typically determined by collective bargaining, while at the higher end of the wage distribution individual wage bargaining is more common (National Media-

---

[21]Notice also that in the working paper version of Lindahl (2005), there is evidence that the impact of a class size reduction is greater for high-income families. This interaction term is not significant at the 10 percent level (the t-value is 1.45); see Lindahl (2001).

tion Office Sweden, 2011). Thus, Swedish wage setting institutions probably contribute to the differential impact across the parental income distribution.

5.5 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. The purpose of this sub-section is to illustrate what we would have concluded had we followed their approach.

To implement the indirect two-step approach, we need estimates of the correlation between cognitive and non-cognitive test scores and long-term wage outcomes. Table A4 reports the results of regressions of ln(Wage) on cognitive and non-cognitive test scores measured at age 13 for various groups.[22] The correlations between the short-term and the long-term outcomes are high. On average, a standard deviation increase in cognitive test scores is associated with a wage increase of 8.4%. Moreover, if cognitive and non-cognitive test scores are included jointly, both are highly significant.

With these estimates in hand we can implement the two-step approach using our data. We find an imputed wage impact of -0.02 × 8.4% = -0.17%. When we add the "imputed" impact of non-cognitive skills, the estimate increases to (-0.02 × 7.7%) + (-0.017 × 2.8%) = -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.[23] All these indirect estimates are substantially below the estimate of -0.7% per pupil that we find when we estimate the wage effect directly.

Since observed cognitive and non-cognitive ability measure limited dimensions of the skills that are priced on the labor market, we think it is natural that these imputation procedures yield a lower estimate than the direct wage impact. In fact, in our data one can impute a better estimate of the wage impact by multiplying the short run estimate (-0.02) with the standard deviation of the log wage distribution (0.27). Using this approach we would impute that the long-run wage impact is -0.54% per pupil, which is much closer to our direct estimate.

---

[22]Table A3 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.

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

Cost-benefit analysis The ultimate question is 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 labor 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.[24] 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 labor 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 not hold. But even if we double 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.

---

[24]Notice that by making the assumption that the average teacher wage equals the average future wages of those subjected to the policy, we abstract from productivity growth. This contributes to a downward bias in our rate of return calculations.

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 (Krueger, 2003; Schanzenbach, 2007; Bingley et al., 2010). In standard deviation terms, 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 have long-lasting effects on wages. 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 correlation in class size across the stages of compulsory school in Sweden. Moreover, reducing class size is a worthwhile investment even if we double the costs. There is also no evidence that the return to skill is higher in Sweden than elsewhere.[25]

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 effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442.

---

[25]Using 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.

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.

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).

Dee, T. and West, M. (2008). The non-cognitive returns to class size. Working paper 13994, NBER.

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

Firpo, S., Fortin, N. M., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3):953–973.

Fredriksson, P., Öckert, B., and Oosterbeek, H. (2011). The devil is in the (institutional) detail – on the perils of the regression discontinuity design. Unpublished manuscript, Stockholm University.

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. (2003). Economic considerations and class size. *Economic Journal*, 113:F34–F63.

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.

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

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.

Lindahl, M. (2001). Home versus school learning: A new approach to estimating the effect of class size on achievement. Discussion paper 261, IZA.

Lindahl, M. (2005). Home versus school learning: A new approach to estimating the effect of class size on achievement. *Scandinavian Journal of Economics*, 107(2):375–394.

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.

Meghir, C. and Rivkin, S. G. (2010). Econometric methods for research in education. Working paper 16003, NBER.

National Mediation Office (2011). Summary of the annual report for 2010. National mediation office.

Schanzenbach, D. W. (2007). What have researchers learned from Project STAR? *Brookings Papers on Education Policy*, 2006/2007:205–228.

Stock, J. and Yogo, M. (2005). Testing for weak instruments in linear iv regression. In Andrews, D. W. K. and Stock, J. H., editors, *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg*. Cambridge University Press.

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	1.00	0.00	0.49	0.49
	(0.50)	(0.00)	(0.00)	(0.50)	(0.50)
Mother's years of education	10.97	10.99	10.96	10.08	11.87
	(2.70)	(2.71)	(2.69)	(2.36)	(2.71)
Father's years of education	10.70	10.70	10.70	9.67	11.73
	(2.97)	(2.98)	(2.96)	(2.43)	(3.11)
Cognitive ability, age 13	0.00	0.03	-0.03	-0.19	0.18
	(1.00)	(0.99)	(1.01)	(1.00)	(0.96)
Non-cognitive ability, age 13	0.00	-0.03	0.03	-0.09	0.09
	(1.00)	(1.00)	(1.00)	(1.05)	(0.94)
Academic achievement, age 16	0.00	0.10	-0.10	-0.20	0.20
	(1.00)	(0.97)	(1.02)	(1.00)	(0.96)
Cognitive ability, age 18	.	.	0.00	-0.21	0.20
	.	.	(1.00)	(1.01)	(0.95)
Non-cognitive ability, age 18	.	.	0.00	-0.17	0.16
	.	.	(1.00)	(0.96)	(1.01)
Years of schooling, age 27-42	13.48	13.87	13.11	12.98	13.97
	(2.60)	(2.61)	(2.53)	(2.50)	(2.60)
Bachelor's degree, age 27-42	0.27	0.32	0.21	0.20	0.33
	(0.44)	(0.47)	(0.41)	(0.40)	(0.47)
Earnings, age 27-42	241730	202548	279127	220101	263350
	(177644)	(135533)	(203183)	(163461)	(188310)
P(Earnings>0), age 27-42	0.91	0.91	0.91	0.89	0.93
	(0.28)	(0.28)	(0.29)	(0.31)	(0.25)
ln(Wage), age 27-42	10.15	10.09	10.23	10.11	10.19
	(0.27)	(0.23)	(0.30)	(0.24)	(0.29)
N	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 labor 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	Income parents High
Class variables					
Class size in grade 4	23.66	23.67	23.65	23.28	24.04
	(4.61)	(4.57)	(4.64)	(4.66)	(4.52)
Class size in grade 4>30	0.02	0.02	0.03	0.02	0.03
	(0.11)	(0.11)	(0.11)	(0.10)	(0.12)
Average class size grades 4-6	23.84	23.85	23.84	23.49	24.20
	(4.13)	(4.11)	(4.15)	(4.19)	(4.04)
School district variables					
Enrollment 4th grade	106.58	106.32	106.84	105.09	108.08
	(42.42)	(42.62)	(42.23)	(41.26)	(43.50)
Expected class size 4-6 grade	26.12	26.12	26.13	26.07	26.18
	(2.84)	(2.84)	(2.84)	(2.89)	(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: Specification test: IV estimates of class size on different background variables

	(1)	(2)	(3)	(4)	(5)	(6)
Background variable:						
Woman	-0.000	0.001	0.003	0.002	0.003	0.002
	(0.003)	(0.008)	(0.004)	(0.004)	(0.004)	(0.004)
Immigrant	-0.010**	0.002	0.001	-0.000	-0.002	-0.001
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
Month of birth	-0.007	-0.001	-0.001	-0.010	-0.009	-0.035
	(0.022)	(0.028)	(0.028)	(0.028)	(0.030)	(0.031)
Parents' years of education	0.032	-0.018	-0.026	-0.007	0.010	0.019
	(0.040)	(0.048)	(0.047)	(0.046)	(0.046)	(0.049)
Parents' labor income	3240	-822	44	-249	-12	-625
	(2407)	(3017)	(3061)	(3118)	(3402)	(3676)
Enrollment controls						
Polynomial:						
- 1st order		√			√	
- 2nd order			√			√
- 3rd order				√		
Interacted with break-points					√	√
N	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. **=the estimates are significantly different from zero at the 5 per cent level of confidence.

Table A3: 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	P(Earnings>0), 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	
P(Earnings>0), 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 A4: Cross-sectional correlations between cognitive and non-cognitive scores at age 13 and ln(Wage)

	Dependent variable: ln(Wage)		
	(1)	(2)	(3)
<u>All</u>			
Cognitive ability	0.084***		0.077***
	(0.002)		(0.002)
Non-cognitive ability		0.047***	0.028***
		(0.002)	(0.002)
N	14770	14770	14770
<u>Women</u>			
Cognitive ability	0.078***		0.073***
	(0.002)		(0.002)
Non-cognitive ability		0.037***	0.019***
		(0.002)	(0.002)
N	8032	8032	8032
<u>Men</u>			
Cognitive ability	0.090***		0.082***
	(0.003)		(0.003)
Non-cognitive ability		0.051***	0.029***
		(0.003)	(0.003)
N	6738	6738	6738
<u>Low parental income</u>			
Cognitive ability	0.070***		0.065***
	(0.003)		(0.003)
Non-cognitive ability		0.036***	0.020***
		(0.003)	(0.003)
N	7017	7017	7017
<u>High parental income</u>			
Cognitive ability	0.090***		0.082***
	(0.003)		(0.003)
Non-cognitive ability		0.052***	0.029***
		(0.003)	(0.003)
N	7753	7753	7753

Note: Each column reports estimates from OLS regressions based on representative samples of individuals born in 1967, 1972, 1977 or 1982. The regressions are estimated separately per group and all models control for cohort×municipality fixed effects. ***=the estimates are significantly different from zero at the 1 percent level of confidence.

Publication series published by IFAU – latest issues

Rapporter/Reports

2012:1 Lundin Martin and Jonas Thelander ”Ner och upp – decentralisering och centralisering inom svensk arbetsmarknadspolitik 1995–2010”

2012:2 Edmark Karin, Che-Yuan Liang, Eva Mörk and Håkan Selin ”Jobbskatteavdraget”

2012:3 Jönsson Lisa and Peter Skogman Thoursie ”Kan privatisering av arbetslivsinriktad rehabilitering öka återgång i arbete?”

2012:4 Lundin Martin and PerOla Öberg ”Politiska förhållanden och användningen av expertkunskaper i kommunala beslutsprocesser”

2012:5 Fredriksson Peter, Hessel Oosterbeek and Björn Öckert ”Långsiktiga effekter av mindre klasser”

Working papers

2012:1 Edmark Karin, Che-Yuan Liang, Eva Mörk and Håkan Selin ”Evaluation of the Swedish earned income tax credit”

2012:2 Jönsson Lisa and Peter Skogman Thoursie “Does privatisation of vocational rehabilitation improve labour market opportunities? Evidence from a field experiment in Sweden”

2012: 3 Johansson Per and Martin Nilsson “Should sickness insurance and health care be administrated by the same jurisdiction? An empirical analysis

2012:4 Lundin Martin and PerOla Öberg “Towards reason: political disputes, public attention and the use of expert knowledge in policymaking”

2012:5 Fredriksson Peter, Björn Öckert and Hessel Oosterbeek “Long-term effects of class size”

Dissertation series

2011:1 Hensvik Lena “The effects of markets, managers and peers on worker outcomes”